ENCYCLOPEDIA OF
GEOLOGY

ENCYCLOPEDIA OF GEOLOGY

EDITED BY

RICHARD C. SELLEY
L. ROBIN M. COCKS
IAN R. PLIMER

ELSEVIER
ACADEMIC
PRESS

Amsterdam Boston Heidelberg London New York Oxford
Paris San Diego San Francisco Singapore Sydney Tokyo

Site Classification

A W Hatheway, Rolla, MO and Big Arm, MT, USA

Introduction

Site characterization may be defined as the

> . . .three dimensional engineering geologic description of the surface and subsurface of the location for intended construction of engineered works, for habitation, com merce, resource development, mitigation of natural hazards or conduct of groundwater protection, waste management or environmental remediation.

Site characterization provides the design information necessary for project planning. Site characterization is frequently broken and discontinuous in nature, reflecting the naturally flawed ground that is being evaluated (Figure 1). The broad elements of site characterization are listed in Table 1. No such table can be wholly inclusive, but it serves to establish a format for discovering and predicting the elements that should be included within the scope of construction work.

The broad goals of site characterization include:

- identification of sites likely to experience ground displacement from fault movement;
- identification of the stability of a site‘s configuration;
- establishment of the need to improve the engineering characteristics of the ground so as to resist damaging deformation;
- establishment of the need to control groundwater to prevent excess pore pressure and subsequent failure under the involved structural loads; and
- identification of the potential for creating green space not only for social gain but also to incorporate active faults and existing geohazards.

Purpose and Scope of Site Characterization

Site characterization is a major component of site investigation, the purpose of which is to determine ground conditions at a proposed construction site. The current terminology signifies a broadening of the purpose and scope to encompass the determination of relevant design needs for a wide variety of construction, resource development, hazard mitigation, and environmental goals.

Site investigation matured from the end of World War II through the 1970s, and then expanded in the 1980s, providing broad scope for the work of engineering geologists.

There is some published guidance on site characterization, but it is the responsibility of the engineering geologist in charge of each project to define the scope of the characterization in order to meet the design needs of the planner and engineer. These needs fall into seven basic categories:

- a three-dimensional description of the project site to the depth of influence (i.e. where the maximum applied (live and dead) static load from the intended use of the site has dropped to 10%);
- a description of the length and breadth of the proposed site and of the area beyond the property boundaries to such a distance as will encompass any anticipated effect of the activities contemplated at the site (this distance will be sufficient to cover the origins of geological constraints that may impinge on the contemplated site, including slope instability, groundwater inundation, and loss of ground support);
- the identification of ground that may be affected by displacement, volumetric change (shrinkage or swelling), or other forms of loss of physical support or chemical integrity relating to the intended construction;
- the discovery or prediction of ground instability above, below, on, or adjacent to the site;
- the detection and measurement of groundwater conditions (including perched and transient conditions) that may affect the construction, operation, or maintenance of the proposed works (including the estimation of annual variations and chemical composition, and in particular the potential to harm human health or the environment);
- the discovery or prediction of conditions that may adversely affect the rate, progress, or method of construction of the project; and
- the detection or prediction of ground conditions that may be unsafe or endanger the health of persons engaged in construction, or in work, visitation, or occupancy of the developed site.

Typical Site Characterization

The owner has the primary and ultimate responsibility for ordering and funding a site characterization and for ensuring that any constraints identified by the

Figure 1 Engineering geologists generally work with already flawed earth materials. This weak rock of Pennsylvanian age at Little Rock, Arkansas, needs to be ripped on excavation, yet it is wholly discontinuous and platy.

site characterization are addressed in the design, construction, operation, and maintenance of the engineered works. Table 2 assesses the typical goals and components of the site characterization of a variety of engineered works.

Areas Subjected to Site Characterization

In reality, no two parties engaged in planning and conducting a site characterization will produce identical pieces of work. The quality and applicability of every site characterization reflects the experience and competence of its leadership. Strictly speaking, the basic (minimal) characterization must assess the site within the property boundaries. There are many circumstances, however, in which additional areas of ground should be considered in order fully to address the client‘s needs. Table 3 presents some of these considerations.

Pitfalls of Site Characterization

There are no hard and fast rules or standards that can be applied to all site characterizations. It is incumbent on the engineering geologist to inform the owner of those elements of the work that may best meet the needs of the design engineers and the owner‘s goals for the project.

If the owner chooses to underfund the site characterization or selects characterization activities on a price-sensitive basis, there are numerous ways in which the end product may fail to give adequate design-related information. Where price has not been a determining factor in the scope of the site

Table 1 General elements of site characterization

Element	*Purpose*	*Important considerations*
Stratigraphy	Identify and describe geological formational units expected in design and construction	Individual engineering geological units
Groundwater regime	Define character of groundwater	Perched water Vadose zone Peizometric surface Potentiometric surface (if present) 12 months of observation
Top of rock	Define elevation or surface below which excavation is difficult	Little's grades of weathering will be helpful here
Rock mass characterization	Delimit observable or likely subsurface bounds of each detectable hard rock unit	Contacts Lithological character and variations Geological structure represented by equal area stereographic projection; to include bedding or foliation attitude, discontinuities, faults, fault zones, shear planes, metamorphic foliation Zones of bad ground, by virtue of alteration, weathering, or shearing Application of one of the standard rock mass qualification concepts
Presence of weak rock	Basis of definition, including why the rock is determined to be weak	Recommendations as to how and why such weak rock may pose problems in design and/or construction, operation, and maintenance
Potential problems related to sedimentological, structural, or geomorphological conditions	Portions of surface or subsurface that may be affected by otherwise unanticipated features	Buried valleys Stratigraphical 'holidays' Facies changes Fault displaced blocks of ground Dissolution cavities (karst features) Glacial erratics (boulders) Irregular features of glacial and periglacial origin
Potential instability	Hillsides River banks Ground expanse of site	Unstable slope masses Subsurface mined or dissolution caverns
Likely excavation characteristics	Relative ease or difficulty of moving earth media	Outstanding troublesome departures include glacial, bedrock knobs at top of rock, and some types of hard igneous intrusions such as dykes and sills
Geological constraints	Observed or postulated geological conditions that may be known or felt to place the project concept in jeopardy; may require additional funds for resolution	Otherwise known as 'geological hazards' but more properly known as 'geological constraints' after the life work of Robert F Legget

characterization, there are still categories of characterization flaws that, if not heeded, could endanger the achievement of design goals. Some of these are presented in Table 4.

Conclusion

The majority of the project technical responsibilities accorded to the engineering geologist fall under the remit of site characterization relating to construction and waste disposal. The aim is to ensure that properly funded exploration and evaluation can establish the geological information required by the design engineers. This information is located by spatial coordinates and is usually managed and displayed by computer-assisted drawing and geographical information systems. The geologist thereby develops a database of usable, relevant information in the form of colour-coded perspective drawings that are easily assimilated by an array of (non-geoscience) professionals who can subsequently manipulate the data for parametric analysis or for their own better understanding, and can do so largely without fear of distorting the relevant facts.

Table 2 Elements of site characterization

Construction works	*Design citeria*	*Characterization*
Construction		
Foundations	Performance of foundation Stability of adjacent buildings	Optimal use of foundation geology to install minimum cost foundation
Stability of excavations	Basements and parking spaces Lateral cuts for building space Recognition of groundwater Stability of adjacent structures	Basements are an economic imperative in cities owing to the cost of land
Drainage and dewatering	Keep excavation free of water Floodworks integrity	Construction control of water versus long term maintenance of groundwater and earthworks pore water
Materials of construction	Use on the project Export to urban areas	Dealing with public opposition Preservation of water quality
Land takings	Provide value to owner	Possible mineral resource value
Transportation		
Highways and rapid transit	Stable roadway Borrow for above ground works Balanced cut and fill	Sufficient to move traffic from A to B with minimal impact on existing infrastructure
Airfields	Runway space Fuel storage Runoff and waste management	Huge land surfaces subject to incidental wastes from operation Anti terrorist considerations
Tunnels and underground caverns	Optimal excavation method Promote ground stability Ground control Management of groundwater	Consider spoil handling and disposal, and stability of overlying build environment
Ports and harbours	Wave protection Operational wastes	Associated dredging may disturb historic contaminated sediment
Navigable rivers	Floodworks and flood control Locks and dammed pools	Associated dredging may disturb historic contaminated sediment
Energy and resource		
Dams	Foundation integrity Seismic withstand behaviour Slope stability Emergency spillways	Locate an inherently stable dam axis position Provide nearby construction materials
Pipelines	Route versus geological costs Erosion protection Seismic stability Stability of adjacent ground	Objective of minimal excavation, siphonage, and above ground support Anti terrorist considerations
Power transmission	Mainly tower stability	Minimize grubbing impacts
Forest resources	Slope stability Haul road location Potential hydraulic damage	Harvest the resource without diminishing the condition of the remaining land
Mineral resources	Minimize exploration damage Opencast Underground workings	Minimize exploration damage Manage acid mine drainage Spoil management Protection of water and soil values
Mineral beneficiation	Concentrate ore value Wastewater effluent	Manage soil impacts
Reservoirs	Reduce bottom leakage Sedimentation protection Shoreline protection Water quality	Requires a considered evaluation of geomorphology and the primary and secondary porosities of geological units in contact with stored water
Groundwater recharge	Use of storm waters Use of treated wastewater	Introduce flood and grey waters to the groundwater regime before their loss to oceans
Environmental		
Waste disposal facility siting	Minimize adverse impacts Optimize geological conditions Locate sufficient daily cover	Control of leachate Protection of water quality
Water quality protection	Suspended solids Dissolved solids Discharge into governed water	Cation exchange leads to degraded water quality (pollution), which can lead to health effects (contamination)

Table 2 Continued

Construction works	*Design citeria*	*Characterization*
Military impacts	Manoeuvre and exercise grounds	Vehicular disturbance of ground leads to sedimentation
	Munitions impact zones	
	Battle damage	Spilled fuels and discharged munitions lead to environmental degradation
Derelict land	Waste dumps Industrial brownfields Abandoned mined lands Battle damage	Ground receives contamination and surface waters and groundwaters transport this to human and natural life receptors
Environmental remediation	Uncontrolled hazardous waste	Characterization of site and waste is essential to formulation of remedial actions
		Without this actual risks become unacceptable
Saline intrusion	Control freshwater withdrawal Subsurface barriers	A mature art requiring only the application of sufficient funding
Sedimentation	Soil particles	Results from disturbed ground
	Cation exchange	Diminishes reservoir capacity when inflow ceases
	River and stream transport	
		Damages wildlife habitat by turbidity
Hazards		
Flood hazards	Flood channels No build zones in floodplains	Depends on geomorphological evidence of past flooding, integrated with runoff character of the related watershed
Rock falls Retain or devise protection from dislodgment or trajectory impact	Gravitational dislodgment of rock and rock masses above inhabited or frequented areas	Delimit rock or masses of rock above endangered populations or transportation routes
Snow avalanches	Recurrence intervals, and conditions and timing of recurrence	A sideline use of geological expertise in geological image interpretation
Seasonal advice on mitigation by explosive intervention or shed deflectors over railroad and highway routes		
Active faults	Ground trace	Dealing with vertical splaying of related fault splays in unconsolidated units
	Sense of displacement	
	Recurrence interval	
	Likely nature of strong motion	Blind thrust faults as unknowns
Unstable ground	Collapsing soils (water inflow)	Soil properties and character make them susceptible to damage
	Expansive soils (water inflow)	
	Liquefaction (earthquake strong motion in the presence of near surface groundwater)	
		Result damages foundations
Shore protection	Wave energy impact analysis Susceptibility of geological units to erosion and toppling	Evaluation requires geomorphological assessment in terms of the structural character of the shore rock
		Wave mitigation usually only transfers wave energy to new targets
Volcanic	Lava flows	Termination or preclusion not yet possible
	Ash falls	Geological characterization can delimit sources, routes, and probable areas of impact
	Pyroclastic bombs	
	Glowing avalanches	
	Mudflows (lahars)	Mitigation involves avoidance or channelization

Continued

Table 2 Continued

Construction works	*Design citeria*	*Characterization*
Defence		
Military posts and installations	Characterization of ground for optimal use in training and staging, while striving for optimal environmental protection of military bases	World economic situation currently favours diminishing grounds allotted to military forces, while increasing environmental controls over lands used for military training
Use of terrain	Defence	Trafficability
	Offense	Obstacles and fields of fire
		River crossings
Potable water	Exploration and development	Alternative to use of water in the field of operations or to importation in competition with high priority food and munitions
	Replacement of damaged wells and reservoirs	
		Supplies for indigenous peoples and prisoners of war
Anti terrorism actions	Defence from attack	Use or exclusion of use of such features as caves, excavated tunnels, and unstable hillsides
	Actions against terrorists	

In all cases, the characterization should reflect the owner's objectives, available funding, the normal and standing considerations for public health, welfare and safety, and applicable and relevant regulations and codes, and should be conducted in compliance with or with reflection on relevant codes and standards, and be completed by or under the direct supervision of a registered (chartered) professional geologist. All specialization categories should consider the presence of geological constraints and natural hazards that may affect the design, construction, operation, and maintenance of the project in a manner that meets the client's objectives and applicable national, regional, and local laws and regulations.

Table 3 Key considerations of site characterization

Construction works	*Key considerations*	*Other considerations*
Construction		
Foundations	To property boundary	Adjacent ground that may be affected by the project or have conditions that may affect the project
Stability of excavations	Host ground to include likely active mass of wall or face subject to possible failure	Possible adversely orientated planes or wedges of geological weakness
Drainage and dewatering	Maximum radius of cones of depression	Search for possible hidden (geomorphological) pathways of high hydraulic transmissibility
Materials of construction	Secure projected quantities of acceptable material	Locate superior materials at somewhat greater distance
Land takings	Natural resource value	Impact on groundwater resources
Transportation		
Highways and rapid transit	Right of way plus additional width affecting existing infrastructure	Minimum of one boring or geophysical traverse for each suspected adverse geological possibility Minimize groundwater impact
Airfields	Footprints of load bearing areas	Toxic waste hotspots in areas of runway expansion
Tunnels and underground caverns	Route characterization for estimated ground conditions Excavation Ground support Groundwater management Spoil management	Geormorphologic and structural hazards such as intercepted bad ground or groundwater flow pathways
Ports and harbours	Toxic dredge spoils	Locations of toxic hotspots
Navigable rivers	Lock and dam foundations	Stability of caving river banks
Energy and resource		
Dams	Stability of embankment	Consider adjacent ground hazards
Reservoirs	Minimize bottom leakage	Minimize incoming sediment
Groundwater recharge	Maximize storativity	Use to protect groundwater regime
Pipelines	Assess geological excavation cost versus alternative routes	Minimize environmental impact on the traversed ground

Table 3 Continued

Construction works	*Key considerations*	*Other considerations*
Power transmission	Cost effective tower locations	Minimize impacts from unstable ground on or adjacent to route
Forest resources	Minimize hydraulic damage	Minimize sediment produced
Mineral resources	Minimize exploration and development impacts	Devise safe spoil disposal strategies and minimal cation exchange leaching
Mineral beneficiation	Proper management of spoil	Plans for plant closure
Environmental		
Waste disposal facility siting	Isolate from groundwater	Provide adequate cover sources
Water quality protection	Remove toxic hotspots	Control entry of freshwater
Military impacts	Remove explosive relics	Grade land for minimal impact
Derelict land	Remove toxic hotspots	Control entry of freshwater
Environmental remediation	Remove toxic hotspots	Control entry of freshwater
Saline intrusion	Control groundwater pumping	Inject flood or treated waters
Sedimentation	Reduce land disturbance	Define interception features
Hazards		
Flood hazards	Establish likely inundation	Evidence of catastrophic flooding
Rock falls	Unstable rock masses	Impact zones on adjacent ground
Snow avalanches	Photogeological evidence of pathways	Foundations and anchorage of structures to deflect snow
Active faults	Establish likely rupture location	Susceptibility to lateral spread or liquefaction
Unstable ground	Mainly related to regional physiography	Avoidance first; mitigation second
Shore protection	Delimit susceptible ground	Other shoreline influences, such as groundwater seepage
Volcanic	Pathway of impact	Methods to deflect impact
Defence		
Military posts and installations	Footprints of structures	Areas of intense use by vehicles
Use of terrain	Obstacles Trafficability	Interdiction of enemy movement Neutralization of underground space
Potable water	Exploration for wells	Water supply basin
Anti terrorism actions	Features for approaches	Unstable surrounding ground Wellfield susceptibility to sabotage

In all cases, the characterization should reflect the owner's objectives, the level of available funding for exploration and characterization, and the normal geotechnical, hydrological, and environmental practices employed in the region.

Table 4 Some pitfalls of site characterization

Potential pitfall	*Nature*
Failure to consider owner's goals and design conditions	Represent the stated truths of project impact on the geological environment. Owner should be informed of the consequences of these considerations, though it is the owner's responsibility to act on such information
Failure to separate factual from interpreted geological information	Discrimination should be made in order to represent the owner's rights and position, in which the construction contractor bears the normal risks associated with construction under the conditions reported in the design documents and construction specifications. Contractors are entitled to additional payments for geological conditions not adequately defined in the contract documents. Separation of factual from interpreted information more clearly defines the information available for consideration by the contractor in bid formulation
Failure to apply site conceptual geological model	Essentially a schematic drawing (cartoon) of known and possible geological conditions at the site, superimposed on one or two (perpendicular) sketch geotechnical profiles, from ground surface to maximum depth of expected impact on site design
'Geological possibilities'	It is proper to consider the possibility of certain sedimentological, stratigraphical, structural, and geomorphological anomalies at the site, resulting from its general regional physiographical affinity. These possibilities should be assessed in terms of the site development plan

Continued

Table 4 Continued

Potential pitfall	*Nature*
Regionally important geological material	Failure to recognize that the subject earth material in the physiographical province has special or unique characteristics that affect or its geotechnical behaviour
Geomorphological and structural geological features	Failure to consider the potential for anomalous subsurface zones, pockets, or buried valleys of materials that very significantly from those of the general host ground
Neglect to consider discontinuities	Failure to recognize that discontinuities represent preformed surfaces of weakness and therefore have the potential for displacement and sometimes groundwater or contaminant migration
Assessment of ‘zones’	‘Zones’ should be invoked whenever evidence suggests the existence of pockets, bands, or other bounded bodies of earth material that could affect the design or performance of the project. ‘Zones’ are labelled and delimited by the characterizing engineering geologist
Deleterious minerals	Ground that has been geochemically altered should be expected to potentially produce *in situ* conditions or construction material that may not perform adequately in terms of project design. Often this occurs along the borders of intrusive bodies
Misassessment of karst	Karst terrain typically shows a concentration of dissolution along major joints and at their intersections. Much of the intervening ground may be devoid of dissolution features of sufficient magnitude to affect the project. Use of geophysical techniques is recommended.
Minimal time sensitive hydrogeological data	Only minimally accurate assessments of the character and seasonal behaviour of the groundwater can be made with a full 12 month record of hydrogeological conditions at and around the site
Inapplicable hydrological borehole testing	A prime example is the conduct of downhole packed hydraulic conductivity testing in a string of end to end positions, in which the conductivity is then averaged over the packed interval, and sometimes over the entire tested interval, whereas the true impact on project design and performance is more properly assigned to individual layers or zones, which themselves should be subjected to packed testing. On occasion this pitfall is used by litigation opposition in order to impart a condition more favourable to the client’s interests
Belief in the ‘pristine’ brownfield (derelict) site	All brownfield sites should be considered contaminated until shown otherwise; brownfield sites cannot be adequately characterized for their toxic wastes without knowledge of site technical and operational history and an understanding of the waste generation nature of the industrial flow path most likely to have been present during operation
Notion of ‘non aqueous’ phase liquids	Most are toxic to people and the environment. ‘Non aqueous’ is a chemist’s term; these compounds have the potential to dissolve in surface water and groundwater in concentrations inimical to health and the environment
Significance & VOCs and SVOCs	Volatile organic compounds (VOCs) migrate generally and in the hydrogeological system; semi volatile organic compounds (SVOCs) more commonly reside at their location of origin or dumping unless transported in sediment; SVOCs are characteristically recalcitrant to natural or induced degradation

See Also

Engineering Geology: Codes of Practice; Aspects of Earthquakes; Geological Maps; Natural and Anthropogenic Geohazards; Site and Ground Investigation; Subsidence. **Geological Field Mapping**.

Further Reading

Anon (1970) Logging of cores for engineering purposes. Working Party Report. *Quarterly Journal of Engineering Geology* 3: 1 24.

Bell FG (1975) *Site Investigations in Areas of Mining Subsidence*. London: Newnes Butterworths.

Bell FG (1992) Description and classification of rock masses. In: Bell FG (ed.) *Engineering in Rock Masses*, pp. 54 77. London: Butterworths Heinemann.

Bell FG, Culshaw MG, Cripps JC, and Coffey JR (eds.) (1990) *Field Testing in Engineering Geology*. Engineering Geology Special Publication No. 6. London: Geological Society.

British Standards Institution (1999) *Code of Practice on Site Investigations, BS 5930*. London: British Standards Institution.

Cabrera JG and Hatheway AW (1991) Investigation of preferred sites for selection and design. In: Kiersch GA (ed.) *Heritage of Engineering Geology The First Hundred Years*, pp. 395 428. Centennial Special Volume 3. Colorado: Geological Society of America.

Clayton CRI, Matthews MC, and Simons NE (1996) *Site Investigation*. 2nd edn. Oxford: Blackwell Science Limited.

Craig C (ed.) (1996) *Advances in Site Investigation Practice*. London: Thomas Telford Press.

Culshaw MG, Bell FG, Cripps JC, and O‘Hara M (eds.) (1987) *Planning in Engineering Geology*, pp. 151 154.

Engineering Geology Special Publication No 4. London: The Geological Society.

Department of the Environment (1976) *Reclamation of Derelict Land: Procedure for Locating Abandoned Mine Shafts*. London: Department of the Environment.

Griffiths JS (ed.) (2001) *Land Surface Evaluation for Engineering Practice*. Engineering Geology Special Publication No. 18. London: Geological Society.

Hatheway AW (2002) Geoenvironmental protocol for site and waste characterization of former manufactured gas plants; worldwide remediation challenge in semi volatile organic wastes. *Engineering Geology* 64: 317 338.

Hawkins AB (ed.) (1986) *Site Investigation Practice: Assessing BS 5930*. Engineering Geology Special Publication No. 2. London: The Geological Society.

Hempen GL and Hatheway AW (1992) *Geophysical Methods for Hazardous Waste Site Characterization*. Special Publication number 3. Lakewood CO: Association of Engineering Geologists.

Hudson JA, Brown ET, Fairhurst C, and Hoek E (eds.) (1993) *Comprehensive Rock Engineering: Principles, Practice and Projects*. Oxford: Pergamon Press.

International Society for Testing and Materials. (1995 to present) *Various Site Characterization Standards Developed by Committee D 18; Soil and Rock and Committee E 50: Environmental Assessment*. West Conshohocken, PA: International Society for Testing and Materials

International Society for Testing and Materials (1995) *Guide to Site Characterization for Environmental Purposes with Emphasis on Soil, Rock, the Vadose Zone and Ground Water*. Standard D 5730. West Conshohocken PA: The International Society for Testing and Materials.

Lee LT, Davios WM, Goodson RA, Powell JF, and Register BA (1994) Site characterization and analysis penetrometer system (SCAPS) field investigation at the Sierra Army Depot, Herlong, California. In: Report WES/TR/GL 94 4, U.D. NTIS order number AD A277887. Vicksburg, MS: US Army Engineering and Waterways Experimental Station.

Little AL (1969) The engineering classification of residual tropical soils. In: Moh Z C (ed.) *Proceedings of the 7th International Conference on Soil Mechanics and Foundation Engineering: Specialty Session on Engineering Properties of Lateritic Soils 28 29 August, 1969, Mexico City, DF, Mexico*. pp. 1 10. Bangkok, Thailand: Asian Institute of Technology.

McCann DM, Eddleston M, Fenning PJ, and Reeves GM (eds.) (1998) *Modern Geophysics in Engineering Geology*. Engineering Geology Special Publication No. 12. London: Geological Society.

McDowell PW, Barker RD, Butcher AP, *et al.* (2002) *Geophysics in Engineering Investigations*. CIRIA C562. London: Construction Industry Research and Information Association.

Olson O (1992) *Site Characterization and Validation Final Report*. Report STRIPA TR 92 22, (US NTIS order number DE93 603603). Stockholm: Swedish Nuclear Fuel Supply Company (SKB; SvenskKaernbraenslefoersoerjning AB).

Sara MN (1993) *Standard Handbook for Solid and Hazardous Waste Facility Assessments*, 2nd edn. Boca Raton, FL: Lewis Publishers.

Sara MN (2003) *Site Assessment and Remediation Handbook*, 2nd edn. Boca Raton, FL: Lewis Publishers, CRC Press.

Simons NE, Menzies B, and Matthews M (2002) *A Short Course in Geotechnical Site Investigation*. London: Thomas Telford Press.

Weltman AJ and Head JM (1983) *Site Investigation Manual*. CIRIA Special Publication 25. London: Construction Industry Research and Information Association.

Subsidence

A B Hawkins, Charlotte House, Bristol, UK

Introduction

Ground subsidence that results in settlement or collapse of the ground surface is grouped into four main categories: (1) subsidence due to man-made voids and natural voids relatively close to the surface (e.g., due to coal mining, stone mining, or karstic features), (2) subsidence due to the removal of fluids from depth (e.g., water/oil extraction) and the consequential change in effective stress conditions, (3) subsidence due to the removal of soluble minerals (salt, gypsum) in groundwater, and (4) subsidence due to the removal of fines in suspension (piping). In addition, subsidence may occur related to alluvial deposits, shrink/swell, volcanism, and thermokarst.

Major Voids – Man-Made and Natural

Mining

The extraction of stone for building construction, and coal for energy, has taken place since time immemorial. Early stone mining invariably extended from adits excavated into a quarry face, either when moist stone was required for carving and/or the overburden was prohibitively thick. On the other hand, coal mining generally extends from shafts, except where coal seams outcrop on valley sides in areas such as

South Wales. There are three main methods of mining, long wall, pillar and room, and linear. In long-wall mining, ground subsidence is expected, whereas in pillar-and-room/stall mining, long-term subsidence is considered possible. Subsidence is not anticipated in linear mineral mining.

Long-wall mining is a relatively recent development. Up to 80% of the horizon may be removed, generally using machinery to win material from a long face. During the extraction process, the roof is supported; behind the advancing face, the roof is allowed to collapse. Although some bulking will occur as the material collapses, the depth of settlement at the surface is commonly one-third to two-thirds the thickness of the worked seam. Settlement is likely to extend over a larger area than that mined because the failure of the ground is controlled by the angle of draw, commonly 25–45° from the vertical, depending on the nature of the strata. This method of mining is applicable only when the material won can be fragmented, and hence in Britain has been used mainly for coal extraction and the working of the Fuller's Earth Bed at Bath.

Pillar-and-room/stall mining has been carried out for many hundreds of years. It is the main method of stone extraction and an early method of winning coal and salt. The size and shape of the support pillars depend very much on the nature of the material being won and the depth of the resource. In stone mines, the pillars may be narrower because the material is stronger, but because it is also more brittle, failure may occur by surface spalling. In weak materials such as coal, salt, and gypsum, creep within the material would be anticipated and hence the pillars are invariably bigger, the height/width ratio being an important consideration. As a consequence of the mining, the ground above the rooms/stalls is de-stressed and spalling of the roof material may occur, resulting in void migration. Important considerations are the dimensions of the rooms/stalls and the nature of the overlying strata. Mudrocks may experience stress-release fissuring and spall or may act as a competent material. Over shallow mines, despite bulking, the voids may extend through to the surface to form a crown hole collapse. Alternatively, the upwards migration of the void may be intercepted by a sandstone horizon with fewer discontinuities, with the result that the original height/width ratio of the pillar may be greatly changed. In this situation, the sandstone acts as a bridge, but if overstressing of the elongated pillars results in the collapse of one or more supports, the adjacent columns must take the additional weight, creating a scenario whereby larger areas may collapse as a result of the domino effect. Typical examples are in the Coal Measures at Bathgate, south-west of Edinburgh, and the Heidegroeve area in the Netherlands, where the roof of a calcarenite mine failed in June 1988. On both occasions, the sudden collapse of the ground produced 'earthquake' shocks.

In addition to the failure of the roof over pillar-and-room workings, collapses frequently occur close to adits and/or mine shafts where the ground experiences more intense climatic change than it does deep in the mine. Ground failure also commonly occurs when shafts or bell pits have been inadequately filled and/or capped.

Linear mineral mining is generally related to the extraction of lodes of metalliferous minerals. Such narrow workings rarely cause ground subsidence because the nature of the rock does not facilitate void migration and the rocks are sufficiently competent to arch over the void. However, subsidence may occur related to old shafts, unless they have been appropriately filled or capped.

Subsidence manifested through a particular rock horizon may be created by failure of mining undertaken in lower strata. Such a phenomenon is common in the West Midlands, where the Silurian limestones have been mined as a flux for the iron and steel industry but the subsidence is seen in the overlying Coal Measure rocks.

Mining frequently involves the necessity to drain the strata. This may result in consolidation of the overlying sediments and/or dissolution as groundwater passes down to much greater depths than would otherwise occur. The consolidation resulting from a modification of the groundwater regime in the area may result in structures above and adjacent to the workings experiencing settlement. If dissolution of cements occurs, the percolating groundwaters may carry a fine fraction in suspension, which can again result in settlement.

Increasingly, tunnels and other areas of underground space are now being developed. During the driving of a tunnel, the ground above the work first experiences extension, and then, as some settlement occurs, the stress regime changes to one of compression. Even when every attempt is made to reduce subsidence and its effect on transport routes and property above a tunnel, including the use of compensation grouting, a settlement trough is a common feature.

Karst

Calcareous rocks are prone to dissolution by slightly acidic water, whether the acidity is carbonic acid from the atmosphere, humic acids from the near-surface organic-rich horizons, or sulphuric acids associated with the weathering of iron sulphides. Although the older limestones (such as the Carboniferous limestones), having experienced greater tectonic stress and have more pronounced fractures/discontinuities, younger limestones (such as the Chalk) also have cracks and hence significant secondary permeability. The high permeability results in limited near-surface erosion and hence the calcareous rocks invariably form uplands with low groundwater levels. Where the calcium carbonate goes into solution, it moves through and may go out of the limestone mass, leaving extensive openings/cave systems, particularly in the older limestones. These underground karstic features may develop at any level, depending mainly on the depth of groundwater in the geological past. Subvertical chimneys are commonly found connecting the semihorizontal caves. These low-pressure zones result in preferential percolation, above which swallow holes and other large surface features are common. In stronger limestones, the arching effect usually prevents collapse of the rock, although spectacular hollows have occurred where the roof of an underground passage has suddenly given way. The infill of wide solution hollows, such as grykes, swallets, or pipes, may be subject to sudden washout and may lead to collapse.

In both the older limestones (such as the Carboniferous Limestone) and the younger Chalk materials, man-made excavations occur. Although in the stronger older limestones it is easy to create stable mine passageways, where Chalk has been exploited for lime burning or to remove flints for building, it is much more difficult to assess the stability of the host rock. Quite spectacular subsidences have taken place related to old Chalk workings, such as occurred at Reading in the 1990s.

Fluid Extraction

Overburden pressure is supported by the volume and structure of the underlying material and its contained fluids. If the condition is confined, with no egress of fluids, the ground will be stable. However, if there is a reduction in the pore pressure due to a net egress of fluid (egress minus ingress), then the pressure experienced by the soil/rock supporting the overburden will be greater, the increase being [in] related to the [reduction in] stress previously taken by the liquid. [With the increase in] as a consequence of the increased amount of stress taken by the soil mass, some internal failure or distortion takes place. This causes non-elastic subsidence and, due to the rearrangement and/or decrease in volume of the material mass, settlement may be manifest at the ground surface.

Groundwater Extraction

Groundwater extraction has taken place under a number of large cities and has resulted in subsidence. Notable subsidence occurred in central London between 1865 and 1931 due to water abstraction from the Chalk. Beneath Mexico City, there is an aquifer between 50 and 500 m below ground level. Progressive pumping from this aquifer for over 100 years has resulted in much of the old city settling by 4 m, and in the north-east of the city, settlement of 7.5 m has been recorded.

Problems with settlement have been noted in the Shanghai area, where up to 300 m of Quaternary deposits are present in the Yangtze delta. These deposits contain five aquifers that were extensively pumped during the twentieth century. In the early years (1921–48), only some 9 million m^3 of water was pumped, but it was noted that a ground settlement of 23 mm per annum occurred and a 19-km^2 area experienced settlement of more than 0.5 m. With the increase in population of the city, between 1949 and 1956, the extraction rose to 140 million m^3 per year, resulting in an average settlement of 43 mm and an area of 93 km^2 experiencing in excess of 0.5 m subsidence. Between 1957 and 1961, the abstraction rose to 200 million m^3 per year and the ground settlement was 100 mm per year. Having recognized the problem that the water extraction was creating, the rate of pumping was reduced to 72 million m^3 per year between 1966 and 1989, and the settlement dropped to 2 mm per year. However, in the period 1990 to 2001, the rate of extraction increased to 113 million m^3 per year and the settlement increased to 16 mm per year. As a result of the groundwater extraction, the ground in the central part of Shanghai has settled by over 2.5 m. As the urban sprawl has extended, the settlement has expanded out such that now the whole of Shanghai has effectively suffered some ground settlement due to water extraction. This has caused major problems and damage to the infrastructure-sewers, roads, subway tunnels, and buildings.

Groundwater extraction near Pixley in California caused some 0.75 m of settlement between 1958 and 1963, the surface depression being mainly over the area pumped, with the effect decreasing with distance. In the Houston/Galveston region, the presence of faults has restricted the area affected by the dewatering such that a face up to a metre high has developed over a length of almost 17 km.

Oil and Gas Extraction

Oil extraction became important in the twentieth century and has also been associated with ground subsidence. In the Wilmington Oil Field of California, subsidence was first noted late in the 1930s, within 3 years of the commencement of oil production. By 1947, the rate of subsidence was some 0.3 m per year, and by 1951, when 140 000 barrels a day was being extracted, the rate of subsidence had reached 0.7 m per year. In 1957, a decision was made to inject water into the sediments in an attempt to repressurize them, and by 1962, subsidence had effectively ceased over most of the field. Nevertheless, by 1966, up to 9 m of settlement had occurred over an elliptical area of more than 75 km^2.

The Ekofisk Oil Field in the Norwegian sector of the North Sea is a series of dome-like structures within the Cretaceous and Paleocene Chalk, sealed from other formations by a shale/low-permeability chalk cap and flanked by low-porosity chalks. The hydrocarbon is extracted from the Chalk reservoir some 3 km below the seabed. With hydrocarbon extraction, there was little natural replenishment of the pore fluids and hence the 3 km of overburden pressure caused breakdown of some of the Chalk within the reservoir rocks, resulting in a loss of support for the overlying strata and surface settlement. The initial subsidence was reportedly in the order of 0.5 m per year, but analyses carried out subsequently have indicated that a seafloor settlement in the order of 7 m has occurred. The exploration/production platform was raised by 6 m and recharge of the reservoir was undertaken by injecting fluids.

Settlement has been reported in the Groningen Gas Field, where there is exploitation of between 70 and 240 m of Permian sandstones at a depth of some 3 km. In this area it is estimated that the subsidence will exceed 0.25 m by 2025. Although this is not a major subsidence, in the low-lying Polder region of Holland, the effect of the ground settlement is significant.

Subsidence Related to Dissolution of Salts

The most important halite deposits in Britain occur between Birmingham and Manchester, where notices frequently draw attention to the effect of salt subsidence. Subsidence related to the halite of Triassic age beneath the plains of the north-west Midlands has developed as a consequence of natural events and human activity. In many areas, particularly around Northwich and Winsford, the salt subcrops beneath the Quaternary deposits, which are frequently quite granular in character. Percolating groundwaters dissolve the salt until the waters become saturated, producing a phenomenon referred to as ‘wet rock head’. The dissolution of the halite removes support from the material above, leading to subsidence of the overlying materials and settlement at the ground surface.

Although mining of the salt has taken place in the north-west Midlands, most salt is now won by brine pumping. As the liquid being drawn to the low-pressure zones around the brine pumping well passes along the natural discontinuities within the rock mass, there is no predictable size or shape of the zone from which the salt has been removed. It may be roughly circular or almost linear, depending on the structure/permeability of the strata. Clearly the dissolution of the salt will be at its maximum where the incoming water is fresh, whereas around the extraction point, provided the rate of extraction is not too fast, the water will be saturated and the brine can be brought to the surface for the salt to be crystallized out. Large subsidence features have been formed as a consequence of brine pumping, such as in Bottom Flash, near the village of Winsford. On many occasions, the main collapse takes place after extraction in an area has ceased. Subsidence up to 1.6 m has been recorded near Hengelo in The Netherlands, with some of the surface settlements occurring several years after pumping ceased.

In other areas, salt has been mined by pillar-and-stall mining, which creates large underground cavities. Because the salt is prone to dissolution and creep, these mines become unstable in time; serious problems have occurred in areas such as Northwich, where the British Government has provided grant funding for the infilling of some of the mines.

Gypsum is another mineral that is prone to dissolution. A number of subsidence features have developed in the Ripon area of Yorkshire due to gypsum mining. General areas have sunk and/or hollows have developed, in part through collapse into voids created in these soluble mineral beds.

Flowing Water

The natural flow of water through granular/noncohesive material is capable of leaching the finer fraction. When the texture of the soil is not controlled by the interaction between the coarser particles, the loss of fines removes support to the overlying material such that the coarser fraction is repositioned and densified. When this takes place, the overlying material, if uncemented, loses [the] some support, resulting in settlement, sometimes referred to as hydrocompaction.

Rainfall penetrating vertically into the ground will be deflected at the groundwater level or at the depth where the soil becomes less permeable, commonly at the top of the bedrock. Flowing down-slope, the rainwater develops a series of very small channelettes that progressively merge to form more definite channels, and then pipes, at this water level. These soil pipes are generally 75 to 100 mm wide but may be up to 200 mm across. When the flow becomes interrupted, resulting in eddying, the water within the pipes may create underground cavities, which subsequently collapse. Such subsidence hollows may be up to a metre or so deep and in the order of 1 to 3 m across.

Alluvium

Most of the alluvial deposits in the world are estuarine in nature, although fluviatile sediments occur in valleys. Generally formed in the past 8000 years associated with the rising sea-level, alluvial deposits are normally consolidated, i.e., they have never had an overburden load greater than that existing today. For this reason the sediments often have a high moisture content and high porosity.

The main types of subsidence that occur in alluvial deposits are related to densification, when the deposits are either drained and/or are loaded in excess of the natural overburden pressure. In both situations, water is squeezed out through the pore throats of the sediment, particles are rearranged, and clay minerals are bent. Where the sediments have sufficient permeability that the water can be easily squeezed out, the settlement is relatively quick and is referred to as 'primary' consolidation, whereas finer sediments with smaller pore throats and lower permeabilities will experience a steady consolidation over a longer period, known as 'secondary' consolidation.

Where organic-rich sediments occur, the initial moisture content may be in the order of 1000%, with water accumulating both between and within the structure and cells of the various plants forming the marsh/peat material. When such deposits are loaded by a linear structure such as a road or railway, or by buildings, the overburden pressure squeezes out water from the organic-rich horizon, resulting in settlement. When the Holme Post was installed near Peterborough in Great Britain in 1851, to assess the shrinkage of the peat due to drainage, the top of the post was level with the ground surface. Within 10 years, the ground level had fallen by 1.5 m, and it had dropped by some 4 m by 1970.

Cuttings Resulting in Change in Groundwater Regime

Cuttings that create a modification to the groundwater regime will invariably result in subsidence as the zone of saturation is lowered and densification occurs.

Shrink/Swell

Groundwater levels vary in depth related to wet and dry periods. In areas where evaporation is comparatively strong, the evaporation front in the dry period may be relatively low such that the ground above loses most of its free water and some of the adsorbed water. Consequently, the ground shrinks and stiffens during the dry periods. Where hedges or trees exist, the nutrient-seeking roots remove considerable quantities of groundwater from the soils, producing a bowl-shaped zone of non-saturated soil. This frequently reaches 3 or 4 m in depth, and the lateral limit of dry soils may extend beyond the canopy of the tree. In the wet season, some of this effect is reversed and, if the trees are removed, rehydration may take place over the course of 5–10 years.

Subsidence Related to Volcanism

At Pozzuoli, on the coast adjacent to Mount Vesuvius in Italy, there is evidence that the ground has experienced both subsidence and uplift as a consequence of the change in ground stresses created by the build-up and release of pressures associated with volcanic activity. The uplift is clearly evidenced by the borings of marine molluscs found in columns some 2 m above sea-level whereas the subsidence is manifest by the presence of ancient floor slabs beneath the present sea-level.

Thermokarst

In Arctic areas, subsidence occurs when frozen ground is thawed. Near Edmonton, Canada, when a power station cooling lake was created, the perimeter dams subsided by more than 1.5 m when the ice lenses within the frozen ground thawed and the overburden pressure displaced the water.

Glossary

collapse The sudden falling in or falling down of the ground.

compaction Densification of the ground induced by human activity, such as the rolling of a road pavement, whereby the particles are forced together by compactive effort to stabilize the material and to reduce interparticle voids.

consolidation A natural densification generally accompanied by a loss of water and/or the precipitation of a cement; may or may not involve particle rearrangement and/or the flexing of particles such as clay minerals.

settlement The surface manifestation of sinking ground.

subsidence The sinking of the ground.

See Also

Engineering Geology: Natural and Anthropogenic Geohazards; Problematic Rocks; Problematic Soils.

Further Reading

Bell FG and Stacey R (1992) Subsidence in rock masses. In: Bell FG (ed.) *Engineering in Rock Masses,* pp. 246 271. London: Butterworth & Heinemann.

Chai J C, Shen S L, Zhu H h, and Zhang X L (2004) Land susbidence due to groundwater drawdown in Shanghai. *Geotechnique, LIV* 2: 143 148.

Cooper AH (1989) Airborne multispectral scanning of subsidence caused by Permian gypsum dissolution at Ripon, North Yorkshire. *QJEG* 23: 219 239.

Evans WB, Wilson AA, Taylor BJ, and Price D (1968) *Geology of the Country around Macclesfield, Congleton, Crewe and Middlewich*. London: Institute of Geological Sciences, HMSO.

Leddra MJ and Jones ME (1990) Influence of increased effective stress on the permeability of chalks under hydrocarbon reervoir conditions. In: Proceedings International Chalk Symposium, Brighton, Thomas Telford, London, pp. 253 260.

Lofgren BN (1979) Changes in aquifer system properties with groundwater depletion. In: Saxene SK (ed.) *Evaluation an Prediction of Subsidence*. Proc. Speciality Conf. Am. Soc. Civil Engineers, Gainsville, pp. 26 47.

Price DG (1989) The collapse of the Heidegroeve: a case history of subsidence over abandoned mine workings in Cretaceous calcarenites. In: Proceedings International Chalk Symposium, Brighton, Thomas Telford, London, pp. 503 509.

Wassmann TH (1980) Mining subsidence in Twente, east Netherlands. *Geologie en Mijnbouw* 59: 225 231.

Ground Water Monitoring at Solid Waste Landfills

J W Oneacre and D Figueras, BFI, Houston, TX, USA

Introduction

Subpart E of the Subtitle D regulations in the USA require substantial ground water monitoring at municipal solid waste landfills [MSWLF]. Owners and operators must adequately define the geological and hydrogeological conditions and must develop a sampling and analysis plan that includes statistical analysis of the geochemical data. These regulations and causes of ground water variability at MSWLFs are discussed in the Further Reading Section at the end of this article. Causes for variability include inadequate site characterization, improper well design, drilling, development, and sampling, laboratory artefacts, and misapplication of statistical methods.

Under 258.54(a)(1)(2) of the Federal regulations and Texas Administrative Code [TAC] 330.234.(a)(1)(2), the executive director may delete any constituent that the owner can document is not reasonably expected to be in or derived from the waste. Also, the director can establish an alternative list of inorganic indicator constituents in lieu of some or all of the heavy metals if the alternative constituents provide a reliable indication of inorganic releases from the MSWLF unit to the groundwater.

This paper presents and proposes several alternative and innovative methods for groundwater monitoring specific to MSWLFs, that the authors believe can meet the regulatory criteria and be cost-effective. These methods include free carbon dioxide determination for landfill gas impact, thermal surveys for gas and/or leachate migration, stable isotope analyses for fingerprinting discrete sources, and passive sampling using dedicated sondes.

Free Carbon Dioxide [CO_2] Determination

The various phases of gas formation in MSWLFs are presented in Figure 1. Four phases exist and are designated as: Phase I, Aerobic; Phase II, Anaerobic Non-Methanogenic; Phase III, Anaerobic Methanogenic Unsteady; and Phase IV, Anaerobic Methanogenic Steady. During Phase II, CO_2 blooms to 50% and 90% by volume. From this peak, CO_2 decreases to about 30% to 50% during Phase IV. Although CO_2 concentration in water is less than 1 mg l at one atmosphere, enrichment to groundwater occurs due to the concentration of CO_2 in landfill gas. CO_2 is much more soluble in water than methane [CH_4], 1700 mgl compared to about 50 mg l; therefore, as the CO_2 gas

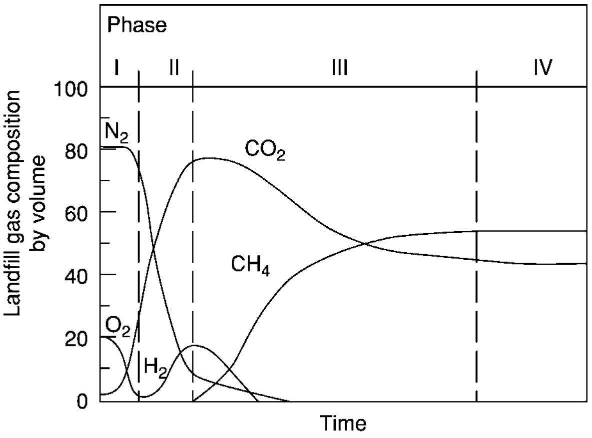


Figure 1 Phases of landfill gas formation (after Farquhar and Rovers (1973)).

pressure increases, CO_2 goes into solution, and formation of carbonic acid occurs which lowers the pH.

Landfill gas will travel through porous media via two processes; diffusion in response to a concentration gradient and convection from a pressure gradient. Gas will move more readily along paths of least resistance (zones of high hydraulic conductivity) and may move outward significant distances if upward movement is hindered by impervious layers such as frost, clay, concrete, etc. The depth to which landfill gas can penetrate is a function of site soil properties and soil moisture content. The groundwater surface represents a good lower limit of movement, particularly for methane; however, CO_2 will readily go into solution.

At MSWLFs with landfill gas migration, a full pH unit drop is typical in impacted wells as shown in Figure 2. Wells with no gas impact have pH values of about six whereas wells with most significant gas impact have a pH of five or lower. Since a relationship exists between pH, total alkalinity, and free CO_2; having values for two parameters allows the calculation of the third. For this site, the owner measures pH and total alkalinity and then determines the free CO_2 concentration. Background CO_2 values are below 50 mg l; however, gas impact has caused the CO_2 values to increase in excess of 400 mg l. Also, as CO_2 values increase, volatile organic compounds (VOCs) begin to appear and the percentages of sample with VOCs increases as CO_2 increases. In contrast, wells with background concentrations of CO_2, nearly 75% of all samples do not show VOCs (Figure 3).

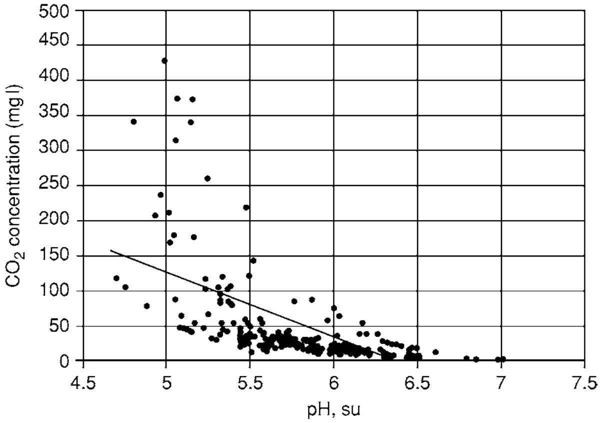
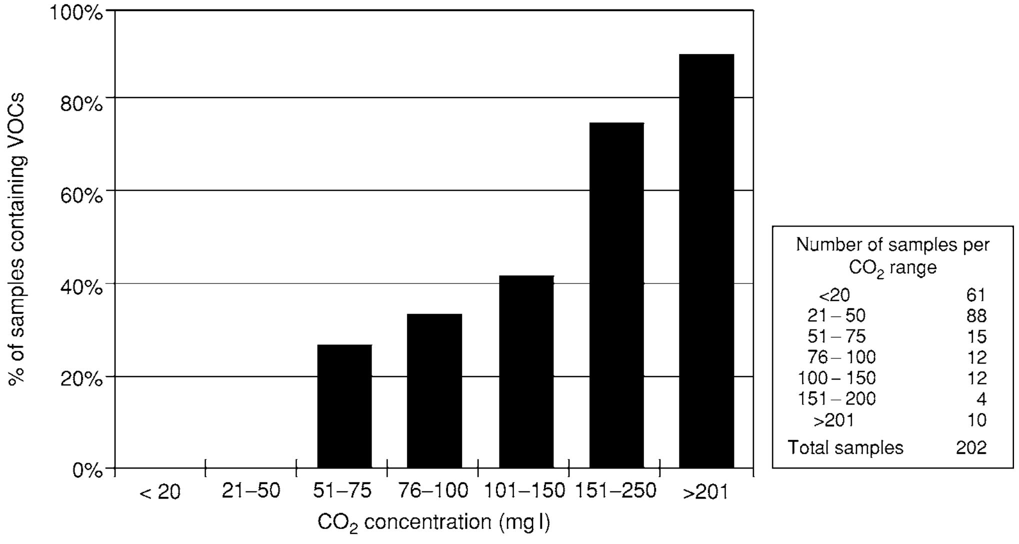


Figure 2 pH and CO_2 relationship in groundwater.

From the above discussion, there is a good correlation between elevated CO_2 values and the occurrence of VOCs. There are 149 samples with CO_2 concentrations less than 50 mg l and no detections of VOCs. In contrast, 47% of samples with CO_2 greater than 50 mg l contained VOCs and the likelihood of a VOC occurrence increases to 90% with increasing CO_2 concentration. Using free CO_2 concentrations as an indicator parameter could reduce the need for routine analyses of VOCs.

Thermal Surveys

The technical approach for using thermal surveying is based upon the fact that, in the absence of local influences, the Earth's temperature increases about 0.2 to 1.2°C with every 100 feet of depth below the ground surface. This temperature increase is due to conductive heat flow; heat moving through a medium such as



Number of samples per CO_2 range	
<20	61
21 – 50	88
51 – 75	15
76 – 100	12
100 - 150	12
151 – 200	4
>201	10
Total samples	202

Figure 3 VOC and CO_2 relationship.

soil, rock, or groundwater as a result of temperature differences between the inner Earth and ground surface. A simple example of heat conduction is heat transferred through a knife held over an open flame; the handle will become hot to the touch as heat moves from the knife tip to the handle. At shallow depths, temperature is also affected by the diurnal, or day–night cycle, and seasonal, or summer–winter cycle. Diurnal influences extend only a few feet below the ground surface but seasonal influences can reach to depths of about 50 feet.

Heat in the Earth may also move due to advection. Gas or water can convey heat through porous media and thus create temperature differentials that may reflect subsurface media processes. A highly permeable gravel bed will carry away the interior Earth's heat more rapidly and therefore show a temperature anomaly that is lower than surrounding, less permeable material. Conversely, an anomaly high will exist where the groundwater is locally affected by heat above the normal geothermal gradient. MSWLFs generate heat as a byproduct of decomposition of wastes; this is demonstrated by the fact that the optimum temperature for methane generation is about 35°C (95°F). Once heated, water temperature will change slowly since it has a high specific heat capacity compared to most natural materials and is therefore a potentially useful tracer. In addition, increased temperature of water will alter the density and viscosity such that groundwater at 40°C will travel twice as fast as water at 5°C. Researchers have used temperature to trace thermal anomalies as much as five miles from the source. Traditional application of thermal surveys include location of groundwater supplies, identification of geothermal areas, and delineation of groundwater seepage around dams.

Precisely calibrated thermistors measure electrical resistance which varies proportionally with temperature. Thermistor probes are set at depths of about 10 feet to eliminate diurnal effects. Equilibrium requires about twenty-four hours; readings are taken with a resistance bridge and values converted to temperatures using calibration tables prepared for each sensor.

Analyses of δD and δ^{18}O help identify the probable source of groundwater. If the isotopic composition plots close to the meteoric water line in a position similar to that of present-day precipitation for a particular geographic region, the groundwater is almost certainly meteoric.

For the southern California area near a landfill test site with elevated chlorides, the expected δD and δ^{18}O values typically fall in the range of $-50‰$ and $-7.5‰$, respectively. Due to shifts in climatic patterns during the Tertiary, the time period associated with the bedrock at the test site, precipitation from the Tertiary differed isotopically from the isotopic composition of present-day precipitation. This compositional difference varies based upon the specific stratigraphic horizon within the Tertiary. The significance of the isotopic differences is discussed below.

D/H and ^{18}O in Landfill Leachates

It has been observed that significant deuterium enrichment in landfill leachates are due in part to the anaerobic decomposition of organic wastes within landfills. In addition, it was noted that increased δD and δ^{18}O in groundwater were contaminated by an adjacent landfill. Leachates from landfills in Illinois have δD values enriched about 30% to 60% relative to δD of local precipitation.

A significant portion of the hydrogen in landfill methane comes from the landfill leachate where microbes preferentially use the lighter hydrogen isotope, thereby enriching the leachate with deuterium. As the landfill ages and more methane is generated, more enrichment of deuterium occurs. Thus deuterium enrichment makes deuterium a very useful tracer for detecting groundwater impact from landfills.

^{13}C in Landfills

The major components of landfill gas, methane, and carbon dioxide (CH^4 and CO^2), will show unique isotopic signatures compared to the surrounding environment. Methanogenesis will cause the CO^2 portion of the landfill gas to become enriched relative to δ^{13}C; whereas the CH^4 portion will become depleted relative to δ^{13}C. δ^{13}C values of around $+20‰$ for CO^2 and $-50‰$ for CH^4 have been reported. Values for δ^{13}C will increase as the landfill ages since the microbes will prefer to use the isotopically light carbon to produce CH^4, thereby enriching the CO^2 at the expense of the CH^4.

Because CO^2 is much more soluble in water than CH^4 (around 1700 mg l compared to 50 mg l), landfill leachates will also become enriched with δ^{13}C due to increased alkalinity. Using the isotopes of hydrogen and carbon, researchers have been able to differentiate various sources for CH^4.

Environmental Isotopic Analyses

General Principles

Isotopes of specific elements, such as hydrogen, carbon, and oxygen, have the same number of protons but different number of neutrons in the nucleus. This gives a different atomic weight for a specific element, even though the element maintains the same atomic number. As an example, hydrogen occurs as 1H with an atomic weight of 1 or as 2H (deuterium) or 3H (tritium) with atomic weights of 2 and 3, respectively. Stable isotopes do not undergo any natural radioactive decay and provide an understanding to the source of water or a process affecting water after release from the atmosphere. Any process that causes different isotopic ratios in different media is the result of isotopic fractionation. The most important process is vapour-liquid fractionation caused by evaporation and condensation. For instance, $^{18}O/^{16}O$ in rainwater is different from ocean water, which is different from oil brines.

For convenience, the δ (delta) notation is used to describe isotope ratios, which is defined as:

$$\delta \times [\text{sample}] = (R[\text{sample}] - R[\text{standard}])/R[\text{standard}] * 1000$$

where δX represents the relative difference in parts per thousand (called per mil (‰)) between the isotope ratio (R) in a sample and the ratio in some specified standard. For $\delta^{18}O$, the reference standard is SMOW, Standard Mean Ocean Water; $\delta^{13}C$ uses the PDB scale, a belemnite from the Pee Dee Formation of South Carolina; and $\delta^{34}S$ uses the Canyon Diablo troilite as the reference standard. Stable isotope values associated with landfill leachates and gases demonstrate pronounced signatures that are separate from the surrounding environment.

$^{18}O/^{16}O$ and D/H in Groundwater

The isotopic composition of seawater, is by definition, zero per mil for both $\delta^{18}O$ and δD. Upon evaporation from seawater, the water vapour has an isotopic concentration of about -80‰ for δD and -9‰ for $\delta^{18}O$. Fractionation processes affect the isotopes such that rain becomes progressively lighter in both δD and $\delta^{18}O$, as it moves farther from the ocean source. Data from modern rainfall plots on a straight line defined as:

$$\delta D = 8\delta^{18}O + 10 \qquad [2]$$

It has been suggested that an upper limit of $+18$‰ $+/-2$‰ exists for $\delta^{13}C$ in landfills due to eventual steady state conditions that develop between CO^2 input and CH^4 production.

^{13}C in Oilfield Water

Data regarding $\delta^{13}C$ in oilfield water are not abundant; however, study of $\delta^{13}C$ values for Miocene petroleum source rocks in California showed high $\delta^{13}C$ values, greater than $+5$‰ PDB. $\delta^{13}C$ studies conducted on Miocene rocks in California showed $\delta^{13}C$ values typically above $+5$ and as high as $+27.8$‰ PDB. Thus, the very high $\delta^{13}C$ values were the result of degradation of organic acid anions by sulphate-reducing bacteria. Decarboxylation of short-chain aliphatic acid anions, principally acetate, is a probable major source for CO_2 in the oilfield waters.

Isotopic Data Interpretation

Isotopic data from the California landfill site have been plotted in various combinations in Figures 4 and 5. Figure 4 depicts $\delta^{18}O$ ratio versus δD ratio and shows the meteoric water line for rainfall in the southern California area. If data points plot off this line, this would indicate that the water (fluid) tested had been formed in an environment significantly different and unique with respect to normal precipitation. Site groundwater samples tend to plot on or below the meteoric water line, showing that they are strongly related to modern precipitation. The brine samples plot well above the line, indicating a different source or time period for this water. The leachate samples are far removed from either brine or groundwater, showing a totally different and distinct process of formation. This plot clearly shows that leachate, groundwater, and brine have distinct isotopic signatures. MW-1, the highest chloride well, and the 'probe', (completed in highly fractured bedrock) are strongly shifted toward the Pico #1 brine sample. There is a strong probability based on this plot, that there is some mixing between the oilfield brine being produced from the Pico #1 (1500 foot zone) and the water samples obtained from MW-1 and the probe. Absolutely no mixing of groundwater and leachate are indicated based on this plot.

Figure 5 is a simple plot that illustrates the wide variance in carbon isotopic analysis of samples from groundwater, brine, and leachate sources. The dissolved inorganic carbon 13/12 ratios have distinct and definitive isotopic signatures. This indicates that the carbon isotopes are extremely useful in delineating any mixing between brine, leachate, and groundwater. Based upon the simple comparison of carbon isotopes, there is no relationship between the groundwater samples and landfill leachate.

Passive *In Situ* Sampling

Much research and discussion have occurred in recent years regarding appropriate groundwater sampling practices.

It has been shown that variable speed centrifugal pumps reduce purge time, aid in slug and pump tests,

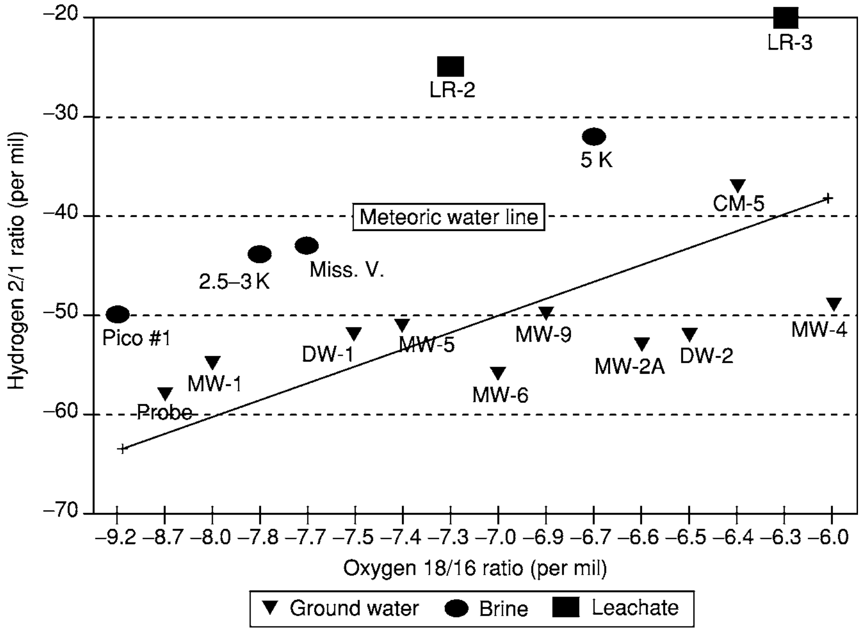


Figure 4 ^{18}O vs. D for brine, leachate, and groundwater for Californian landfill site.

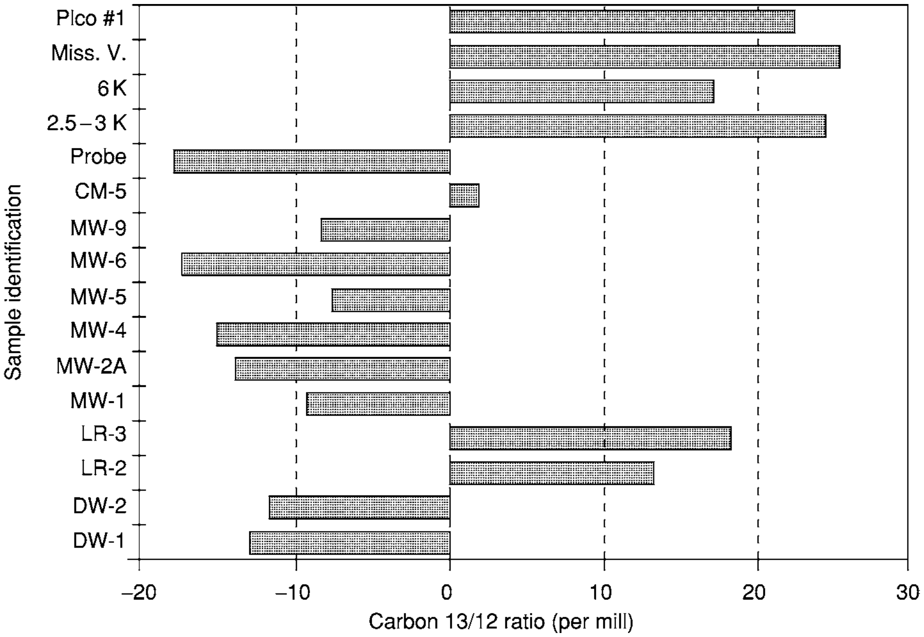


Figure 5 Carbon 13/12 ratios for brine, leachate, and groundwater for Californian landfill site.

and consistently provide low turbidity. Recent studies suggest that dissolved oxygen (DO) and Redox, key field parameters for groundwater sampling, are the last field parameters to stabilise. Unfortunately, these parameters are infrequently measured, even though DO and Redox play an important role in geochemical equilibrium of trace constituents such as heavy metals. Figure 6 shows upgradient and downgradient DO and Redox data from a MSWLF in New England with a defined leachate plume. Due to the anaerobic nature of leachate, impacted downgradient wells are depleted of DO and show a strongly reduced condition. Preliminary findings suggest that purging a nominal one well volume is necessary to establish equilibrium of DO and Redox and an average is two well volumes. In addition, atmospheric conditions

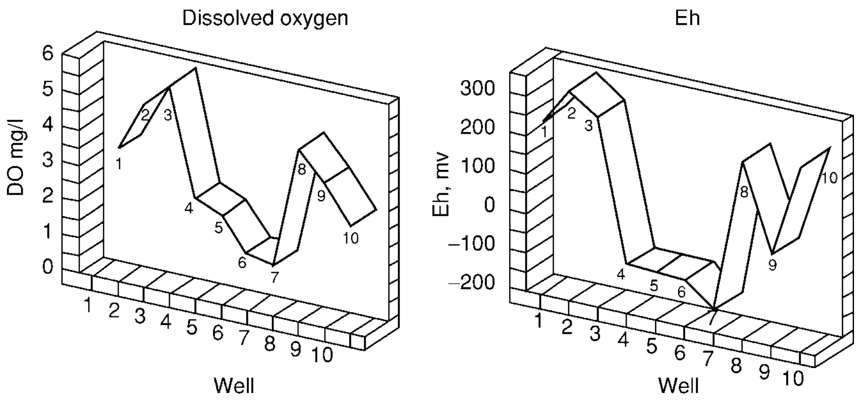


Figure 6 Reduction of DO and Eh in groundwater from leachate impact.

apparently influence DO and Redox in a standing column of well water to depths of 30 feet.

Low-flow purging that removes a minimal volume of water may not fully remove this atmospheric artifact. Figure 7 shows readings for DO and Redox, demonstrating an apparent 'stabilization' of the two parameters occurring at low-flow purging. Based upon USEPA guidelines for sampling, one would conclude that representative formation water is present in the well. However, increasing the flow rate shows a dramatic decrease in Redox of about 60%; a better representation of true formation water. This has significance for parameters sensitive to reducing conditions such as arsenic. Low-flow purging may have benefits for sites with known contamination, but for routine detection monitoring, may have limitations.

In situ passive sampling may have the benefit of eliminating purging while measuring key parameters specific to MSWLFs. *In situ* passive sampling can also provide continuous monitoring rather than semi-annual sampling. The owner can set the sampling intervals to weekly, daily, hourly, or by the minute. Maintenance is low, with only periodic recalibrations and battery replacement. Such *in situ* passive sampling devices can fit into 2 inch wells and monitor as many as eight parameters. For MSWLFs, key parameters would include: water level, temperature, pH, Specific Conductance, DO, Redox, chloride, ammonium, and CO_2. As stated previously, landfill gas impact is characterised by increased CO_2 and depressed pH. Conversely, leachate impact is characterized by increases in temperature, Specific Conductance, chloride, and ammonium, while DO and Redox will show dramatic decreases.

By obtaining data on a daily or weekly basis, short-term temporal or long-term seasonal changes can be noted and filtered from the data. For example, Figure 8 shows seasonal temperature data for the New England MSWLF.

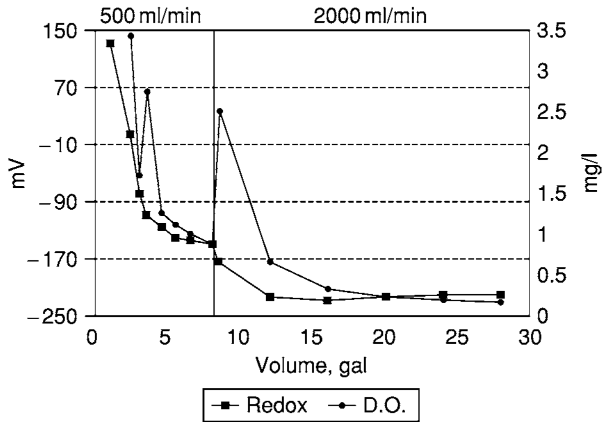


Figure 7 Changes in Eh and DO due to pumping rates in well MW 1 3D.

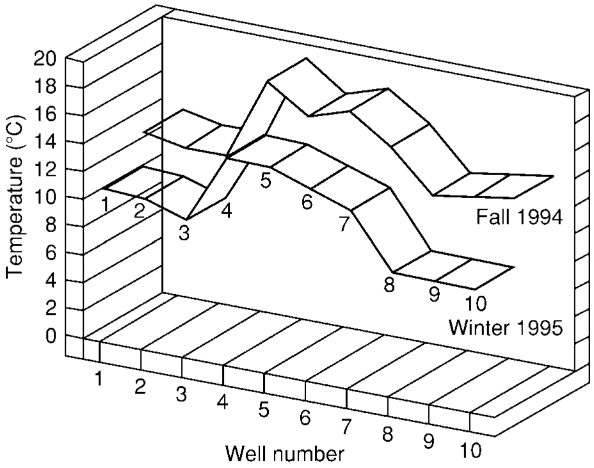


Figure 8 Seasonal temperature anomalies from leachate impact.

Note that seasonal fluctuations occur for temperature; however, the most impacted wells the GHR-11, maintains their high temperature anomaly from early autumn to winter. By utilising *in situ* passive samplers, owners and regulatory agencies could realize several key benefits:

- *Parameters Unique to MSWLFs* – The Subtitle D parameter list has inherent problems. Firstly, most of the heavy metals are not critical to MSWLFs; secondly, some are soluble under oxidising conditions and/or low pH, a situation not pertinent due to the strongly reducing and buffering nature of MSWLF leachate; thirdly, the list implicitly assumes leachate impact only and does not address impact from landfill gas. Alternative parameters, such as those presented in this article, not only serve as better indications of groundwater impact but also will better delineate the means of impact, leachate, or landfill gas.
- *Continuous Monitoring of Ground Water* – Rather than quarterly or semi-annual monitoring, *in situ* passive sampling can provide much more data that depicts temporal, seasonal, or long-term trends in groundwater quality.
- *Reduced Frequency of Sampling* – Owners could reduce sampling frequency by using the data from *in situ* monitoring to determine if and when a sample should be taken. By establishing confidence intervals or other statistical criteria, owners would take samples only when exceedences of such criteria occurred.
- *Cost Control* – Reduction of groundwater monitoring costs, particularly post-closure accruals, would assist owners with small buckets, such as municipalities.
- *Error Reduction* – Data is simply downloaded from a datalogger to a laptop computer, thereby reducing a significant source of error, manual transcription.
- *Better Public Relationship* – Continuous monitoring could reduce public concern and increase confidence that a MSWLF is operated properly.

Conclusion

This article presents options to traditional groundwater monitoring at MSWLFs. By incorporating these options, owners can better delineate leachate versus gas impact; fingerprint definitive sources, and generate more and higher-quality groundwater data specific to MSWLFs. *In Situ* passive sampling may provide several benefits that include sampling frequency reduction, understanding of temporal trends, cost control, transcription error reduction, and better public confidence.

See Also

Engineering Geology: Natural and Anthropogenic Geohazards; Liquefaction; Made Ground. **Quarrying**.

Further Reading

Baedecker MJ and Back W (1979) Hydrogeological Processes and Chemical Reactions at a Landfill. *Ground Water* 17: 429 437.

Barcelona Michael, *et al.* (1990) *Contamination of Ground Water: Prevention, Assessment, Restoration.* Park Ridge, New Jersey: Noyes Data Corporation.

Beluche R (1968) *Degradation of Solid Substrate in a Sanitary landfill.* Ph.D. Thesis, University of Southern California, Los Angeles, California.

Birman J (1990) *Handbook of Ground Water Development by Roscoe Moss Company.* New York: John Wiley and Sons.

Carothers WW and Kharaka YK (1980) Stable Carbon Isotopes Of HCO_3 in Oil field Waters Implications for the Origin of CO_2. *Geochimica et Cosmochimica Acta* 44: 323 332.

Coleman DD, *et al.* (1993) Identification of Landfill Methane using Carbon and Hydrogen Isotope Analysis. *Proceedings of 16th International Madison Waste Conference*, pp. 303 314. University of Wisconsin Madison.

Drever JI (1982) *The Geochemistry of Natural Waters.* Englewood Cliffs: NJ Prentice Hall.

Farquhar GJ and Rovers FA (1973) *Gas Production During Refuse Decomposition, Water, Air and Soil Pollution.* Vol. 2.

Matthess FPG and Brown RM (1976) Deuterium and oxygen 18 as indicators of Leachwater Movement from a Sanitary Landfill. In: *Interpretation of Environmental Isotope and Hydrochemical Data in Groundwater Hydrology.* Vienna International Atomic Energy Agency.

Games LM and Hayes JM (1977) Carbon Isotopic Study of the Fate of Landfill Leachate in Groundwater. *Journal of Water Pollution Control Federation* 49: 668 677.

Hackley KC, Liu CL, and Coleman DD (1996) Environmental Isotope Characteristics of Landfill Leachates and Gases. *Ground Water* 34(5): 827 834.

Liu, *et al.* (1992) *Application of Environmental Isotopes to Characterize Landfill Gases and Leachate.* Geological Society of America Abstracts with Programs, Cincinnati, Ohio, p A35.

Ludwig H (1967) *Final report: In Situ Investigation of Gases Produced from Decomposing Refuse.* California State Water Quality Control Board.

Murata KJ, Friedman I, and Madsen BM (1969) *Isotopic Composition of Diagenic Carbonates in marine Miocene Formations of California and Oregon.* US Geology Survey Professional Paper 614 B.

Oneacre JW (1992) *Solid waste Principles and Practice in the United States.* International Conference on Environmental Protection and Control Technology, Enserach, Kuala Lumpur, Malaysia, pp. 661 674.

Oneacre JW (1993) Subtitle D Regulations Impact on Ground Water Monitoring. *Geotechnical News* 11(3): 49 52.

Oneacre JW and Figueras D (1996) *Ground Water Variability at Sanitary Landfills: Causes and Solutions, Uncertainty in the Geologic Environment*. Proceedings, ASCE, Madison, WI, pp. 965 987.

Ramawsamy JN (1970) *Effects of Acid and Gas Production as Sanitary Landfills*. Ph.D. Dissertation,West Virginia University.

Rank, *et al.* (1992) *Environmental Isotopes Study at the Breitenqu Experimental Landfill (Lower Austria). Tracer Hydrology.* In: Hotzl and Werner (eds.) Proceedings of the 6th International Symposium on Water Tracing, Karlsruhe, Germany, Sept 21 26, pp. 173 177. Rotterdam: Balkema.

Whiticar MJ and Faber D (1985) Methane Oxidation in Sediment and Water Column Environments Isotope Evidence. *Advances in Organic Geochemistry* 10: 759 768.

ENVIRONMENTAL GEOCHEMISTRY

W E Dubbin, The Natural History Museum, London, UK

Introduction

Soils and sediments occupying the Earth's surface lie at the interface of the lithosphere, atmosphere, biosphere, and hydrosphere. Within the weathered, complex, and porous milieu of these Earth-surface materials, myriad biogeochemical processes govern the movement of both nutrients and pollutants from the lithosphere to biota, where they are incorporated into plant and animal tissues, or to groundwaters, where they may be transported great distances to streams, rivers, and oceans. This chapter describes the most notable pollutants, both organic and inorganic, and the dominant processes governing their mobility and bioavailability in terrestrial and aquatic environments.

Trace Elements

Trace elements are those that occur in the lithosphere at concentrations typically less than 1 g kg ^{1} (**Table 1**). Among the trace elements are the micronutrients (e.g., Cu, Ni, Zn), which are essential for the growth and development of micro-organisms, plants, and animals, and also the metalloids, which have characteristics of both metals and non-metals (e.g., As, B) (*see* **Minerals:** Arsenates). Heavy metals are defined as those trace elements with densities $>5.0\,\mathrm{g\,cm}$ ^{3}. Virtually all trace elements, even the micronutrients, exhibit toxicity to animals and plants when present at excessive concentrations. Radionuclides are a separate but important class of inorganic contaminant that may occur naturally (e.g., ^{222}Rn, ^{226}Ra, ^{238}U) or as a consequence of nuclear fission related to atomic weapons testing and nuclear power generation (e.g., ^{90}Sr, ^{137}Cs, ^{239}Pu).

Although trace elements occur naturally in all terrestrial environments, anthropogenic inputs may increase these concentrations considerably. The principal anthropogenic sources of trace elements are mining and smelting activities, fossil fuel combustion, chemical and electronics industries, as well as the addition of fertilisers and biosolids arising from agricultural operations. One notable example of severe trace metal pollution caused by smelting activities is found near Karabash, in the south Ural Mountains region of Russia, where for decades smelting operations have deposited metals on the surrounding landscape, destroying much of the vegetation and so contributing to widespread soil erosion (*see* **Environmental Geology**).

Toxic levels of trace elements (e.g., Cr, Ni) may also occur naturally, as in soils derived from serpentinitic rocks, which can lead to phytotoxicity and the consequent lack of vegetation over large areas of the landscape where the serpentine soils occur (*see* **Clay Minerals**). Alternatively, these elevated trace element concentrations may induce metal tolerance among certain plant species (e.g., *Thlaspi* spp.) as biological communities adapt to these metal rich environments.

Trace Element Bioavailability and Speciation

It is well established that total trace element content in soil or sediment is a poor indicator of toxicity. A more reliable measure of ecotoxicity is trace element bioavailability. A trace element is considered bioavailable if it can be utilized by biota. Bioavailability is therefore broadly equated with solubility, although some plants and micro-organisms are able to extract metals from solid phases normally considered insoluble. The solubility, and hence bioavailability, of a particular trace element is determined largely by its solid-phase speciation and mode of surface complexation. The main parameters

Oneacre JW (1993) Subtitle D Regulations Impact on Ground Water Monitoring. *Geotechnical News* 11(3): 49 52.

Oneacre JW and Figueras D (1996) *Ground Water Variability at Sanitary Landfills: Causes and Solutions, Uncertainty in the Geologic Environment*. Proceedings, ASCE, Madison, WI, pp. 965 987.

Ramawsamy JN (1970) *Effects of Acid and Gas Production as Sanitary Landfills*. Ph.D. Dissertation,West Virginia University.

Rank, *et al.* (1992) *Environmental Isotopes Study at the Breitenqu Experimental Landfill (Lower Austria). Tracer Hydrology.* In: Hotzl and Werner (eds.) Proceedings of the 6th International Symposium on Water Tracing, Karlsruhe, Germany, Sept 21 26, pp. 173 177. Rotterdam: Balkema.

Whiticar MJ and Faber D (1985) Methane Oxidation in Sediment and Water Column Environments Isotope Evidence. *Advances in Organic Geochemistry* 10: 759 768.

ENVIRONMENTAL GEOCHEMISTRY

W E Dubbin, The Natural History Museum, London, UK

Introduction

Soils and sediments occupying the Earth's surface lie at the interface of the lithosphere, atmosphere, biosphere, and hydrosphere. Within the weathered, complex, and porous milieu of these Earth-surface materials, myriad biogeochemical processes govern the movement of both nutrients and pollutants from the lithosphere to biota, where they are incorporated into plant and animal tissues, or to groundwaters, where they may be transported great distances to streams, rivers, and oceans. This chapter describes the most notable pollutants, both organic and inorganic, and the dominant processes governing their mobility and bioavailability in terrestrial and aquatic environments.

Trace Elements

Trace elements are those that occur in the lithosphere at concentrations typically less than 1 g kg $^{-1}$ (Table 1). Among the trace elements are the micronutrients (e.g., Cu, Ni, Zn), which are essential for the growth and development of micro-organisms, plants, and animals, and also the metalloids, which have characteristics of both metals and non-metals (e.g., As, B) (*see* **Minerals:** Arsenates). Heavy metals are defined as those trace elements with densities $>5.0\,\mathrm{g\,cm^{-3}}$. Virtually all trace elements, even the micronutrients, exhibit toxicity to animals and plants when present at excessive concentrations. Radionuclides are a separate but important class of inorganic contaminant that may occur naturally (e.g., ^{222}Rn, ^{226}Ra, ^{238}U) or as a consequence of nuclear fission related to atomic weapons testing and nuclear power generation (e.g., ^{90}Sr, ^{137}Cs, ^{239}Pu).

Although trace elements occur naturally in all terrestrial environments, anthropogenic inputs may increase these concentrations considerably. The principal anthropogenic sources of trace elements are mining and smelting activities, fossil fuel combustion, chemical and electronics industries, as well as the addition of fertilisers and biosolids arising from agricultural operations. One notable example of severe trace metal pollution caused by smelting activities is found near Karabash, in the south Ural Mountains region of Russia, where for decades smelting operations have deposited metals on the surrounding landscape, destroying much of the vegetation and so contributing to widespread soil erosion (*see* **Environmental Geology**).

Toxic levels of trace elements (e.g., Cr, Ni) may also occur naturally, as in soils derived from serpentinitic rocks, which can lead to phytotoxicity and the consequent lack of vegetation over large areas of the landscape where the serpentine soils occur (*see* **Clay Minerals**). Alternatively, these elevated trace element concentrations may induce metal tolerance among certain plant species (e.g., *Thlaspi* spp.) as biological communities adapt to these metal rich environments.

Trace Element Bioavailability and Speciation

It is well established that total trace element content in soil or sediment is a poor indicator of toxicity. A more reliable measure of ecotoxicity is trace element bioavailability. A trace element is considered bioavailable if it can be utilized by biota. Bioavailability is therefore broadly equated with solubility, although some plants and micro-organisms are able to extract metals from solid phases normally considered insoluble. The solubility, and hence bioavailability, of a particular trace element is determined largely by its solid-phase speciation and mode of surface complexation. The main parameters

Table 1 Abundance, speciation, and toxicity of trace elements in soil and sediment

Element	*Median soil content*[a] *(mg kg* ^{1}*)*	*Dominant solution species*[b]	*Function; Toxicity*
Ag	0.05	Ag^{+}, AgCl	None known; Plant and animal toxin
As	6	$As(OH)_3$, AsO_3^{3} , H_2AsO_4 , $HAsO_4^{2}$	None known; Plant and animal toxin
B	20	H_3BO_3, H_2BO_3	Plant nutrient; Phytotoxin
Be	0.3	Be^{2+}, $Be(OH)_3$, $Be(OH)_4^{2}$	None known; Plant and animal toxin
Bi	0.2	Bi^{3+}	None known; Plant and animal toxin
Cd	0.35	Cd^{2+}, $CdSO_4$, CdCl , $CdHCO_3^{+}$	None known; Animal toxin
Cl	100	Cl	Plant nutrient; Phytotoxin
Co	8	Co^{2+}, $CoSO_4$, $Co(OH)_2$	Plant and animal nutrient; Plant and animal toxin
Cr	70	$Cr(OH)_2^{+}$, $Cr(OH)_4$, $HCrO_4$	Animal nutrient; Cr(VI) is a plant and animal toxin
Cs	4	Cs^{+}	None known; None known
Cu	30	Cu_2^{+}, CuCl , $CuCO_3$, $CuHCO_3^{+}$, $Cu(OH)^{+}$	Plant and animal nutrient; Plant and animal toxin
Hg	0.06	Hg^{2+}, $HgCl_2$, CH_3Hg^{+}, $Hg(OH)_2$	None known; Plant and animal toxin
Mo	1.2	H_2MoO_4, $HMoO_4$, MoO_4^{2}	Plant and animal nutrient; Plant and animal toxin
Ni	50	Ni^{2+}, $NiSO_4$, $NiHCO_3^{+}$, $NiCO_3$	Plant and animal nutrient; Plant and animal toxin
Pb	19	Pb^{2+}, $PbSO_4$, $PbHCO_3^{+}$, $PbCO_3$, $PbOH^{+}$	None known; Plant and animal toxin
Rb	67	Rb^{+}	None known; Phytotoxin
Sb	1	$Sb(OH)_2^{+}$, $Sb(OH)_3$, $Sb(OH)_4$, $Sb(OH)_6$	None known; Plant and animal toxin
Se	0.4	$HSeO_3$, SeO_3^{2} , SeO_4^{2}	Animal nutrient; Plant and animal toxin
Sn	4	Sn^{4+}	None known; Plant and animal toxin
Tl	0.2	Tl^{+}	None known; Plant and animal toxin
U	2.7	$UO_{2(s)}$, UO_2^{2+}, UO_2CO_3, $UO_2(CO_3)_3^{4}$, $(UO_2)_3(OH)_7$	None known; Animal toxin
V	90	VO^{2+}, VO_2^{+}, $VO_2(OH)_2$, $VO_3(OH)^{2}$	Plant and animal nutrient; Plant and animal toxin
W	1.5	WO_4^{2}	None known; None known
Zn	70	Zn^{2+}, $ZnSO_4$, $ZnHCO_3^{+}$, $ZnCO_3$, $Zn(OH)^{+}$	Plant and animal nutrient; Plant and animal toxin

[a] Bowen (1979).

[b] Hayes and Traina (1998).

governing trace element speciation are pH and p*e*, the so-called 'master variables', as well as the total concentration of the ion in solution. These variables determine whether the ion in question will undergo hydrolysis, precipitation, redox reactions, or any number of complexation reactions at surfaces or in aqueous solution. Generally, metal cations are most soluble and available at low pH, principally pH <5, where hydrolysis is minimal and sorption to layer silicates and metal oxides is limited. Conversely, oxyanions such as chromate (*see* **Minerals:** Chromates) ($HCrO_4$) and selenite ($HSeO_3$) show greatest solubility at high pH, where electrostatic repulsion with negatively charged colloids minimizes surface complexation.

Redox-sensitive elements such as As, Cr, Fe, and Mn are subject to oxidation state changes under Earth surface conditions, with important implications for solubility and toxicity. Of the two As oxidation states predominating under Earth surface conditions (i.e., As(III) and As(V)), As(III) is more problematic because of its greater mobility and bioavailability. Given the ubiquity of manganese oxides (e.g., δ-MnO_2) in soils and sediments, these minerals have been proposed as key agents in the natural attenuation of As(III) contamination, as the oxidation of As(III) to As(V) by Mn(IV) is both rapid and thermodynamically favourable. The human health implications of As bioavailability is illustrated most clearly on the deltaic plains of Bangladesh and West Bengal, where groundwater As concentrations may exceed 400 μg l ^{1} as a consequence of biologically mediated reductive dissolution of Fe(III) oxides, leading to the release to solution, in bioavailable form, of the previously sorbed As. Manganese(IV) oxides can also mediate the oxidation of Cr, from Cr(III) to Cr(VI). However, this redox reaction is undesirable because, unlike As, the oxidized form of Cr, Cr(VI), is the form of Cr most mobile and toxic.

Organic Contaminants

The environmental geochemistry of organic contaminants primarily concerns the sources, movement, and fate of petroleum hydrocarbons and their byproducts, as well as the halogenated hydrocarbons, the group to which many pesticides belong. Petroleum hydrocarbons and associated compounds such as the oxygenates (e.g., methyl *t*-butyl ether (MTBE)) constitute a significant environmental risk by virtue of their widespread occurrence, mobility, and ecotoxicity (*see* **Geochemical Exploration**). Following release of hydrocarbons to the environment by multiple pathways, the low molecular weight volatile fraction ($<C_{15}$), often containing the carcinogenic benzene and polycyclic aromatic hydrocarbons, is largely lost to the atmosphere through volatilisation, leaving a relatively small but environmentally important portion of the light hydrocarbon pool to react with soil or sediment, or to enter groundwater. The less soluble, more chemically inert higher molecular weight hydrocarbons ($>C_{14}$) are potentially more disruptive to ecosystems, as illustrated by the spill in 1989 of crude oil from the *Exxon Valdez* oil tanker, which contaminated nearly 1750 km of Alaskan shoreline.

Halogenated hydrocarbons have both natural and anthropogenic origins, and many belong to the class of contaminant known as persistent organic pollutants (POPs), recalcitrant organic compounds that bioaccumulate and exhibit animal toxicity. POPs are dominated by the chlorinated hydrocarbons, which include many of the pesticides, such as aldrin, atrazine, chlordane, DDT, heptachlor, and the polychlorinated biphenyls. Despite the relative recalcitrance of POPs, their degradation can be mediated both by abiotic processes (e.g., oxidation by δ-MnO_2) as well as by the native soil microflora, particularly the fungi, which employ hydrolytic, reductive, or oxidative reactions to induce molecular dehalogenation. The hydroxylated compound so produced is thus rendered more susceptible to the degradation reactions of other soil microbiota.

The tendency of certain POPs to migrate from tropical and temperate climates to the colder polar regions has been the subject of study and debate for decades. A model of redistribution described as ‘global distillation’, involving POP evaporation followed by transport and condensation in colder regions, has received wide acceptance. Fractionation of POPs, during redistribution to higher latitudes, is driven by differential migration rates arising from variable POP vapour pressures and partition coefficients, with POP transport occurring in distinct jumps which are closely coupled to diurnal and seasonal temperature cycles.

Acidification of Terrestrial and Aquatic Environments

An important aspect of environmental geochemistry is acid deposition and the related acidification of Earth surface environments through both anthropogenic and natural processes. The burning of fossil fuels releases SO_2 and NO_x compounds which combine with atmospheric water to yield H_2SO_4 and HNO_3 that may be carried great distances before deposition as rain, mist, fog, or snow. Deposition of these acidic materials impacts negatively on soils, vegetation, and water bodies, particularly lakes which are poorly buffered and whose aquatic organisms are therefore at risk of increased soluble Al concentrations following a significant decrease in lake pH. Monuments and buildings constructed of limestone and marble, and which frequently represent much of our cultural heritage, are also at risk from acid deposition through dissolution of their constituent carbonate minerals (*see* **Minerals:** Carbonates). The effects of acid deposition are not entirely negative, however, as the additions of N and S to soils are beneficial, and these added nutrients frequently comprise a significant portion of the available soil N and S in highly industrialized regions.

An aspect of acidification of growing environmental importance concerns the oxidation of mining waste rich in reduced S, principally in the form of pyrite (FeS_2) (*see* **Environmental Geology**, **Minerals:** Sulphides). Oxidation of this S can yield vast amounts of H_2SO_4, giving rise to highly acidic waters, known as acid mine drainage (AMD) waters, containing toxic levels of soluble metals (Figure 1). The environmental significance of these acidic, metal-rich waters was emphasised in dramatic fashion with the collapse in 1998 of the Aznalcóllar mine tailings dam in southwest Spain. Failure of the dam led to the release of 1.3 million cubic metres of AMD waters, laden with Ag, As, Cd, Cu, Pb, and Zn, and the subsequent flooding of nearly 4,600 hectares of land with this toxic effluent.

Environmental Restoration

Decontamination of terrestrial environments is often costly and time-consuming, owing to the frequent occurrence of multiple pollutants, as well as the complexity of the contaminated matrices (e.g., soil or sediment). Strategies employed to remediate contaminated environments involve both *in situ* and *ex situ* techniques (Table 2), with the former generally receiving wider acceptance because of greater efficacy and lower implementation costs. *In situ* remediation may simply involve introducing a liming material (e.g., CaO, $Ca(OH)_2$, $CaCO_3$) to raise pH and thus decrease metal bioavailability through sorption and precipitation. The addition of smectites or zeolites (*see* **Minerals:** Zeolites) serves to remove pollutants by means of ion exchange reactions, whereas apatite addition induces the formation of sparingly soluble metal phosphate precipitates.

Figure 1 Acid mine drainage waters near the Kristineberg Cu and Zn mine, Sweden (photo courtesy L. Lovgren, Umea University).

Phytoremediation encompasses a variety of *in situ* strategies involving the use of plants to remove or render environmental pollutants harmless. The most common of these strategies is phyto-extraction, which involves growing plants capable of hyperaccumulating the contaminant of interest, followed by plant harvest to remove both plant and contaminant from the site. Among the hundreds of hyperaccumulators recently identified are the brake fern (*Pteris vittata*), an effective As accumulator, and also the basket willow (*Salix viminalis*), which effectively sequesters both Zn and Cd. Biodegradation is a related *in situ* remediation strategy that utilises native microbial populations to degrade multiple organic contaminants. This microbial degradation is optimal within aerobic rhizosphere communities at circumneutral pH, particularly in the presence of abundant nutrients, which may be augmented with external sources to enhance microbial activity.

The more costly *ex situ* techniques are generally applied only to the most high-value contaminated sites. The simplest of the *ex situ* remediation strategies involves direct excavation of the contaminated material followed by landfill disposal. Soil washing (*see* **Soils:** Modern) is more complex, involving excavation followed by mechanical screening to obtain coarse ($>50\,\mu m$) and fine ($<50\,\mu m$) fractions which are subsequently treated with a series of surfactants and chelates to remove both organic and inorganic contaminants, including radionuclides. The aim of soil washing is to concentrate the pollutants within the fine fraction, which is then treated with a solidification/stabilizing agent (e.g., Portland cement, lime, fly ash) prior to landfill disposal.

Table 2 Summary of *in situ* and *ex situ* environmental restoration techniques

Method	*Contaminant*	*Processes*	*Limitations*
In situ			
Lime addition	Metals, radionuclides	pH increase causing sorption, precipitation	Ineffective for oxyanions
Smectite, zeolite, apatite addition	Metals, radionuclides	Ion exchange, sorption, precipitation	Selective, short term remediation
Phytoremediation	Metals, organics	Phytoaccumulation, phytodegradation	Unsuitable for highly contaminated sites
Biodegradation	Organics	Microbial degradation	Long term remediation
Volatilization	Volatile organics	Evaporative loss of volatile pollutants	Contaminant must be volatile
Ex situ			
Excavation and disposal	All	Removal and disposal	Costly, risk of pollutant dispersal
Soil washing	Metals, organics, radionuclides	Excavation, leaching of contaminants	Costly
Solidification/stabilization	Metals, radionuclides	Addition of stabilising agent	Volatile organics not immobilized

See Also

Analytical Methods: Geochemical Analysis (Including X-Ray). **Clay Minerals**. **Environmental Geology**. **Geochemical Exploration**. **Minerals:** Arsenates; Carbonates; Chromates; Sulphides; Zeolites. **Soils:** Modern. **Weathering**.

Further Reading

Adriaens P, Gruden C, and McCormick ML (2004) Biogeochemistry of halogenated hydrocarbons. In: Lollar BS (ed.) *Environmental Geochemistry,* vol. 9, pp. 511 539. *Treatise on Geochemistry* Holland HD and Turekian KK (eds.). Oxford: Elsevier Pergamon.

Adriano DC (1986) *Trace Elements in the Terrestrial Environment.* New York: Springer Verlag.

Alexander M (1999) *Biodegradation and Bioremediation,* 2nd edn. London: Academic Press.

Alloway BJ (1995) *Heavy Metals in Soils,* 2nd edn. London: Blackie Academic and Professional.

Blowes RJ, Ptacek CJ, Jambor JL, and Weisener CG (2004) The geochemistry of acid mine drainage. In: Lollar BS (ed.) *Environmental Geochemistry,* vol. 9, pp. 149 204. *Treatise on Geochemistry* Holland HD and Turekian KK (eds.). Oxford: Elsevier Pergamon.

Bowen HJM (1979) *Environmental Chemistry of the Elements.* London: Academic Press.

Cozzarelli IM and Baehr AL (2004) Volatile fuel hydrocarbons and MTBE in the environment. In: Lollar BS (ed.) *Environmental Geochemistry,* vol. 9, pp. 435 474. *Treatise on Geochemistry* Holland HD and Turekian KK (eds.). Oxford: Elsevier Pergamon.

Hayes KF and Traina SJ (1998) Metal ion speciation and its significance in ecosystem health. In: Huang PM (ed.) *Soil Chemistry and Ecosystem Health,* pp. 45 84. Madison, USA: Soil Science Society of America.

Schwarzenbach RP, Gschwend PM, and Imboden DM (1993) *Environmental Organic Chemistry.* New York: Wiley.

Sposito G (1996) *The Environmental Chemistry of Aluminum,* 2nd edn. London: CRC Press.

ENVIRONMENTAL GEOLOGY

P Doyle, University College London, London, UK

Introduction

Environmental geology has grown in stature as a discipline over the past 40 years, as considerations of economic geology have moved away from the simple exploitation of the late nineteenth and early twentieth centuries. Today, environmental geology is a broad area of geological endeavour and a major industry in its own right.

Defining Environmental Geology

Environmental geology may be defined as the interaction of humans with their – fundamentally geological – environment. The environment can be considered to consist of both the constituents of the Earth itself (rocks, sediments, and fluids) and its surface and the processes that operate to change it through time.

Environmental geology is a subset of environmental science, which is the study of the interaction of humans with all aspects of their environment – physical, atmospheric, and biological – and is linked directly with engineering geology (*see* **Engineering Geology:** Overview). This definition clearly indicates that it is the introduction of the human element to the equation that defines the concept of the environmental sciences – and, therefore, environmental geology – and it is a consideration of both the debits (impacts) and credits (benefits) of our existence. Environmental science is a way of managing our existence so as to maximize human success while minimizing the negative aspects. At the heart of environmental geology, as with all environmental sciences, is the concept of sustainable management – working with natural systems to sustain development but not at an unacceptable environmental cost.

Scope of Environmental Geology

It has been estimated that around 50% of the Earth's population live in urban centres, centres that cover just 1% of the Earth's surface. Given that environmental geology is the interaction of humans with their environment and given that the majority of humans live in cities, it follows that environmental geology can be considered as primarily an urban issue, with the most challenging problems occurring within the immediate hinterland of urban centres.

Environmental geology has been defined as an urban concept, a 'machine' that balances inputs, outputs and maintenance, most if not all of which have a geological component (**Figure 1**). The inputs include water (derived locally or from more distant locations, but feeding into the city), raw materials (in the form of

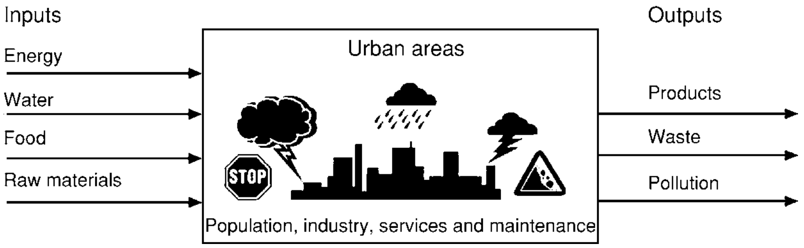


Figure 1 The urban machine. (Reproduced with permission from Bennett MR and Doyle P (1997) *Environmental Geology: Geology and the Human Environment.* Chichester: John Wiley.)

mineral resources for industry and construction), food resources (both locally produced and transported), and energy (the result in many cases of mineral resources such as coal, gas, and uranium). Outputs from the 'machine' include products from industry, wastes (in the form of worn-out materials, by-products of industry, and day-to-day wastes from domestic, industrial, and commercial sources), and pollution from poor waste-management strategies. The urban machine needs constant maintenance to replace its infrastructure and foundations, and to protect it from natural hazards.

The demands of the machine drive our demands on the environment, with the need for exploration and exploitation of mineral, water, and soil resources to provide the inputs; the need to combat environmental stress and natural hazards in protecting the workings of the machine, and the requirement to deal with the outputs, particularly wastes and their potential to pollute if not handled with appropriate sensitivity. All these aspects (and more) fall within the broad scope of environmental geology, and they can be distilled into the following themes.

1. Geology of resource management, including exploration and exploitation of resources (e.g. fuels, industrial minerals, and water) and the mitigation and limitation of associated adverse environmental impacts.
2. Geology of the built environment, particularly the constraints that ground conditions place on development.
3. Geology of waste management, specifically the disposal of wastes in the physical environment.
4. Geology of natural hazards.

Geology of Resource Management

Resources may be defined in the geological context as naturally occurring solids, liquids, or gases that are known or thought to exist in or on the Earth's crust in concentrations that make extraction economically feasible either at present or at some time in the future. This definition includes those geological materials that can be extracted with currently available technology (reserves) as well as those that are thought to exist but that will require further technological development to remove them.

In the short term, resource-management options focus on reserves, and environmental considerations have more power than at any other time in the past to reduce available reserves by altering the economics of their exploitation. Long-term planning must, of course, be based on the probability of new resources being discovered. The size of the available reserves relative to the total resources can be altered by several factors, such as commodity price, exploration, increasing the extent of the known resource, technological developments, and changes in regulation.

Economic Mineral Resources

Economic mineral resources are varied and can be defined in their broadest sense as any geological material that is of commercial value to society. This broad definition includes such diverse materials as fuels (e.g. coal, gas, uranium), construction materials, industrial minerals, metals, and precious minerals. Extraction of these materials varies from bulk extraction to specialist mining, and, although the basic concepts of exploration and exploitation may have many parallels, the environmental impacts posed by each one may well differ strongly.

Environmental Impacts of Mineral Extraction

The environmental impacts of mineral extraction vary according to the type of mineral and the extent of its deposit, with impacts varying throughout the working life of a mine or quarry, and with the issues often continuing long after the deposits worked are no longer economically viable. Typical issues are aspects of mine operation, mining subsidence, tackling mine wastes, and quarry or mine restoration.

There are a wide range of issues associated with mining. Quarries are troubled by blast noise and vibration, which can lead to increased rock-fall and landslip activity. In deep mines, the emission of methane and other gases is a problem, while wastewaters may be charged with iron and may be acidified by the breakdown of pyrite and other sulphides. Acid mine drainage is a growing problem in many disused metal mines and has adverse affects on groundwater supplies and water bodies, such as estuaries, rivers, and lakes.

Mining subsidence is a result of the removal of geological materials underground, creating a void space, which subsequently collapses (Figure 2). Similar situations occur when voids are created for purposes other than mineral extraction. Longwall mining is associated with deep coal extraction and generally involves the planned subsidence of a relatively large area. Pillar-and-stall mining, which dates back centuries, can lead to the differential collapse of the pillars, giving rise to graben-like failure structures. In other cases, removal of deep salt deposits through the process known as brining – the pumping out of brine-charged waters – can lead to severe subsidence. Large-scale subsidence is often difficult to manage, and the possibility of constructing buildings over an unknown area of mining activity is a very live issue, even where the mining is known to date back to antiquity, as in Rome. Raft foundations, grouting, deep piling, and excavation and backfill are all possibilities, depending on the depth and extent of the problem.

Mine wastes are produced at two stages during the mineral-extraction process: during mining, when waste rock or spoil is produced; and during further processing of the materials extracted, which creates a further set of mineral wastes, generally much finer than simple spoil and referred to as tailings. Spoil is of variable grain size and generally accumulates in open tips close to the mine workings. Tailings produced by mineral processing tend to be much richer in minerals and more uniform and of a finer grain size owing to milling. In some cases these materials have a high toxicity, as hazardous chemicals are used in the separation of mineral particles. Tailings are also more likely to be distributed by weathering and erosion, raising the consequent possibility of hazardous air-borne particulate matter. The long-term stability of mine wastes is also of concern, as with the South Welsh Aberfan disaster in 1966, where 111 000 m^3 of debris moved down a 13° slope, enveloping a school with tragic loss of life.

Dealing with waste tips is a matter of concern for environmental geologists, and the solution depends very much on the nature of the material. In the UK, tax on aggregate extraction led to the reuse of some tax-exempt spoil, particularly slate and shale, as poor-quality aggregate. Other mineral spoil may be reused for rock fill or inert landfill or even in dealing with mine subsidence and quarry-restoration projects. Tailings, on the other hand, given the potentially high level of toxicity, have to be treated with caution. Wet tailings have to be ponded, treated, and removed, while dry tailings have to be carefully monitored so that water passing through the wastes does not lead to

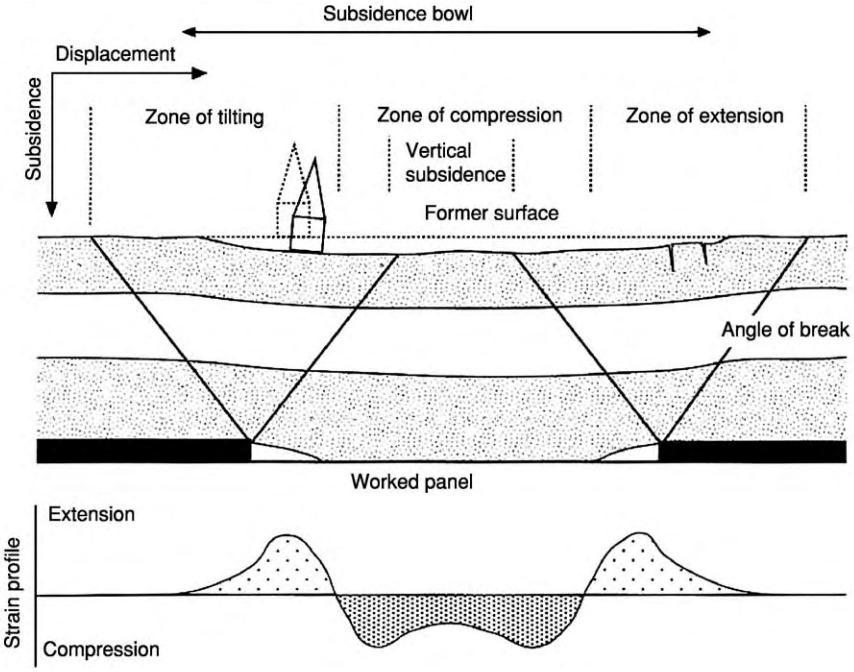


Figure 2 Subsidence associated with longwall coal mining.

groundwater pollution. Restoration and stabilization of wastes is essential, but this can be a tall order when vegetation that could be useful in binding surfaces is inhibited by the toxicity of the waste material itself.

Quarry and mine restoration is a major task. For mines, it is important first of all to ensure the safety of the underground workings themselves and to monitor such issues as the build-up of gases and acid mine drainage. In other cases, the mine workings may present a hazard in the form of future subsidence. Pit-head machinery and processing works, which may have used chemical processes requiring specialist attention, will also be an issue. Finally, the mine wastes themselves must be dealt with. For quarries, restoration issues reflect the nature of the materials quarried, with production-blasted hard-rock quarry faces representing an unsafe and unstable option. Restoration depends very much on the projected end use of the quarry; if left as a void, then faces will require engineering attention. This is particularly so given that many quarries are now being used to house industry or commercial concerns. In other cases, the void will be filled, most often as landfill.

Water Resources

Water is a vital geological resource that sustains life, yet most of the Earth's water is unavailable, with 98% being in the oceans and 1.6% in ice-sheets, leaving only 0.4% for drinking. Suitable drinking-water resources are distributed unevenly across the globe, and this, together with the predictions of environmental change, means that the search for drinking-water supplies is a major preoccupation of environmental geologists. Water-resource management is essentially a combination of three factors: acquisition, redistribution, and the treatment and disposal of wastewaters. The potential to pollute existing supplies through inadequate waste-disposal strategies is also a major issue, and this will be dealt with later when considering waste management.

Acquisition of water is primarily a process of extraction from groundwaters, which are controlled by rainfall and the porosity and permeability of the rocks present in a given area (Figure 3). These can be highly variable, and, even within a small area such as the UK, the percentage of water derived from this source is variable. Where groundwater is not an option, reliance on surface waters – rivers, lakes, and reservoirs – becomes much more significant, with desalination of seawater also possible.

The redistribution of supplies is an enormous undertaking and has lead to major international disputes when some countries have sought to dam rivers to the detriment of others downstream. Such international tensions are unlikely to recede in the future and could lead to 'water wars' if diplomacy breaks down. The creation of dams to contain waters and generate hydroelectric power has been an attractive but often highly controversial approach in some water-poor countries. However, inadequate site assessment has led to the failure of some dam sites, with leakage of waters or failure of the dam structure itself, so the high-cost deployment of such schemes requires a complete understanding and assessment of the issues involved.

Soil Resources

Together with water, soils are a fundamental geological resource, allowing us to grow sufficient crops to supply the food required to sustain the planet's growing population. The distribution and formation of soils across

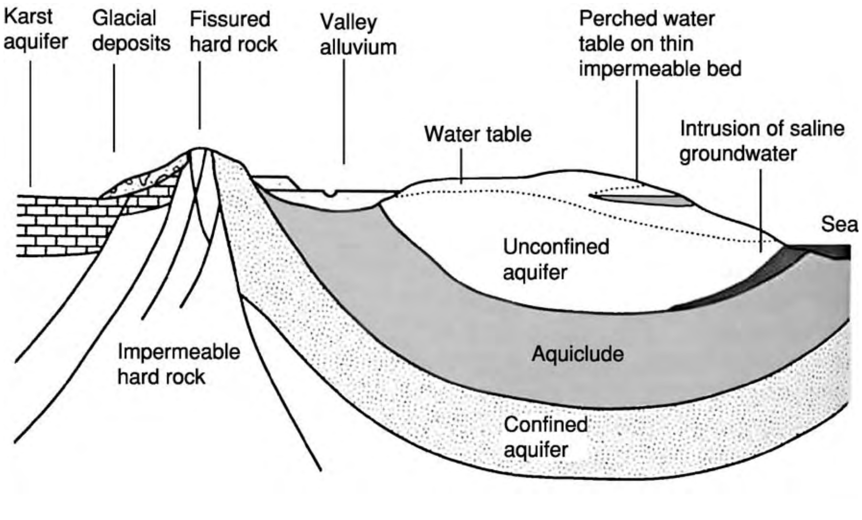


Figure 3 Schematic diagram of the main water bearing units. (Reproduced with permission from Bennett MR and Doyle P (1997) *Environmental Geology: Geology and the Human Environment.* Chichester: John Wiley.)

the world is contingent on climate, rainfall, biological activity, topography, and the underlying geology. Soils vary considerably, with an infinite variety – there are at least 19 000 different types in the USA alone – and widely differing classification schemes across the world (*see* **Soils:** Modern). The potential to erode and denude soils depends on the intensity of use, the ground cover, and so on. Soil conservation is therefore a major issue and can vary with land-use changes, new technologies, and colonization.

In the UK, raising the awareness of soils and the need to conserve them has led to the construction of five basic principles: that soils are an essential part of life support; that soils should enjoy the same level of protection as water and air; that integrated environmental management should include soils; that contaminated soils should be remediated; and that contamination should be avoided. These issues obviously have a global relevance.

Aesthetic and Scientific Resources

Conservation of aesthetic and scientific resources involves the conservation of areas of landscape and/or specific geological sites – known as geodiversity – for future generations (*see* **Geological Conservation**). The concept is based on four basic convictions: that they should be conserved for their own sake; that they form the basis for exploitation; that they form the basis for research and training; and that they have aesthetic and/or cultural value. This approach is growing in stature across the world, leading to a wider awareness of the value of geological features.

Geology of the Built Environment

The built environment – the towns and cities in which the vast majority of people live – is the central pillar of the urban machine, and, in many cases, the cause of many environmental issues. Built from geological resources – geomaterials – towns and cities are in constant need of renovation, renewal, and development, with the consequent need for greater exploration for aggregates, stone, and other materials (Figure 4). Engineering geology is an important factor in the construction of the built environment, particularly in the provision of sound foundations. Many of the older cities of the world are founded on ancient settlement sites, and this in turn creates its own issues of long-term stability.

Geomaterials

Construction materials, known as geomaterials, are quarried and mined from a wide variety of settings. The pressure to discover new resources is increasing, as the pressure to house the world's burgeoning population also increases. Stone and clay for adobe bricks have been the main building resources of the world since antiquity, and concrete has a long and venerable history. These materials are still in demand, and quarrying for them creates much the same environmental issues as any other type of extractive industry (*see* **Aggregates**).

Another important role for environmental geologists is in assessing the appropriateness of geomaterials for the job for which they are intended. This might include, for example: a durable, structurally strong, easily worked, and attractive stone for construction; large relatively uniform blocks of stone that can take a high degree of pounding without breaking down for use as armourstone blocks in coastal defences; or the provision of aggregates for roads that do not polish easily and can take a fair amount of crushing from large vehicles. Durability of stone and the effects of stone cleaning are major issues in some of the major cities of the world, needing constant monitoring (*see* **Building Stone**).

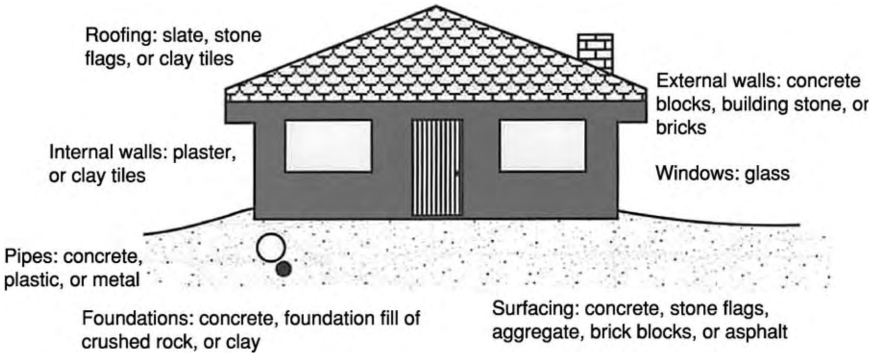


Figure 4 The main geomaterials and their uses. (Reproduced with permission from Bennett MR and Doyle P (1997) *Environmental Geology: Geology and the Human Environment.* Chichester: John Wiley.)

Engineering Geology

Major construction works depend upon sound site investigation, which documents the underlying geological setting and the strengths and behaviours of the rocks and engineering sols and recognizes hazardous ground conditions. In the UK alone, over 30% of all major civil engineering projects are delayed owing to poor ground conditions. It is essential to carry out thorough site investigations, using a range of geological techniques, including mapping and geophysical surveys, as well as committing significant time to material testing. Consideration of climate is a major issue, as in Hong Kong, where the failure of tower blocks has been attributed to the failure of thick tropical soils overlying crystalline basement, coupled with high rainfall (*see* **Engineering Geology:** Overview).

Geology of Waste Management

The urban machine creates an immense amount of waste, which may be defined as any substance that constitutes a scrap material, effluent, or unwanted surplus or any substance or article that requires disposal because it is worn-out or contaminated in some way. Although there is increasing emphasis on waste minimization and recycling, it is accepted that there will always be waste and that a strategy for its disposal will be required.

Wastes can conveniently be divided on the basis of whether they are managed (i.e. there is a management strategy in place to deal with their disposal that is in line with current practice and legislation) or unmanaged (i.e. there is no management strategy). Strategies for dealing with wastes have been based largely on two concepts: dilute and disperse, or concentrate and contain (Figure 5). The failure of these strategies ultimately leads to pollution, which is a by-product of inadequate waste-management strategies.

The dilute-and-disperse strategy is now largely discredited, because it relies on the potential of a given environmental setting to dilute the waste materials and then to disperse them. This has been, and continues to be, an issue for the dispersal of effluents in the sea and of gases in the atmosphere, as these natural systems are becoming choked with inadequately dispersed wastes. Concentrate and contain is now seen as the most viable option, with wastes being concentrated, compacted, or shredded where possible, before being placed in a safe totally enclosed (or ‘sanitary’) facility. At one level this entails the adequate preparation of a landfill site; at another, the containment of radioactive materials in an appropriate repository in perpetuity.

Contaminated Land

Contaminated land is any area that has been contaminated by its past industrial use or by the disposal of wastes. Contamination is therefore an inevitable effect of industrialization, and of waste disposal, as it involves the introduction of materials not naturally present at a given location. Contaminated land is a big issue in developed nations, with the decline in heavy industry and the targeting of former sites of industry as valuable real estate in inner-city locations. Increasingly, there is a need for adequate assessment of a site – assessing the type and level of contamination – before developing a clean-up strategy.

Sanitary Landfill

Employing the principle of concentrate and contain, sanitary landfills (Figure 5) ensure that appropriate voids are chosen for landfilling, appropriate containment strategies are employed in developing the landfill (using liners, separating the void into cells divided by bunds, and capping the site securely), the void is filled in an economical and organized manner, and

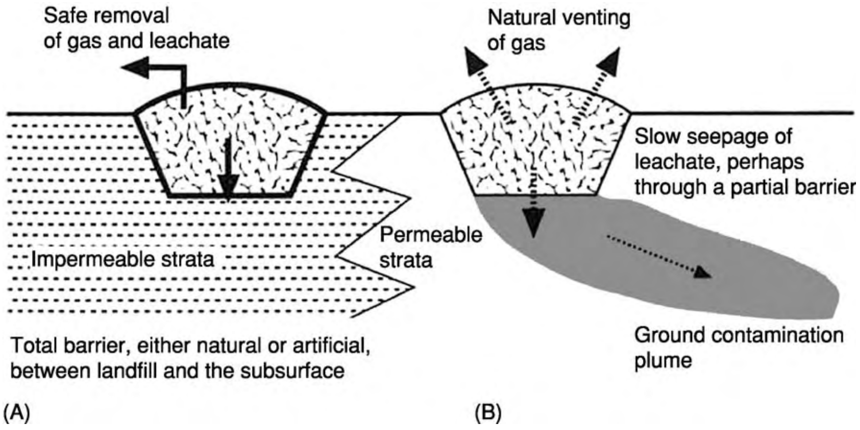


Figure 5 Landfill options: (A) total containment, and (B) slow dispersal. (Reproduced with permission from Bennett MR and Doyle P (1997) *Environmental Geology: Geology and the Human Environment.* Chichester: John Wiley.)

the by-products of the decay of putrescible wastes (i.e. landfill gas and leachates) are removed safely. The decomposition of wastes by bacteria takes place in three phases: first, aerobic bacteria break down the large organic molecules; second, anaerobic bacteria create simple compounds, such as hydrogen, ammonia, carbon dioxide, and organic acids; and, finally, the methane-rich landfill gas is created. In practice, this means that in the early years of a landfill the amount of gas produced is relatively small, rising to a maximum in around 20 years. Uncontrolled migration of this gas can be lethal, as in some cases the results have been explosive, with gases migrating through suitable geological pathways, as happened in 1986 at Loscoe in northern England. Leachates are dealt with by ponding and removal, in most cases, and then treated.

Deep Repositories for Nuclear Wastes

The disposal of nuclear radioactive wastes is a difficult problem, given the 'geological' time-scales of the half-lives of the isotopes of uranium used in the nuclear-power industry. One option is to concentrate wastes in a secure surface repository, which can be monitored to ensure that the multiple walls of the containment are secure. This obviously requires a long-term commitment. Over the last few decades, other options have been considered, including using a uniform densely crystalline deep-crustal rock body as a natural repository, which could be sealed in perpetuity (Figure 6). Lithologies considered have included fine-grained tuffs, basalts, and evaporates, and detailed investigations in the UK and the USA of the feasibility of sites in Cumbria and Texas, respectively, have been carried out. However, technical difficulties and political issues have stalled these geological solutions.

Geology of Natural Hazards

Natural hazards occur on a regular basis, cause disruption and destruction, and annually lead to large losses of life under tragic circumstances (*see* **Engineering Geology:** Natural and Anthropogenic Geohazards). Geological hazards include high-magnitude low-frequency events (earthquakes, volcanoes, and tsunamis) and low-magnitude high-frequency events

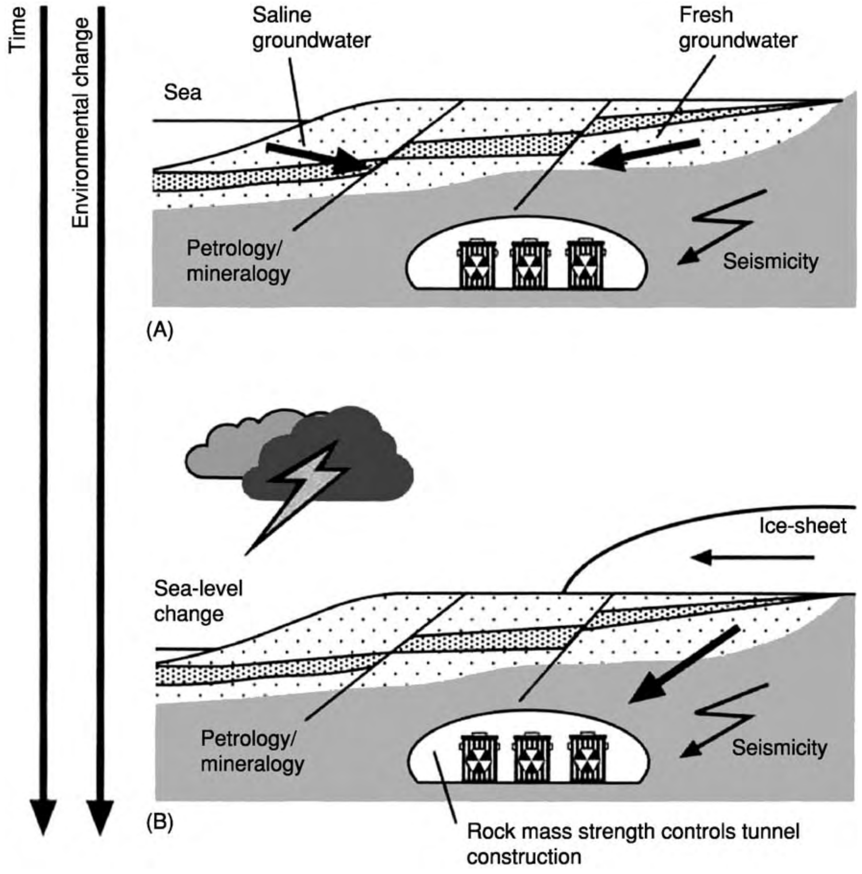


Figure 6 Deep repository option for nuclear wastes, illustrating (A) the current geology, and (B) the geology at some point in the future with the influx of fresh glacial meltwaters (arrow). (Reproduced with permission from Bennett MR and Doyle P (1997) *Environmental Geology: Geology and the Human Environment.* Chichester: John Wiley.)

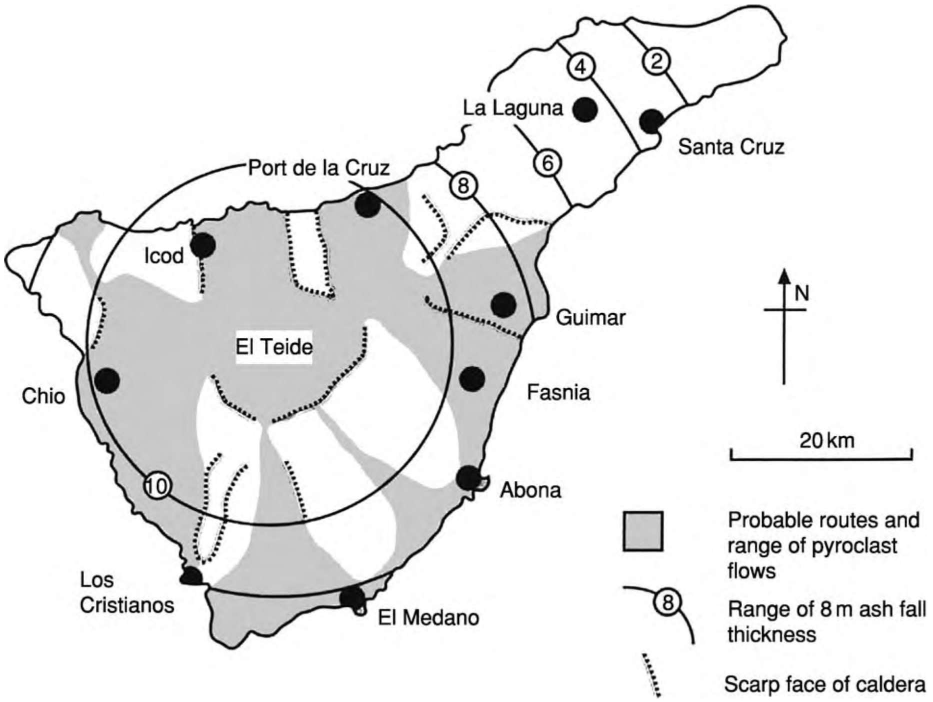


Figure 7 Volcanic hazard map of Tenerife, Canary Islands.

(such as soil, fluvial, and coastal erosion). Dealing with both types of event severely tests the local civil-defence infrastructures, and, with the frequency of disasters apparently on the increase – with attendant casualties and economic costs – planning for disaster management is a major preoccupation of local authorities.

There are three broad hazard types: atmospheric hazards, caused by tropical storms, hurricanes, and droughts; exogenic hazards, such as flooding, coastal erosion, and mass movement; and endogenic hazards, resulting from internal Earth processes, such as volcanoes and earthquakes. In many cases there can be considerable overlap between these events, as one can trigger another. In profiling natural hazards, it is important to gather data on the magnitude and frequency (including temporal spacing) of the events, the duration and speed of onset of a typical event, and the extent and dispersion of the effects. These issues have to be taken into account in civil-defence planning strategies, which will identify the risk to life, property, and the economy.

Risk assessment is an important task and will in many cases involve a predictive element, although prediction of earthquakes and volcanic eruptions is a tricky business, without a simple solution. Hazard mapping is a significant technique and had been used to good effect in identifying patterns of eruption in populated volcanic regions (Figure 7). This, together with a numerical assessment of the vulnerability of a given settlement, are important tools in the armoury of an environmental geologist.

See Also

Building Stone. **Clays, Economic Uses**. **Engineering Geology:** Overview; Natural and Anthropogenic Geohazards. **Geological Conservation**. **Geological Engineering**. **Quarrying**. **Urban Geology**.

Further Reading

Bailleul TA (1987) Disposal of high level nuclear waste in America. *Bulletin of the Association of Engineering Geologists* 24: 207 216.

Bell FG (1992) Salt mining and associated subsidence in mid Cheshire, England, and its influence on planning. *Bulletin of the Association of Engineering Geologists* 29: 371 386.

Bell FG (1993) *Engineering Geology*. Oxford: Blackwell.

Bell FG (1996) Dereliction: colliery spoil heaps and their rehabilitation. *Environmental and Engineering Geoscience* 2: 85 96.

Bell FG, Duane MJ, Bell AW, and Hytiris N (1996) Contaminated land: the British position and some case histories. *Environmental and Engineering Geoscience* 2: 355 368.

Bennett MR and Doyle P (1997) *Environmental Geology: Geology and the Human Environment*. Chichester: John Wiley.

Booth B (1979) Assessing volcanic risk. *Journal of the Geological Society of London* 136: 331 340.

Bryant EA (1991) *Natural Hazards*. Cambridge: Cambridge University Press.

Culshaw MG, Bell FG, Cripps JC, and O'Hara M (eds.) (1987) *Planning and Engineering Geology*. London: Geological Society.

Doyle P, Barton P, Rosenbaum MR, Vandewalle J, and Jacobs K (2002) Geoenvironmental implications of the underground war in Flanders, 1914 1918. *Environmental Geology* 43: 57 71.

Fookes PG and Poole AB (1981) Some preliminary considerations on the selection and durability of rock and concrete materials for breakwaters and coastal protection works. *Quarterly Journal of Engineering Geology* 14: 97 128.

Gray M (2004) *Geodiversity: valuing and conserving abiotic nature*. Chichester: John Wiley.

Gray RE and Bruhn RW (1984) Coal mine subsidence: eastern United States. *Geological Society of America Reviews in Engineering Geology* 6: 123 149.

Joseph JB (2004) Perception or reality waste, landfill and the environment. *Geology Today* 20: 107 112.

Lumsden GI (ed.) (1992) *Geology and the Environment in Western Europe*. Oxford: Oxford University Press.

McFeat Smith I, Workman DR, Burnett AD, and Chau EPY (1989) Geology of Hong Kong. *Bulletin of the Association of Engineering Geologists* 26: 23 107.

Prentice JE (1990) *Geology of Construction Materials*. London: Chapman & Hall.

Price M (1985) *Introducing Groundwater*. London: Allen & Unwin.

Smith K (1996) *Environmental Hazards: Assessing Risk and Reducing Disaster,* 2nd edn. London: Routledge.

Thomas RG (1989) Geology of Rome, Italy. *Bulletin of the Association of Engineering Geologists* 26: 415 476.

Waltham AC (1994) *Foundations of Engineering Geology*. Glasgow: Blackie.

Williams GM and Aitkenhead N (1991) Lessons from Loscoe: the uncontrolled migration of landfill gas. *Quarterly Journal of Engineering Geology* 24: 191 207.

Woodcock NH (1994) *Geology and Environment in Britain and Ireland*. London: UCL Press.

EROSION

See **SEDIMENTARY PROCESSES: Erosional Sedimentary Structures; Aeolian Processes; Fluxes and Budgets**

EUROPE

Contents

East European Craton
Timanides of Northern Russia
Caledonides of Britain and Ireland
Scandinavian Caledonides (with Greenland)
Variscan Orogeny
The Urals
Permian Basins
Permian to Recent Evolution
The Alps
Mediterranean Tectonics
Holocene

East European Craton

S V Bogdanova and R Gorbatschev, Lund University, Lund, Sweden
R G Garetsky, Institute of Geological Sciences, Minsk, Belarus

Introduction

The East European Craton is the coherent mass of Precambrian continental crust in the north-eastern half of Europe. Across a suture zone between the North Sea and the Black Sea, that crust meets the younger, thinner, warmer, and more mobile crust of western and southern Europe.

The overall area of the East European Craton is more than 6.7 million km^2, including the shelf seas.

Within the East European Craton, Precambrian crystalline crust is exposed in the Baltic (also Fennoscandian) and Ukrainian Shields as well as in minor areas in Belarus and the Voronezh Massif of south-western Russia. Elsewhere, the craton is covered by the Late Proterozoic and Phanerozoic sedimentary deposits of the Russian Platform (Figure 1).

The present state of the East European Craton is largely controlled by structures dating back to the time of its formation by the successive collision of three large, once independent, crustal segments – Fennoscandia, Sarmatia, and Volgo-Uralia – in the Palaeoproterozoic, *ca*. 2.1–1.7 Ga ago. This view was first proposed in 1993. Until then, the East European Craton was regarded as a rather uniform region of numerous minor ‘blocks’ of Archaean crust set in a matrix of Proterozoic folded belts.

Margins and Borders

Most of the margins of the East European Craton are characterized by the presence of younger mobile belts (Figure 1). In the north-west, the Scandinavian Caledonides (*see* **Europe:** Scandinavian Caledonides (with Greenland)) form a *ca*. 1800 km long strongly eroded, but still up to 15-km thick, pile of large thrust sheets built up of Early Palaeozoic and Precambrian rocks. These nappes derive from source areas far to the present west and were thrust atop the crystalline basement around 450–400 Ma ago. Autochthonous basement fabrics are still perceptible throughout the Caledonides, while geophysical and palaeogeographical data indicate that the Pre-Caledonian margin of the East European Craton may be as much as 400 km off the Atlantic coast of Scandinavia.

Along the north-eastern margin of the East European Craton, the Timanide Belt extends between the Urals and northernmost Norway (*see* **Europe:** Timanides of Northern Russia). It features initial sedimentation along a passive continental margin, followed by compression, metamorphism, and igneous activity due to subduction and collision 650–600 Ma ago.

The eastern limit of the East European Craton is the Late Palaeozoic Uralide Orogen, which represents the zone where ancient Europe collided with Asian terranes 350–300 Ma ago (*see* **Europe:** The Urals). Farther south-west, the margin of the East

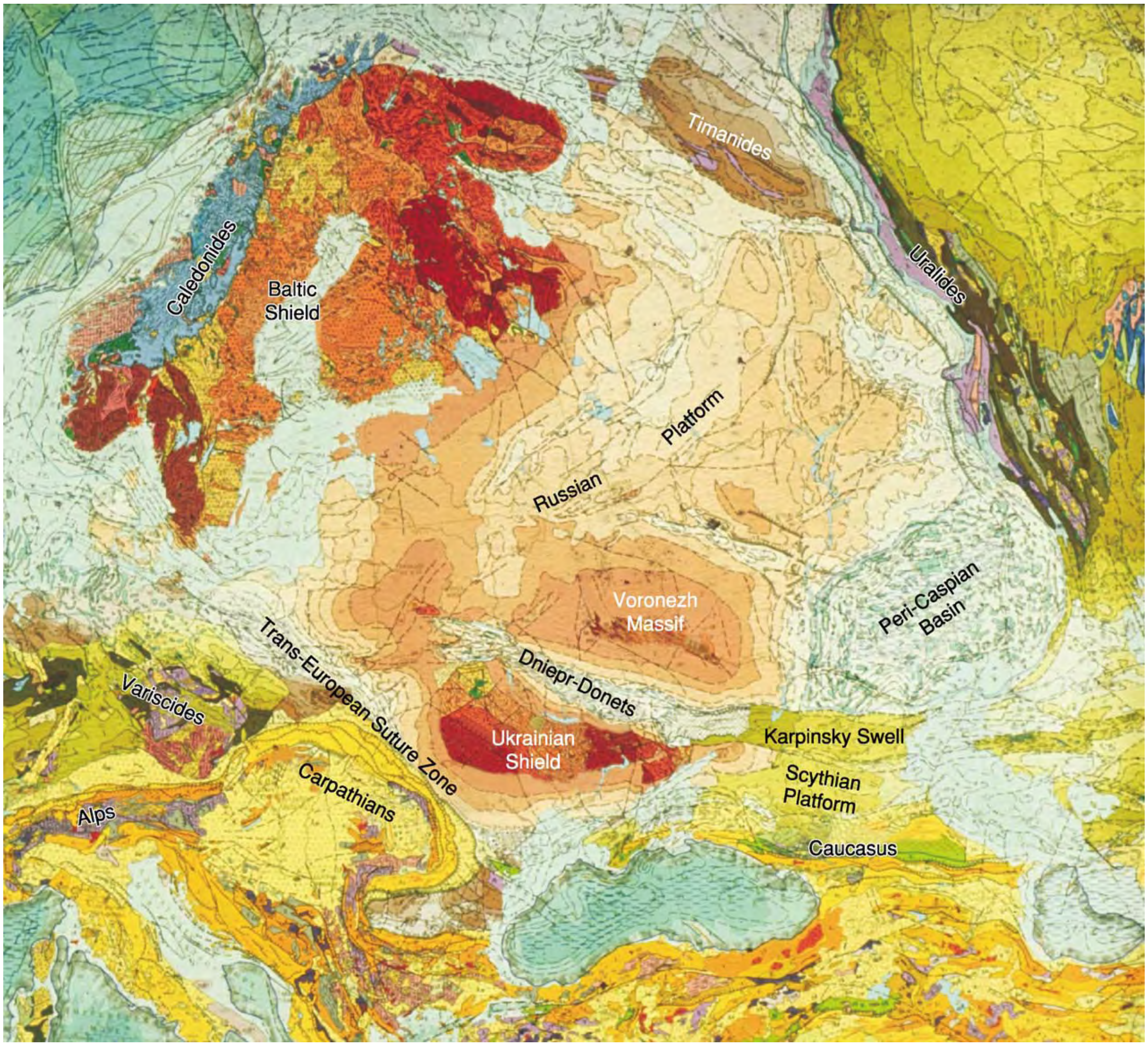


Figure 1 The East European Craton in its crustal setting between central Europe and western Siberia. The craton area is marked in reddish brown colours. The fainter the tint, the thicker the sedimentary cover. The areas of outcrop in the Baltic and Ukrainian Shields and the Voronezh Massif can be clearly seen. The cover is also thin in Belarus, but only in a minor uplift is the craton actually exposed. It can also be seen that the basement is more deeply buried in the eastern part of the Russian Platform than in the west. Between the eastern region and the basin beneath the southern Baltic Sea is the so called 'Scythian Rampart' of the early geologists, which connects the Baltic and Ukrainian Shields. The cover is thickest in the Peri Caspian Basin, where the mottled patterns denote salt dome tectonics. Among the large rifts and aulacogens characteristic of the East European Craton, the originally Proterozoic but subsequently rejuvenated central Russian and Pachelma systems can be clearly recognized. The latter is immediately north north east of the Voronezh Massif. The rift systems to the west of the Urals and in the south western foreland of the Timanides can also be seen. Prominent in the south is the Devonian Dniepr Donets Aulacogen and its inverted eastward continuation (the so called 'Karpinsky Swell') between the Scythian Platform and the Peri Caspian Basin and onwards into Asia. This figure is part of the 'Carte Tectonique de l'Europe et des Régions Avoicinantes,' scale 1 to 10 min. IUGS Commission for the Geological Map of the World, 1975 (by courtesy of the CGMW.)

European Craton was developed in the Neoproterozoic and Early Cambrian due to collision between the Craton and part of Gondwana. That margin was subsequently complicated by the uplifted southeastern continuation of the Devonian Dniepr–Donets Aulacogen (the Karpinsky Swell in Figure 1).

The southern edge of the East European Craton is outlined by the Alpine mountain belts of the Crimea and Caucasus. In that region, the Scythian Platform marks a slab of East European cratonic crust involved in the Alpine–Mediterranean orogenic process.

In the south-west, the Trans-European Suture Zone is the limit of the East European Craton. Across that boundary, the Craton abuts terranes successively formed during the Caledonian, Variscan, and Alpine–Mediterranean orogenies. Along the central course of the Trans-European Suture Zone, palaeontological data indicate that some terranes in Poland represent detached slices of the Craton. In northern Germany, a wedge of East European cratonic crust has been traced seismically far to the south-west of the Trans-European Suture Zone.

Crustal Thickness and Magnetic and Gravity Fields

The crust of the East European Craton is mostly around 35 km to somewhat more than 40 km thick. Moho depths of up to 50 km occur particularly in a wide central area near Moscow, in the south-west, and in the Ukrainian and Baltic Shields. Local maxima may even exceed 60 km. Under some of the Archaean parts of the East European Craton, cratonic roots in the lithospheric mantle reach down to 200–250 km.

The most pronounced Moho uplift, to a level of around 30 km, is associated with the northern part of the Peri-Caspian Basin, where a thick Proterozoic to Phanerozoic cover is also present. In consequence, the thickness of the Precambrian crystalline crust is only 10–15 km in that region.

At the borders of the East European Craton, Moho depths increase along the Uralides and the Caucasus collisional belts as well as in the central and southern parts of the Trans-European Suture Zone. Further north, continental crust thins markedly towards the North Atlantic Ocean.

In detail, steep gradients of Moho depth commonly follow one-time collisional and accretionary plate and terrane boundaries and, in general, the boundaries between Archaean and Proterozoic crust. However, in places, the original continental crust has been thickened by later mafic underplating or thinned by extension and magmatism.

The East European Craton differs from the neighbouring parts of Europe and western Siberia in featuring numerous belts of strong magnetic anomalies. These outline the boundaries of the different crustal units as well as later rifts and some major belts of granitic and high-grade metamorphic rocks (Figure 2). The patterns of the gravity field are similar, but strongly negative anomalies occur only along the Scandinavian Caledonides and just outside the limits of the craton along the Carpathian, Crimean, and Caucasian Alpine–Mediterranean mountains.

Morphology, Topography, and Sedimentary Cover

Morphologically, most of the East European Craton forms a vast low-lying plain. Some small-scale broken topography exists in the shield areas, particularly in the recently glaciated Baltic Shield. The largest flats, in contrast, extend towards the south-east, where the Peri-Caspian depression is largely below global sea-level (Figure 1). Pronouncedly mountainous areas occur only in western and northern Scandinavia, where the Caledonides and their Precambrian basement were uplifted and topographically rejuvenated during the opening of the North Atlantic.

The cover of the Russian Platform mostly ranges between some tens of metres and 2 km in thickness (Figures 1 and 3A). Several rifts and basins, however, contain sedimentary piles 3–5 km thick, while the Dniepr–Donets Aulacogen in the Ukraine and south-western Russia has a fill exceeding 15 km in thickness (Figure 3B). A general depression of the basement–cover boundary occurs towards the marginal mountain belts of the Timanides, Uralides, and Caucasus, the greatest cover thickness of approximately 20 km being reached in the area to the north of the Caspian Sea.

The covered parts of the East European Craton comprise several large basins of sedimentation, e.g. the Moscow, Baltic, and Peri-Caspian basins. These were largely formed in response to recurrent cycles of rifting, subsidence, and compression. Maxima of basin formation and filling occurred during the Riphean (Meso- to Neoproterozoic), the Early Vendian (terminal Neoproterozoic), the Late Cambian-Ordovician, the Middle-to-Late Devonian, the Carboniferous–Permian transition, and the Triassic (Figures 3–5). Generally, these maxima were related to orogenies, major plate-tectonic movements, rotations of the East European Craton, and its interaction with other proto-continents such as Laurentia, Greenland, and Avalonia.

Large rifts and aulacogens are characteristic elements of the East European Craton (Figure 6). Many of the oldest, i.e. of Meso- to Neoproterozoic age, are associated with the system of Palaeoproterozoic sutures that arose as a result of the formation of the craton from several colliding crustal segments. To this group belong the Pachelma, Volyn-Orsha, and Central Russian rifts and aulacogens (Figure 6). Other rifts of that age follow age-province boundaries, while several Neoproterozoic troughs in the Peri-Urals region were developed on a Mesoproterozoic passive continental margin. Unlike the Precambrian rifts, many Phanerozoic rifts, for instance the Dniepr-Donets Aulacogen, cut sharply across the grain of the crust.

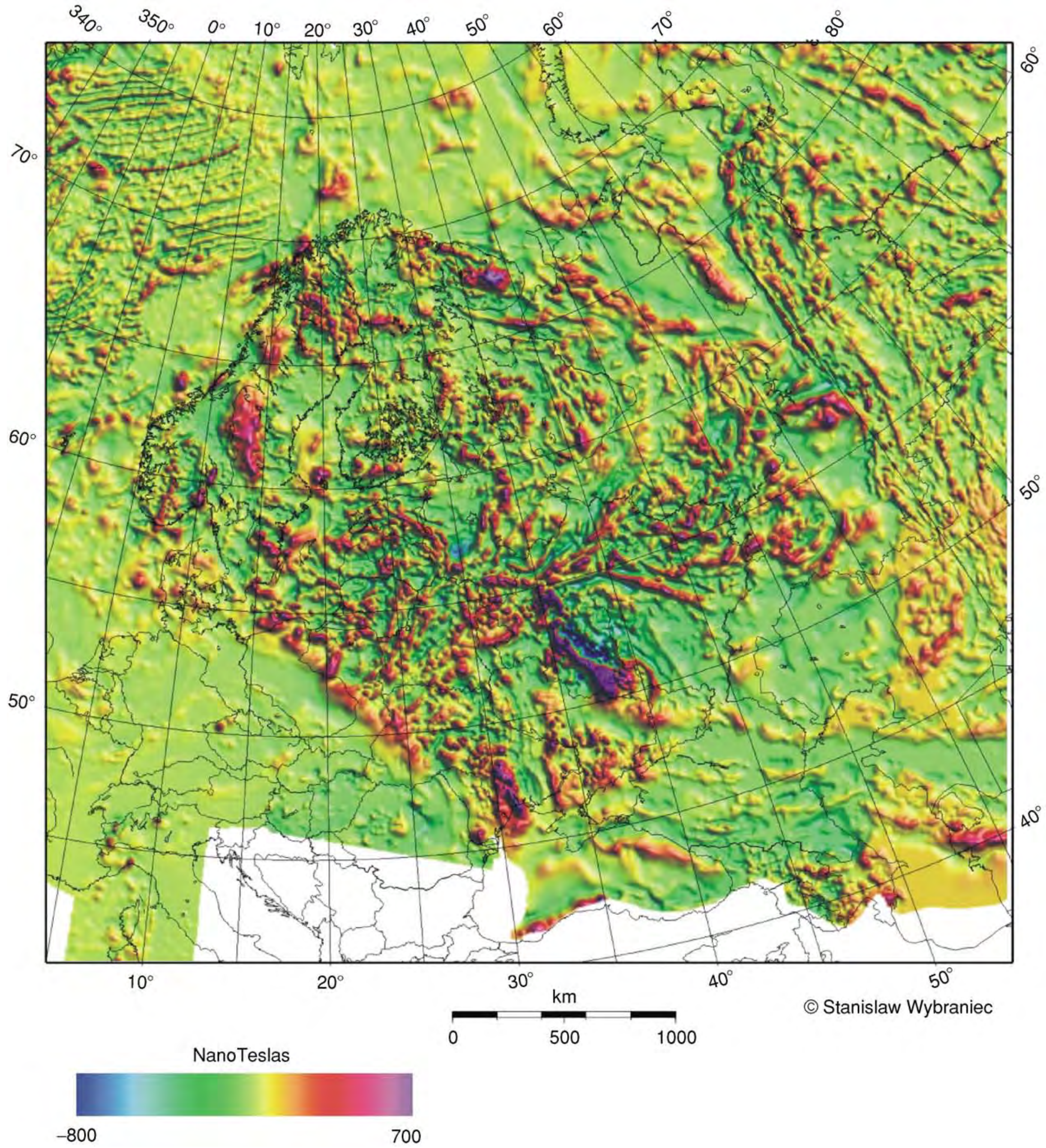


Figure 2 Magnetic anomaly patterns in the East European Craton. In general, these patterns describe very well the tripartite crustal segment structure of the East European Craton (cf. **Figure 8**). In addition, the strongest positive anomalies mark major occurrences of magnetic iron ores, for example the Kursk magnetic anomaly in the Voronezh Massif and the anomalies indicating the ores of northernmost Sweden (Kiruna, etc) and the Kriviy Rog ores in the Ukraine. Where sedimentary fillings are very thick in rifts and basins, ribbon shaped negative or alternating negative positive anomaly patterns prevail (e.g. in the Dniepr Donets and Pachelma Aulacogens and the Peri Caspian Basin). Moderately positive anomalies are associated with amphibolite to granulite facies metamorphic belts and some granite provinces, such as the rapakivi massifs and parts of the Trans Scandinavian Igneous Belt. However, other rapakivi plutons (Viborg, for instance) have negative anomalies. Similar diversity is also found for the Archaean protocratons. Map by courtesy of S Wybaniec of the Polish Geological Institute.

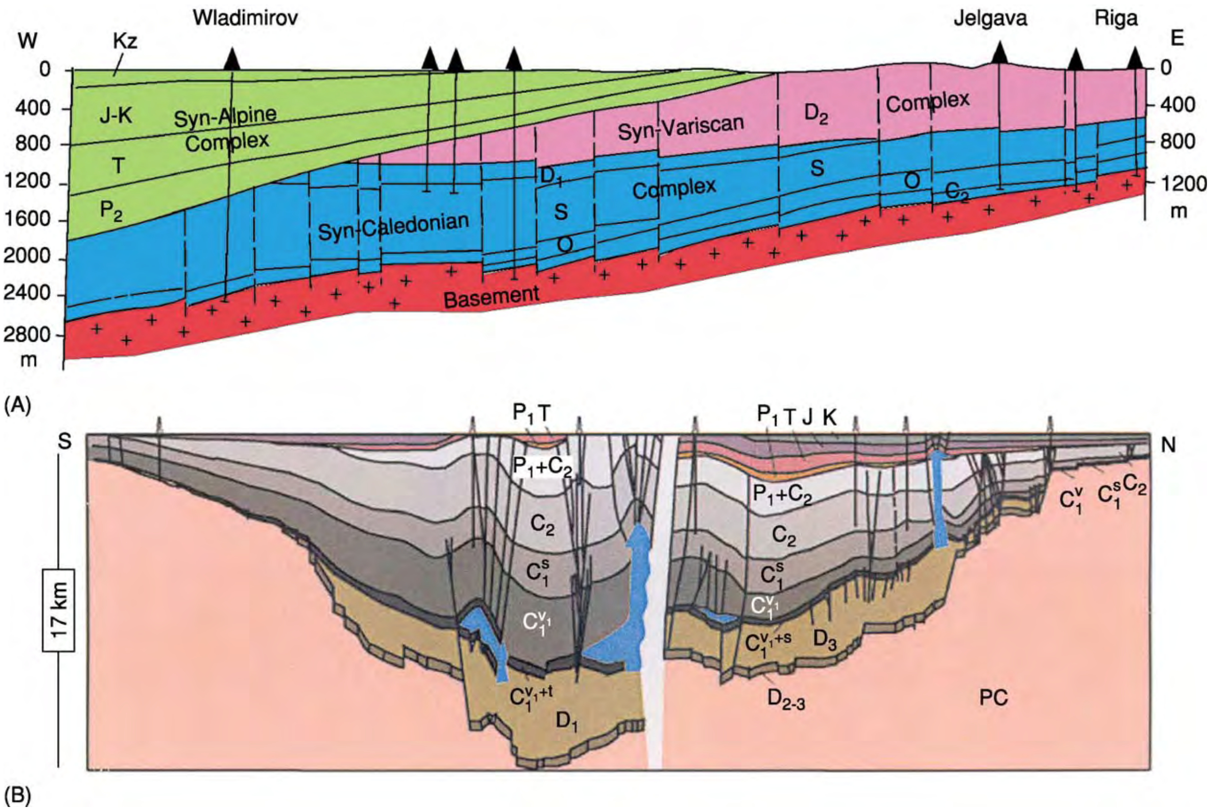


Figure 3 Sections across the Phanerozoic sedimentary cover of the Russian Platform and adjoining areas. (A) Profile of the eastern part of the Baltic Sea Depression, showing the disconformities between three different tectonostratigraphic complexes. These correspond roughly to the three Phanerozoic orogenies that shaped adjoining central Europe and changed the regimes of sedimentation and erosion within the East European Craton. C, Cambrian; O, Ordovician; S, Silurian; D, Devonian; P, Permian; T, Triassic; J, Jurassic; K, Cretaceous; Kz, Cenozoic. (B) General seismic section across the Dniepr–Donets Rift Basin. The ends of the section are approximately 170 km apart, but the middle portion of the profile has been omitted from the diagram. Salt diapirs are shown in blue. Pc, Precambrian; D, Devonian; C, Carboniferous; P, Permian; T, Triassic; J, Jurassic; K, Cretaceous. Seismic section by courtesy of the Secretariat of the EUROPROBE Programme.

Rifting in the East European Craton was commonly accompanied by within-plate mafic magmatism, which, during some periods, e.g. the Devonian, may have been associated with mantle-plume upwelling. Among the products are kimberlites (*see* **Igneous Rocks:** Kimberlite) found, for instance, in the Ukrainian Shield, Voronezh Massif, and Kola Peninsula region.

Crustal Segments of the East European Craton

The Meso- and Neoproterozoic Volyn-Orsha, central Russia, and Pachelma rifts and aulacogens divide the East European Craton into three different parts (Figures 6 and 8). Research during the last 15 years has demonstrated that these correspond roughly to three Precambrian crustal segments, each of which has its own Archaean–Palaeoproterozoic history. The three segments are Fennoscandia (including the Baltic-Fennoscandian Shield) in the north-west and north, Sarmatia (comprising the Ukrainian Shield and the Voronezh Massif) in the south, and Volgo-Uralia in the east.

Fennoscandia and Sarmatia both feature Archaean cores and accreted Palaeoproterozoic juvenile continental crust, whereas Volgo-Uralia is almost entirely Archaean but was strongly reworked in the Palaeoproterozoic.

Fennoscandia

The Fennoscandian crustal segment consists of two principal parts (Figure 9). In the north-east there is a large domain of Archaean crust, while the rest of the segment is made up of Palaeo- and Mesoproterozoic, mostly juvenile, mantle-derived rocks. These belong to several successively formed orogenic belts.

The demarcation between the Archaean and Proterozoic regions is fairly sharp, but the edge of the Archaean has been reworked tectonically and

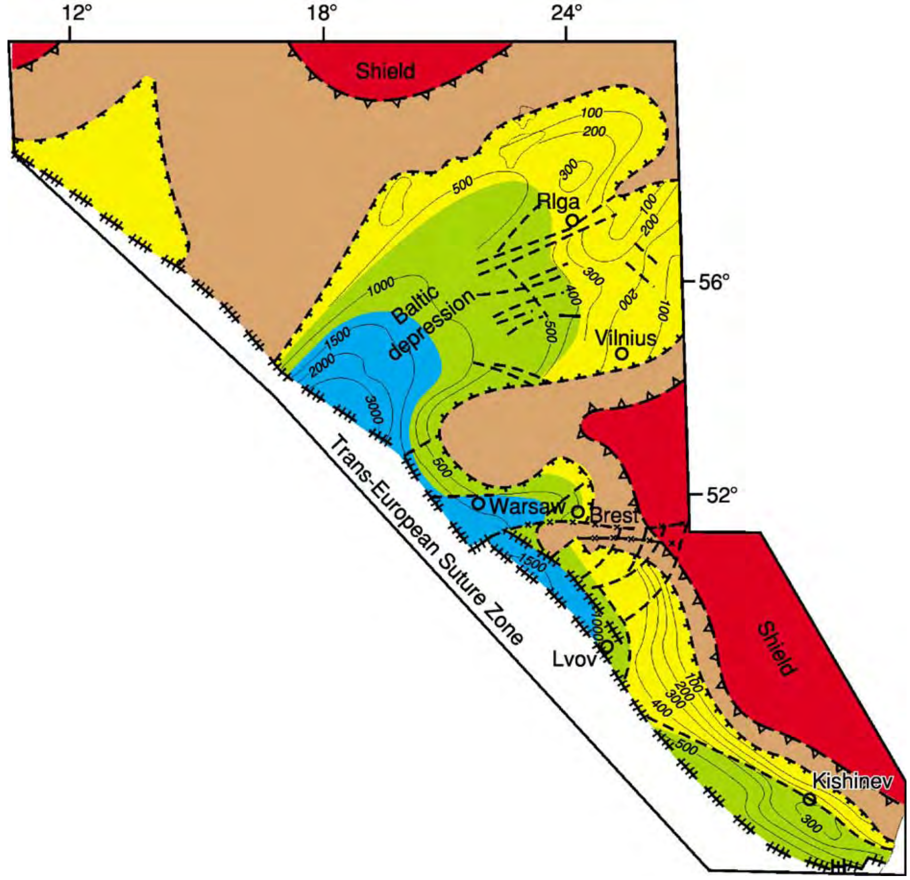


Figure 4 Lithologies of the Cambrian to Silurian rocks along the south western margin of the East European Craton. The lithofacies zones are: blue, deep water pelites; green, lower shelf calcareous and pelitic rocks; and yellow, shallow water calcareous rocks. The brown fields represent areas from which these and similar lithologies have been removed by erosion; red indicates other parts of the exposed Precambrian crystalline crust. © Svetlana Bogdanova.

magmatically, and is overlain or overthrusted by Palaeoproterozoic rock units. Inside the Archaean domain, however, is the north-west–south-east trending Late Palaeoproterozoic collisional Lapland–Kola Orogen, while isotopic data suggest that parts of the Palaeoproterozoic region in the south-west are underlain by blocks or wedges of Archaean crust.

The Archaean domain comprises the Karelian Protocraton, the Belomorian Belt along the White Sea, several different terranes in the Kola Peninsula, and a western province extending all the way to the Lofoten Islands of Norway but partly hidden beneath the Caledonide allochthon.

The Karelian Protocraton is the largest province of Archaean crust. It contains a sizable area and several minor occurrences of igneous rocks older than 3.1 Ga, the oldest zircon ages being about 3.5 Ga. These rock units are set in terranes dominated by 3.0–2.8 Ga greenstone belts rich in komatiites (*see* **Igneous Rocks:** Komatiite) and juvenile tonalite–trondhjemite–granodiorite (TTG) granitoids. The greenstone belts are of several generations and have been attributed to a number of different geodynamic settings such as continental rifts, oceanic arcs, and active continental margins. Remnants of approximately 2.8 Ga ophiolites are found in the north-eastern part of the Protocraton. In the literature, block mosaics as well as large gently dipping crustal wedge structures are indicated. This suggests a complex but still largely undeciphered story of Mesoarchaean continent assemblage. In the Neoarchaean, i.e. after 2.8 Ga, volcanism and granitoid magmatism continued, producing rocks ranging from TTG complexes to granites of collisional regimes and late post-kinematic intrusions.

Similar Neoarchaean developments characterize the adjacent Kola and Belomorian provinces, but these lack rocks older than 3.0 Ga. A notable feature of the Central Kola sub-province is the presence of the thick but subsequently highly metamorphosed 2.7 Ga Keivy cover sequence, which is known for its giant crystals of kyanite, staurolite, and other minerals.

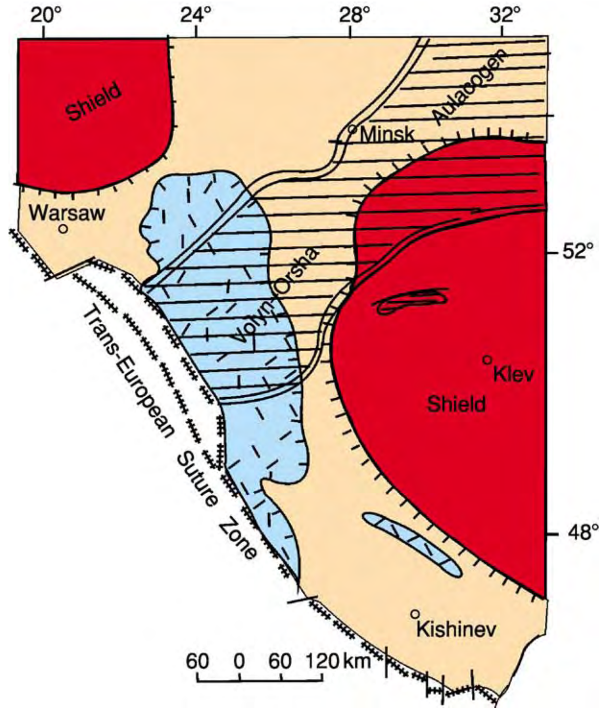


Figure 5 The south western margin of the East European Craton in the Neoproterozoic. The ruled area shows the Meso to Neoproterozoic Volyn Orsha Aulacogen, which is filled by mostly terrigenous sedimentary rocks. In the vicinity of the Trans European Suture Zone, however, the approximately 550 Ma Volynian flood basalts (blue markings) form a large outcrop set in Vendian and Early Cambrian sedimentary rocks (the sand coloured areas). The basalts fill a basin that trends across the Aulacogen and runs roughly parallel to the Trans European Suture Zone. This indicates that the south western margin of the craton was passive at the end of the Neoproterozoic. © Svetlana Bogdanova.

The Belomorian typically features extensive belts of mafic metavolcanics, continental-margin turbidites, and TTG-type granitoids, all arranged in large folded thrust sheets overriding the edge of the Karelian Protocraton. Thrusting here occurred at about 2.70–2.65 Ga, when the different terranes and crustal provinces – except possibly the western Archaean province – collided and merged to form a Neoarchaean continent. Concomitantly, parts of the Karelian province underwent high-grade metamorphism and migmatite formation. Along its boundary with the Belomorian Belt, 2.7 Ga eclogites are present. Exhumation and cooling then followed. Cratonization became complete at around 2.5 Ga.

Between 2.5 Ga and 2.0 Ga, Archaean Fennoscandia underwent extension, rifting, and even breakup, leading to the opening of small oceanic basins. Intermittently there was strong dominantly mafic but also komatiitic igneous activity. Most conspicuous are large greenstone rifts and belts, which tend to follow the boundaries between the different Archaean terranes. Ophiolites with ages of about 1.96 Ga occur in places.

The latter half of the Palaeoproterozoic was a period of intense orogenic activity across most of Fennoscandia. In the north, the Lapland–Kola collisional orogeny at about 1.96–1.85 Ga welded together the crustal units that constitute the present Archaean domain. Some of these appear to have been parts of a Neoarchaean continent that was broken up in the Early Palaeoproterozoic. In the case of other crustal units, an exotic provenance has been suggested. Altogether, the Lapland–Kola Orogen is a tectonic collage of reworked Late Archaean terranes and intervening belts of Palaeoproterozoic juvenile crust and turbidites (Figure 10). Prominent among the latter is the Lapland Granulite Belt, made famous by Eskola's work on the facies of metamorphism. Arc magmatism has been dated at 1.96 Ga; southwards thrusting and collision occurred at 1.91–1.90 Ga, while post-collisional magmatism, exhumation, and cooling lasted until 1.85 Ga.

To the south-west of the Archaean domain, the Proterozoic crust of central Fennoscandia was also created around and soon after 1.90 Ga in a series of orogenic events collectively named the Svecofennian Orogeny. Initial subduction between about 1.93 Ga and 1.91 Ga appears to have been towards the present south-west, creating a primitive volcanic arc, but soon its direction reversed to plunge beneath the Archaean Protocraton in the north-east.

In simplified outline, the Svecofennian Orogen consists of two volcanic belts and an intervening turbidite basin (the Bothnian Basin). In south central Finland there is a very large complex of granitoid plutons. The northern volcanic belt follows the edge of the Archaean Protocraton in Scandinavia but then swings south-eastwards to western and southern Finland, skirting the complex of granites. The southern belt extends between southernmost Finland and south central Sweden. Continuations of these terranes beyond Lake Ladoga into Russia are conceivable (Figure 9).

The ages of volcanism in the two belts overlap within the 1.90–1.88 Ga bracket; dominantly TTG-type plutonism lasted until about 1.87 Ga. The pressure–temperature ratios of Svecofennian metamorphism were mostly low to moderate, which attests to the absence of continent–continent collision. Between them, the two volcanic belts house the majority of the renowned Finnish and Swedish sulphide and iron ores. By 1.85 Ga, the Svecofennian Orogen had been consolidated; however, 1.86–1.84 Ga plutons of granitoids occur locally at its south-western margin and along some major intra-orogenic province boundaries.

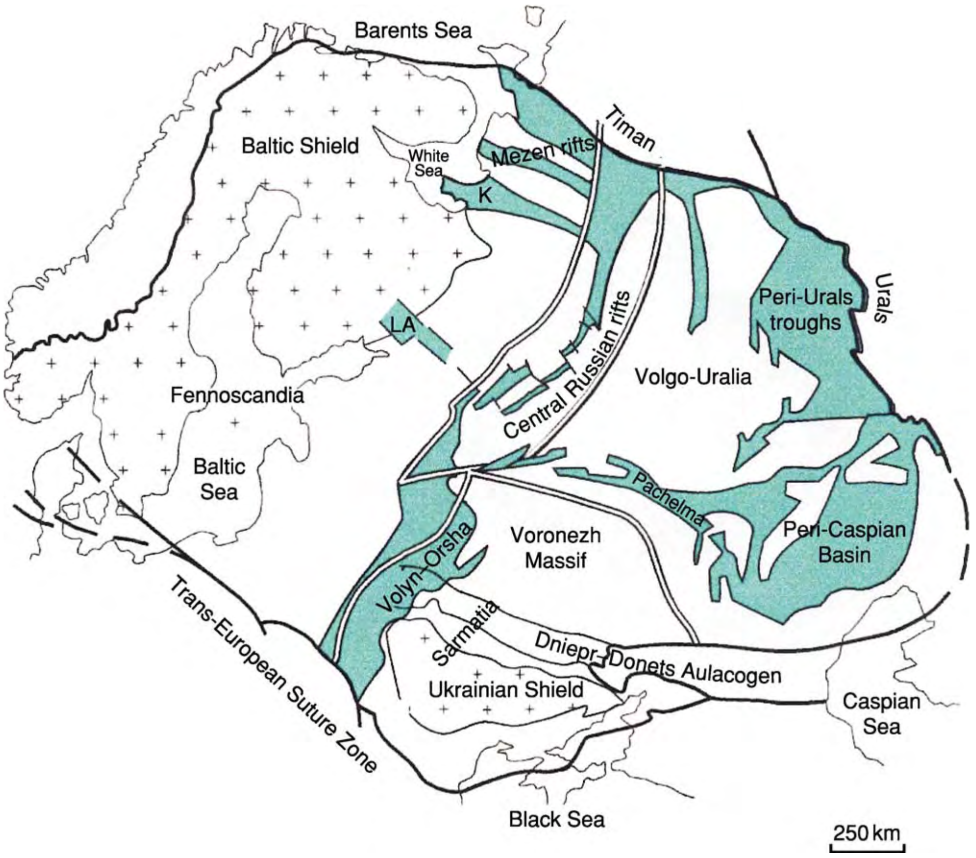


Figure 6 The patterns of the Meso- to Neoproterozoic rifts and aulacogens beneath the sedimentary cover of the Russian Platform conform well with the three-segment fundamental structure of the East European Craton. Thus, the Volyn–Orsha, central Russian, and Pachelma rifts follow the Palaeoproterozoic sutures between Fennoscandia, Sarmatia, and Volgo-Uralia. The Mezen rifts in the north align with the Lapland–Kola collisional orogen and delimit the Timanides, while Volgo-Uralia is completely surrounded by troughs, aulacogens, and the Peri-Caspian Basin. © Svetlana Bogdanova.

Figure 7 The Mesoproterozoic (Riphean) cover sequence in the western foreland of the southern Urals was folded in the Palaeozoic. The sequence is approximately 1500 m thick. Its components are interbedded conglomerates, sandstones, siltstones, pelites, limestones, and dolostones, with numerous deposits of siderite ore. The picture shows rocks of the Bakal Formation in the Irkutskan siderite pit. © Svetlana Bogdanova.

Subsequently, granitic magmatism associated with metamorphism and migmatite formation but lacking connection with mafic rocks took place at various sites and times within the 1.84–1.78 Ga bracket. Current interpretation sees this as a response to compression and thickening of the Svecofennian crust, largely caused by orogeny outside the limits of the Svecofennian domain.

After 1.85 Ga, growth of new crust continued towards the south and the west, outwards from the cratonized Svecofennian territory. Several accretionary orogenic events at around 1.84–1.83 Ga and 1.81–1.78 Ga, and again during the Gothian Orogeny at 1.71–1.67 Ga, have been distinguished in a belt that stretches from Lithuania across southern and western Sweden to central Norway. Outside the limits of Svecofennian crust, the granitoid rocks are TTG-type calc-alkaline, whereas within a marginal western zone of the Svecofennian Orogen the 1.84–1.67 Ga

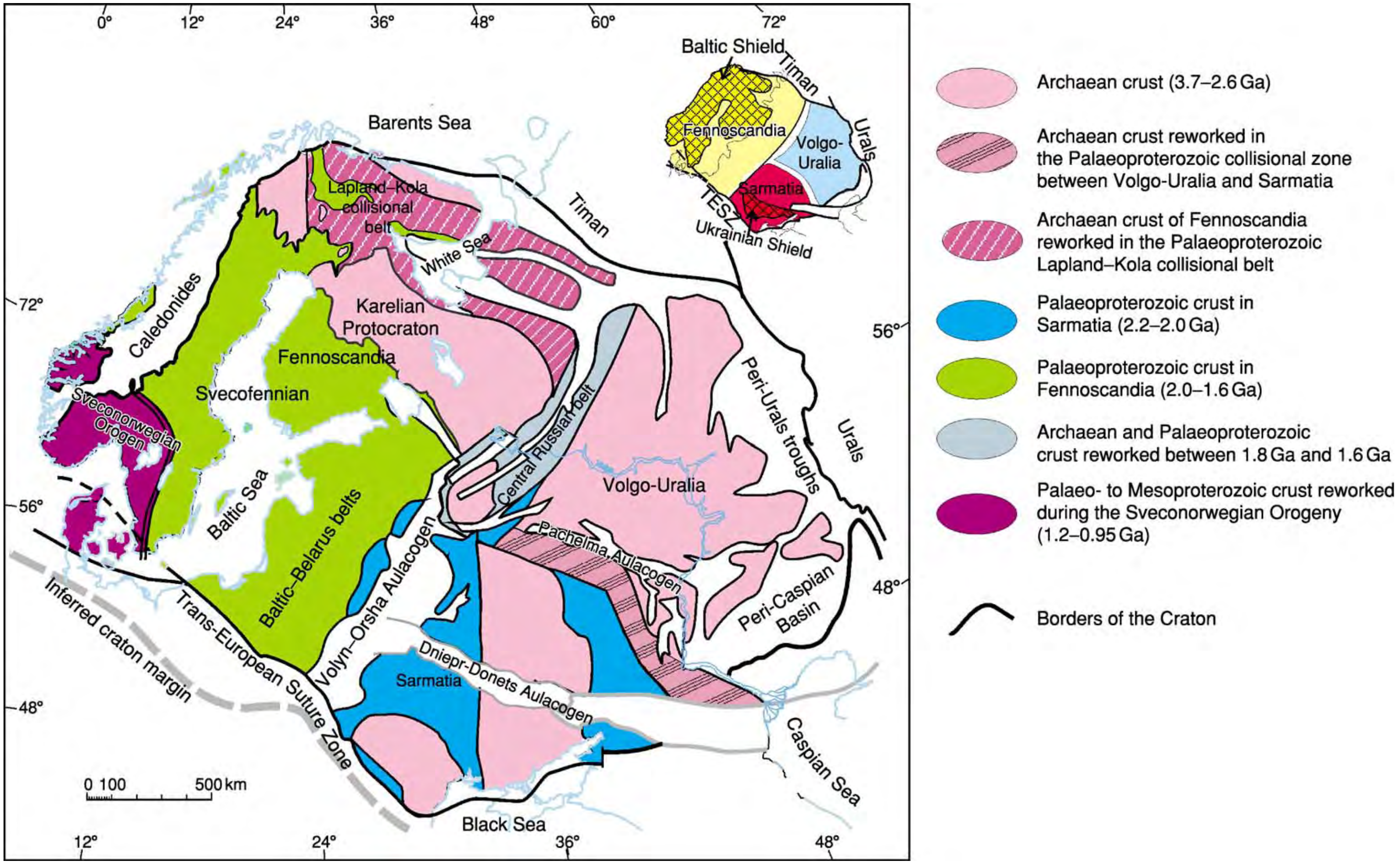


Figure 8 Major crustal provinces of the East European Craton. Apart from Palaeoproterozoic domal structures and some sedimentary and igneous rocks, Volgo Uralia is almost completely Archaean. Sarmatia, in contrast, comprises several large Archaean protocratonic units that were cemented together mostly during the Palaeoproterozoic. In Fennoscandia, Archaean crust is concentrated in the north eastern part, where it forms a separate domain. Sarmatia and Fennoscandia contain relics of approximately 3.7 3.5 Ga crust, while Volgo Uralia is wholly Neoarchaean. Another important difference is that the dominant 2.2 2.0 Ga Palaeoproterozoic crust in Sarmatia is substantially older than the 1.92 1.85 Ga crust of central Fennoscandia. Thus, the age patterns and structural arrangements in each of the three crustal segments are distinct. They are also largely discordant at the segment boundaries. The boundary between Sarmatia and Volgo Uralia is marked by belts of reworked Archaean crust, while the boundary between Fennoscandia and the other two crustal segments is different in its north eastern, largely Archaean, and south western, dominantly Proterozoic, parts. In the north east, the 200 km wide Central Russian Belt is built up of elongated wedged blocks derived from the adjacent crustal segments, whereas in the south west, beneath the Neoproterozoic Volyn Orsha Aulacogen, there are various Palaeoproterozoic terranes ranging from continental margin types to outboard volcanic arcs. These were stacked atop each other between about 1.8 Ga and 1.7 Ga. © Svetlana Bogdanova.

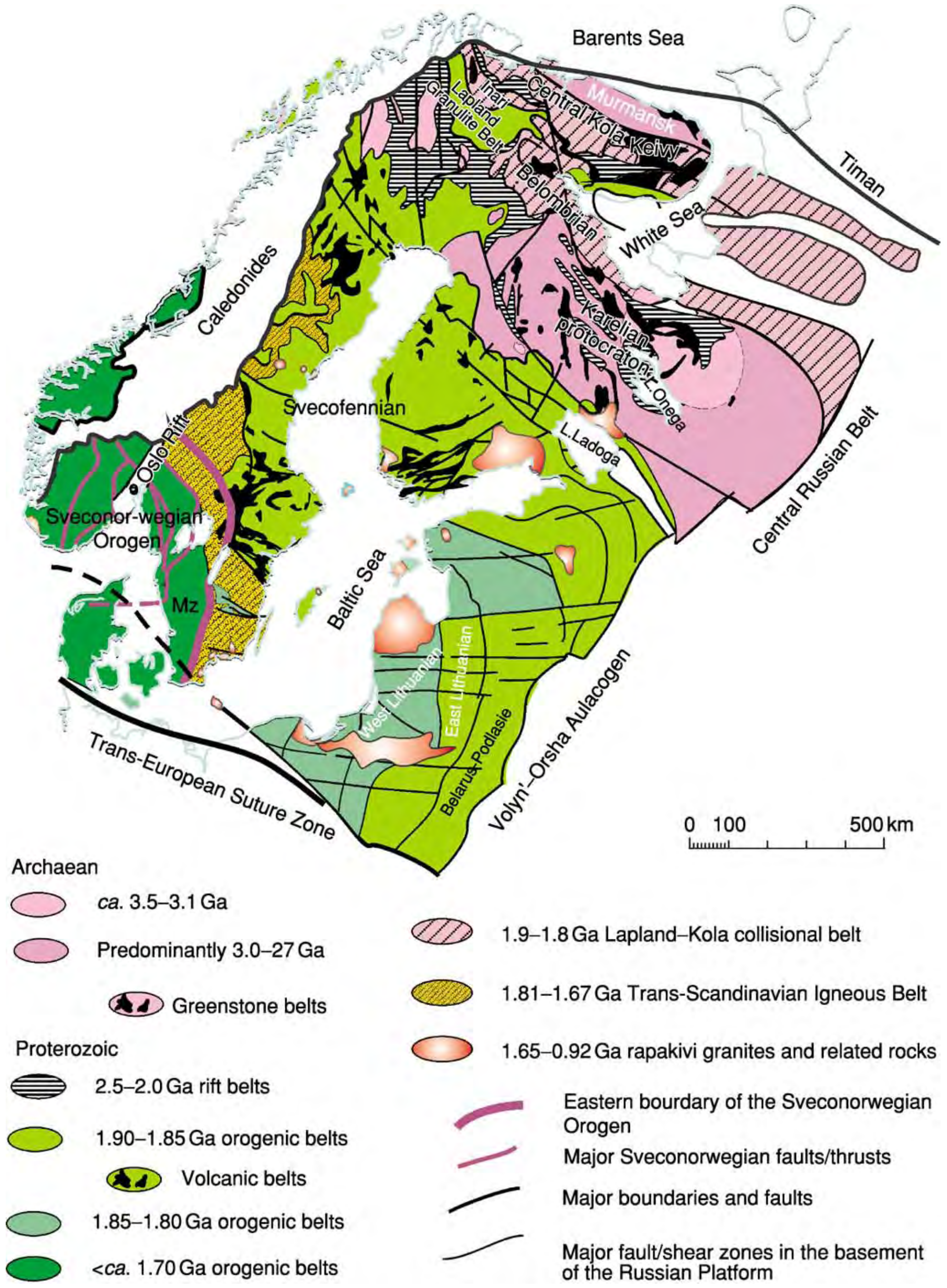


Figure 9 The major crustal provinces of Fennoscandia reflect the subdivision of that crustal segment into largely separate Archaean and Proterozoic parts. The master story of the Archaean part appears to have been the formation of a Mesoarchaean Protocraton in Karelia followed by a Neoarchaean continent. The latter was broken up in the Early Palaeoproterozoic but then brought together again during the Lapland Kola collisional Orogeny. In that process, Archaean terranes exotic to Fennoscandia may also have participated. Essentially coeval with the Palaeoproterozoic Lapland Kola collision in the present north was the first stage of the accretionary formation of Proterozoic crust in the rest of the Fennoscandian crustal segment. In the present west, this process continued episodically until about 1.55 1.50 Ga, apparently without involving major collisions of continental crustal units. In the south west, continental collision is indicated for the period between 1.5 Ga and 1.4 Ga, when interaction with a different craton possibly Proto Amazonia appears to have led to the Danopolonian Orogeny. The subsequent Sveconorwegian orogenic process, between about 1.1 Ga and 0.95 Ga, was pronouncedly collisional and approximately coeval with the Grenvillian Orogeny in Laurentia and the assembly of the Supercontinent Rodinia. Along the south eastern margin of Fennoscandia, the wide Baltic Belarus belts of stacked inboard and outboard terranes mark the Late Palaeoproterozoic (*ca.* 1.8 1.7 Ga) collision zone with Sarmatia. Here, the West Lithuanian Belt and its continuation into south eastern Poland differ sharply from the adjacent East Lithuanian and Belarus Podlasie belts in terms of both the direction of the tectonic grain and the age of crustal formation. In western Lithuania, the crust is nearly coeval with that in south western Sweden. Mz, Mylonite Zone. © Svetlana Bogdanova.

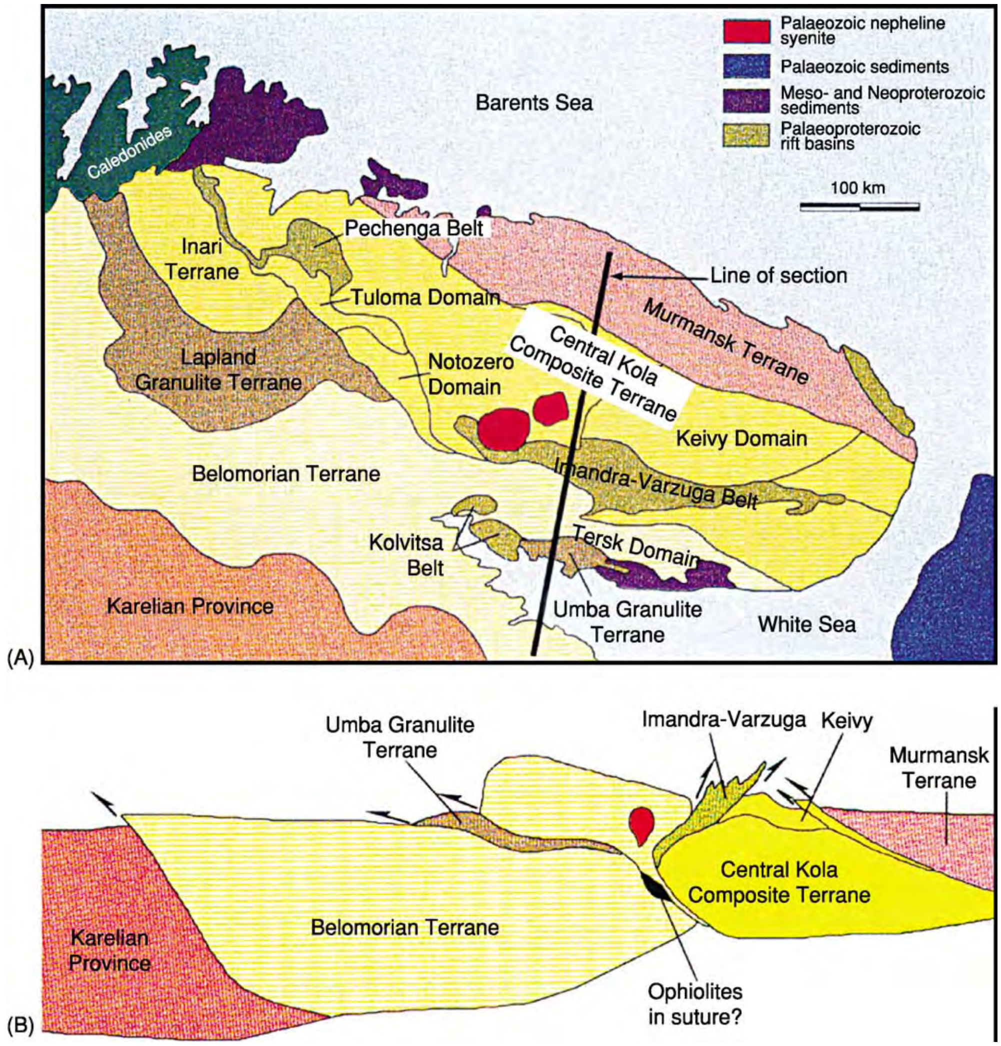


Figure 10 (A) Tectonics of the Lapland Kola collisional belt in the Kola Peninsula. (Map supplied by the Secretariat of the EUROPROBE Programme). (B) Cross section showing the juxtaposition of various Archaean terranes (the Karelian, Belomorian, Murmansk, and Central Kola Keivy) with the Palaeoproterozoic Imandra Varzuga volcanic belt and the Umba Granulite Terrane. Map and cross section by courtesy of the Secretariat of the EUROPROBE Programme.

magmatism was bimodal mafic to granitic and commonly alkali-calcic. As a cumulative result of the different igneous episodes, the large north–south trending Transscandinavian Igneous Belt was formed within the western margin (Figure 9). The Baltic–Belarus belts, some of them with a Palaeoproterozoic evolution similar to that of the Svecofennian in the Baltic Shield, characterize the south-western part of Fennoscandia; they are, however, overlain by phanerozoic sedimentary cover (Figure 3A).

During the Mesoproterozoic, renewed Orogeny between 1.6 Ga and 1.5 Ga created new crust in westernmost Sweden and southern Norway (the Kongsbergian event), etc. In contrast, orogenic events around 1.5–1.4 Ga (the Danopolonian Orogeny) and between 1.1 Ga and 0.9 Ga (the Sveconorwegian Orogeny) were largely collisional. The Sveconorwegian Orogeny involved only limited creation of new crust in south-western Norway but fairly voluminous granitic magmatism and granulitic to eclogitic metamorphism with migmatization in many places. Some of the metamorphic assemblages indicate pressures between 1 GPa and 1.5 GPa. High-grade granulitic rocks accompanied by charnockite, and numerous large granite bodies were also formed at around 1.4 Ga. During both these events, major more-or-less north–south-trending faults, thrusts, and suture zones were developed in most of southern Scandinavia. According to some hypotheses, all of southern Norway and western Sweden west of the so-called ‘Mylonite Zone’ (Figure 9) constitute terranes exotic to Fennoscandia, which were accreted to the East European Craton during the Sveconorwegian Orogeny.

At about 1.64–1.62 Ga, and again at 1.57 Ga, around 1.5 Ga, and 0.93 Ga, numerous large intrusions of rapakivi granites and associated charnockitic, mangeritic, and anorthositic rocks penetrated the Fennoscandian crustal segment. Extensive swarms of mafic dykes and sheets were formed throughout the Proterozoic, with maxima at 2.45–2.1 Ga, 1.6–1.5 Ga, and 1.3–1.2 Ga and 0.97–0.95 Ga. Some of these may be the results of superplume events in the Earth's mantle. In addition there are plugs and plutons of alkaline and carbonatitic rocks (*see* **Igneous Rocks:** Carbonatites), some of which are Precambrian and others Phanerozoic in age. Alkaline magmatism was particularly voluminous in the Kola Peninsula and along rifts such as the Permian Oslo Graben in Norway. Other intrusions formed isolated minor plugs.

Sarmatia

The Sarmatian segment is built up of several Archaean crustal provinces with ages ranging between 3.7 Ga and 2.7 Ga (Figure 11). Belts of Palaeoproterozoic crust intervene between some of these, but accretion and welding appear to have commenced in the Archaean.

The oldest crustal units are the Palaeo- to Mesoarchaean Podolian Block in the extreme southwest and the Oskol (Kursk-)–Azov Block in the east. Ages of 3.65–3.60 Ga have been obtained from ultramafic and tonalitic rocks, while major TTG intrusions are 3.4–3.3 Ga old. Both the Podolian Block and the Oskol–Azov Block represent one-time lower crust but have evolved through several events of high-grade granulitic metamorphism and multiphase magmatism in the Archaean and Palaeoproterozoic.

The Sumy–Dniepr gneiss–granite–greenstone terrane is generally somewhat younger. It was formed mainly between 3.2 Ga and 3.1 Ga and is characterized by numerous belts of principally mafic metavolcanic and sedimentary rocks. These contain basal sequences of komatiites, komatiitic basalts, and tholeiites, reaching thicknesses of 4–5 km. Atop the mafic volcanics rest calc-alkaline felsic volcanic rocks, tuffites, immature sandy metasediments, and banded iron formations with quartzites. These belts were metamorphosed and deformed concomitantly with the development of domal structures containing juvenile TTG-type granitoids and felsic gneisses with granulites. The alteration processes were related either to rifting and the opening of a Mesoarchaean ocean or to the evolution of a preexisting oceanic basin of Palaeo- to Mesoarchaean age.

In the Palaeoproterozoic, between about 2.6 Ga and 2.3 Ga, major belts of supracrustal rocks containing numerous banded iron formations were formed. Some of these outline the margins of Archaean terranes, while others appear to be controlled by rifts in the interior parts of the Oskol–Azov Block. The latter are the cause of the strong Kursk magnetic

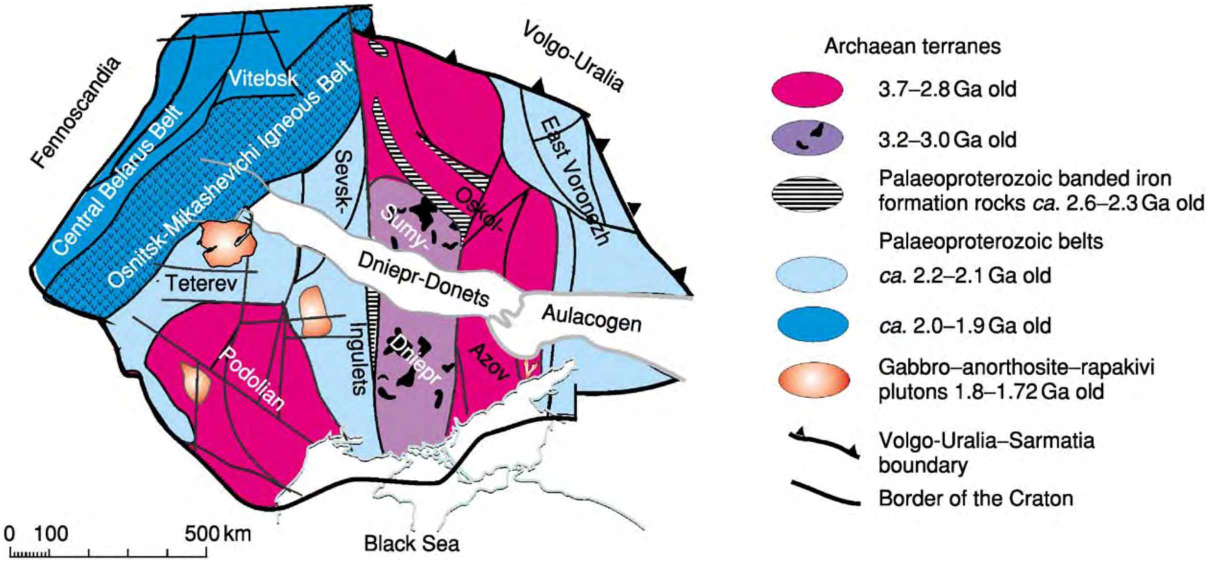


Figure 11 The processes that shaped the internal structure of Sarmatia were the docking of the older Archaean Oskol Azov and the younger Archaean Sumy Dniepr crustal blocks and, during the Proterozoic, the accretion of the Podolian Block to the evolving main mass of the Sarmatian crustal segment. The latter process was coeval with the generation of mantle derived juvenile Proterozoic crust elsewhere in Sarmatia. Directly related to the assembly of the East European Craton were the formation of the East Voronezh belts along the developing boundary between Sarmatia and Volgo Uralia about 2.1 2.0 Ga ago and the formation of the continental margin Osnitsk Mikashevichi Igneous Belt and the outboard Central Belarus Belt at around 2.0 Ga. These two belts faced the ocean basin that was consumed during the collision of united Volgo Uralia and Sarmatia (i.e. 'Volgo Sarmatia') with Fennoscandia. © Svetlana Bogdanova.

anomaly that marks an important iron-ore province. At about the time that these rock belts were formed, an ancient platform cover began to be deposited. The largest sedimentary basin was in the Sevsk-Ingulets (Kirovograd) domain.

In north-western Sarmatia, some Palaeoproterozoic volcanic rocks, e.g. those in the Teterev Belt, have mantle-related geochemical signatures and probably represent island arcs. The most conspicuous complexes of juvenile igneous rocks, however, are found in the East Voronezh belts in eastern Sarmatia, where an approximately 2.1 Ga volcanic arc and back-arc sedimentary basins with numerous ultramafic and mafic intrusions occur. Together with numerous granitoid intrusions of the same age along the edge of the Archaean Oskol–Azov Block, these indicate the presence of an ocean–continent interface. At this time Sarmatia was assembled and its dominant north–south-trending suture zones were created. These have crustal roots approximately 60 km deep.

Along the north-western margin of Sarmatia, the Early Palaeoproterozoic north–south fabric is truncated by several wide north-east-trending belts of Late Palaeoproterozoic rocks. These are the Osnitsk–Mikashevichi Igneous Belt, the Central Belarus Belt, and the Vitebsk granulite domain (Figure 11), all of which were formed between 2.0 Ga and 1.9 Ga, a top south-eastwards dipping subduction zones or zones of arc collision.

In particular, the presence of the extensive Osnitsk–Mikashevichi Igneous Belt suggests the existence at around 2.0 Ga of a unified Sarmatia, complete with a coherent north-western to northern margin. The Osnitsk–Mikashvichi Igneous Belt is 200–250 km wide and comprises volcanic, hypabyssal, and dominantly plutonic igneous rocks. The volcanic and dyke rocks are metabasalts and dolerites, meta-andesites, and meta-keratophyres, all metamorphosed in various ranges of the amphibolite facies. Granodiorites and granites with subordinate gabbros and diorites form large batholithic plutons, which are only weakly deformed and metamorphosed. The metavolcanics and minor interbedded metasedimentary rocks occupy small areas within the plutons or form septa between them.

A striking feature in Sarmatia is the occurrence of several large plutons of gabbro-anorthosites and rapakivi granites of 1.80–1.72 Ga age that are rich in titanium and other ore deposits. Traditionally, these intracratonic plutons have been considered to be anorogenic, but recent discussion tends to connect them with zones of preceding plate collision. New seismic results indicate that the anorthosite–rapakivi magmatism strongly influenced the composition of both the crust and the upper mantle.

In the Devonian, the formation of the Dniepr–Donets Aulacogen subdivided the Sarmatian crustal segment into two parts. The southern part contains the Ukrainian Shield, and the northern part contains the Voronezh Massif. However, no significant displacement along the aulacogen can have taken place, since the terrains on each side correlate excellently.

Volgo-Uralia

The crystalline crust of Volgo-Uralia is mostly Neoarchaean, the highest depleted-mantle Nd model ages being approximately 3.1 Ga. This crust is completely buried beneath a Phanerozoic, mostly Devonian to Triassic, sedimentary cover. The only exception is the Taratash complex (Figure 12), where a basement slice was brought up to the Earth's surface by thrusting connected with the Uralide Orogeny.

Under the Phanerozoic cover, however, there are numerous Meso- to Neoproterozoic aulacogens and troughs filled with 2–10 km thick sedimentary rocks intercalated with rare volcanics. Proterozoic rifting substantially complicated the margins of Volgo-Uralia and penetrated deeply into its interior. Because of the extensive cover, information on the crystalline crust is mostly derived from geophysics and from thousands of drill cores, which are particularly numerous because of the high oil and gas potential of the region.

Volgo-Uralia in general is a realm of high magnetization and dense upper crust. A characteristic feature of its magnetic field is the presence of numerous, in part very large (300 km and more), circular to oval concentric anomalies separated by wide belts of more or less linear anomaly patterns. Traditionally, such patterns have been explained by assuming that the non-linear nearly isometric anomalies represent stable Archaean massifs, while the linear zones relate to Palaeoproterozoic mobile belts. In Volgo-Uralia, however, this interpretation is at complete variance with the presently available geological information, which demonstrates that the circular anomalies represent Palaeoproterozoic domal, relatively uplifted, structures. In contrast, the linear belts consist of relatively well-preserved Archaean rocks. These are 3.0–2.7 Ga old and comprise metasedimentary and subordinate metaigneous granulites, these two types mostly forming separate belts. In addition, there are also greenstone sequences with komatiitic volcanics. Associated with the Archaean belts are zones of strong shearing and mylonitization defining fold–thrust structures and generally trending north-east–south-west to east–west.

Most of the domes in the Volga–Kama megablock (Figure 12) are made up of Archaean granulites reworked during the Palaeoproterozoic. In their

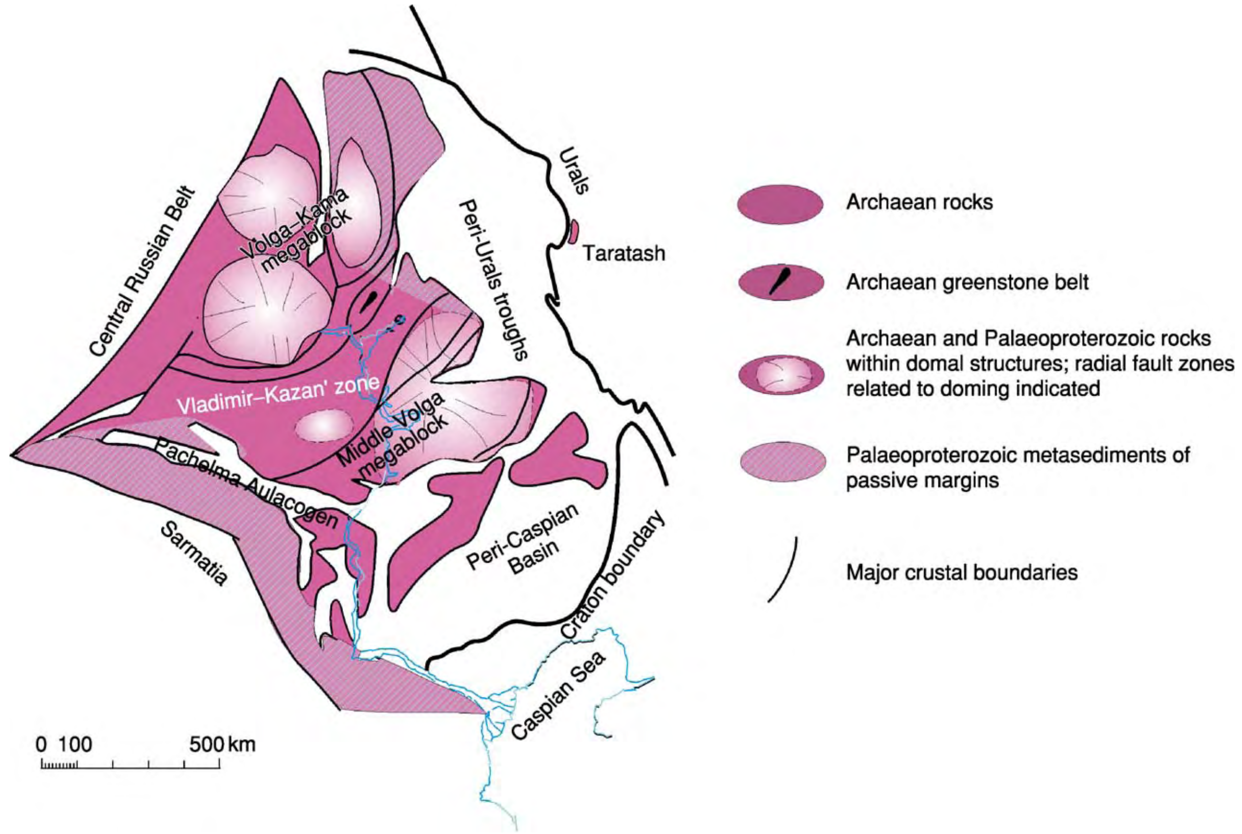


Figure 12 The specifics of Volgo Uralia were largely determined by the development of Palaeoproterozoic domal structures in the Neoarchaean continental crust. At about 2.1 2.0 Ga, the marginal belt facing Sarmatia was formed. This belt is composed of Palaeoproterozoic sedimentary commonly flysh like rocks, which suggest the one time existence of a passive margin. Subsequent Mesoproterozoic rifting and the formation of troughs, aulacogens, and basins influenced the structure of Volgo Uralia more than that of the other two East European crustal segments. The Peri Caspian Basin is unique in the East European Craton in having a relatively shallow Moho and a great thickness of sedimentary cover. © Svetlana Bogdanova.

central parts are Palaeoproterozoic granite and gneiss complexes. In contrast, the very large circular structure in the Middle Volga megablock has Palaeoproterozoic rocks and Palaeoproterozoic reworking on its flanks, while north-east–south-west striking Archaean granulites similar to those outside the dome occupy the centre.

All the domes feature radial and ring-shaped faults related to the deformation of the Archaean crust and associated with rare Palaeoproterozoic metasedimentary and metavolcanic rocks. The domes were formed between 2.3 and 2.0 Ga and may be related to the nearly coeval collision of Volgo-Uralia with Sarmatia. Along that collision zone and in the north-east of Volgo-Uralia there are large areas of Palaeoproterozoic turbiditic and metapelitic mica schists, silts, sandstones, and carbonaceous shales formed on the shelf margins of the crustal segment.

The Archaean and Palaeoproterozoic rocks of Volgo-Uralia are cut by numerous mafic dykes and sills, which were intruded during several stages of Meso- to Neoproterozoic rifting.

The Proterozoic fault systems of Volgo-Uralia strongly influenced the formation and structure of the sedimentary cover and the locations of the major Palaeozoic oil and gas resources.

Assembly of the East European Craton

The assembly of the East European Craton can be followed from palaeomagnetic data, isotopic ages, and analyses of rock formation and igneous activity at the margins of the three involved crustal segments (Figure 13).

First to collide were Volgo-Uralia and Sarmatia. The collision occurred at 2.1–2.05 Ga, when Volgo-Uralia was thrust onto Sarmatia. Geophysically, the suture can be followed for 30 km beneath Volgo-Uralia. At the Earth's surface, the junction zone contains extensive shelf deposits on the Volgo-Uralian

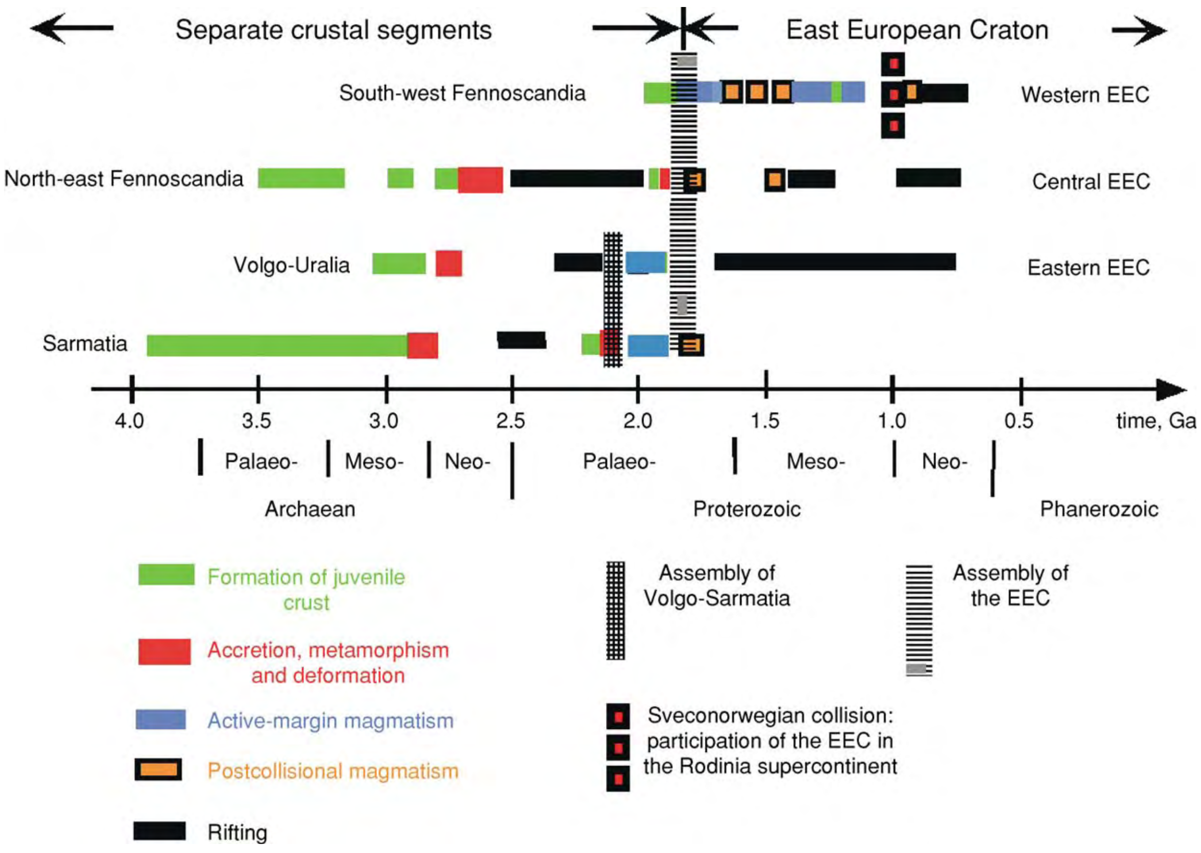


Figure 13 Fennoscandia, Sarmatia, and Volgo Uralia, the three crustal segments of the East European Craton (EEC), evolved differently during the Archaean and Palaeoproterozoic. Sarmatia and Volgo Uralia merged at around 2.1 2.0 Ga, creating a larger continental unit, Volgo Sarmatia, and thereafter shared a common history. Between about 1.8 Ga and 1.7 Ga, Fennoscandia and Volgo Sarmatia gradually amalgamated to form the EEC. The eastern and western margins of the EEC differed during the rest of the Proterozoic: mostly passive in the east and with episodes of activity in the west. In the west, the *ca.* 1.0 Ga Sveconorwegian collisional orogeny marks the beginning of the participation of the EEC in the Supercontinent Rodinia. © Svetlana Bogdanova.

side, while the so-called East Voronezh belts mark the edge of Sarmatia.

The south-westernmost belt represents a volcanic island arc, while the north-eastern belts preserve back-arc volcanic rocks and turbidites. Voluminous granitic magmatism occurred particularly in Volgo-Uralia. During the Mesoproterozoic, the Pachelma Aulacogen developed along the Sarmatia–Volgo-Uralia junction.

After that collision, it was several hundred million years before the combined Sarmatia and Volgo-Uralia ('Volgo-Sarmatia') approached Fennoscandia. Meanwhile, accretionary crustal growth occurred separately along the edges of both Fennoscandia and Sarmatia. At approximately 2.0 Ga, subduction of oceanic crust beneath the presently north-western edge of Volgo-Sarmatia commenced. This process eventually consumed all the oceanic crust between the protocraton and Fennoscandia. Within the continental margin of Sarmatia, the Osnitsk–Mikashevichi Belt began to develop, while the Central Belarus island arcs were formed outboard. Further west, the Middle Lithuanian Suture Zone, separating east–west trending rocks of the West Lithuanian Terrane from a belt of north-north-east-striking rocks in eastern Lithuania, was created. It may either coincide with the actual Fennoscandia–Sarmatia suture or be one of several components within a swarm of intersegment terrane boundaries. Palaeomagnetic data indicate that the colliding segments were still at least 2000 km apart between 1.80 Ga and 1.75 Ga.

The Palaeoproterozoic suture between Fennoscandia and Volgo-Sarmatia became the site where the Volyn–Orsha and central Russian systems of aulacogens, rifts and troughs, and sedimentary basins were developed in the Late Mesoproterozoic and Early Neoproterozoic.

See Also

Europe: Timanides of Northern Russia; Scandinavian Caledonides (with Greenland); The Urals. **Igneous Rocks:**

Carbonatites; Kimberlite; Komatiite. **Shields**. **Tectonics:** Mountain Building and Orogeny.

Further Reading

Bogdanova SV (1986) *Zemnaya Kora Russkoy Plity v Rannem Dokembrii* [The Earth's Crust of the Russian Platform in the Early Precambrian]. Transactions of the Geological Institute of the USSR Academy of Sciences, vol. 408. Moscow: Nauka. [In Russian.]

Bogdanova S, Gorbatschev R, Stephenson RA, and Guterch A (eds.) (2001) EUROBRIDGE: Palaeoproterozoic accretion of Fennoscandia and Sarmatia. *Tectonophysics* 339: 1 237.

Gaál G and Gorbatschev R (1987) An outline of the Precambrian evolution of the Baltic Shield. *Precambrian Research* 35: 15 52.

Gee DG and Zeyen HJ (eds.) (1996) EUROPROBE 1996, Lithosphere Dynamics: Origin and Evolution of Continents. Published by the EUROPROBE Secretariate, Uppsala University, p. 138.

Glebovitsky VA (1997) *Early Precambrian of Russia*. London: Taylor and Francis Books Ltd.

Gorbatschev R and Bogdanova S (1993) Frontiers in the Baltic Shield. *Precambrian Research* 64: 3 21.

Gower CF, Rivers T, and Ryan B (eds.) (1990) *Mid Proterozoic Laurentia Baltica*. Special Paper 38. St Johns: Geological Association of Canada.

Khain VE (1985) *Geology of the USSR. First Part: Old Cratons and Paleozoic Fold Belts*. Berlin: Gebrüder Borntraeger.

Koistinen T, Stephens MB, Bogachev V, *et al.* (2001) *Geological Map of the Fennoscandian Shield, Scale 1 : 2 000 000*. Moscow: *Geological Surveys of Finland, Norway, and Sweden and North West Department of Natural Resources of Russia*.

Leonov YG and Khain VE (eds.) (1996) *International Tectonic Map of Europe and Adjacent Areas,* 3rd edn, *Scale 1 : 5 000 000, 5 sheets*. Paris: IUGS/UNESCO.

Nironen M (1997) The Svecofennian Orogen. *Precambrian Research* 86: 21 44.

Peive AV, Khain VE, Muratov MV, and Delany F (eds.) (1979) *International Tectonic Map of Europe and Adjacent Areas, Scale 1 : 2 500 000, 20 sheets*. Moscow: IUGS/UNESCO.

Stephenson RA, Wilson M, De Boorder H, and Starostenko VI (eds.) (1996) EUROPROBE: intraplate tectonics and basin geodynamics of the East European Platform. *Tectonophysics* 268: 1 309.

Thybo H, Pharaoh T, and Guterch A (eds.) (2002) Geophysical investigations on the Trans European Suture Zone II. *Tectonophysics* 360: 1 314.

Windley BF (1995) *The Evolving Continents,* 3rd edn. Chichester: John Wiley & Sons Ltd.

Ziegler P (1990) *Geological Atlas of Western and Central Europe*. Bath: Shell International Petroleum Mij BV and Geological Society Publishing House.

Timanides of Northern Russia

D G Gee, Uppsala University, Uppsala, Sweden

Introduction

Much of eastern and northern Europe, from the Ural Mountains to the high Arctic of Novaya Zemlya and westwards into the eastern Barents Sea and Pechora Basin (Figure 1), has a bedrock that is dominated by Palaeozoic and Mesozoic successions, which unconformably overlie deformed and metamorphosed Neoproterozoic rocks. The Late Neoproterozoic (Vendian) orogeny of the latter is referred to as Timanian, based on type areas in the Timan Range. Because it is largely covered by Phanerozoic successions and, towards the east in the Ural Mountains, much influenced by Late Palaeozoic to Early Mesozoic orogeny, the Timanide Orogen is less well known than many of the other old exhumed European mountain belts. Nevertheless, it dominates a large part of the Earth's crust in northeastern Europe, comprising the basement to thick hydrocarbon-bearing successions in the overlying Phanerozoic cover.

The concept of a Timanide mountain belt and related orogeny goes back at least a hundred years ago, based on type areas in the Timan Range. Other names that are widespread in the literature, for example, Baikalian, Cadomian and Assyntian, derived from contemporaneous orogenies in other parts of Eurasia, are less appropriate (and may even be misleading) for this East European Orogen.

The sedimentary rocks involved in the Timanide Orogen, exposed in the Timan Range and further south, in the western foreland of the Ural Mountains, were deposited along the eastern margin (present day coordinates) of the East European Craton (EEC) (*see* **Europe:** East European Craton) in the Neoproterozoic (1000–545 Ma); in some areas, sedimentation started in the Mesoproterozoic (1600–1000 Ma). The underlying EEC, dominating the European bedrock from Poland to the Uralian foreland and forming the ancient core of northern Europe, is composed of Archaean and Palaeoproterozoic crystalline complexes, mostly of

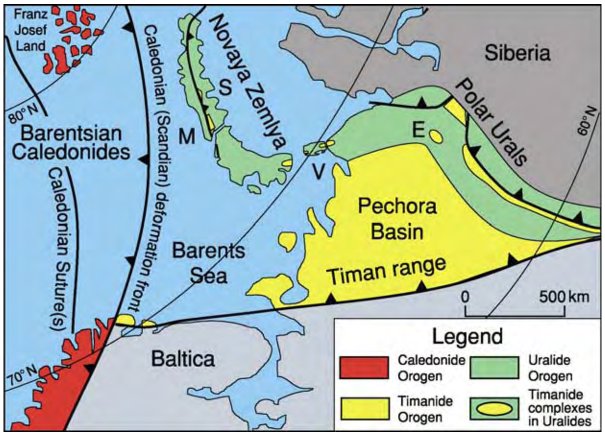


Figure 1 North western Russia: tectonic relationships. M: Mitushev Bay, S: Sulmenev Bay, V: Vaigach Island.

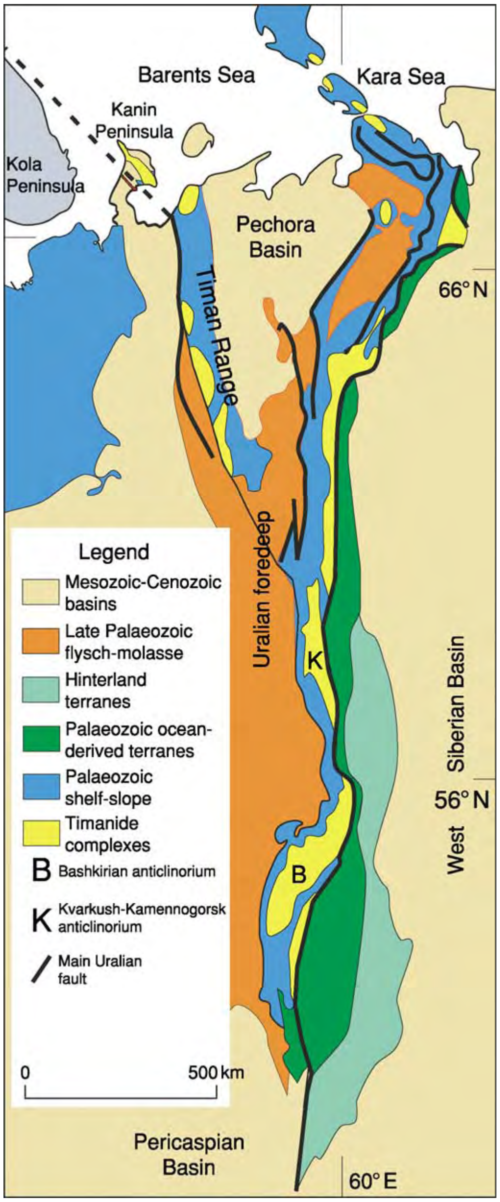


Figure 2 Timanide Complexes of the Urals and Timan Range.

highly metamorphosed sedimentary and igneous rocks. From about 1650 Ma, the EEC was subject to repeated episodes of rifting, with the development of intracratonic basins (aulacogens). This rifting also controlled the structure of the eastern margin of the craton, with general subsidence leading to widespread deposition. The latter continued until the onset of Timanian Orogeny in the Vendian. Thereafter, uplift and deep erosion reduced most of the orogen to a peneplain which, early in the Palaeozoic, developed as a rifted margin to the craton-cored continent; the latter is generally referred to by palaeogeographers as Baltica. Deposition along the eastern edge of Baltica continued, mostly in passive margin shelf facies, until the beginning of the Uralian Orogeny in the Late Palaeozoic.

The name Baltica is applied to both a palaeocontinent (with a core of the EEC) and a plate (i.e. the continent with surrounding oceanic crust, reaching out to the plate boundaries) which existed in the Late Precambrian and Early Palaeozoic. Characteristic endemic faunas defined its independence from other major continents in the Early Cambrian. Baltica became a part of Laurussia during the mid-Palaeozoic Caledonian Orogeny. Baltica came into existence as an independent plate as the result of Neoproterozoic rifting and fragmentation of the supercontinent Rodinia; exactly when is not well defined. However, the eastern (Timanian) oceanic margin of Baltica was clearly established earlier in the Neoproterozoic than the north-western Baltoscandian margin.

The foreland folding and thrusting of the Timanide Orogen (Figure 2) extends for a length of at least 3000 km from the southern Urals (where it disappears beneath younger successions), northwards along the Uralian deformation front and then northwestwards via the Timan Range and northern edge of the Kola Peninsula to the Varanger Peninsula of northernmost Norway. Here, the Timanide Orogen is truncated by the mid-Palaeozoic, Scandian thrust front of the Scandinavian Caledonides and the latter is inferred to continue northwards through the Barents Sea to the northern edge of the Eurasian Shelf (Figure 1).

How far east the Timanides extended is not known. Timanian deformed and metamorphosed rocks occur throughout the western flank of Ural Mountains and into southern and central Novaya Zemlya. They dominate the Baltica margin complexes which underlie the main Palaeozoic allochthons of ocean-derived ophiolites and island-arc volcano-sedimentary assemblages, and the latter, towards the east, are covered by Mesozoic successions of the West Siberian Basin.

The timing of the Timanian Orogeny is constrained by the ages of the youngest sedimentary successions involved in the Precambrian deformation and the oldest unconformably overlying sediments. The latter are generally Ordovician, but locally (beneath Kolguev Island and the Pechora Basin and on Novaya Zemlya) reach back into the Cambrian, the oldest strata (late Early Cambrian) occurring in central Novaya Zemlya. The youngest strata involved in the Timanian deformation include tillites in southern areas and contains Vendian acritarchs. Late Vendian successions in some Timanian foreland areas are developed in molasse facies. Thus, it can be inferred that the Timanian Orogeny occurred in the Vendian (Late Vendian in the front of the orogen; perhaps earlier in the hinterland) and lasted at least until the end of the Vendian, during uplift and erosion of the orogen, prior to Early Cambrian peneplanation and then the start of Palaeozoic platform deposition along the Baltica margin.

The different parts of the Timanide Orogen are summarized below, starting in the foreland fold-and-thrust belt and then continuing eastwards beneath the Pechora Basin to the metamorphic complexes of the Subarctic and Polar Urals and Novaya Zemlya.

Timanian Foreland Fold and Thrust Belt

Two very different Neoproterozoic sedimentary successions are preserved along the western front of the Timanide Orogen, the one in the Southern and Middle Urals and the other further to the north-west.

In the Timan Range and north-westwards to the Varanger Peninsula, thick turbidite-dominated successions occur in thrust sheets emplaced south-westwards onto platform facies, shallow marine carbonate (often stromatolite-bearing) and siliciclastic formations. These turbidites were apparently deposited along the margin of the East European Craton in continental slope and rise environments. These sedimentary rocks occur in upright to SW-vergent folds and are generally well preserved, with excellent characteristic sedimentary structures (graded bedding and Bouma sequences) and low grade of metamorphism (low greenschist facies). Only locally in the exposed thrust belt, on the Kanin Peninsula and northernmost Timan, are more deformed and metamorphosed sedimentary rocks present in the Timanide thrust sheets, providing evidence of the influence of regional high amphibolite facies metamorphism at depth within this part of the orogen. Blueschists have also been reported in the north-easternmost parts of the Kanin Peninsula. The turbidite-dominated successions are extensively intruded by pre-tectonic dolerite dykes and, in northern Timan, by an alkaline suite of gabbros, granites, and syenites, often nepheline-bearing, yielding zircon U/Pb ages of *ca*. 615 Ma.

To the south-west of the main Timanide deformation front, Late Vendian generally non-marine siliciclastic successions in the Varanger Peninsula, the Mezen Basin, and further south-east in the orogen were derived from the Timanide hinterland to the north-east, and are inferred to be molasse.

In the southern Urals (*see* **Europe:** The Urals), the Neoproterozoic turbidite facies is not exposed, and thick (up to 15 km) intracratonic successions were deposited from the beginning of the Mesoproterozoic to the Late Neoproterozoic, dominated by shallow-water siliciclastic and carbonate formations with some rift volcanics and sub-volcanic intrusions. These successions are called Riphean and are overlain by Early Vendian tillites and Late Vendian molasse, the latter a response to Timanian Orogeny further east.

The type area for Riphean stratigraphy is located in the core of a major Late Palaeozoic fold, the Bashkirian Anticlinorium, in the foreland fold-belt of the Urals (Figure 2). Beneath a major Ordovician unconformity, the folded and faulted Riphean and Vendian sedimentary rocks are overlain by a more deformed and, in part, more metamorphosed (amphibolite facies, locally with eclogites) allochthon, the Beloretsk Terrane. Neoproterozoic turbidites, similar to those in the Timan Range, may have been deposited in the areas at present occupied by the hinterland of the Southern Urals; if this was the case, they are now in the unexposed footwall beneath the Beloretsk Terrane.

Further north, in the middle Urals (Figure 2), another major Uralian fold, the Kvarkush-Kamennogorsk Anticlinorium, like the Bashkirian Anticlinorium (above), also contains Proterozoic successions below a post-Timanian, Ordovician unconformity. The siliciclastic and carbonate formations of this region are thought to be Neoproterozoic in age, reaching up into Early Vendian tillites and, probably, Late Vendian molasse. Of particular interest is the occurrence in the Kvarkush area of blue schists of Late Vendian (or possibly earliest Cambrian) age, thrust on to the other less metamorphosed, low greenschist facies sedimentary rocks. Ordovician quartzites overlie these Precambrian rocks and structures with major unconformity.

Basement of the Pechora Basin

From the Timan Range, the Timanide Orogen extends eastwards beneath the Palaeozoic and younger strata of the Pechora Basin (Figure 2) to the foreland fold and thrust belt of the Polar Urals. The Phanerozoic cover thickens eastwards into the Uralian foredeep, where the unconformably overlying successions are 10–15 km thick and the character of the underlying basement is unknown. However, in some central parts of the Pechora Basin, along NW-trending axes of uplift, the basement reaches to within a few kilometres of the surface and deep drilling (about seventy holes) has succeeded in penetrating the entire cover succession to sample the basement. On the basis of this drilling, and also regional potential field (gravity and magnetic) anomaly maps, it has been possible to reconstruct a simple geological map of the pre-Palaeozoic rocks (Figure 3) beneath the Pechora Basin.

To the east of the Timan Range, three main tectonostratigraphic zones have been recognised. The western is referred to as the Izhma Zone and is a direct easterly extension of the Timan Range turbidite assemblages. Towards its eastern contact to the adjacent Pechora Zone, it is intruded by granites, mainly of calc-alkaline composition. The Pechora Zone is well seen on the potential field maps, it being well defined by a broad belt of NW-trending magnetic anomalies. Drillcores show that the Pechora Zone is dominated by volcanic and volcaniclastic rocks, extensively intruded by plutons, mainly of mafic and intermediate composition. Calc-alkaline granites also occur in this zone; together with those in the Izhma Zone, they yield ages of 550–560 Ma and apparently date the end of Timanian deformation and metamorphism in this part of the hinterland (Figure 4). Further east, the Bolshezemelskaya Zone also is preserved locally in structural highs, accessible to the drilling-rig. Acid volcanic rocks and granites have been sampled here from below the Ordovician unconformity, and somewhat older zircon ages of *ca* 570 and 620 Ma have been obtained from the latter, along with Grenville-age xenocrysts.

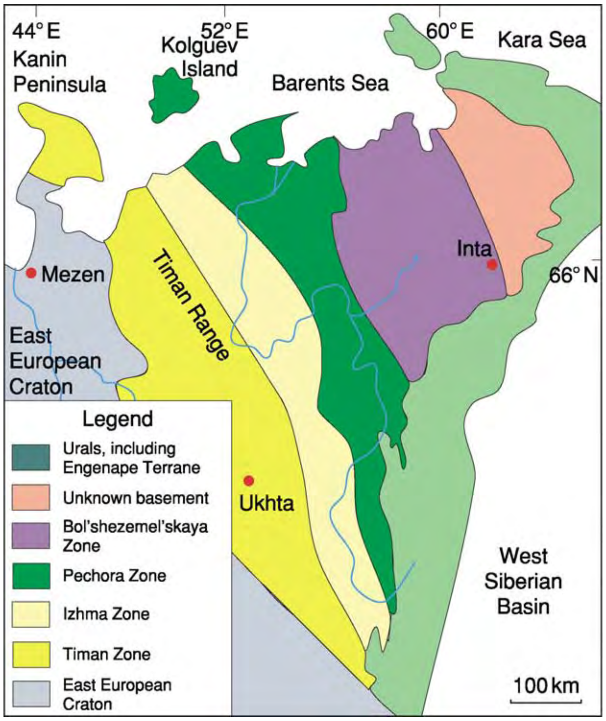


Figure 3 Pre Palaeozoic basement of the Timan Pechora region.

Previous literature, especially in Russia, contains several different names for the Bol'shezemel'skaya Terrane (e.g., Khoreyer and Novozemel'skaya) of continental blocks inferred (mostly on geophysical evidence) to exist at depth below the Pechora Basin and southern Barents Sea.

Pre-Ordovician Complexes of the Subarctic Urals

The Neoproterozoic complexes of the Pechora and Bol'shezemel'skaya zones, identified at depths of several kilometres beneath the Pechora Basin, strike south-eastwards into the mountain front of the Subarctic Urals. Within the Uralian frontal folds and lower thrust sheets, pre-Ordovician formations are exposed. Andesitic volcanic rocks dominate these Precambrian rocks in the northern Urals, along strike to the south from the Pechora Zone. Further to the north, in the subarctic Urals, a variety of granitic gneisses, granites, and metasediments occur, apparently as a southern continuation of the Bol'shezemel'skaya Zone. Isotopic ages are less reliable in these associations, but the presence of Mesoproterozoic metasediments and granites, and unconformably overlying Neoproterozoic siliciclastic (mainly quartzites) and carbonate formations, and acid volcanic rocks, are well documented. Early Ordovician conglomerates and quartzites overlie the Neoproterozoic successions with marked unconformity. Thus, it can be inferred that a Precambrian complex, differing in age and character from the Palaeoproterozoic and Archaean rocks of the EEC, occurs within the Timanide hinterland, outboard of the Pechora Zone calc-alkaline volcanic suites.

Polar Urals

North of the Arctic Circle in the Uralide Orogen, Neoproterozoic complexes comprise major components in the footwall to the Palaeozoic ocean-derived allochthons. In the core of foreland folds, beneath Ordovician quartzites, a fragmented ophiolite (Enganepe Terrane) has been dated by the U/Pb zircon method to 670 Ma, and is associated with arc

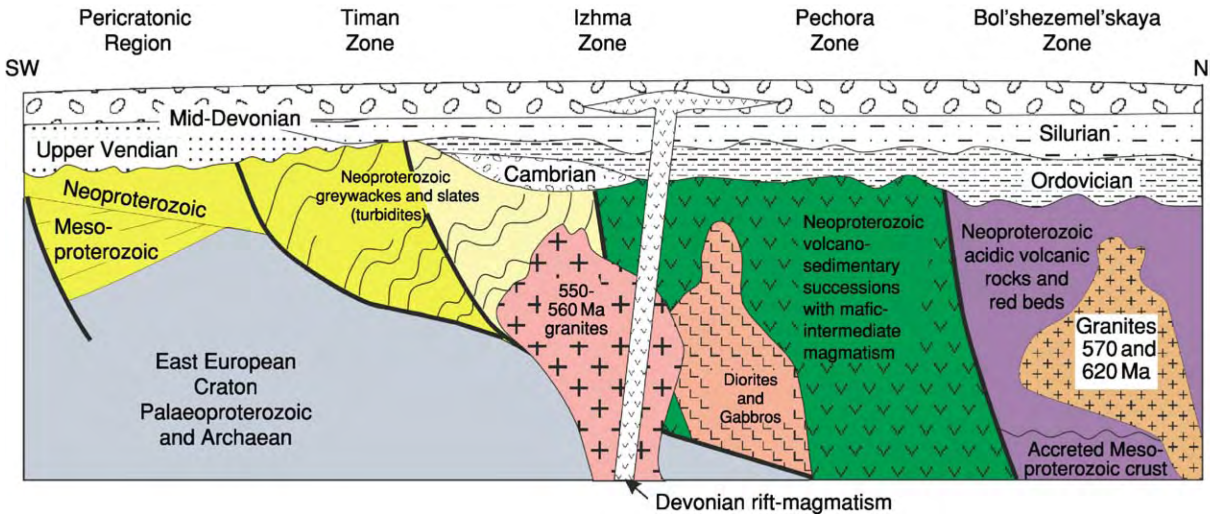


Figure 4 Diagramatic profile (SW NE) through the Timan Pechora region in the Devonian, prior to the Uralian Orogeny.

volcanites and a variety of volcaniclastic sedimentary rocks. Further east, in the Uralian thrust sheets, where the Palaeozoic metamorphic grade is higher, mafic island-arc related igneous complexes contain 580 Ma old gabbros, and, within the overlying Palaeozoic eclogite-bearing associations, similar ages characterise the igneous protoliths. Thus, in the most easterly exposed hinterland of the Timanide Orogen, the main Neoproterozoic component is derived from oceanic domains.

Novaya Zemlya

Northwards and north-eastwards from the Polar Urals, the Uralide Orogen continues via the Pai Khoi Peninsula, Vaigach Island, and the Kara Strait into Novaya Zemlya (Figure 1). Only the foreland folds and thrusts of the Uralide Orogen are preserved on land in Novaya Zemlya, the hinterland of the orogen being buried beneath the Mesozoic successions of the Kara Shelf.

As in the Polar Urals, major foreland anticlines and thrusts expose Neoproterozoic complexes beneath Lower Palaeozoic unconformities. On southern Novaya Zemlya and Vaigach Island, Lower and Middle Ordovician quartzites and limestones overlie greenschist facies metaturbidites. Further north, in the central parts of the archipelago (Figure 1), in the vicinity of the Mitushev Bay and Sulmenev Bay, Vendian (*ca* 600 Ma old) granites have been found, in the latter case intruded into schists, marbles, quartzites, and amphibolites. Whereas in the Mitushev Bay area, Silurian conglomerates rest unconformably on the granites, further north at Sulmenev Bay, both Ordovician and Cambrian strata overlie the unconformity, the oldest strata reaching back into the late Early Cambrian. Thus, in these northernmost hinterland areas, the unconformity may have been established in the Vendian. The amphibolite facies complex near Sulmenev Bay contains Mesoproterozoic detrital zircons; some authors infer that Precambrian orogeny there may have been older than Timanian (e.g., Grenville-age) and subsequently influenced by Timanian magmatism.

Barents Shelf

The Timanian rock units in the basement below the Pechora Basin strike north-westwards into the southern Barents Shelf and their associated geophysical (potential field) signatures can be followed towards the central parts of the Barents Sea, where they fade beneath the thickening Palaeozoic and Mesozoic successions of the Barents Shelf basins. Seismic studies, both near-vertical reflection and wide-angle refraction, have identified a thick wedge of Timanian metasediments along the southern margin of the Barents Sea towards the Kola Peninsula. Further out into the Barents Shelf, the identification of Timanian rocks is controversial and mainly based on the interpretation of seismic velocity data. Only on Franz Josef Land (Figure 1), is there clear evidence of a Timanide bedrock and even this is limited to a single deep drillhole located on the westernmost island, which penetrated a Lower Carboniferous unconformity to reach low greenschist facies, small folded Vendian meta-turbidites at 2–3 km depth.

Timanide Tectonic Evolution

Evidence of the Timanian Orogeny occurs over a wide area of north-eastern Europe. It provides a basis for reconstructing the Neoproterozoic to Early

Palaeozoic tectonic evolution of this part of Baltica; however, its fragmentary character and the extensive Phanerozoic cover of the Pechora and Barents Shelf basins, imply that these reconstructions are, at best, rudimentary. A major dilemma is the lack of evidence of the character of the Timanide hinterland to the east of the Urals – from the Palaeozoic ocean-derived allochthons and eastwards beneath the Mesozoic cover of the West Siberian Basin. In addition, it is probable that the Timanide Orogen extended both further south-eastwards, prior to the Uralian Orogeny, and north-westwards prior to the Caledonian Orogeny, but evidence is lacking.

There follows here a tectonic synthesis (Figure 5), based on the evidence presented above.

By the Early Neoproterozoic, after a long period (Mesoproterozoic) of intracontinental rifting, a passive margin had been established along the eastern edge of the EEC, at least from northern Norway to the Middle Urals. The change from an extensional to a compressional regime in the Vendian, with the onset of the Timanian Orogeny, is well defined in the foreland fold-and-thrust belt along this part of the EEC margin and extending southwards into the southern Urals. Deposition of a Late Vendian molasse facies in foreland basins along the length of the orogen provides evidence of the timing of hinterland uplift and foreland deformation.

The Timanian Ocean existing along the eastern margin (present-day coordinates) of the EEC in the

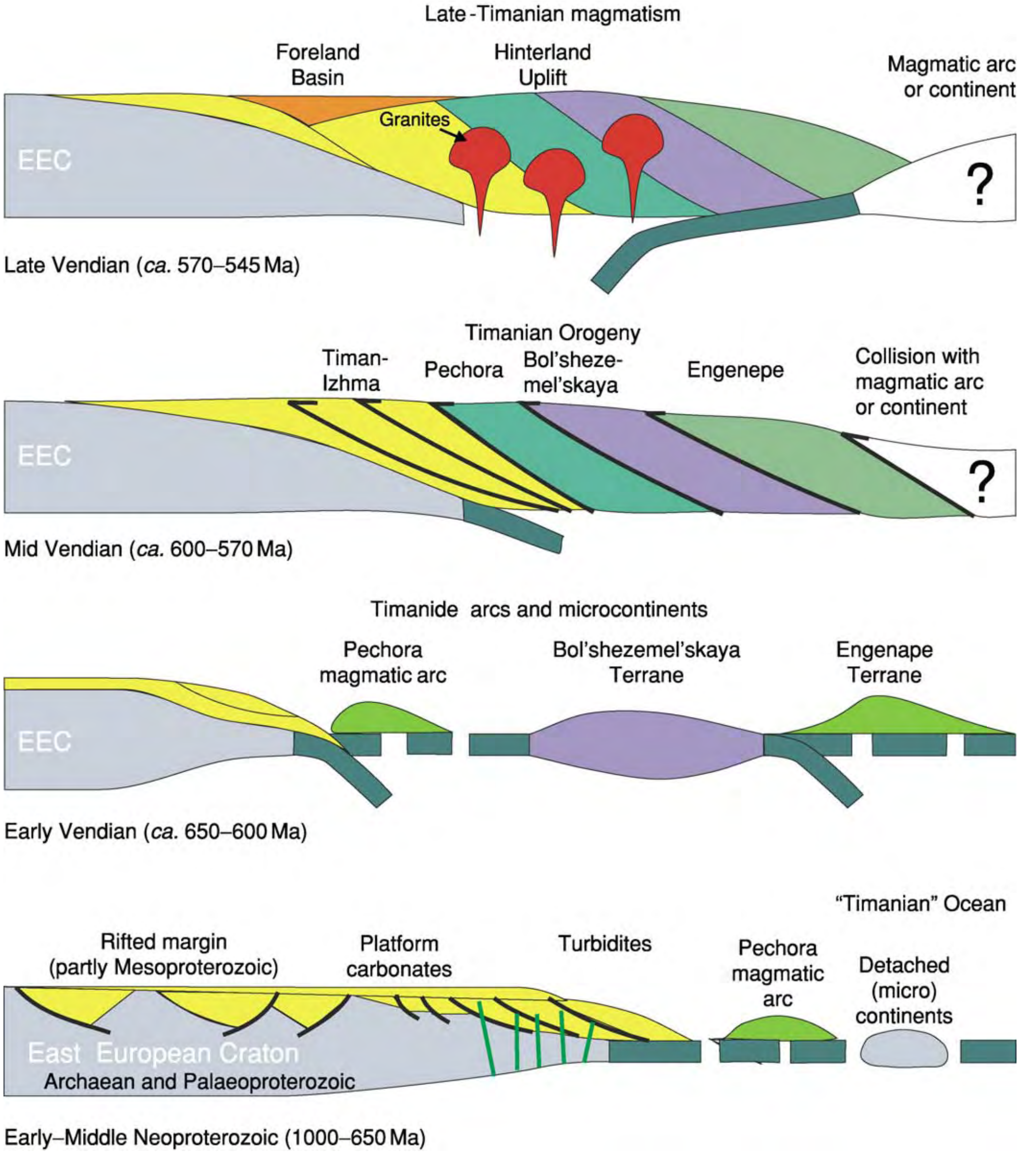


Figure 5 Timanian tectonic evolution, from the Early Neoproterozoic to the Late Vendian.

Neoproterozic, during deposition of the Timanian continental margin turbidites, development of the Pechora magmatic arc and Engane-Pe ophiolites, has been referred to in previous literature by several names, e.g., Palaeo-Asian, Turkistan, and Aegir, based on different interpretations of continent–ocean relationships. The latter are not well constrained and the informal term Timanian Ocean is preferred here.

Towards the hinterland, the dominance in the Pechora zone of mafic to intermediate calc-alkaline igneous suites and volcanogenic sedimentary rocks has been interpreted to imply thrusting of oceanic domains onto the thick passive margin turbidites of the EEC margin (Izhma Zone). The western contact of the Pechora Zone is not well defined, but the presence of Timanian blue schists along strike, both to the southeast, in the northern Urals (Kvarkush) and to the northwest, in the easternmost Kanin Peninsula, suggests that subduction-related arc magmatism culminated in thrust emplacement onto the EEC margin.

The existence of microcontinental terranes (blocks) in the Timanian Ocean has been widely inferred (indeed, some authors regard the Timanides as a number of NW-trending aulacogens with oceanic crust separating continental blocks). The microcontinental terranes differ from the EEC by having Grenville-age signatures with a Mesoproterozoic metasedimentary succession intruded by *ca.* 1000 Ma granites. The oceanic character of the adjacent Pechora Zone favours the interpretation that these outboard terranes (e.g., the Bol'shezemel'skaya Zone, and perhaps the Novaya Zemlya metamorphic rocks at Sulmenev Bay) were not accreted to the EEC until the Vendian.

To the east of the Bol'shezemel'skaya Zone, Late Neoproterozoic ophiolites in the front of the Polar Urals (Engane-Pe) and calc-alkaline intrusions in some of the Uralian allochthons further east, provide the only evidence of the oceanic character of the most internal parts of the Timanide hinterland. Taken together with the lack of evidence of the character of the Timanides east of the Urals, this has allowed the widely accepted hypothesis that Timanian Orogeny resulted from the subduction and accretion of Neoproterozoic ocean floor and island-arc assemblages, along with some fragments of continents (microcontinents), but without the involvement of a major Asian continent (e.g., Siberia). Continent–continent collision has not been favoured and the Neoproterozoic Timanian Ocean has been regarded as a forerunner of the Palaeozoic Uralian Ocean. This attractive hypothesis remains largely unconstrained.

Timanian orogenesis was contemporaneous with the passive margin evolution of north-western Baltica's Baltoscandian margin. It was also approximately contemporaneous with orogeny in many other parts of the world; for example, the Baikalian along the margin of the Siberian Craton, the Cadomian of western Europe and the Pan-African of many parts of Africa and the Middle East. Palaeomagnetic evidence has indicated that Baltica rotated about 120° counterclockwise between the Vendian and the Mid-Ordovician, prior to Caledonian collisional orogeny in the Silurian. This evidence has prompted a variety of palaeogeographic reconstructions relating the Timanides to these other orogens. It also implies that the rotation of Baltica may have accompanied the Timanian Orogeny, implying a significant component of dextral transpression during terrane accretion.

See Also

Europe: East European Craton; The Urals.

Further Reading

Belyakova LT and Stepanenko Vya (1991) Magmatism and geodynamics of the Baikalide Basement of the Pechora Syneclise. *Doklady Akademii nauk SSSR, (geologiya)* 106 117 (in Russian).

Bogatsky VI, Bogdanov NA, Kostyuchenko SL, Senin BV, Sobolev SF, and Khain VE (1996) *Tectonic map of the Barents Sea and the northern part of European Russia: explanatory notes*. Moscow: Institute of the Lithosphere, Russian Academy of Sciences Moscow.

Cocks LRM and Fortey RA (1998) The Lower Palaeozoic margins of Baltica. *Geologiska Foreningens Forhandlinger* 120: 173 179.

Cocks LRM and Torsvik TH (2002) Earth geography from 500 to 400 million years ago: a faunal and palaeomagnetic review. *Journal of the Geological Society, London* 159: 631 644.

Dushin VA (1997) Magmatism and geodynamics of the palaeocontinental sector of the northern part of the Urals. *Moscow Nedra*. pp. 211 (in Russian).

Gee DG and Pease VL (2005) The Neoproterozoic Timanide Orogen of Eastern Baltica. *Geological Society of London, Memoir.*

Gee DG, Belyakova LT, Pease V, Larionov AN, and Dovzhikova E (2000). New, single zircon (Pb evaporation) ages from Vendian intrusions in the basement beneath the Pechora Basin, northeastern Baltica. *Polarforschung* 68: 161 170.

Korago EA, Kovaleva GN, Ilin VF, and Pavlov LG (1992) *Tectonics and metallogeny of the early Kimmeridgian of Novaya Zemlya,* pp. 1 196. St Petersburg: Nedra (in Russian).

Lopatin BG, Pavlov LG, Orgo VV, and Shkarubo SI (2001) Tectonic structure of Novaya Zemlya. *Polarforschung* 69: 131 135.

Olovyanishnikov VG, Roberts D, and Siedlecka A (2000) Tectonics and sedimentation of the Meso to Neoproterozoic Timan Varanger Belt along the northeastern margin of Baltica. *Polarforschung* 68: 269 276.

Puchkov VN (1997) Structure and geodynamics of the Uralian orogen. *Geological Society of London Special Publications* 121: 201 236.

Roberts D and Siedlecka A (2002) Timanian orogenic deformation along the northeastern margin of Baltic, Northwest Russia and Northeast Norway, and Avalonian Cadomian connections. *Tectonophysics* 352: 169 184.

Siedlecka A (1975) Late Precambrian stratigraphy and structure of the northeastern margin of the Fennoscandian Shield (East Finnmark Timan region). *Norges geologiske undersokelse* 316: 313 348.

Torsvik TH and Rehnström EF (2001) Cambrian palaeomagnetic data from Baltica: implications for true polar wander and Cambrian palaeogeography. *Journal of the Geological Society, London* 158: 321 329.

Willner AP, Ermolaeva T, and Stroink L (2001) Contrasting provenence signals in Riphean and Vendian sandstones in the SW Urals (Russia): constraints for a change from passive to active continental margin conditions in the Neoproterozoic. *Precambrian Research* 110: 215 239.

Caledonides of Britain and Ireland

J F Dewey, University of California Davis, Davis, CA, USA, and University of Oxford, Oxford, UK
R A Strachan, University of Portsmouth, Portsmouth, UK

Introduction

The term 'Caledonian' is derived from the Latin word for Scotland and has been used widely with reference to the Early to Mid-Palaeozoic orogenic belt that is exposed in the British Isles, eastern Greenland, and Scandinavia (**Figure 1**). When these areas are restored to their relative positions prior to the Cretaceous–Tertiary opening of the North Atlantic (**Figure 1**), it can be seen that they were formerly continuous, comprising the northern part of the Mesozoic supercontinent Pangaea. The Caledonian Orogeny is the oldest of three main Palaeozoic collisional events that formed Pangaea, the other two being the Variscan and Uralian orogenies.

The Caledonian orogenic belt has three main arms (**Figure 1**): the North Atlantic Caledonian Belt and its southern extension, the Appalachian Belt, which were both formed by the closure of the Iapetus Ocean, and the Tornquist Belt, which resulted from the closure of the Tornquist Sea. These belts separate areas of Archaean and Proterozoic crust, which formed the palaeocontinents of Laurentia (North America, Greenland, and the northern British Isles), Baltica (Scandinavia and the Baltic), and Eastern Avalonia (the southern British Isles, Belgium, and the adjacent crust). Eastern Avalonia is the northernmost of a group of continental fragments that were originally derived from Gondwana and include Western Avalonia, Iberia, and Armorica. The collision of these fragments through the Mid- to Late Palaeozoic resulted in the Variscan orogenic belt (*see* **Europe:** Variscan Orogeny), which just impinges on the south of the British Isles.

Palaeogeographical and Tectonic Framework

The Caledonian plate-tectonic cycle began in the Mid- to Late Neoproterozoic with the rifting and break-up of the supercontinent Rodinia and the subsequent dispersion of continental blocks to form the Iapetus Ocean. Palaeomagnetic studies indicate that, in the Late Cambrian to Early Ordovician, Laurentia occupied an equatorial position to the north of Baltica and Avalonia (the latter was still attached to Gondwana at this stage). The distance across the Iapetus Ocean between the locations of present-day Scotland and England was probably about 5000 km, and the width of the Tornquist Sea, which separated Gondwana from Baltica, was probably about 1300 km. Each continent was characterized by a distinctive trilobite fauna. The initial closure of Iapetus was associated with the development of a southwards-dipping intraoceanic subduction zone and volcanic arc off the eastern Laurentian margin. The collision during the Early to Mid-Ordovician between the Laurentian margin and the volcanic arc resulted in the Grampian orogenic event (**Figure 2**). Subduction polarity then reversed to dip northwards beneath the Laurentian margin.

Eastern Avalonia rifted from Gondwana in the Arenig and drifted rapidly northwards between 480 Ma and 460 Ma to reach a similar latitude to Baltica during the Caradoc. A southwards-dipping subduction zone was located along the northern Iapetan margin of Eastern Avalonia. As on the Laurentian margin, a suprasubduction zone ophiolite–arc

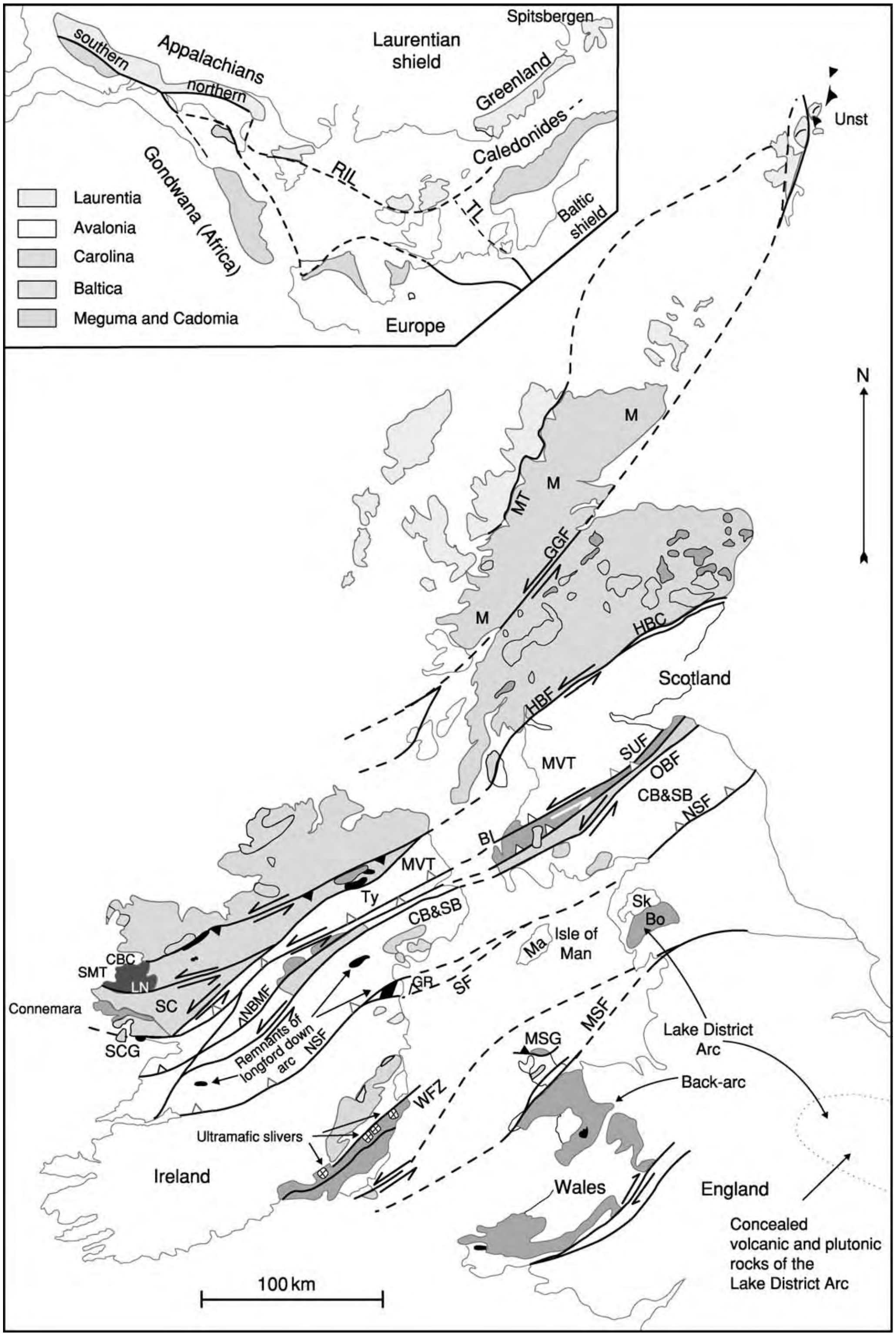


Figure 1 Simplified Caledonian tectonic map of the British Isles with their Appalachian continuation shown in the inset. BI, Ballantrae Ophiolite; BO, Borrowdale Volcanics; CB, Central Belts; CBC, Clew Bay Complex; GGF, Great Glen Fault; GR, Grangegeeth Terrane; HBC, Highland Border Complex; HBF, Highland Boundary Fault; LL, Leadhills Line; M, Moine; Ma, Manx Group; MSF, Menai Straits Fault; MSG, Monian Supergroup; MT, Moine Thrust; MVT, Midland Valley Terrane; OBF, Orlock Bridge Fault; RIL, Red Indian Line; SC, South Connemara Terrane; SCG, South Connemara Group; Sk, Skiddaw Group; SL, Slane Fault; SMT, South Mayo Trough; SUF, Southern Uplands Fault; TL, Tornquist Line; Ty, Tyrone Ophiolite; WFZ, Wicklow Fault Zone.

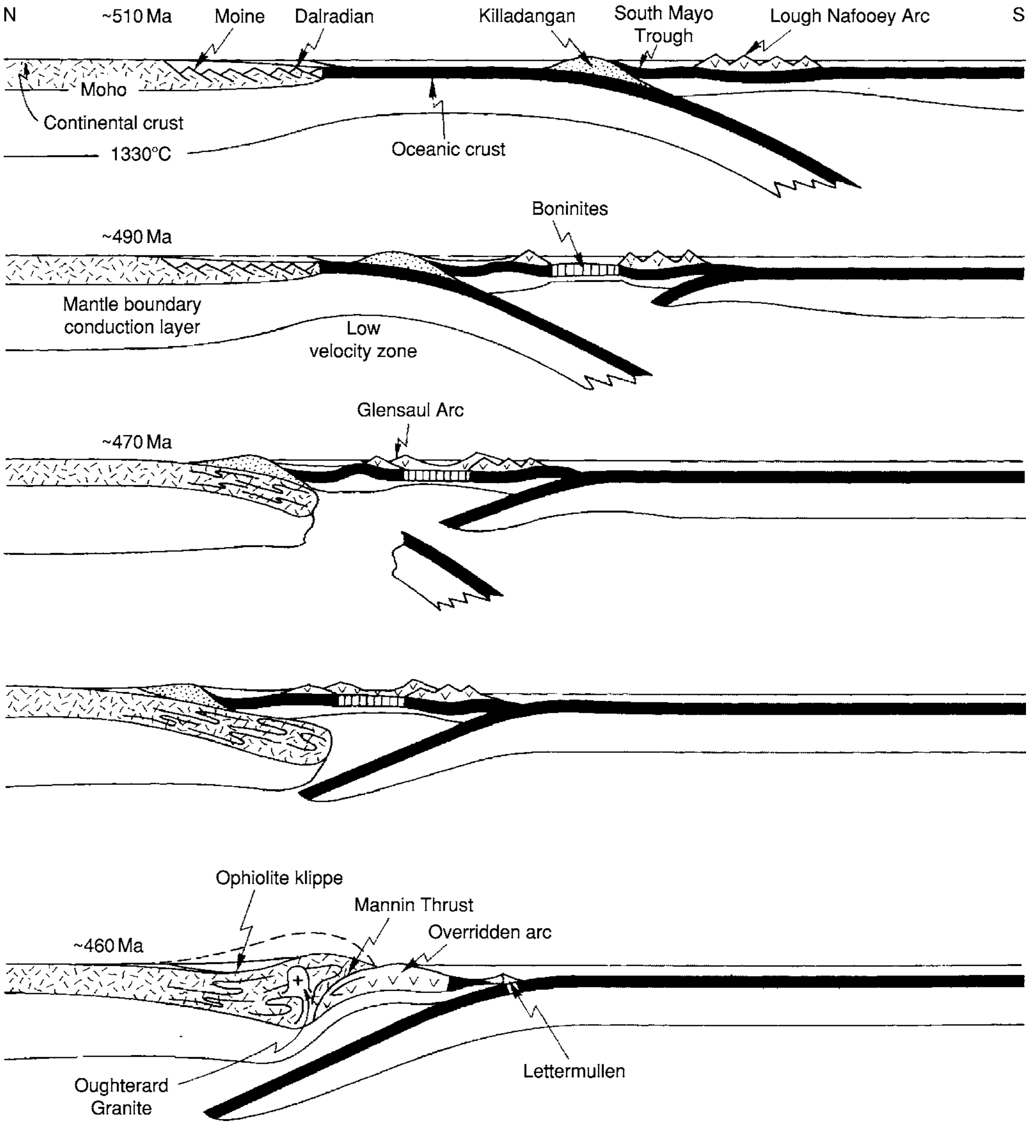


Figure 2 Schematic sections illustrating the Ordovician tectonic evolution of the Grampian Zone.

complex was obducted onto the Eastern Avalonian margin during that margin's collision with a northwards-dipping subduction zone, followed by polarity flip to southwards-directed subduction. The oceanic tract that opened between Eastern Avalonia and Gondwana is known as the Rheic Ocean. At 460 Ma, Eastern Avalonia was probably about 1000 km west of Baltica, and the width of the Iapetus Ocean was reduced to approximately 2000 km across the British sector. The intermediate position of Avalonia during its transit across the closing Iapetus Ocean is reflected in the changing affinities of its brachiopods, from Gondwanan to a mixture of genera with ancestors from Baltica and Laurentia, through the Ordovician. By Late Ordovician–Early Silurian times (440 Ma), the Iapetus Ocean and the Tornquist Sea had narrowed sufficiently to allow the interchange of the larvae of most benthic animals. Eastern Avalonia collided with Baltica during the Ashgill. By this time, the width of the Iapetus Ocean across the British sector had been reduced to around 1300 km.

Sinistrally oblique collision between the combined Avalonia–Baltica landmass and Laurentia occurred between 435 Ma and 425 Ma to form the continent Laurussia. The 'hard' collision of Baltica with the eastern Greenland–northern Scotland segment of Laurentia resulted in the Scandian orogenic event. In contrast, the 'soft' collision of Eastern Avalonia with Laurentia was not associated with regionally significant shortening, suggesting that Eastern Avalonia and Baltica may have been only loosely coupled during the final collision. The Iapetus Suture is referred to as the Solway Line, because it is apparently located in the Solway Firth, albeit obscured entirely by unconformably overlying Carboniferous strata.

Between 425 Ma and 410 Ma, sinistral relative motion between Laurentia and Baltica–Avalonia became orogen-parallel and was associated mainly with displacement along the Great Glen, Highland Boundary, and Southern Uplands faults. From about 410 Ma, the Caledonides went into regional sinistral transtension, resulting in the formation of a series of transtensional Old Red Sandstone basins. These events were contemporaneous with the rapid contraction of the Rheic Ocean as Gondwana approached Laurussia. During the Early to Mid-Devonian, Eastern Avalonia was affected by widespread Acadian deformation. Although this has traditionally been assigned to the Caledonian orogenic cycle and the closure of Iapetus, it postdates the latter by *ca.* 20 Ma. Alternatively, the Acadian event could be the result of the collision of Armorica with the southern margin of Avalonia. The collision of Armorica with Avalonia and the collision of Iberia with the southern margin of Armorica during the Mid-Devonian could be viewed as the early stages of the protracted Variscan Orogeny, which ultimately resulted in the assembly of Pangaea.

The along-strike continuation of the Caledonides in Newfoundland provides, arguably, the best-exposed section across the orogen, which can be used as a template to aid the understanding of other sectors such as the British Isles. The Caledonides of the British Isles are subdivided by major faults into a series of terranes, each with a coherent internal history. The correlation of geological units and events across terrane-bounding faults is commonly problematic. These structures display evidence of long and complex reactivation histories, although the last major displacement for most faults north of the Iapetus Suture was during Silurian to Devonian time. Although the detail of the tectonic template is extremely complex, there are three main groups of terranes: those north of the Fair Head–Clew Bay Line and Highland Boundary Fault, which have Laurentian affinities; those south of the Solway Line, which have Gondwanan affinities; and the intermediate terranes, which represent fragments of continental margin, island arc, and oceanic lithosphere within a composite complex suture zone. Seismic imaging indicates that the Iapetus Suture dips moderately to the north, separating reflective Avalonian crust from overthrust non-reflective Laurentian crust.

Terranes and their Geotectonic Affinities

From north to south, the main terranes are characterized as follows.

Hebridean Terrane

In north-west Scotland, the Hebridean Terrane forms the foreland to the Caledonian belt and is limited to the east by the Moine Thrust Zone. The oldest component is the Archaean–Palaeoproterozoic Lewisian Gneiss Complex, which is correlated with Laurentian basement in south-east Greenland. The unconformably overlying continental Torridonian sedimentary rocks were deposited from 1200–1000 Ma and comprise two successions separated by an angular unconformity. The Lewisian and Torridonian rocks are overstepped by a highly distinctive succession of Cambrian–Ordovician shallow-marine quartzites and limestones, which accumulated on the Laurentian margin of Iapetus and correlate closely with similar sequences in north-western Newfoundland and eastern Greenland.

Northern Highland Terrane

The Northern Highland Terrane is bounded in north-western Scotland by the Moine Thrust Zone and the Great Glen Fault. The terrane is dominated by the Early Neoproterozoic Moine Supergroup, a thick succession of strongly deformed and metamorphosed marine sediments that were deposited in an intracratonic rift within the Laurentian sector of the Rodinia Supercontinent at 1000–870 Ma. Early deformation and metamorphism during a Knoydartian tectonothermal event at 800 Ma may have resulted from rift closure. Inliers of Archaean orthogneisses, variably reworked at 1000 Ma during the Grenville orogeny, represent the basement on which the Moine sediments were unconformably deposited.

Grampian Terrane

In Scotland and north-western Ireland, the Grampian Terrane occurs between the Great Glen Fault and the Fair Head–Clew Bay Line or Highland Boundary Fault. The terrane is dominated by the Dalradian Supergroup, a thick succession of variably deformed and metamorphosed metasandstones, metasiltstones, and metalimestones with minor contemporaneous mafic igneous rocks. The duration of sedimentation is uncertain, but it probably spanned at least the period 750–520 Ma. Tillites have been correlated with the widespread Sturtian and Varangerian glacial events. The Dalradian succession shows a transition from shallow-water to deep-water sedimentation, reflecting the progressive thinning of the Laurentian crust during rifting of the Rodinia Supercontinent and the evolution of a passive continental margin adjacent to the developing Iapetus Ocean. Inliers of Palaeoproterozoic gneisses in the Inner Hebrides and north-western Ireland can be linked with similar units in south-eastern Greenland and probably represent the Laurentian basement on which the Dalradian sediments were unconformably deposited. Shear zone-bounded inliers of Early Neoproterozoic Moine-like rocks occur in the northern Grampian Highlands and may provide linkage across the Great Glen Fault.

Midland Valley Terrane

The Midland Valley Terrane is bounded to the north by the Fair Head–Clew Bay Line and the Highland Boundary Fault and to the south mainly by the Southern Upland Fault (in the west of Ireland it lies north of the Doon Rock Fault). Much of the pre-Late Palaeozoic geology, particularly in the Scottish Midland Valley, is obscured by Devonian and Carboniferous cover. However, the exposed Early Palaeozoic rocks have oceanic affinities. Early Ordovician ophiolites occur on Shetland, along the Highland Boundary Fault, and at Ballantrae, Tyrone, and Clew Bay. In western Ireland, remnants of an accretionary prism are present in Clew Bay and on Achill Island, locally associated with blueschists. Immediately to the south are the fore-arc Ordovician sediments of the South Mayo Trough and the mafic to intermediate calc-alkaline volcanics of the Lough Nafooey Group. Whether the unexposed continental basement that underlies the terrane has Laurentian affinities or is entirely exotic is unknown.

Connemara Terrane

The Connemara Terrane is located in western Ireland between the Doon Rock and Skirds Rock faults. It consists of metasedimentary rocks that have been correlated with the Dalradian Supergroup, and thus the terrane is interpreted as a tectonic slice of the Laurentian margin that became detached and interleaved with units of the Midland Valley Terrane during strike-slip displacements.

Southern Uplands Terrane

Located between the Southern Uplands Fault and the Solway Line, the Southern Uplands Terrane is characterized by thick sequences of Ordovician and Silurian sediments, which were deposited in deep-marine environments and subjected to complex deformation and low-grade metamorphism. The terrane itself comprises three major units bounded by strike-slip faults: the Northern, Central, and Southern belts. The Northern Belt comprises Ordovician sediments and rare volcanics; the Central Belt is formed of Ordovician and Silurian sediments, and the Southern Belt is entirely Silurian in age. Although opinions vary as to the tectonic setting of sedimentation, the balance of evidence suggests that the sediments represent an accretionary prism developed above the northwards-subducting Iapetus Ocean. In eastern Ireland, the Iapetus suture is drawn between the Ordovician inliers of the Grangegeeth and Bellewstown terranes, which show Laurentian and Avalonian affinities, respectively, in their brachiopod and trilobite faunas.

Lake District–Leinster Terrane

The Lake District Terrane is situated between the Solway Line and an unexposed boundary thought to lie to the north-west of Anglesey. It contains Cambrian–Arenig clastic rocks, mid- to Late Ordovician basic and acid volcanic and volcaniclastic rocks, such as the Borrowdale Volcanic Group of the English Lake District, and the unconformably overlying Silurian sediments of the Windermere Supergroup of northern England. The Lake District represents the northern edge of Eastern Avalonia, and the volcanic rocks are thought to have originated in a volcanic arc created by southwards-directed subduction on the Gondwanan side of the Iapetus Ocean.

Monian Terrane

The Monian Terrane is bounded to the south-east by the Menai Strait Fault System and comprises the Mona Complex of Anglesey and the Llyn Peninsula in north-western Wales and the Rosslare Complex of south-eastern Ireland. The terrane is highly segmented by shear zones and brittle faults. Proven Precambrian rocks include high-grade metasedimentary gneisses, a calc-alkaline granite pluton, and a belt of blueschist facies metamorphic rocks, which may represent an accretionary prism. Thick low-grade metasedimentary sequences, including a major melange unit, have been viewed as Precambrian but may be Cambrian, albeit very different from the Cambrian rocks of the adjacent Welsh Basin Terrane. The Monian Terrane may be an Allochthonous Gondwanan Terrane that was amalgamated with Eastern Avalonia by the Early Cambrian at the latest, after which it formed part of the southern foreland of the orogen.

Welsh Basin Terrane

Bounded to the south-east by the Welsh Borderland Fault System, the Welsh Basin Terrane comprises Cambrian–Silurian sedimentary successions, largely deposited in deep-water environments, and thick sequences of Ordovician tholeiitic to calc-alkaline volcanics. Volcanism and sedimentation probably occurred in a back-arc basin floored by extended Avalonian basement, behind the Lake District–Leinster volcanic arc. A range of Ordovician calc-alkaline volcanic and plutonic rocks underlie the Mesozoic and younger cover of eastern England and may represent continuations of the Welsh Basin and/or Lake District volcanic belts.

Midlands Terrane

The Midlands Terrane is a triangular terrane that represents the southern foreland of Eastern Avalonia during the Caledonide Orogen and acted as a rigid indentor during the Acadian event. Late Neoproterozoic calc-alkaline igneous rocks and associated low-grade metasedimentary rocks, which developed along an active plate margin of Gondwana, are overlain unconformably by relatively thin sequences of Cambrian–Silurian shallow-marine platform sediments.

Tectonic Evolution of the Caledonides

The tectonic evolution of the Caledonides began in the Early to mid-Ordovician, with the commencement of closure of the Iapetus Ocean, and finished with the Early Devonian Acadian event. The main stages are as follows.

Early to mid-Ordovician Rifting of Eastern Avalonia and Arc–Continent Collision on the Laurentian Margin

Eastern Avalonia rifted away from Gondwana in the Early Ordovician (*ca*. 475 Ma), initiating closure of the Iapetus Ocean and opening, to the south, of the Rheic Ocean. A major southwards-dipping (in the present reference frame) subduction zone developed beneath the leading northern margin of Eastern Avalonia, and arc volcanism began in the Welsh Basin in the Early Ordovician. Sedimentary sequences deposited on Eastern Avalonia vary from shallow-marine on upstanding blocks such as the Midlands terrane and the Irish Sea landmass (effectively the Monian terrane) to deeper-marine sequences in the Welsh Basin.

By the Late Cambrian–Early Ordovician, oceanwards-dipping subduction zones had developed along the margins of Laurentia and Baltica, to form intraoceanic volcanic arcs. Localized orogenic events reflect the collision of these volcanic arcs with the continental margins. These include the Finnmarkian event (505 Ma) along the Baltica margin and the Grampian event (470–460 Ma), which affected the Northern Highlands and Grampian terranes. In western Ireland, the Grampian Arc is represented by the Lough Nafooey Volcanic Group, and the existence of a similar arc beneath the Late Palaeozoic cover of the Scottish Midland Valley is indicated by the presence within the Ordovician sediments of clasts of contemporaneous calc-alkaline plutonic rocks derived from a proximal source. By analogy with a similar-aged event in Newfoundland (the Taconic event), collision of the Laurentian margin with the arc at the subduction zone was associated with northward obduction of a major ophiolite nappe onto the Laurentian margin. Ophiolitic remnants of this nappe lie along the north-western margin of the Midland Valley Terrane, which effectively represents the Grampian Suture; the Ballantrae Ophiolite was probably obducted at about the same time. Within the Grampian Terrane, ophiolite obduction resulted in regional deformation and Barrovian metamorphism at greenschist–amphibolite facies of the Dalradian Supergroup. In north-eastern Scotland and Connemara, this was accompanied by the emplacement of major gabbros and granites. Late-stage backfolding and underthrusting of the Dalradian Supergroup formed south-eastwards-directed regional-scale fold nappes such as the Tay Nappe in Scotland and the Ballybofey Nappe in north-western Ireland. Within the Northern Highland Terrane, similar-aged folding and high-grade metamorphism of the eastern parts of the Moine Supergroup is assigned to the Grampian event. Following arc–continent collision, the polarity of subduction reversed to northward.

Mid-Ordovician–Silurian: Collision of Eastern Avalonia, Baltica, and Laurentia

On the Avalonian side of Iapetus, south-eastwards-directed subduction beneath Avalonia continued into the Late Ordovician. A major calc-alkaline volcanic arc lay through south-east Ireland and the Lake District, with the Welsh Basin being a volcanically active back-arc basin. Volcanism and subduction ceased during the Late Ordovician, probably because the Avalonian margin overran the ridge system in the Iapetus Ocean. Throughout the Late Ordovician and Silurian, the stable Midland Platform and Irish Sea landmass hosted shallow-marine or emergent conditions surrounded by deeper-marine basins, which were often filled with turbidites. Eastern Avalonia collided with Baltica during the Late Ordovician (440 Ma) as shown by palaeomagnetic data and evidence of progressive faunal integration. South-western Baltica records evidence of a Late Ordovician thermal event, contemporaneous with open folding and uplift in the Welsh Borderlands (the Shelveian event). This event is also synchronous with low-grade metamorphism in the North Sea, which was probably also related to the collision between Baltica and Eastern Avalonia.

On the northern side of Iapetus, the uplands created by the Grampian Orogeny remained emergent through the Late Ordovician and Silurian. In the Midland Valley Terrane, the former volcanic arc was mostly buried under a blanket of marine and nonmarine sediments, most of which were sourced from the emergent Grampian Highlands to the north. Sedimentation occurred in a series of localized sinistrally transtensive basins. Outboard, to the south-east, the Southern Uplands Terrane is dominated by thick fault-bounded sequences of complexly deformed trench sediments, which define an accretionary prism. The dominant folds, cleavages, and associated detachments within these sediments formed progressively as an integral part of the development of the accretionary prism, rather than as the result of a distinct collisional event. A switch from early approximately orthogonal shortening to sinistral transpression within the accretionary prism occurred in the Early Silurian at 430 Ma.

The earliest linkage of Laurentian-derived and Gondwanan-derived crustal blocks occurred in Newfoundland in the mid-Ordovician, at approximately 450 Ma. However, the earliest closure events in the British Isles appear to be Silurian (*ca.* 425 Ma). With progressive collision of the two margins of Iapetus through the Silurian, deep-water basins tended to shallow whilst shallow-marine basins became nonmarine or emergent. In the Lake District, the Silurian Windermere Supergroup was deposited in a flexural foreland basin that resulted from overthrusting of the Southern Uplands Terrane onto the leading edge of Eastern Avalonia during the terminal stages of collision. The oblique collision between Eastern Avalonia and Laurentia along the Solway Line, which resulted in the final eradication of the Iapetus Ocean, was complete by the end of the Silurian. The collision here was relatively ‘soft’ and did not result in major folding or cleavage development in the immediately adjacent terranes. In western Ireland, the collision was ‘harder’ and was associated with sinistral transpression, which deformed the Ordovician and Silurian strata of the South Mayo Trough.

Silurian: Oblique Collision of Laurentia and Baltica, and Closure of Northern Iapetus

The Northern Highlands Terrane was probably located 700 km along strike and to the north-east of the Grampian Terrane when this part of Laurentia collided obliquely with Baltica at 430–425 Ma to result in the Scandian orogenic event. This was associated with regional nappe stacking in Scandinavia and eastern Greenland, with intense folding, ductile thrusting, and metamorphism of the Moine rocks in the Northern Highlands Terrane, and with amalgamation of the Northern Highlands and the Hebridean terranes along the Moine Thrust. A total north-west–south-east shortening displacement across the Moine Thrust of at least 150 km seems likely, and an equivalent total amount of shortening may also have occurred along internal ductile thrusts within the Moine. Analysis of Scandian structures in Scandinavia, eastern Greenland, and north-west Scotland is consistent with overall sinistral transpression, which became progressively partitioned into orogen-orthogonal components and orogen-parallel left-lateral strike slip.

Regional-scale nappe stacking was therefore followed by major sinistral strike-slip displacements from 425 Ma onwards along the Great Glen, Highland Boundary, and Southern Uplands faults. The magnitude of the displacements is uncertain, although movement of 700 km along the Great Glen Fault and minimum movements of 200 km along both the Highland Boundary Fault and the Southern Uplands Fault seem likely. Strike-slip faulting was largely synchronous with the emplacement of the Newer Granite Suite, a series of calc-alkaline mainly I-type plutons probably derived from the melting of lithospheric mantle and lower-crustal sources above the north-westwards-dipping subduction zone. Major plutons crop out extensively in the Grampian Highlands and in north-western Ireland, and eruption of coeval lavas and volcanics occurred in the area between Glencoe and Lorne in the south-west Grampian Highlands. Plutons were mainly emplaced in transtensional pull-aparts along relatively minor faults that were splays of, or Reidel shears related to, the major displacement faults.

Early to mid-Devonian: the Final Caledonian (Acadian) Collision – Closure of the Rheic Ocean?

Sinistral relative displacement of the terranes north of the Solway Line and subduction-related plutonism and volcanism continued into the Early Devonian. By this time, the marine basins on both margins had been largely uplifted following crustal shortening. In the Scottish Highlands, Early Devonian continental sediments were deposited unconformably on deeply eroded Moine and Dalradian rocks. The Early Devonian sediments of the Moray Firth–Orkney–Shetland area and eastern Greenland probably accumulated in transtensional basins, reflecting a change from orogen-parallel displacements to sinistrally oblique divergence between Laurentia and Baltica–Eastern Avalonia between 410 Ma and 395 Ma.

A similar transtensional setting may also apply to the Early Devonian basins of the Midland Valley and the Lake District. Fluvial and alluvial sedimentation also occurred on the Irish Sea landmass and in the Anglo-Welsh Basin, with major rivers sourced from uplifted segments of the Caledonides to the north and north-east.

In southern Britain, the Cornubian basins record south-deepening Devonian marine sedimentation on the southern rifted passive margin of Avalonia, which faced Armorica across an already very narrow arm of the Rheic Ocean. East–west trending Early Devonian growth faults have been identified in South Wales and North Devon. In southern Cornwall, Devonian sediments are characterized by thick sequences of distal turbidites, which were deposited either on highly thinned continental crust at the edge of the shelf or on oceanic crust. The ultramafic and associated rocks of the Lizard Complex in Cornwall have been interpreted as an ophiolite derived from the Rheic Ocean basin, which is inferred to have existed to the south.

During the Early to Mid-Devonian, widespread Acadian deformation and low-grade metamorphism occurred south of the Solway Line, forming the slate belts of the Lake District, south-eastern Ireland, and north-western Wales. In contrast to the area north of the Solway Line, where the Caledonian structural grain is mainly north-east–south-west, the trend of folds and cleavage in Avalonia is much more variable, showing a distinctly arcuate form. This structural pattern is apparently strongly controlled by the triangular shape of the old rigid continental block that underlies the Midlands terrane, which did not experience Caledonian deformation. The Welsh Borderlands Fault System broadly defines the south-eastern limit of Acadian deformation. There is a common tendency in the slate belts for the steeply dipping Acadian cleavage to cut gently across associated folds in a clockwise-deflected sense, indicating that the collision direction was sinistrally oblique. Although the Acadian phase of the Caledonian Orogeny has been ascribed to the closure of Iapetus, it is not clear why the shortening and uplift associated with this event should occur as late as the Early to Mid-Devonian. It is unlikely that forceful northward collision of Avalonia continued for a further 20 Ma after the last remnant of Iapetan crust was subducted, especially as there is growing evidence for sedimentation in sinistrally transtensive basins at this time, both south and north of the suture. Another possibility is that the Acadian deformation resulted from the collision of Armorica with Avalonia in the Early Devonian. This is supported by isotopic data that indicate that initial exhumation and northward-directed thrusting of the Lizard Complex in southern Cornwall occurred between 400 Ma and 380 Ma, overlapping in time with cleavage development in the slate belts to the north.

See Also

Europe: Scandinavian Caledonides (with Greenland); Variscan Orogeny; The Urals. **North America:** Northern Appalachians; Southern and Central Appalachians. **Palaeozoic:** Ordovician; Silurian; Devonian. **Pangaea**.

Further Reading

Dewey JF and Mange MA (1999) Petrology of Ordovician and Silurian sediments in the western Irish Caledonides: tracers of short lived Ordovician continent arc collision orogeny and the evolution of the Laurentian Appalachian Caledonian margin. In: MacNiocaill C and Ryan PD (eds.) *Continental Tectonics:* Geological Society, London, Special Publication 164, pp. 55 108.

Dewey JF and Shackleton RM (1984) A model for the evolution of the Grampian tract in the early Caledonides and Appalachians. *Nature* 312: 115 121.

Draut AE and Clift PD (2001) Geochemical evolution of arc magmatism during arc continent collision, South Mayo, Ireland. *Geology* 29: 543 546.

Soper NJ, Ryan PD, and Dewey JF (1999) Age of the Grampian Orogeny in Scotland and Ireland. *Journal of the Geological Society, London* 156: 1231 1236.

Van Staal CR, Dewey JF, MacNiocaill C, and McKerrow WS (1998) The Cambrian Silurian tectonic evolution of the northern Appalachians and British Caledonides: history of a complex, west and southwest Pacific type segment of lapetus. In: Blundell DJ and Scott AC (eds.) Lyell: the past is the key to the present: Geological Society [London] Special Publication, 143, pp. 199 242.

Scandinavian Caledonides (with Greenland)

D G Gee, University of Uppsala, Uppsala, Sweden

Introduction

The Caledonide Orogen of the North Atlantic region reaches northwards from type areas in the British Isles, along the eastern edge of Greenland and western Scandinavia, to the Barents Shelf and the Svalbard Archipelago. Prior to the opening of the Norwegian and Greenland seas and the Eurasian Basin in the Tertiary (Figure 1), this *ca*. 3000 km long segment of the orogen was about 1000 km wide. A substantial part of this width, perhaps as much as 30–40%, was the result of a long period of post-orogenic extension, lasting from the Late Palaeozoic into and through the Mesozoic, which was accompanied by the deposition of thick sedimentary successions; these now compose the continental shelves and host the main oil and gas resources of northern Europe. Thus, at the time of orogeny in the mid-Palaeozoic, this northern part of the Caledonides was a long and relatively narrow (*ca*. 600–700 km wide) mountain belt, similar in dimensions and majesty to today's Himalayas, separating the low-lying old cratons of eastern Europe and Greenland.

The Caledonian Orogeny, referred to as the Scandian Orogeny in these northern regions, resulted from the collision of two continents, Baltica and Laurentia. The former was much smaller than the latter and, during collision, played a similar role to that of India in the present-day Himalayan context. The Scandian Orogeny began in the Silurian and extended into and through the Early Devonian; the name is derived from the Scandes, the mountains of Norway and western Sweden. However, it is worth remembering that today's mountains, along the coasts of Scandinavia and eastern Greenland, are the result of Tertiary uplift during the opening of the Greenland and Norwegian seas; they are not the relics of Palaeozoic mountains, although they are dominated by Caledonian rocks.

Before the collision of Baltica and Laurentia and the Scandian Orogeny, these two continents were separated by the Iapetus Ocean. The closure of the Iapetus Ocean occurred over a period of about 80 Ma and involved the development of subduction systems along the margins of both Laurentia and Baltica and a wide range of tectonothermal activity. This complex situation, involving magmatism and sedimentation, deformation and metamorphism, was an essential part of the Early Caledonian evolution, prior to the final collision of the continents and a Devonian change in global stress regimes: compression and lateral shortening gave way to regional extension.

The tectonic evolution of the northern part of the Caledonide Orogen is discussed below, after a presentation of each of the three major regions of development – western Scandinavia, eastern Greenland, and the Barents Shelf.

Western Scandinavia

The Scandian mountains, with many peaks reaching a little over 2000 m, extend for nearly 2000 km along the length of Norway; they include substantial regions of western Sweden and the westernmost highest parts of Finland. The Caledonide Orogen, on land, is up to 300 km wide and extends off the Norwegian coast for a further 200–300 km beneath the shallow shelf areas of the Norwegian Sea. The orogen (Figure 2) is dominated by thrust sheets transported from west-north-west to east-south-east onto the Palaeozoic platform successions of the Baltoscandian margin of Baltica. The front of the orogen is generally marked by a prominent thrust scarp, clearly indicating that these allochthonous (i.e. transported) rocks originally, in the Devonian, extended much further eastwards, perhaps as far as 100 km, onto the platform.

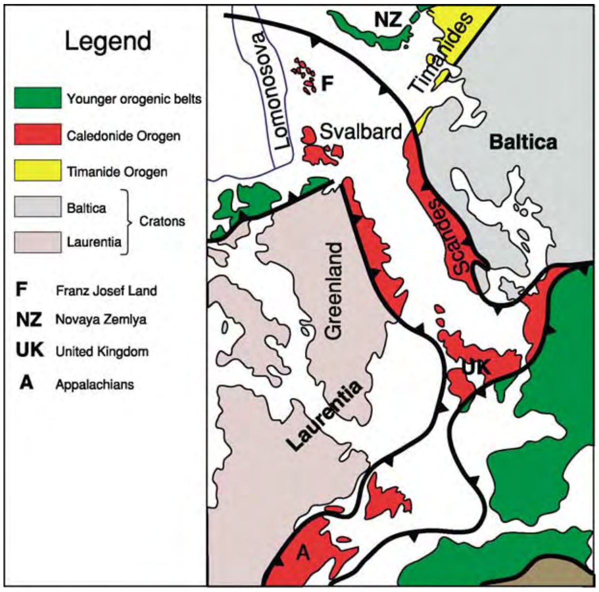


Figure 1 The North Atlantic Caledonides, from eastern Canada to the high Arctic Barents Shelf, in the Late Mesozoic.

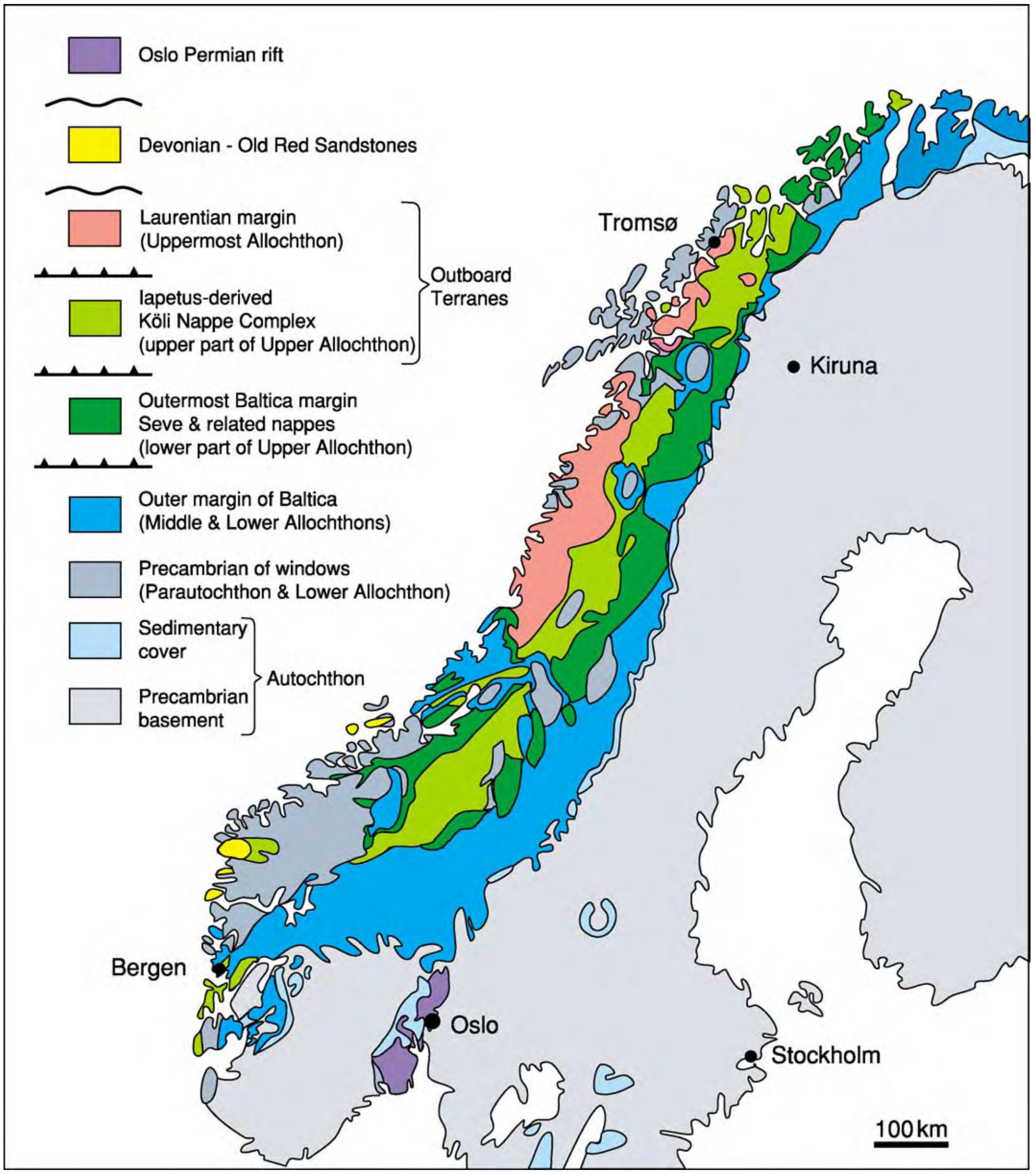


Figure 2 The Scandinavian Caledonides.

The Caledonide Orogen of Scandinavia is one of the world's classic areas for demonstrating the importance of very large scale near-horizontal thrusting of rocks during orogeny. Nearly 120 years ago, Alfred E Törnebohm demonstrated that the lateral transport of major allochthons from Norway into Sweden must have exceeded 100 km; subsequent work has shown that the total displacements amount to many hundreds of kilometres.

Along the eastern front of the orogen, the lowermost Scandian thrust sheets rest on thin (usually a few tens of metres) Cambrian sandstones and shales; locally, these autochthonous successions include underlying Vendian tillites and, in some central and southern areas, Ordovician and Silurian limestones and shales may also be present. These mainly Palaeozoic cover sediments were deposited on the deeply eroded and peneplained surface of the Fennoscandian Shield, which forms the basement along the entire Caledonian front. The shield is dominated by crystalline rocks, ranging in age from Late Mesoproterozoic (*ca*. 1000 Ma) in the south to Palaeoproterozoic and Archaean in the far north. The contrast in physical properties between these ancient granites and gneisses and the overlying sedimentary cover, especially the Cambrian alum shales, played an important role in controlling the character and geometry of the Scandian Orogeny. The alum shales are highly organic-rich (locally over 10%) and even today sometimes smell of bitumen. During orogeny, these beds, up to a few tens of metres thick, provided a well-lubricated surface along the entire length of the mountain belt, over which the allochthons could easily be transported. These Cambrian shales are also characterized by unusually high levels of particular trace elements, such as uranium, vanadium, and molybdenum, allowing them to be identified further west, towards the hinterland of the orogen, where they are more metamorphosed and lack fossils.

The Scandian thrust sheets contain rocks that are derived both from the Baltoscandian margin (basement and cover) and from a variety of terranes that existed in the Early–mid-Palaeozoic outboard of Baltica. The latter are largely composed of ocean-floor and island-arc assemblages that formed in the Iapetus Ocean and were thrust onto the Baltoscandian margin during the Scandian Orogeny. Identification and correlation of the different allochthons across the orogen from the foreland, where they are relatively well preserved, to the hinterland, where they are generally highly attenuated and metamorphosed, has been possible only because all the thrust sheets are influenced by major late-orogenic folding with orogen-parallel axes. Within the cores of the anticlines, in tectonic windows, basement–cover complexes and overlying thrust sheets have been mapped, and allochthons have been correlated both along and across the orogen.

The many Scandian thrust sheets can be subdivided into four major groups of allochthons – Lower, Middle, Upper, and Uppermost. Where the rocks below the Lower Allochthon are disturbed but still comparable with the autochthon, the term parautochthon has been widely applied. The Lower and Middle Allochthons are composed of rocks derived from typical Baltoscandian-margin sedimentary-cover successions and their underlying crystalline basement. Mafic dykes intrude the uppermost units of the Middle Allochthon (Särv Nappes). The Upper Allochthon is more varied and is readily divisible into two parts: a lower group of thrust sheets (Seve Nappe Complex) that are compositionally similar to some of the underlying sedimentary cover but are much more highly metamorphosed, and an upper group (Köli Nappe Complex) of lower metamorphic grade that includes a wide range of igneous and sedimentary rocks derived from oceanic environments and including ophiolites and island-arc and back-arc assemblages. The Uppermost Allochthon has also proved to be complex, containing both continental-margin lithologies and ophiolites, intruded by major granite batholiths; these highest nappes were derived from the margin of Laurentia.

Baltoscandian Platform to Outer Margin (Lower and Middle Allochthons)

The Baltoscandian-platform successions of the Autochthon generally dip 1–2° westwards and reappear towards the hinterland in the antiformal windows. The overlying thrust sheets of the Lower Allochthon, riding on a décollement surface above the Cambrian alum shales, are dominated by a Cambro-Silurian stratigraphy that can be readily correlated with that in the Autochthon, but in which the formations are generally thicker. Prominent facies changes occur in the Ordovician, with the platform carbonates of the Autochthon giving way westwards to basinal shales and westerly derived turbidites. A return to shallow-marine environments characterizes the Late Ordovician and Early Silurian, before Late Llandovery deepening and the influx of Mid-Silurian turbidites (flysch). In southern Norway, in the Oslo graben and, locally, in the Swedish Caledonides, Early Silurian shallow-marine limestones and deeper-water turbidites shallow upwards in the Wenlock to a nonmarine siliciclastic Old Red Sandstone facies indicative of the development of a foreland molasse-filled basin. This basin is inferred to have existed along the entire Caledonian front, prior to Late Palaeozoic and Mesozoic erosion of the mountain belt.

In foreland parts of the orogen, the Lower Allochthon is dominated by low-grade sedimentary successions, with only subordinate slices of Precambrian crystalline rocks. Further west, towards the hinterland, the thrusts cut deeper into the basement. The overlying Middle Allochthon contains extensive sheets of highly mylonitized Precambrian granites and gneisses, along with metasedimentary successions that are generally of Neoproterozoic age. Prominent mylonites separate the Middle Allochthon from the Lower Allochthon, and the metamorphic grade (low greenschist facies) is generally somewhat higher in the former. Precambrian complexes (Jotun Nappe) of the Middle Allochthon comprise the highest mountains in Scandinavia.

Included in the upper part of the Middle Allochthon are thrust sheets that have a stratigraphy similar to that in the underlying nappes, including thick Neoproterozoic siliciclastic and carbonate successions, overlain by Vendian tillites and then sandstones. However, these allochthons (Särv Nappes) are remarkable because of the widespread occurrence of *ca.* 600 Ma old mafic dyke swarms, which are often composite and even occur as sheeted complexes. Igneous rocks of Neoproterozoic or Palaeozoic age are notably absent from all the underlying tectonic units in the Scandes, except where they are exposed in the deep hinterland along the Norwegian west coast.

Thus, it can be inferred that these allochthons were originally located off the coast of western Norway and, along with the overlying nappes, have been transported at least 300 km eastwards onto the Baltoscandian Platform.

Outermost Margin of Baltica (lower part of the Upper Allochthon)

The Särv Nappes are overthrust by a higher grade allochthon, which is itself dominated by psammites, pelites, and subordinate marbles, with numerous amphibolited dolerites and gabbros and occasional dunites and serpentinites. These rock units comprise the Seve Nappe Complex. Many of the lithologies are comparable with, but more heterogeneous than, the underlying Särv Nappes; they are always of higher metamorphic grade, reaching granulite and eclogite facies in parts of the mountain belt. The eclogite-facies metamorphism is of latest Cambrian to earliest Ordovician age (*ca*. 500–490 Ma), and at least some of the granulites are younger and related to the Scandian collisional orogeny.

The Seve Nappe Complex has been treated as a separate so-called suspect terrane, which was probably derived from the outermost edge of Baltica, where the extensive mafic magmatism and solitary ultramafites were concentrated in the transition zone between continental and oceanic crust.

Iapetus Ocean Terranes (the Köli Nappe Complex of the Upper Allochthon)

Well-preserved ophiolites, derived from the Early Palaeozoic ocean floor, along with volcanic island-arc igneous complexes and associated thick, mainly volcaniclastic, sedimentary successions characterize the outboard terranes of the Scandinavian Caledonides. The ophiolites are mostly of Early Ordovician (perhaps partly Late Cambrian) age, and the arc complexes date from throughout the Ordovician, providing evidence of the subduction systems controlling the closure of the Iapetus Ocean. The rock units are generally metamorphosed in greenschist facies and, within the sedimentary successions, fossils are locally preserved, providing evidence of both age and faunal affinity. Of particular interest has been the identification of Early Ordovician faunas (mainly trilobites and brachiopods, but also molluscs) with North American affinities, indicating that some of the ophiolite-bearing allochthons were derived from the Laurentian side of the Iapetus Ocean; others, apparently, were more centrally placed in the ocean.

These Scandian Iapetus-derived outboard terranes occur in the Köli Nappe Complex, which, in central parts of the mountain belt, can be readily divided into three parts – Lower, Middle, and Upper. Fragmented ophiolites occur locally in the base of the Lower Köli Nappes (e.g. at Otta and Handöl). Associated black-shale formations often host solitary ultramafites. The latter are sometimes mantled by detrital serpentinite and at one remarkable location (Otta) these sedimentary rocks are highly fossiliferous, giving good control of age (Llanvirn) and ocean-island affinity. Early Ordovician island-arc and rifted-arc volcanites dominate the lower parts of the Lower Köli Nappes, which are overlain by turbidites and sandstones; these shallow upwards in the Late Ordovician into continentally derived quartz sandstones and then limestones. The latter host a coral- and brachipod-bearing fauna with *Holorhynchus*, providing evidence of proximity to Baltica by Ashgill times. These shallow-marine environments deepen into basinal black shales, sometimes with pillow basalts, in the mid–Late Llandovery, after which the basins are filled with turbidites.

The Middle Köli Nappes are dominated by igneous complexes with calc-alkaline volcanic-arc affinities. Major plutons range in composition from gabbro and gabbro–diorite to tonalite, granodiorite, and trondhjemite, and the host rocks are mainly basalts and volcaniclastic formations, all generally metamorphosed in greenschist facies. Ordovician turbidites are conspicuous, and black shales have locally yielded Llandovery graptolites. The igneous complexes have provided mid-Ordovician uranium–lead zircon ages and evidence of Ordovician deformation; their chemistry indicates an origin in an oceanic setting, and most authors have considered them to be related to Laurentian-margin subduction systems.

The Upper Koli Nappes contain a major basal ophiolite (e.g. at Støren) and a range of mid-Ordovician calc-alkaline plutons. The ophiolite is directly overlain by Arenig limestones containing trilobites and brachiopods of unambiguous North American affinities; it has therefore been inferred that initial obduction of this ocean floor was onto the outer continental margin of Laurentia. Turbidites and fanglomerates dominate the overlying successions, which are also thought to be Ordovician in age; Silurian fossils have not been found in the Upper Koli Nappes.

Laurentian Continental Margin (Uppermost Allochthon)

Highest in the Scandian nappe pile, located along the west coast of central Norway, is a characteristic allochthon dominated by major granitoid batholiths, intruded into amphibolite facies sedimentary successions of continental affinity – marbles, schists, and psammites. Within this allochthon (e.g. in the

Helgeland Nappe Complex) there are fragmented ophiolites that may have affinities with those in the underlying Støren Nappe. The batholiths are of Ashgill to Llandovery age; thus they were intruded immediately prior to Scandian collision.

Scandian Collision

The Scandian collision, with thrusting of major allochthons many hundreds of kilometres onto the Baltoscandian platform, resulted in deep depression of the western edge of Baltica. Eclogite facies (high pressure and temperature) metamorphism, locally with mineralogical evidence of very high pressures (coesite and microdiamonds), was widespread in mafic rocks in the Precambrian basement of south-western Norway (Western Gneiss Region); it also occurs further north along the coast at a similar structural level. Exhumation of these deeply depressed rocks occurred during both Late Silurian thrusting and the extensional collapse of the orogen, with particularly rapid uplift in the Early Devonian. Crustal thickening during continental collision led to the build-up of the Caledonian mountain belt, and this was kept in balance during continued collisional compression by upper-crustal extension. This gravitational collapse resulted both in further migration of the nappe pile onto the Baltoscandian platform and in major westerly directed detachments in the hinterland, the latter accompanying the development of Old Red Sandstone basins.

Eastern Greenland

The Caledonian mountains of eastern Greenland, reaching from 70°–82° N, flank the eastern edge of the Greenland ice dome. As in Norway, the coastline is deeply penetrated by fjords, and the steep mountain sides, many rising vertically for 1000–2000 m out of the sea, provide spectacular outcrops for geological analysis. The continental shelf reaches only about 50 km offshore in the south, but widens northwards to about 300 km in north-east Greenland. Beneath these shelf areas of the Greenland Sea, the Caledonian bedrock is covered by Late Palaeozoic and younger successions. Tertiary mafic volcanics and intrusions, related to the opening of the Greenland Sea, are extensive south of Scoresby Sund (*ca.* 70° N).

The outcrop of Caledonian bedrock in eastern Greenland (Figure 3) is about 300 km wide in the south, at 70° N, and narrows northwards to about 100 km, striking obliquely offshore into the northern Greenland continental shelf. It forms the western part of the Caledonide Orogen and is dominated by thrust sheets, which were emplaced west-north-westwards onto the Laurentian Craton with its Early Palaeozoic cover. Characteristic of the autochthonous cover of the Laurentian margin of the entire Caledonides, from eastern Canada in the south, via north-western Scotland and eastern Greenland, to Svalbard, is a Cambro-Ordovician carbonate succession that contrasts markedly with the coeval succession of Baltica in terms of both lithology and fauna. In southern parts of the eastern Greenland autochthon, these formations, with underlying quartzites and, locally, tillites, rest on a deeply eroded and peneplained basement of Archaean and Palaeoproterozoic age. In northern areas, thick Mezoproterozoic sandstones and basalts separate this old basement from the Palaeozoic cover, and tillites are absent.

Most of the orogen front in eastern Greenland is covered by ice, and the relationships between the major Caledonian thrust sheets and the underlying cover are best seen in antiformal windows, similar to those in the Scandes. The allochthons have been divided into two groups and are referred to as ‘thin-skinned’, overlain by ‘thick-skinned’. Whereas in the far north the former is a fold-and-thrust belt dominated by sedimentary successions, further south it incorporates extensive slices of Palaeoproterozoic and Archaean crystalline rocks. The upper ‘thick-skinned’ complex differs greatly from the underlying thrust sheets in the area south of 76° N, by incorporating latest Mesoproterozoic successions (Krummedal and Smallefjord groups), which are several kilometres thick and intruded by *ca.* 930 Ma granites; these are overlain, probably unconformably, by Neoproterozoic siliciclastic and carbonate successions (Eleonore Bay Supergroup), Vendian tillites, and platform Cambro-Ordovician carbonate formations. Thus the entire Caledonian mountain belt of eastern Greenland is composed of rocks derived from the Laurentian margin, including both Archaean and Palaeoproterozoic crystalline basement and younger generally shallow-marine sedimentary successions deposited on it. Outboard (e.g. Iapetus-related) terranes are notable by their absence. West-north-west-directed thrust transport of the allochthons has been estimated to be at least 200 km, along with some sinistral strike displacements; thus, prior to the Caledonian Orogeny, the platform margin of Laurentia was nearly twice as wide as it is today.

Only in north-easternmost Greenland do the Early Palaeozoic carbonate-dominated successions continue upwards through the Ordovician into the Silurian, to be replaced by easterly derived turbidites in the mid-Silurian. This evidence, together with the presence of mid–Late Devonian Old Red Sandstones in extensional basins in southern areas and many mid-Palaeozoic isotopic ages, indicates that, although

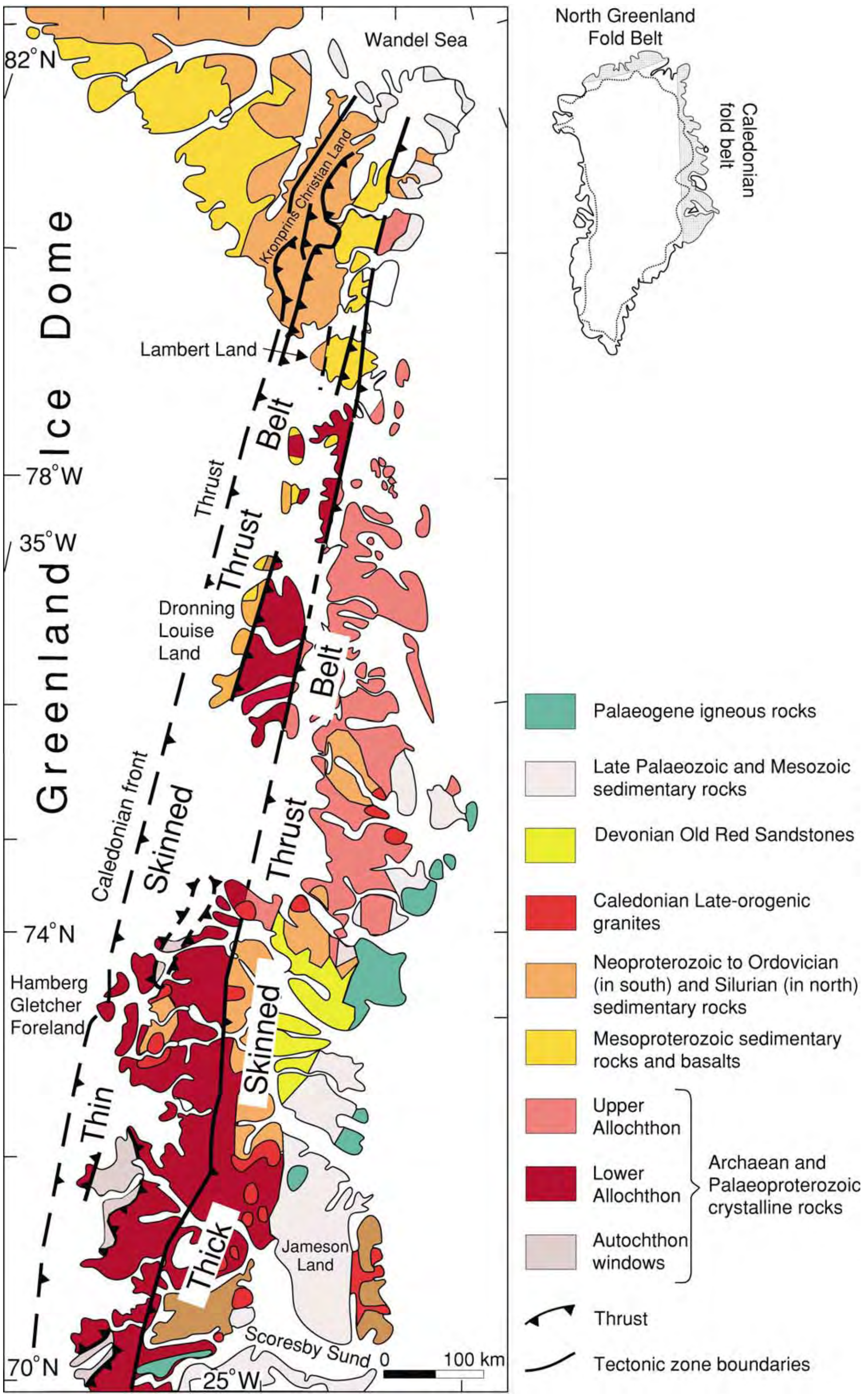


Figure 3 The Eastern Greenland Caledonides.

thrusting may have started in the Late Ordovician, the timing of the collisional orogeny in eastern Greenland coincided with that in the Scandes. In marked contrast to the Scandes, however, major post-kinematic granite plutons are present.

Svalbard

The Svalbard Archipelago (Figure 4), located on the north-western corner of the Barents Shelf, is dominated by a larger island, Spitsbergen, which is flanked to the east by Nordaustlandet, Barentsøya, and Edgeøya. Numerous other smaller islands are part of the archipelago, which reaches as far east as Kvitøya and south to Bjørnøya. The western and northern parts of the archipelago are dominated by Caledonian bedrock, including Old Red Sandstones; Carboniferous and younger successions unconformably overlie this 'basement', occurring in a major syncline that dominates the structure of the central and southern parts of the archipelago. These younger successions extend eastwards and southwards over much of the Barents Shelf. Along the western coast of Spitsbergen, a Tertiary east-verging fold-and-thrust belt is superimposed on the older structures.

Prior to the opening of the northernmost Atlantic and the Fram Strait, Svalbard's Caledonian bedrock was a direct northern extension of the East Greenland Caledonides (Figure 1). On Svalbard, the Caledonian rocks generally strike northwards and are split by major north-trending faults and Old Red Sandstone graben. Various provinces, or terranes, have been recognized, and nearly all can be related directly to the East Greenland Caledonides. However, the exception is important and occurs along the west coast of Spitsbergen.

Svalbard's eastern province (Figure 4) is readily divisible into two terranes, Nordaustlandet (including Kvitøya and eastern Ny Friesland) and West Ny Friesland. The Nordaustlandet Terrane is characterized by the typical Laurentian Cambro-Ordovician carbonate platform succession overlying tillites and thick Neoproterozoic dolomites, limestones, sandstones, and shales. Major unconformities separate

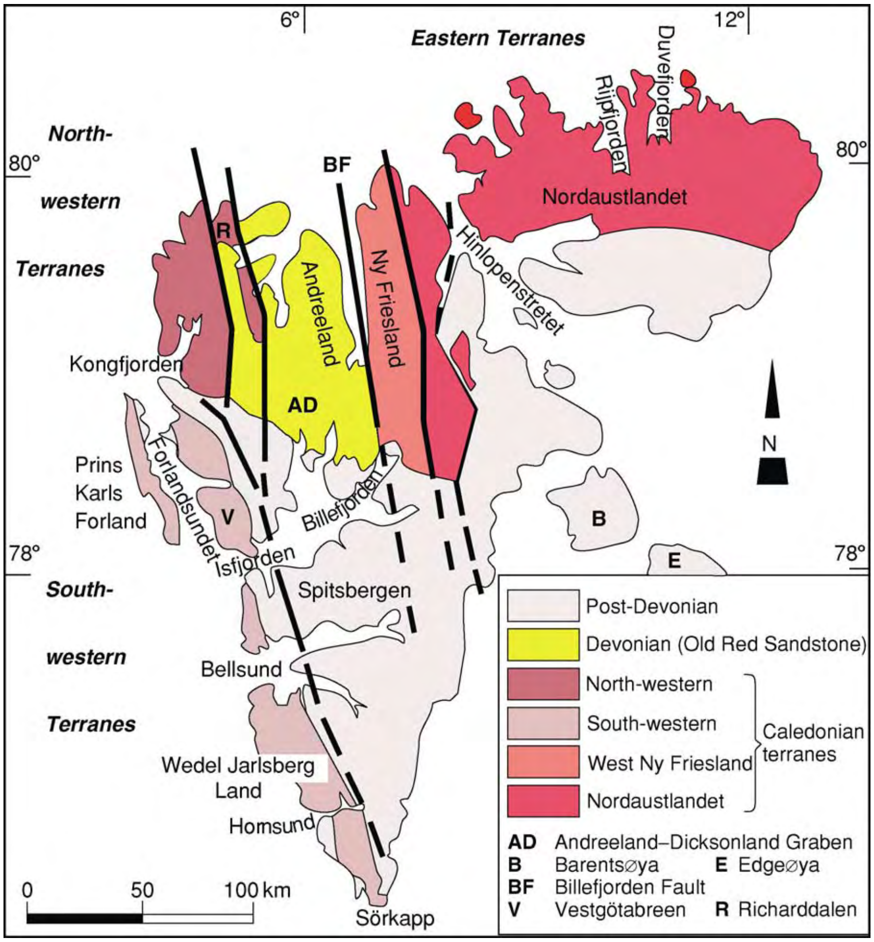


Figure 4 The Svalbard Caledonides.

this succession from a Late Grenvillian basement of andesites and rhyolites and underlying meta-turbidites of latest Mesoproterozoic or earliest Neoproterozoic age, intruded syntectonically by *ca*. 950 Ma granites. Upright to west-verging folding characterizes this easternmost Svalbard Terrane, and the grade of Caledonian metamorphism increases eastwards, with migmatization of the eastern parts of Nordaustlandet and the island of Kvitøya. Late-orogenic granites (*ca*. 420–410 Ma) also characterize this terrane, which, in all respects, compares closely with the highest allochthon of the central East Greenland Caledonides.

The Nordaustlandet Terrane is thrust westwards onto a succession of high amphibolite facies orthogneisses and isoclinally folded metasediments that is nearly 10 km thick – the Atomfjella Complex. The metasediments are dominated by quartzites, but also include marble and schist formations; the orthogneisses are mainly metagranites, dated at 1750 Ma. In both the metagranites and the quartzites, metadolerites are ubiquitous. Dating of detrital zircon has shown that the quartzites are Mesoproterozoic in age, but probably not younger than 1300 Ma (the age of the metadolerites), and that the marble and schist formations are latest Mesoproterozoic or younger. The Late Palaeoproterozoic 'basement' metagranites are repeated by thrusting at least three times in a major north-trending antiform that dominates the structure of western Ny Friesland. Only Caledonian (*ca*. 430–410 Ma) argon–argon ages of metamorphism have been obtained from the Atomfjella Complex; evidence of Grenvillian tectonothermal activity is notable by its absence.

The West Ny Friesland Terrane is closely comparable with the 'thick-skinned' allochthon of north-east Greenland, in terms of both the stratigraphy and the character of the Caledonian deformation and metamorphism. Taken together with the evidence (above) for the comparability of the Nordaustlandet Terrane and the central East Greenland allochthons, there can be little doubt that Svalbard's eastern Caledonian terranes are a direct northerly continuation of the East Greenland Caledonides (Figure 5). In much of

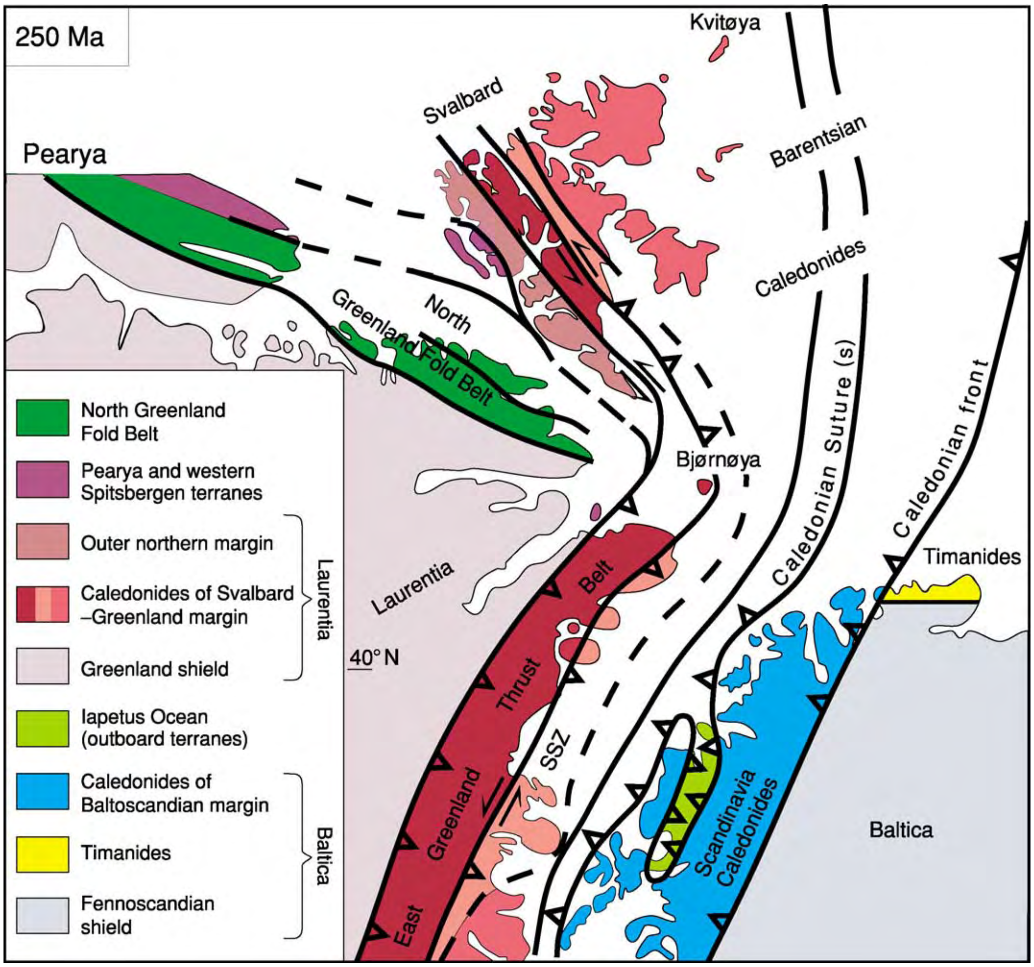


Figure 5 The Arctic Caledonides and Laurentia–Baltica relationships in the Early Mesozoic.

the previous literature on the Svalbard Caledonides, a hypothesis has been favoured that these eastern terranes are derived from an off-shore area to the east of central East Greenland, implying sinistral transport over a distance of at least 1000 km; this hypothesis is not supported by recent investigations of Svalbard and Greenland.

Svalbard's eastern Caledonides are separated from the western terranes by the 50 km wide Andreéland–Dicksonland Old Red Sandstone graben. The Caledonian bedrock, from north-western Spitsbergen southwards via the west-coast Tertiary fold-and-thrust belt to southernmost areas and further south to Bjørnøya, provides evidence of both Laurentian-platform affinities and an outboard subduction-related terrane. Unambiguous Laurentian lithological and faunal signatures are found on Bjørnøya and southernmost Spitsbergen, where Cambro-Ordovician carbonate successions are closely related to strata of the same age in north-eastern Greenland. The underlying Neoproterozoic carbonate and siliciclastic successions are also comparable with those in this part of Greenland, but, interestingly, the metamorphic complexes unconformably underlying these sedimentary rocks are of Grenvillian age, a characteristic that is unknown in north-eastern Greenland.

In central western Spitsbergen, Neoproterozoic successions, including thick tillites, are overthrust by a blueschist and eclogite assemblage (the Vestgötabreen Complex). The high pressure–low temperature metamorphism occurred in the Early Ordovician and was followed by thrusting to high structural levels, erosion, and deposition of a mid–Late Ordovician conglomerate and limestone succession, passing up into Silurian turbidites. The subduction-related Vestgötabreen Complex is thought to be related to the Pearya Terrane of northernmost Canada (Ellesmere Island), both being foreign to Laurentia.

As shown in Figure 6, the Caledonian terranes of Svalbard comprise the western part of an orogen that extends northwards from the Scandes beneath the Barents Shelf to the edge of the Eurasian Basin. This part of the orogen is called the Barentsian Caledonides. Prior to the opening of the Eurasian Basin in the Tertiary, the Barentsian Caledonides are thought to have crossed Lomonosova (today's Lomonosov Ridge) and may have continued into the continental shelves of the Amerasian Basin.

The eastern side of the Barentsian Caledonides is flanked by the Timanides, a Late Neoproterozoic orogen that dominates the bedrock of north-easternmost Europe. The Timanides strike north-westwards

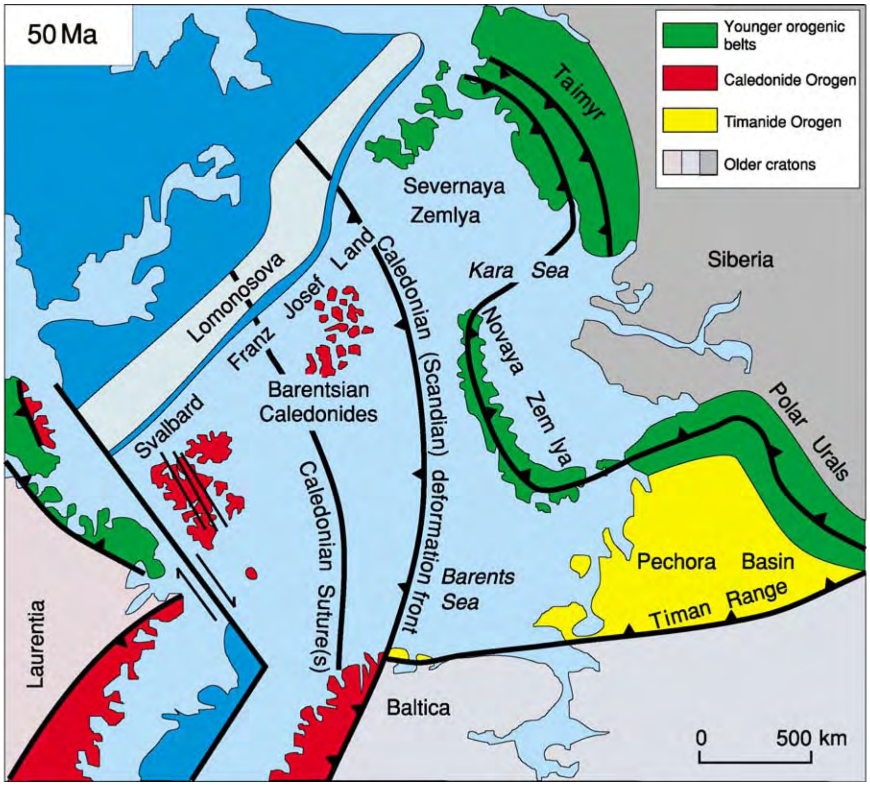


Figure 6 Tectonic elements of the western Eurasian Arctic in the Early Tertiary, during the initial opening of the Eurasian Basin and the Norwegian and Greenland seas (light blue, shelf and ridges; deep blue, ocean floor).

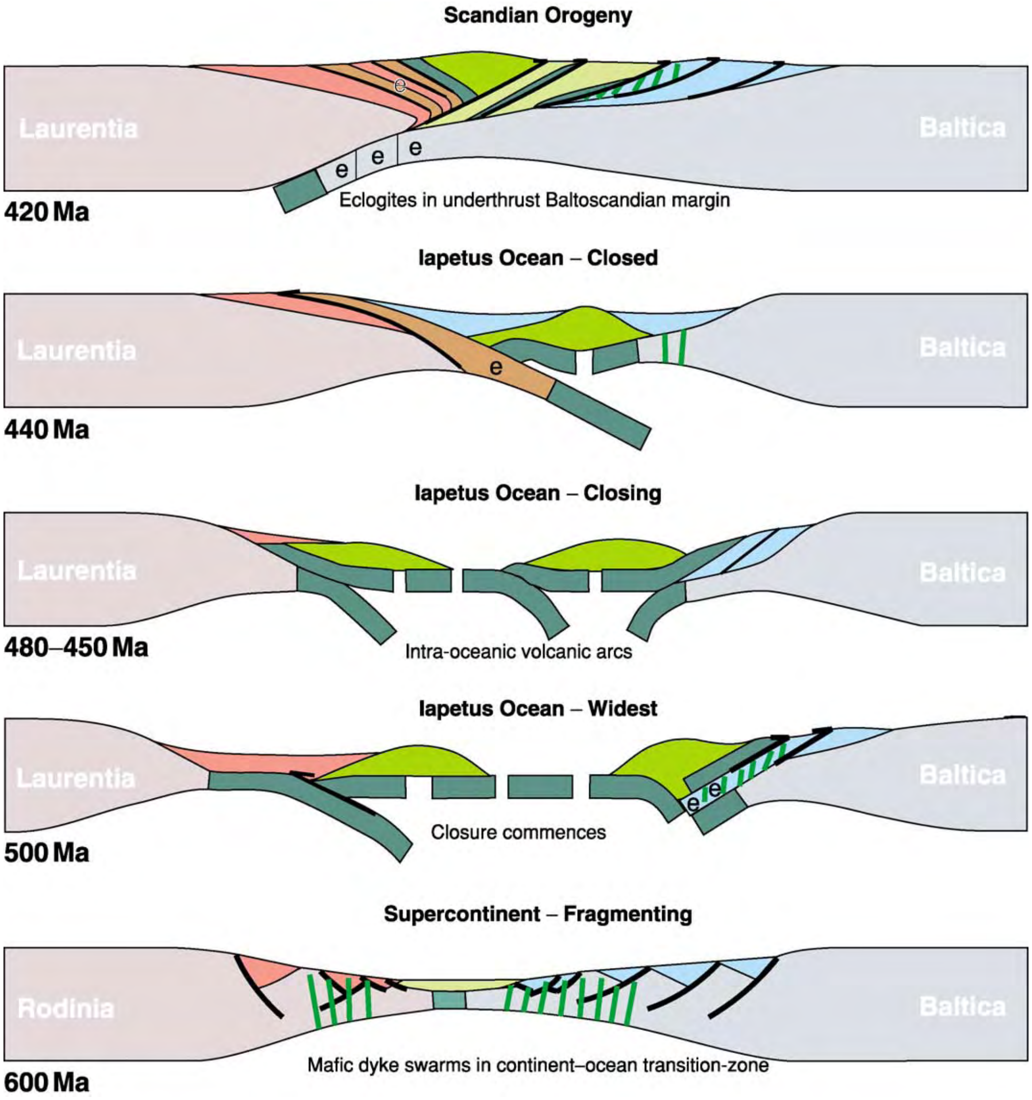


Figure 7 Caledonian tectonic evolution, from the Vendian to the Late Silurian. e; eclogite.

from a type area in the Timan Range, beneath the Pechora Basin, into the Barents Sea. A deep borehole on Franz Josef Land sampled folded Vendian turbidites at depth of 2 km; this evidence, along with geophysical data, suggests that the Barentsian Caledonides occur beneath Late Palaeozoic, Mesozoic, and Tertiary strata over much of the Barents Shelf.

Tectonic Evolution of the Northern Caledonides

The first interpretations of the North Atlantic Caledonides in terms of continental drift and plate tectonics inferred that Laurentia and Baltica were part of a single supercontinent (Rodina) in the Late Neoproterozoic. The Caledonian cycle was thought to involve rifting and separation of these continents, with the opening of a substantial ocean by seafloor spreading, followed by closure, resulting in a mid-Palaeozoic configuration that was similar to that which existed before ocean opening. Palaeomagnetic evidence has subsequently suggested that Baltica rotated about 120° between the Vendian and the Middle Ordovician; thus the model of simple opening and closing (the so-called 'Wilson cycle') is not generally accepted today. However, the structure of the Scandinavian Caledonides does provide evidence of a tectonic history that is readily divisible into three parts:

1. a period of Neoproterozoic rifting, leading to the separation of Baltica from a larger continental assemblage by the Early Cambrian;

2. opening of the Iapetus Ocean in the Cambrian and closing during the Ordovician, with subduction along the margins of both Laurentia and Baltica; and
3. Scandian collision of Baltica and Laurentia in the mid-Silurian and Early Devonian, with underthrusting of the latter by the former.

The Caledonides of eastern Greenland and Svalbard provide complementary evidence on the character of the Laurentian margin and allow the reconstruction of the tectonic evolution of the whole orogen, which is summarized schematically in Figure 7.

The vast Himalayan-type thrust systems of western Scandinavia and eastern Greenland, with many hundreds of kilometres of crustal shortening across the orogen, testify to more-or-less orthogonal collision of Laurentia and Baltica. Support for this conclusion is found in the kinematic evidence of west-north-west and east-south-east transport of the allochthons and the correlation of thrust complexes over distances of up to 1000 km along the orogen. This northern segment of the Caledonides differs markedly from the type areas of the Caledonide Orogen in the UK and further to the south-west, where sinistral transpression appears to have dominated the Caledonian collision of Avalonia and Laurentia.

See Also

Europe: Timanides of Northern Russia; Caledonides of Britain and Ireland; Permian to Recent Evolution. **Plate Tectonics**. **Tectonics:** Faults; Folding; Mountain Building and Orogeny.

Further Reading

Andreasson PG (1994) The Baltoscandian margin in Neoproterozoic early Palaeozoic times. Some constraints on terrane derivation and accretion in the Arctic Scandinavian Caledonides. *Tectonophysics* 231: 1 32.

Birkenmajer K (1981) The geology of Svalbard, the western part of the Barents Sea, and the continental margin of Scandinavia. In: Nairn AEM, Churkin M Jr, and Stehli FG (eds.) *The Arctic Ocean*, pp. 265 329. The Ocean Basins and Margins, Part 5. New York: Plenum.

Fortey RA and Bruton DL (1973) Cambrian Ordovician rocks adjacent Hinlopenstretet, north Ny Friesland, Spitsbergen. *Geological Society of America Bulletin* 84: 2227 2242.

Gee DG (1975) A tectonic model for the central part of the Scandinavian Caledonides. *American Journal of Science* 275A: 468 515.

Gee DG and Sturt BA (eds.) (1985) *The Caledonides Orogen Scandinavia and Related Areas*. Chichester: Wiley.

Gee DG and Tebenkov AM (2005) Svalbard: Fragments of the Laurentian Caledonian Margin. In: Gee DG and Pease VL (eds.) *The Neoproterozoic Timanide Orogen of Eastern Baltica*. London: Geological Society Memoir.

Grenne T, Ihlen PM, and Vokes FM (1999) Scandinavian Caledonide Metallogeny in a plate tectonic perspective. *Mineralium Deposita* 34: 422 471.

Harland WB (1997) *The Geology of Svalbard*. Memoir 17. London: Geoogical Society of London.

Henriksen N, Higgins AK, Kalsbeek F, and Pulvertaft TC (2000) Greenland from Archaean to Quaternary. Description text to the Geological map of Greenland 1 : 2 500 000. *Geology of Greenland Survey Bulletin* 185: 1 96.

Higgins AK and Leslie AG (2000) Restoring thrusting in the East Greenland Caledonides. *Geology* 28: 1019 1022.

Stephens M (1988) The Scandinavian Caledonides: a complexity of collisions. *Geology Today* 4: 20 26.

Stephens MB and Gee DG (1989) Terranes and polyphase accretionary history in the Scandinavian Caledonides. *Geological Society of America Special Papers* 230: 17 30.

Trettin HP (1989) The Arctic Islands. In: Bally AW and Palmer AR (eds.) *The Geology of North America An Overview*, pp. 349 370. Boulder, Colorado: Geological Society of America. V. A.

Variscan Orogeny

W Franke, Johann Wolfgang Goethe-Universität, Frankfurt am Main, Germany
P Matte, University of Montpellier II, Montpellier, France
J Tait, Ludwig-Maximilians-Universität, München, Germany

Introduction

Western and Central Europe were consolidated during an Upper Palaeozoic orogenic event, which has been named the 'Variscan' orogeny after a legendary Germanic tribe in north-eastern Bavaria, mentioned by the Roman author *Tacitus*. The term 'Variscan' should be given preference over 'Hercynian', which is often used for the same orogeny. The latter name is derived from the Harz Mountains in northern Germany, which do contain Variscan basement, but which represent a fault block thrust up during the Late Cretaceous.

Variscan rocks can be traced from Portugal to Poland and from the British Isles to the Mediterranean. The Variscan basement is exposed in 'massifs', which emerge from under younger Late Palaeozoic or Mesozoic cover. The most important massifs are the Iberian Massif in Spain, the Armorican Massif, the Massif Central, and the Maures Massif in France, parts of Corsica and Sardinia, the Vosges and the Black Forest, the Ardennes, the Rhenish Massif and the Harz Mountains in west-central Europe, and the Bohemian Massif further east. In the Alps and Carpathians, as well as in the Mediterranean realm, the Variscan basement has been much reworked by the Mesozoic–Cenozoic 'Alpine' orogenic processes (*see* **Europe:** The Alps).

The extraordinary width of the Variscan belt is the result of a complex palaeogeographical situation. The Variscan orogen is actually a collage of major and minor continental plates (Figures 1 and 2), which were once separated by at least three oceanic areas. These plates first spread apart (520–450 Ma), then converged, and eventually collided with each other (*ca*. 420–300 Ma). Sequential collision produced the huge landmass of Pangaea. By the Early Permian, the orogenic edifice was largely eroded, and was subsequently covered by Permian, Mesozoic, and Cenozoic deposits. The angular unconformity between Variscan basement and the cover is one of the key features of European geology (Figure 3).

There are two complementary ways to reconstruct such large-scale and long-term processes. Studies of palaeomagnetism (*see* **Palaeomagnetism**) can reveal fossilized magnetic fields in ancient rocks. The palaeodeclination reveals rotation of the sampling area with respect to the present-day magnetic meridian. The palaeoinclination records the palaeolatitude: inclination is vertical at the magnetic poles and horizontal at the magnetic equator. Since the Palaeozoic oceans in Europe were orientated more or less east–west, their closure implies changes in palaeolatitude, which are well documented in the palaeomagnetic record. Palaeomagnetic data can be compared with palaeoclimatic indicators, such as evaporites, coral reefs, tropical forests, or glacial deposits, and biogeography (areal distribution of fossil faunas and floras). With errors of ±500 km, palaeomagnetism, palaeoclimatology and biogeography yield estimates of palaeolatitude and can therefore be used to deduce the larger-scale movements of the plates. It is important to note, however, that palaeomagnetism provides no information about palaeolongitude.

The second approach is geology. Even in ancient mountain belts, it is possible to identify characteristic elements of the plate-tectonic cycle, such as continental rifts, volcanic belts (magmatic arcs), remnants of oceanic lithosphere (ophiolites), and belts of high-grade metamorphic rocks, some of which have been subducted into the mantle before reascending to the surface. These features allow us to identify ancient active and passive plate margins, to establish a relative sequence of geological processes, and to date events. While the amount of ocean floor lost by subduction can be assessed only by palaeomagnetic methods, the deformation of the continental crust must be reconstructed by unravelling the polyphase tectonic deformation and metamorphism. In most orogenic belts, collisional deformation has reduced the width of the colliding blocks to 50% or even less.

In this article, we first address the large-scale migration of the continental blocks that are now welded together into the 'united plates of Europe', and then proceed to sketch out the 'ground truth' revealed by geological studies.

Palaeomagnetic and Biogeographical Record

The Palaeozoic geography of the crustal segments now contained in the continents bordering the Atlantic was dominated by the large plates of Laurentia, Gondwana, and Baltica. Between these plates there were a number of smaller microplates, which are now incorporated along the eastern margin of North America and

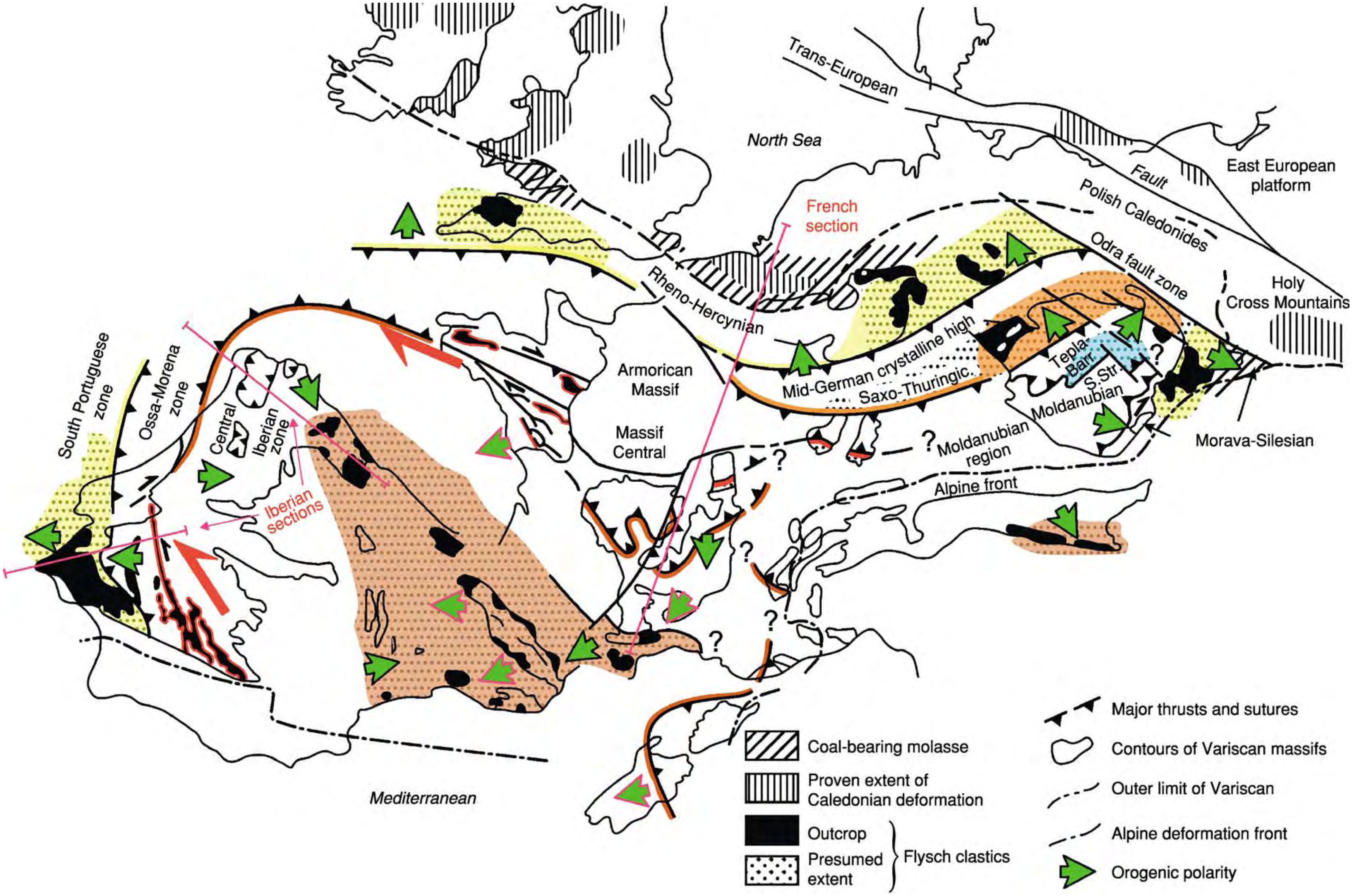


Figure 1 Tectonic and palaeogeographical subdivision of the Variscan basement in Europe, after Frank W (2000) The mid European segment of the Variscides: tectonostratigraphic units, terrane boundaries and plate tectonic evolution. In: Franke W, Haak V, Oncken, O, and Tanner D (eds.) *Orogenic Processes: Quantification and Modelling in the Variscan Belt*, pp. 35 62. Special Publication 179. London: Geological Society.

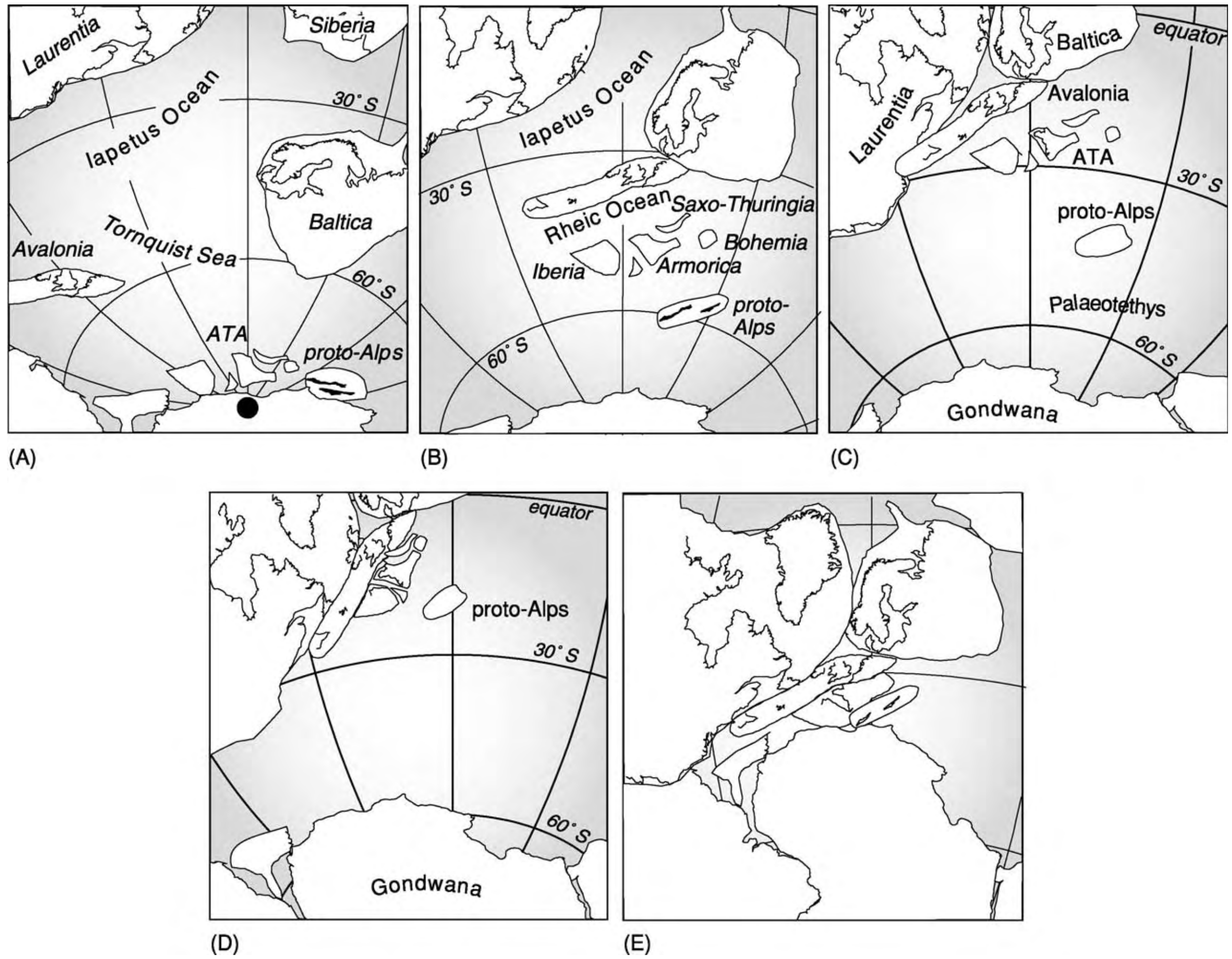


Figure 2 Positions of the continents and microplates bordering the Atlantic in (A) the Early Ordovician, (B) the Late Ordovician, (C) the Late Silurian, (D) the Middle Devonian, and (E) the Permian based on palaeomagnetic data.

Figure 3 Angular unconformity between folded Late Carboniferous deep water sandstones and Triassic fluvial deposits. Telheiro beach near Cabo Sao Vicente, Algarve, southern Portugal. Photograph taken by P. Matte.

the southern border of Eurasia. These terranes include Avalonia, the Armorican Terrane Assemblage, and the proto-Alps, which, in Cambrian to Early Ordovician times, were still attached to the north Gondwanan margin. Subduction of oceanic crust under this margin caused the Cadomian orogen (named after Caën, the Roman *Cadomus*, in Normandy). Cadomian basement is present in nearly all parts of the Variscan belt.

The positions and drift histories of Gondwana, Baltica, Laurentia, Avalonia, the Armorica Terrane Assemblage, and the proto-Alps are now fairly well constrained by palaeomagnetic data, and faunal and palaeoclimatic indicators are generally in good agreement. Palaeomagnetic data suggest the following palaeogeographical evolution (Figure 2).

Cambro-Ordovician

In the Early Ordovician, the northern margin of Gondwana was situated at high southerly palaeolatitudes, Baltica was situated between 30° S and 60° S and was inverted with respect to its present-day position, and Laurentia was in an equatorial position. The Cadomian basement rocks of Avalonia and the Armorican Terrane Assemblage clearly indicate that in Cambrian times these terranes were contiguous with the northern margin of Gondwana, and remained marginal to Gondwana until the Early Ordovician. By the late Early Ordovician (Tremadoc), data indicate that Avalonia had started to drift northwards, away from Gondwana, opening the Rheic Ocean in its wake (*see* **Palaeozoic:** Ordovician). It continued to move northwards throughout the Ordovician, gradually closing the Tornquist Sea and Iapetus Ocean, which separated it from Baltica and Laurentia, respectively. Palaeomagnetic data from different elements of the Armorican Terrane Assemblage indicate a similar, but independent, movement of these microplates, with separation from Gondwana being initiated slightly later in the Ordovician. No palaeomagnetic data are yet available from Ordovician rocks of the proto-Alps, but geological evidence suggests a continued Gondwanan affinity during this period.

Late Ordovician

By the Late Ordovician, Gondwana had moved some 30° northwards, and northern central Africa was situated over the south pole according to palaeomagnetic data from western Australia. Baltica was now in its present-day orientation, and its northern margin was at the equator. Laurentia, which did not move much throughout the Palaeozoic, remained straddling the equator and was separated from Baltica and Gondwana by the Iapetus Ocean. By Ashgillian times, palaeomagnetic and biogeographical data indicate that the Tornquist Sea, separating Baltica and Avalonia, had closed. Collision of Avalonia with Baltica created a narrow belt of deformation and metamorphism (the Polish Caledonides; Figure 1). The Rheic Ocean still separated Avalonia/Baltica from the Armorican Terrane Assemblage, which was situated at more southerly palaeolatitudes in the Ashgillian, based on palaeomagnetic evidence from the Bohemian Massif. The presence of glaciomarine sediments and cold-water faunas throughout the Armorican Terrane Assemblage reflects the Late Ordovician period of global cooling, which enabled the colonization of previously warmer-water realms by cold-water faunas. It has been shown that these glacial sediments of central Europe were deposited by seasonal or floating ice, in agreement with palaeomagnetic data from the Bohemian Massif, which clearly indicate intermediate to low palaeolatitudes. Strong faunal and lithological similarities in the Ordovician–Devonian successions of different massifs of the Armorican Terrane Assemblage indicate similar ecological conditions, demonstrating that they were all part of the same palaeogeographical domain. The proto-Alps were positioned at higher palaeolatitudes, between northern Gondwana and the southern margin of the Armorican Terrane Assemblage. This conclusion is based predominantly on faunal evidence, which indicates separation from northern Gondwana.

Siluro-Devonian

The palaeogeographical position of Gondwana from Silurian to Late Devonian times remains controversial on the basis of palaeomagnetic evidence, as two different models have been proposed in the literature. The more conservative model involves gradual northward movement of northern Gondwana throughout the Palaeozoic, with final closure of the ocean separating northern Africa from southern Europe in the Late Carboniferous. The alternative model is based primarily on palaeomagnetic data from south-east Australia and requires rapid northward movement of Gondwana in the Silurian, followed by rapid southerly movement in the Devonian. However, whether or not it is viable to use palaeomagnetic data from this region of Australia, whose autochthony with cratonic Australia is questioned, remains open to debate. In summary, taking all the palaeomagnetic, palaeoclimatic and biogeographical data into account, the more conservative model, involving gradual northward movement of Gondwana throughout the Palaeozoic, is considered more plausible.

Final closure of the Iapetus Ocean between Baltica/Avalonia and Laurentia occurred in the Siluro-Devonian, after which Baltica and Laurentia (Laurussia) remained in equatorial palaeolatitudes until the end of the Palaeozoic era.

Late Silurian–Early Devonian palaeomagnetic data from a number of different crustal blocks of the Armorican Terrane Assemblage indicate palaeolatitudes of 20°–30° S. This implies gradual migration towards the southern margin of Baltica/Avalonia and closure of the intervening Rheic Ocean.

Late Devonian

By the Late Devonian, the Iapetus Ocean between Laurentia and Baltica/Avalonia had closed. Similarly, the Rheic Ocean between Avalonia and the Armorican Terrane Assemblage closed in the late Mid-Devonian. Closure of the ocean was essentially a longitudinal process and thus cannot be accurately constrained by palaeomagnetic data; however, invertebrate faunal differences between Bohemia and Avalonia persisted until the Emsian or Givetian. The now amalgamated Laurasian landmass moved southwards in Late Devonian times, but its northern border remained equatorial.

In Gondwana, however, the Late Devonian remains one of the more controversial periods. High-quality palaeomagnetic data from Australia clearly place central Africa over the south pole, requiring an ocean between the northern margin of Gondwana and the now amalgamated Laurasia. Faunal data from the southern Alps show poor similarity with the coeval fauna of northern Africa, and the sedimentary sequences of the southern Alps reflect a period of continuous sedimentation until the Late Carboniferous, with little evidence of any major deformation until Carboniferous times. However, the similarity of the fossil fish records of Gondwana (Australia) and Laurasia suggests that there was no oceanic separation of these two continents from late Early Devonian times onwards. These discrepancies in the faunal record, and between biogeographical and palaeomagnetic evidence, remain, as yet, unresolved. Nevertheless, there is general consensus that the collision of Gondwana with Laurasia to form the supercontinent Pangaea occurred in Late Carboniferous to Permian times.

A modern analogue of Palaeozoic plate dispersal might be the Indian Ocean: like Avalonia and the Armorican blocks, India, the Seychelles, and Madagascar have separated from Africa and, at least partly, made their way towards Asia (Laurussia).

Geological Record: Central Europe

Geological observations provide more detailed information about the plate-tectonic processes involved. Figure 4 is a diagrammatic representation of the plate-kinematic evolution of central Europe. Northward migration of the Armorican Terrane Assemblage was accommodated by northward-directed subduction of the Rheic Ocean beneath the southern margin of Avalonia, forming a magmatic arc, which is now preserved in the Mid-German Crystalline High (Figures 1, 4, and 5). By the late Early Devonian, the Rheic Ocean had more or less closed. At the same time, narrow oceans or seaways between the Armorican islands – Saxo-Thuringia and Bohemia – and between Bohemia and Moldanubia (possibly representing northern Gondwana) were being closed. The generally convergent movements were interrupted only by renewed extension along the Avalonia–Armorica boundary, creating a new narrow oceanic basin (the Rheno-Hercynian Ocean) whose remains can be traced from southern Portugal via south-west England and the Rhenish Massif into the Harz Mountains, with an easterly extension into Moravia (Figure 1).

The Rheno-Hercynian Ocean closed from the late Middle Devonian onwards, thus joining in the general convergence. The final collision of Avalonia with Franconia (the northernmost Armorican terrane) occurred in the earliest Carboniferous. Collision between the Armorican terranes further south (Saxo-Thuringia, Bohemia, and Moldanubia) took place in Late Devonian times.

Plate convergence was accommodated by a bilateral array of subduction or collision zones: the narrow Rheno-Hercynian and Saxo-Thuringian oceans were subducted towards the south, while the seaway between Bohemia and Moldanubia was subducted towards the north. This resulted in three collisional belts: the Rheno-Hercynian, Saxo-Thuringian, and Moldanubian (Figure 5), which were originally proposed by F Kossmat in 1927.

Collision resulted in the exhumation of previously subducted oceanic and continental rocks, which were emplaced as thrust sheets over the foreland areas. In each of the three collisional belts, the frontal thrust migrated towards the foreland, thus accreting foreland rocks to the orogenic wedge (Figure 5). Thrusting and folding reduced the widths of the original microplates considerably: collisional shortening in the three belts amounts to a minimum of *ca.* 800 km. This figure does not include the unknown width of oceanic crust subducted between the microplates.

Foreland basins filled with synorogenic clastic debris migrated ahead of the tectonic fronts. In the Rheno-Hercynian Belt, deposition kept pace with subsidence, so that sedimentation occurred in a shallow-marine to fluvial environment. The extensive coastal forests that developed in these tropical regions provided the raw material for economically important coal seams, which are exploited in Wales, the northern part of the Rhenish Massif, and Silesia (Figure 1).

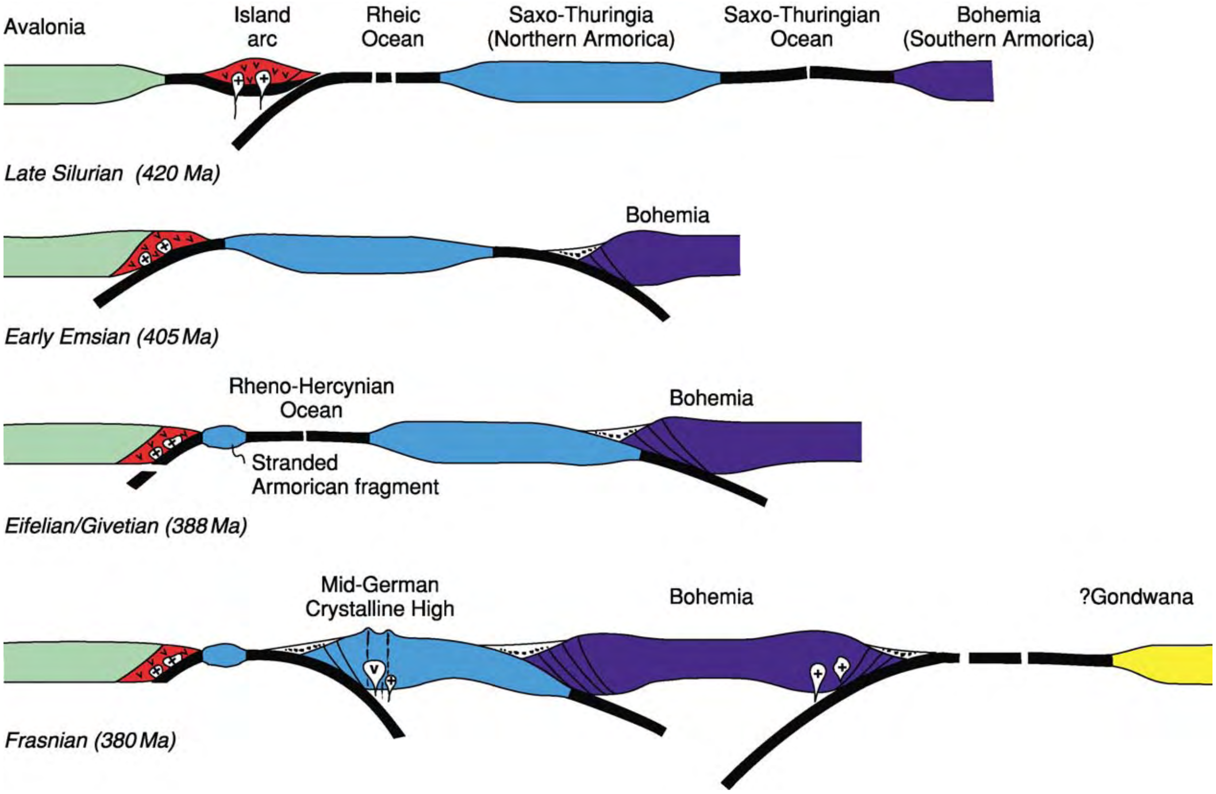


Figure 4 Plate kinematic evolution of central Europe, after Franke W (2000) The mid European segment of the Variscides: tectonostratigraphic units, terrane boundaries and plate tectonic evolution. In: Franke W, Haak V, Oncken, O, and Tanner D (eds.) *Orogenic Processes: Quantification and Modelling in the Variscan Belt*, pp. 35 62. Special Publication 179. London: Geological Society.

The present-day Variscan Belt of central Europe is not a linear feature. In Early Carboniferous times, the paralinear or linear subduction and collision zones were transported north-westwards, and drag along the south-western margin of Baltica caused it to rotate clockwise through approximately 90° (Figures 1 and 5). In about the latest Early Carboniferous, the south-eastern flank of this 'Bohemian Arc' was truncated by a huge transpressional fault zone (the Moldanubian 'Thrust'), which carried the south-eastern blocks south-westwards, possibly for a distance of about 1500 km.

Geological Record: Western Europe

An even more prominent arc structure is observed in the western part of the Variscan belt: if Iberia is rotated into its pre-Mesozoic position by closing the bay of Biscay, Variscan structures define a tight arc with a curvature of 180° in its internal part (the 'Ibero-Armorican' arc) (Figure 1).

The terranes of north-west Iberia are clearly linked with those of Brittany and the Massif Central in France and can be traced further eastwards into the Bohemian Massif, thus forming parts of the Armorican Terrane Assemblage. As in central Europe (see above), cross-sections through the two arms of the Ibero-Armorican virgation show a fan-like orogen with opposite vergences (verging towards the north-east and south-west in Iberia, and towards the south and north in France), i.e. convergent on its concave side and divergent on its convex side.

Two main sutures are visible in the internal metamorphic parts of the belt, and are the roots of nappes with opposing vergences (Figures 6 and 7). The Beja suture resulted from the closure of an oceanic realm between Avalonia (the south Portuguese foreland) and Armorica. There are no pre-Carboniferous rocks exposed in south Portugal, thus it remains uncertain whether this oceanic realm was the early Palaeozoic Rheic Ocean or a younger Devonian feature (as in south-west England and Germany). In any case, the Beja suture can be correlated with the Rheno-Hercynian ophiolites of the Lizard complex in south-west England and the Gießen-Harz Nappe in Germany. The second suture resulted from the closure of the Galice-Brittany ocean and can be traced from the French Massif Central, through the southern Vosges and Black Forest, into the Moldanubian belt of the Bohemian Massif (Figure 1). In contrast to

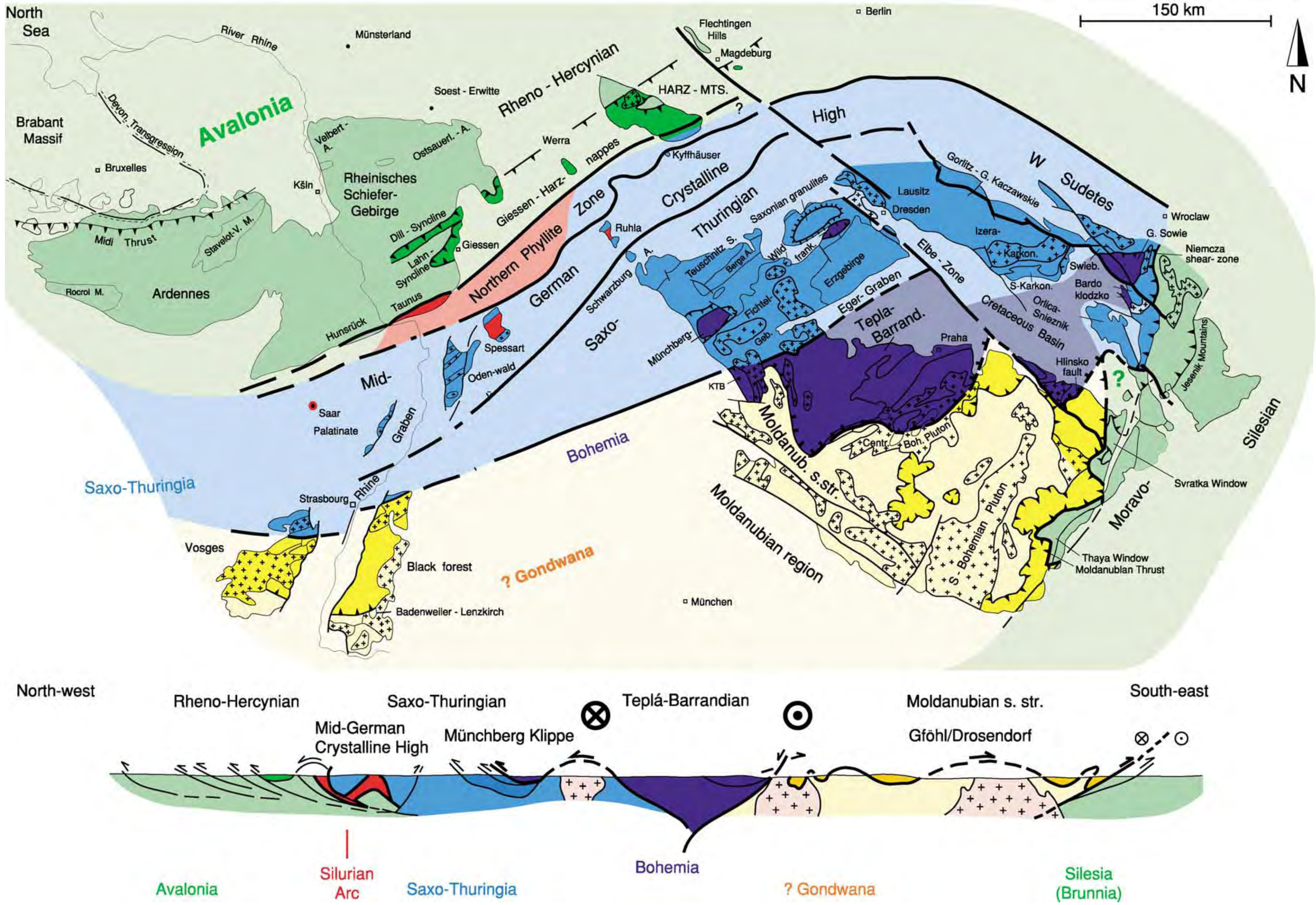


Figure 5 (A) Tectonic and palaeogeographical subdivision of the Variscan basement in central Europe. (B) Simplified tectonic cross section. A, Anticline; KTB, German Continental Deep Drilling site; M, Massif; S, South (as in S Bohemian Pluton); Teplá Barrand, Telpá Barrandian; Centr. Boh. Pluton, Central Bohemian Pluton; S karkon, South Karkonasze.

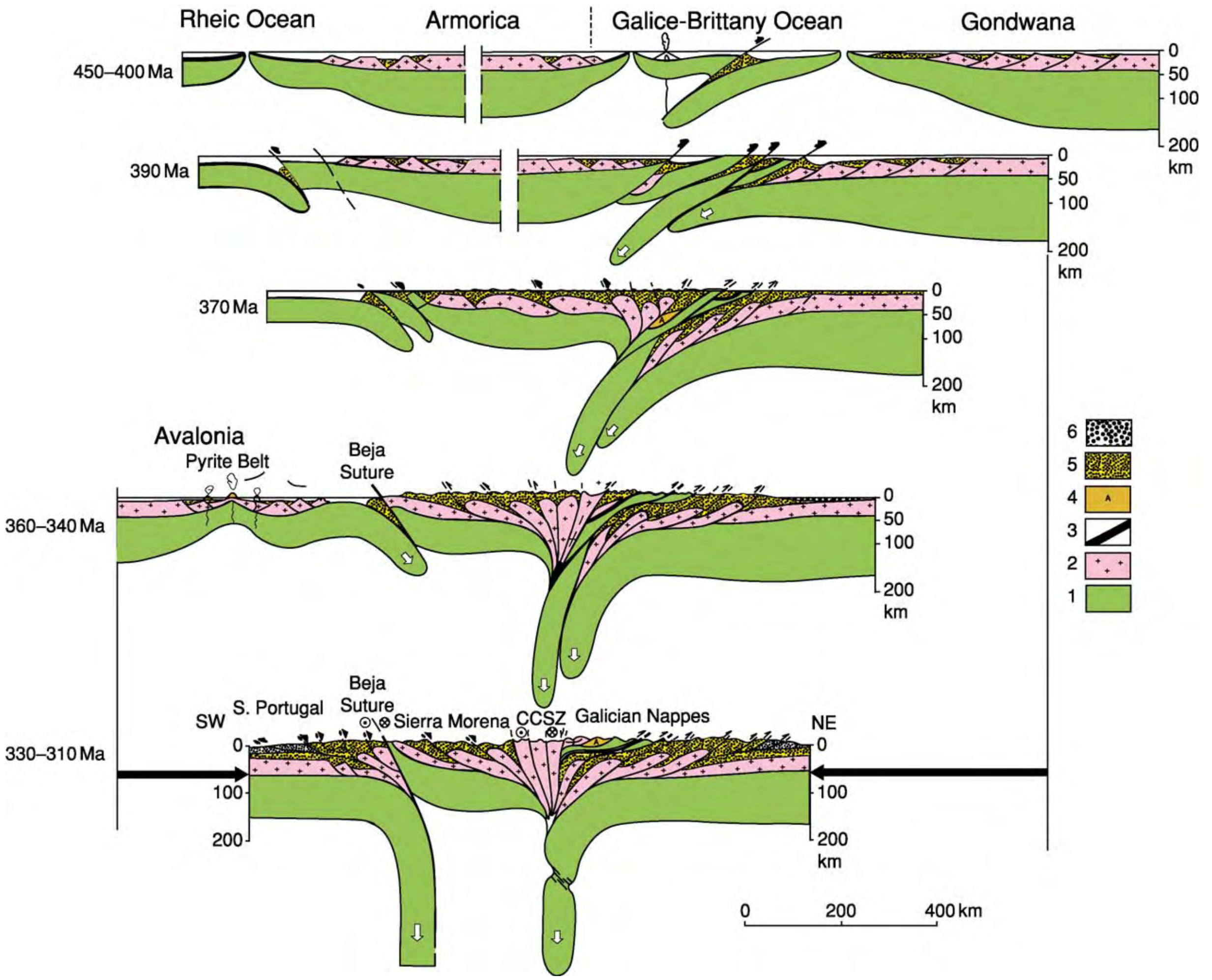


Figure 6 Plate kinematic evolution of the Variscan belt in Iberia: 1, mantle; 2, continental crust; 3, oceanic crust; 4, volcanic island arc; 5, Early Palaeozoic sediments; 6, Carboniferous foreland sediments. CCSZ, Coimbra Cordoba Shear Zone.

the situation in the central European section, there appears to be only one Armorican microcontinent (instead of Saxo-Thuringia and Bohemia). Nappes and sutures of the Ibero-Armorican domain are characterized by oceanic lithospheric rocks and early (420–370 Ma) ultrahigh-pressure, high-pressure and medium-pressure metamorphism. The most external parts of the belt comprise low-grade to non-metamorphic Carboniferous marine to paralic basins, which were deformed between 320 Ma and 290 Ma. The structures and tectonothermal histories of the Iberian and French sections of the Variscan Ibero-Armorican arc are described below (Figures 6 and 7). The Iberian section (Figures 6 and 7A) combines a southern segment from the South Portuguese Zone to the Central Iberian Zone and a northern segment based upon observations in north-western Spain (Figure 1). The northern segment originated from the closure of the Galicia-Brittany–Massif Central–Moldanubian ocean. Westward-directed subduction (in present-day coordinates) under an Armorican terrane affected first oceanic and then continental rocks between about 400 Ma and 300 Ma. Collision produced an eastward-facing accretionary wedge, in which deformation migrated from west to east. The lower autochthonous part consists of Early Palaeozoic shallow-water sediments deposited on the margin of Gondwana. The upper allochthonous part includes, from bottom to top, rocks of a passive margin thinned during the Ordovician, ophiolitic rocks (400–480 Ma) of arc or oceanic mantle, representing the remnants of the Galicia-Brittany–Massif Central ocean, and an uppermost ultranappe, which may represent the extended eastern margin of Armorica.

The southern segment of the Iberian section shows an opposite orogenic polarity, with south-west-facing folds and thrust sheets. Four main units are separated

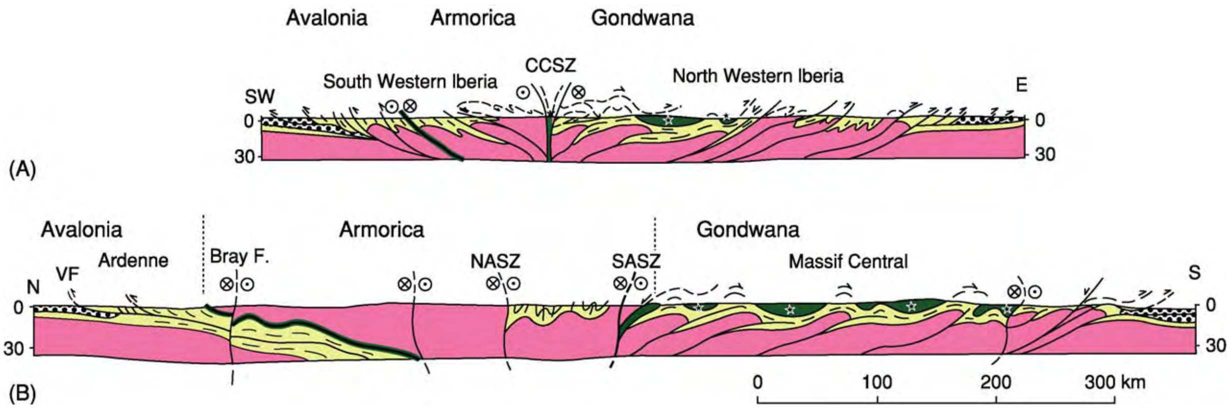


Figure 7 Simplified tectonic cross sections through the Variscan belt in (A) the Iberian peninsula and (B) France: pink, pre Palaeozoic basement and early Palaeozoic granitoids (undifferentiated); yellow, Early Palaeozoic sediments; green, thrust sheets with Early Palaeozoic ophiolites; dots, Carboniferous foreland basin deposits; VF, Variscan Front; Bray F, Bray Fault, NASZ, North Armoricn Shear Zone; SASZ, South Armorican Shear Zone; CCSZ, Coimbra Cordoba Shear Zone.

by major faults and/or sutures and are discussed below.

The Central Iberian Zone is the southern continuation of the autochthonous series of the northern Iberian section, with Early Palaeozoic rocks unconformably overlying low-grade Proterozoic sediments. A very important intracontinental sinistral shear zone separates the Central Iberian Zone from the Ossa Morena Zone and marks the boundary between tectonic domains with north-east and south-west vergences. This shear zone either represents or overprints the suture zone in which the north-west Iberian allochthon is rooted. There is evidence of very peculiar Ordovician magmatism with alkaline to peralkaline gneisses.

The Ossa Morena Zone shows a relatively complete fossiliferous Early Palaeozoic record from the lowermost Cambrian through to the Silurian and Devonian. Lower Cambrian sandstones and limestones unconformably overlie Upper Proterozoic rocks that are characterized by the presence of black cherts metamorphosed and intruded by Cadomian granitoids (550–500 Ma) as in northern Brittany. In some places (e.g. Cordoba) calc-alkaline andesites erupted during the latest Proterozoic/Early Cambrian. The Ossa Morena Zone shows large south-west-facing recumbent folds emplaced before the Early Carboniferous (which is, in part, coal-bearing). The Variscan metamorphism (epizonal to catazonal) is of a high-temperature–low-pressure type, and an important Variscan magmatism (diorites and granodiorites) is present. The present-day southern boundary of the Ossa Morena Zone is considered to be a suture, based on the existence of a belt of oceanic amphibolites (Beja-Acebuches) thrust over a possibly Devonian accretionary prism with slices of oceanic metabasalts. This north-west–south-east-trending boundary with the South Portuguese Zone has been reworked as a sinistral wrench fault.

The South Portuguese Zone exposes only Devonian and Carboniferous sediments and is characterized by important bimodal Tournaisian volcanic deposits, which contain the largest copper ore bodies in Western Europe (the South Portuguese Pyrite Belt). Time-equivalent volcanic rocks are known from south-west England and are widespread in the Rheno-Hercynian Belt of Germany. The Rheno-Hercynian and, hence, Avalonian affinities of the South Portuguese Zone are also suggested by a thick turbidite sequence deposited in a foreland basin ('Culm' facies), with a change towards paralic deposits in the south-west. It comprises the Viséan to Westphalian D. The South Portugese Zone is characterized by south-west-facing folds and thrusts. Deformation occurred during the Late Carboniferous.

The French section (see Figure 1 for location) can be subdivided into three segments. The Massif Central segment is the most complete section on the southern flank of the Variscan belt and is exposed over a distance of 400 km from the Montagne Noire to the southern margin of the Paris Basin. The Paris Basin segment is buried beneath Mesozoic sediments, but has been sampled in rare drillings and imaged at depth by a wide-angle reflection profile. Potentially equivalent rocks are exposed to the west in central Brittany. The Ardennes section is partly buried beneath the sediments of the Paris Basin, but is well known at depth from drilling, coal mining, and a deep seismic-reflection profile. Approximately 100 km of section are well exposed along the Meuse River in the Ardennes.

In the Massif Central segment, three main tectonostratigraphical units may be distinguished. First, the Montagne Noire at the southern extremity of the Massif Central consists of fossiliferous Palaeozoic sediments including Lower Cambrian shallow-water sandstones and limestones, Lower Ordovician shales and sandstones, Devonian limestones, and a thick turbiditic syntectonic Visean–Namurian series. The whole of this lithostratigraphical sequence is involved in large (10 km) recumbent folds facing south to south-west. The lowermost para-autochthonous unit (Zone Axiale) consists of Upper Proterozoic sediments that were intruded by Ordovician granites at 450–460 Ma. Deformation and low-pressure metamorphism occurred between 330 Ma and 310 Ma.

Second, the 'Schistes des Cévennes-Albigeois' are a very thick (possibly around 4000 m) greenschist-grade series, probably derived from both Early Palaeozoic and Late Proterozoic protoliths. They were intruded by Cambrian–Ordovician diorites and granites at 540–460 Ma. The Cévennes series shows a northwards gently dipping slaty cleavage related to a pervasive southward or south-westward shearing. Barrovian-type metamorphism increases towards the top of the pile (inverted). Deformation and metamorphism occurred between 350 Ma and 340 Ma. Large granitic plutons were emplaced between 330 Ma and 305 Ma.

Third, the complex Leptyno-Amphibolitic Group consists of mafic and ultramafic rocks characterized by high-pressure to ultrahigh-pressure metamorphism (in places they consist of coesite-bearing eclogites). It is a very large (300 km) nappe with ophiolite fragments. The probable root zone in the southernmost part of the Paris Basin is marked by a significant positive gravity anomaly. The Leptyno-Amphibolitic Group is overlain by pelitic gneisses with slices of high-pressure granulites and peridotites. Metamorphism has been dated at 420 Ma (high-pressure) to 380 Ma (Barrovian). All these units are intruded by various types of Variscan granites, which were emplaced between 360 Ma and 300 Ma.

The Ardennes segment is well documented from drilling, coal-mining, and surface outcrops in the Ardennes. Like the more easterly Rhenish Massif, this segment represents the northern flank of the European Variscan Belt. Palaeozoic fossiliferous sediments with a thick Devonian clastic wedge are involved in a large north-west-facing fold and thrust belt, which has been well imaged by ECORS (France) and DEKORP (Germany) deep reflection profiles. The main frontal thrust (Faille du Midi), exposed at the France–Belgium border, carries Devonian rocks over the Late Carboniferous coal basin. Deformation occurred at around 300 Ma.

Features Characteristic of the Variscan Belt

The central parts of the Variscan belt are intruded by huge volumes of granite. In contrast to the situation in many other orogenic belts, most of these granites were not formed over oceanic subduction zones but originated from the melting of metasediments in the continental crust. Heating and melting of the crust probably occurred initially when the crust was thickened (two mica and cordierite leucogranites) and then later when subducted parts of the subcontinental lithospheric mantle became detached and sank back into the asthenospheric mantle. This permitted the upward ascent of hot asthenospheric mantle and advection of heat to the crust.

The Variscan crust is rich in unstable isotopes (mainly potassium, uranium, and thorium), whose decay may produce up to 30% of the heat flow observed in continental rocks. These heat-producing isotopes were extracted from the mantle by repeated magmatic episodes, shortly before and during Variscan plate convergence (Cadomian subduction magmatism, Cambro-Ordovician rift magmatism, and subduction- and collision-related magmatism in the Devonian and Carboniferous). These elements are also contained in mica and feldspar, two main constituents of the thick Early Palaeozoic shelf deposits, which piled up during the collision of the major and minor plates.

For these reasons, the Variscan orogen was 'hot' in comparison with 'cold' orogens such as the Alps (*see* **Europe:** The Alps), the Caledonides (*see* **Europe:** Scandinavian Caledonides (with Greenland); Caledonides of Britain and Ireland), and the Urals (*see* **Europe:** The Urals). The high temperatures prevailing during continental collision effected mechanical weakening of the crust. This is documented by the 'squeezing out' of melts or low-viscosity metamorphic rocks towards the forelands. The same effect is responsible for the rapid destruction of the orogenic 'root'. When heated, the deeper parts of the thickened crust spread laterally like oil on water. Therefore, the base of the continental crust (the Mohorovicic discontinuity) had already levelled out at a depth of about 30–35 km by Late Carboniferous or Permian times, i.e. shortly after the termination of crustal thickening (*see* **Moho Discontinuity**). This process was aided by the ascent of mantle-derived melts during the Permian, which initiated the break-up of Pangaea (*see* **Pangaea**).

See Also

Europe: Caledonides of Britain and Ireland; Scandinavian Caledonides (with Greenland); The Urals; The Alps. **Moho Discontinuity**. **Palaeomagnetism**. **Palaeozoic:** Ordovician. **Pangaea**. **Tectonics:** Convergent Plate Boundaries and Accretionary Wedges; Mountain Building and Orogeny.

Further Reading

Burg JP, Leyreloup A, Marchand J, and Matte P (1984) Inverted metamorphic zonation and large scale thrusting in the Variscan Belt: an example in the French Massif Central. In: Hutton DHM and Sanderson PJ (eds.) *Variscan Tectonics of the North Atlantic Region*. pp. 47 61. Special Publication 14. London: Geological Society.

Fortey RA and Cocks LRM (2003) Palaeontological evidence bearing on global Ordovician Silurian continental reconstructions. *Earth Science Reviews* 61: 245 307.

Franke W (2000) The mid European segment of the Variscides: tectonostratigraphic units, terrane boundaries and plate tectonic evolution. In: Franke W, Haak V, Oncken O, and Tanner D (eds.) *Orogenic Processes: Quantification and Modelling in the Variscan Belt,* pp. 35 62. Special Publication 179. London: Geological Society.

Franke W and Stein E (2000) Exhumation of high grade rocks in the Saxo Thuringian Belt: geological constraints and geodynamic concepts. In: Franke W, Haak V, Oncken O, and Tanner D (eds.) *Orogenic Processes: Quantification and Modelling in the Variscan Belt,* pp. 337 354. Special Publication 179. London: Geological Society.

Franke W and Zelazniewicz A (2002) Structure and evolution of the Bohemian Arc. In: Winchester JA, Pharaoh TC, and Verniers J (eds.) *Palaeozoic Amalgamation of Central Europe,* pp. 279 293. Special Publication 201. London: Geological Society.

Kossmat F (1927) Gliederung des varistischen Gebirgsbaues. *Abhandlungen des Sächsischen Geologischen Landesamtes, Neue Folge* 1: 1 39.

McKerrow WS, MacNiocaill C, Ahlberg PE, *et al.* (2000) The late Palaeozoic relations between Gondwana and Laurussia. In: Franke W, Haak V, Oncken O, and Tanner D (eds.) *Orogenic Processes: Quantification and Modelling in the Variscan Belt,* pp. 9 20. Special Publication 179. London: Geological Society.

Martinez Catalan JR, Arenas R, Diaz Garcia F, and Abati J (1997) Variscan accretionary complex of northwestern Iberia: terrane correlation and succession of tectonothermal events. *Geology* 25: 1103 1106.

Matte P (1998) Continental subduction and exhumation of HP rocks in Paleozoic belts: Uralides and Variscides. *Journal of the Geological Society of Sweden* 120: 209 222.

Matte P (2001) The Variscan collage and orogeny (480 290 Ma) and the tectonic definition of the Armorica microplate: a review. *Terra Nova* 13: 122 128.

Owen AW, Harper DAT, and Rong Jia Yu (1991) Hirnantian trilobites and brachiopods in space and time. In: Barnes CR and Williams SH (eds.) *Ordovician Geology,* pp. 179 190. Ontario, Canada: Geological Survey of Canada.

Robardet M, Verniers J, Feist R, and Paris F (1994) Le Paléozoique anté varisque de la France, contexte paléogéographique et géodynamique. *Géologie de la France* 3: 3 31.

Schönlaub HP (1992) Stratigraphy, biogeography and paleoclimatology of the Alpine Paleozoic and its implications for plate movements. *Jahrbuch der Geologischen Bundesanstalt* 135: 381 418.

Scotese CR, Boucot AJ, and McKerrow WS (1999) Gondwanan palaeogeography and palaeoclimatology. *Journal of African Earth Sciences* 28: 99 114.

Simancas JF, Carbonell R, Gonzalez Lodeiro F, *et al.* (2003) The crustal structure of the transpressional Variscan orogen of SW Iberia: the IBERSEIS deep seismic reflection profile. *Tectonics* 22. DOI 10.1029/2002TC001479.

Stampfli GM (1996) The intra alpine terrain: a Paleotethyan remnant in the Alpine Variscides. *Eclogae Geologicae Helvetiae* 89: 13 42.

Suess E (1888) *Das Antlitz der Erde*, vol. IV. Prague, F. Tempsky.

Tait J, Schätz M, Bachtadse V, and Soffel H (2000) Palaeomagnetism and Palaeozoic palaeogeography of Gondwana and European terranes. In: Franke W, Haak V, Oncken O, and Tanner D (eds.) *Orogenic Processes: Quantification and Modelling in the Variscan Belt,* pp. 21 34. Special Publication 179. London: Geological Society.

Torsvik TH, Smethurst MA, Meert JG, *et al.* (1996) Continental break up and collision in the Neoproterozoic and Palaeozoic: a tale of Baltica and Laurentia. *Earth Science Reviews* 40: 229 258.

Van der Voo R (1993) *Paleomagnetism of the Atlantic, Tethys and Iapetus Oceans.* Cambridge: Cambridge University Press.

Wegener A (1915) *Die Entstehung der Kontinente und Ozeane*. Braunschweig: Vieweg.

The Urals

D Brown, Instituto de Ciencias de la Tierra 'Jaume Almera', CSIC, Barcelona, Spain
H Echtler, GeoForschungsZentrum Potsdam, Potsdam, Germany

Introduction

The Uralide Orogen was one of the main mountain belts built during the Palaeozoic assembly of the supercontinent Pangaea. Since the breakup of Pangaea in the Mesozoic, the Uralides have remained intact and are today located in the interior of the Eurasia Plate. The current extent of the orogen is seen by its roughly north- and south-oriented magnetic signature, which abruptly interrupts that of Baltica, Kazakhstan, and Siberia, the tectonic plates that collided to form the Uralides (Figure 1A). Most of what is known about the Uralide orogen is confined to the present-day Ural Mountains, a narrow range of low to moderate topography extending for nearly 2500 km from near the Aral Sea in the south to the islands of Novaya Zemlya in the Arctic Ocean (Figure 1B). East and south of the Ural Mountains, much of the orogen is buried beneath Mesozoic and Cenozoic sediments of the West Siberian and Precaspian basins and is not well known. For descriptive purposes, the geology of the Uralide Orogen has been divided into a number of longitudinal zones that are largely based on the ages and palaeogeography of the dominant rocks within them. From west to east, these zones are the Pre-Uralian zone, the West Uralian zone, the Central Uralian zone, the Magnitogorsk–Tagil zone, the East Uralian zone, and the Trans-Uralian zone (Figure 1C). Additionally, the Uralides have been divided geographically into the South, Middle, North, Cis-Polar, and Polar Urals; the following discussions focus on the South and Middle Urals (Figure 2).

The Pre-Uralian, West Uralian, and Central Uralian zones, which together make up the western foreland thrust-and-fold belt, contain Late Carboniferous to Early Triassic sediments of the foreland basin, Palaeozoic platform and slope sediments of the Baltica margin, and Archaean and Proterozoic rocks of the East European Craton (that part of the cratonic nucleus of Baltica that took part in the Uralide orogeny) (*see* **Europe:** East European Craton). The Magnitogorsk–Tagil zone is made up of two volcanic arcs, the Magnitogorsk Arc (South Urals), which is composed of Lower Devonian to Middle Devonian basalts that are overlain by Upper Devonian volcanoclastic sediments, and the Tagil Arc (Middle Urals), which is composed of Silurian to Lower Devonian basalts and volcanoclastic sediments that are locally overlain by Lower and Middle Devonian sediments. The East Uralian zone is composed predominantly of deformed and metamorphosed volcanic arc fragments with minor amounts of Precambrian and Palaeozoic rocks thought to represent continental crust. The East Uralian zone was extensively intruded by Carboniferous and Permian granitoids, forming the 'main granite axis' of the Uralides. The Trans-Uralian zone is composed of Devonian and Carboniferous volcanic and plutonic complexes overlain by terrigenous redbeds and evaporites. Ophiolitic material (oceanic crust) and high-pressure rocks have also been reported.

Tectonic Evolution

The tectonic evolution of the Uralides (Figure 3) began during the Devonian as intra-oceanic subduction formed the Magnitogorsk and Tagil island arcs (island arcs form along convergent margins when one oceanic plate subducts beneath another oceanic plate). This was followed by the entry of the continental margin of Baltica into the subduction zone and the emplacement of an accretionary complex over the continental margin, and the exhumation of high-pressure rocks (the process of bringing the rocks in question from depth toward the surface) along the arc–continent collision boundary. At the same time, subcontinental subduction (in which oceanic crust subducts beneath continental crust) and volcanic arc formation appear to have been taking place along the margin of Kazakhstan, and to have continued throughout the Carboniferous. Throughout much of the Carboniferous, there was a deformation hiatus along the Baltica margin, which by then included the accreted volcanic arcs, and shallow-water platform margin sedimentation continued undisturbed. By the latest Carboniferous to Early Permian, the Uralian ocean basin had closed completely and the continent–continent collision between the Kazakhstan and Baltica plates had begun. As this collision progressed through the Early Permian to the Early Triassic, the western foreland thrust-and-fold belt and foreland basin of the Uralides developed, while widespread strike–slip faulting, with exhumation of lower crustal material accompanied by melt generation and granitoid emplacement, took place in the interior part of the orogen. An episode of Early Triassic extension and volcanism followed in the eastern

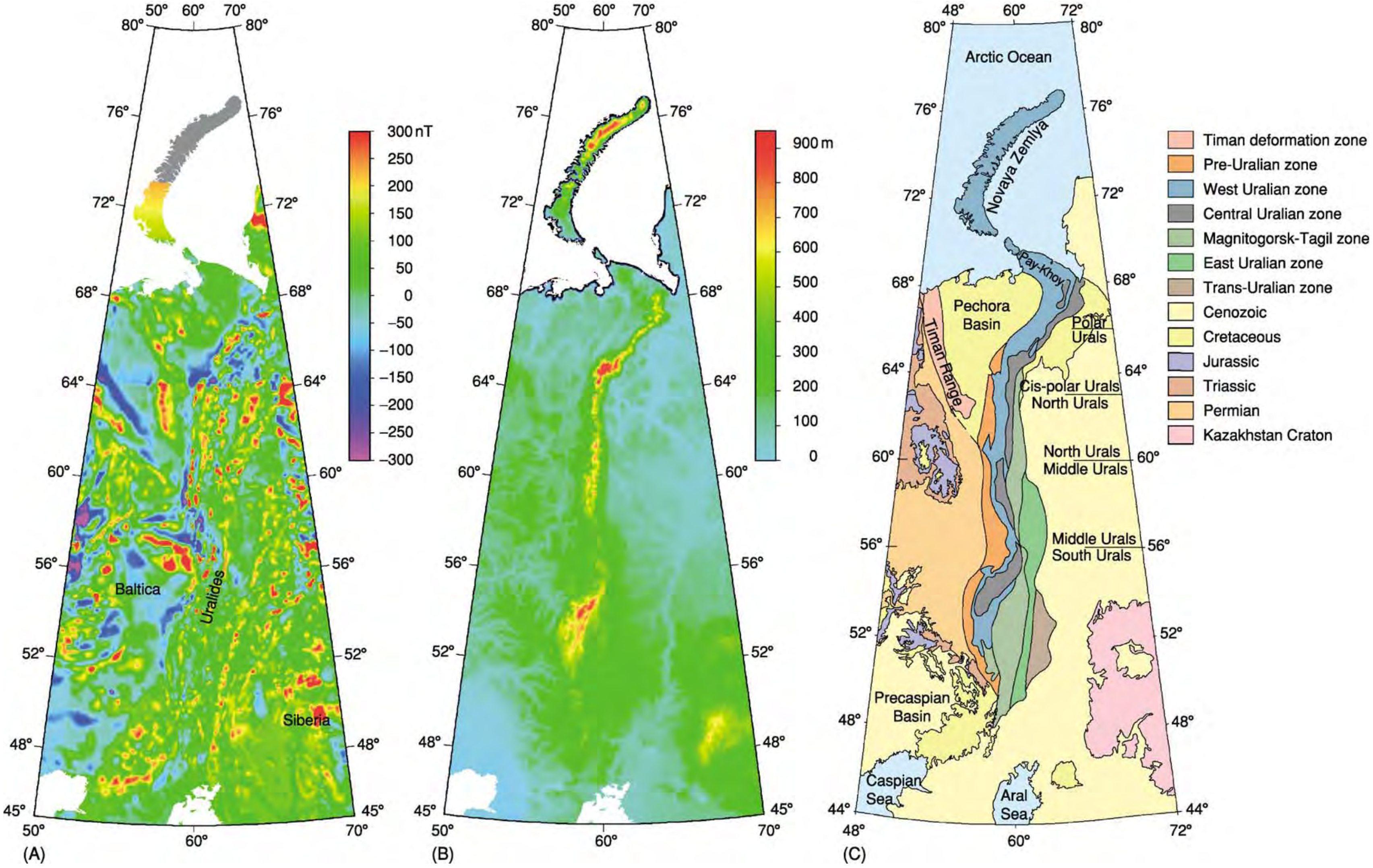


Figure 1 (A) Total field aeromagnetic map. The north east trending short wavelength anomalies represent the Uralides. Courtesy of National Geophysical Data Center. (B) Topography of the Ural Mountains. Note that the topography corresponds only to the western part of the orogen. Courtesy of National Geophysical Data Center. (C) Map of the Uralide Orogen outlining the extent of the different zones and geographical areas discussed in the text.

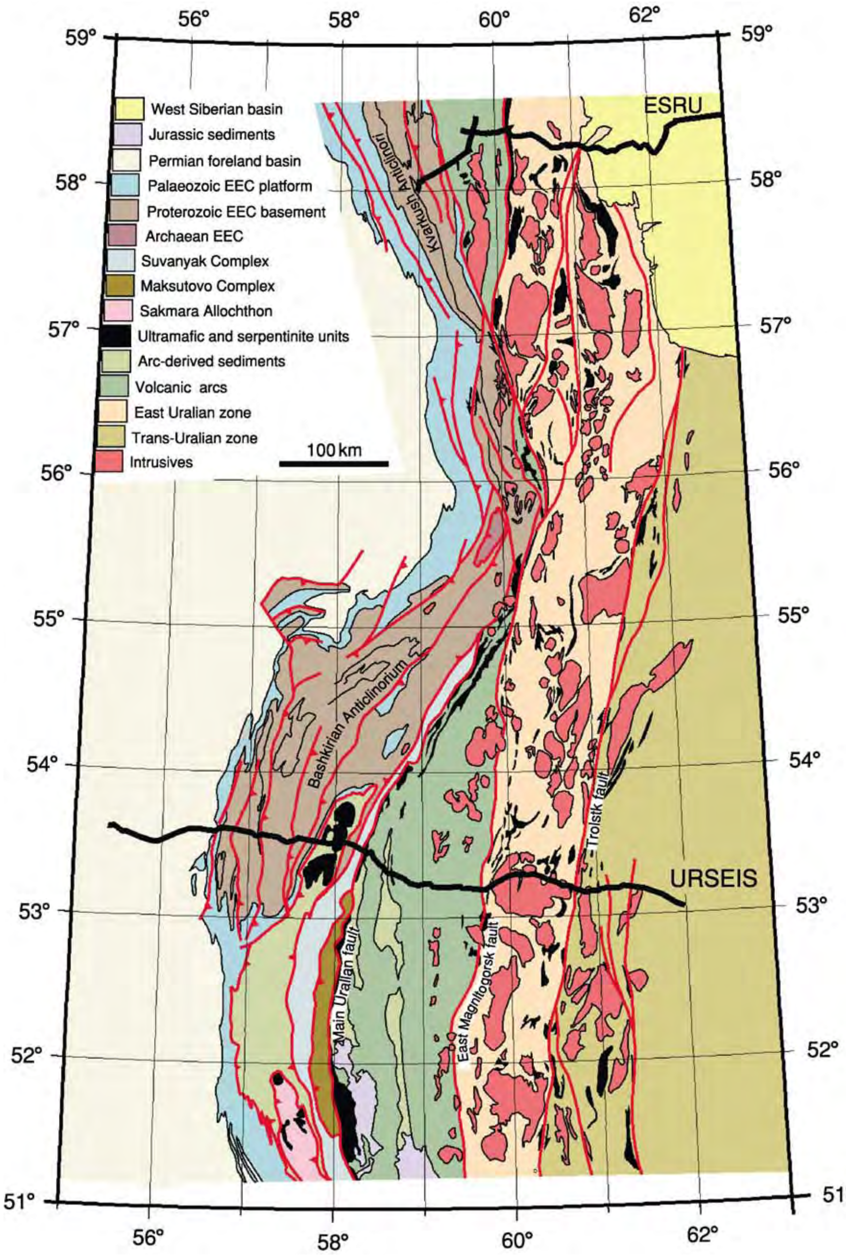


Figure 2 Geological map of the South and Middle Urals. The location of the Europrobe Seismic Reflection profiling in the Urals (ESRU) and Urals Seismic Experiment and Integrated Studies (URSEIS) transects are shown.

part of the Middle Urals and northward; by the mid-Jurassic, the intraplate Old Cimmerian deformation event led to the uplift of the Timan Range and formation of the Pay–Khoy–Novozemelian foldbelt in the northernmost Uralides (Figure 1C), and localized the deformation southward.

Arc–Continent Collision

Throughout geological time, intra-oceanic island arc development and its subsequent collision with a continental margin have been important processes in collisional orogenic belts, and among the most important means by which Earth's continental crust has grown. In the case of the Uralides, the Tagil and Magnitogorsk island arcs developed from the Silurian (Tagil) and Early Devonian (Magnitogorsk) and began to collide with the margin of Baltica in the Middle Devonian (Magnitogorsk) and the Early Carboniferous (Tagil) (Figure 3 shows this evolution schematically for the Magnitogorsk Arc). The Tagil Arc was pervasively deformed and metamorphosed to lower greenschist facies. Deformation in the

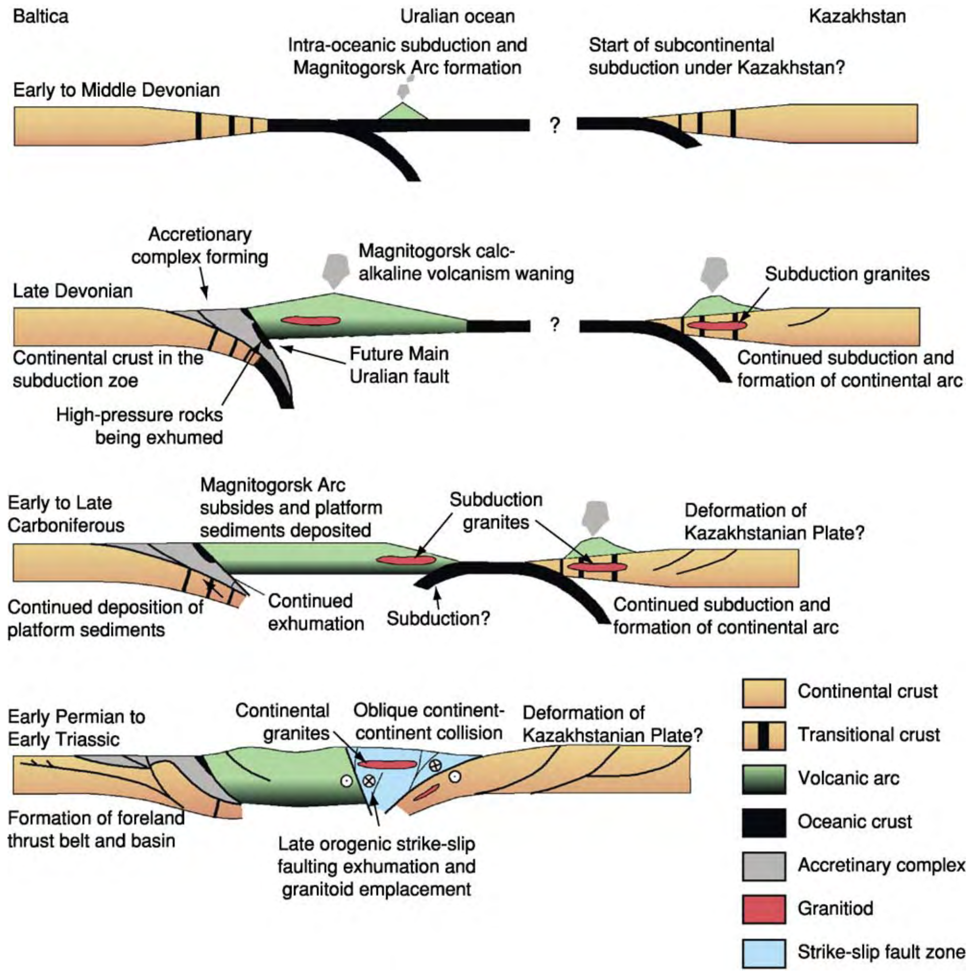


Figure 3 A simplified model for the tectonic evolution of the Uralides from the Early Devonian to the Early Triassic along the latitude of the Magnitogorsk Arc, highlighting the geodynamic processes that were active at each stage.

Magnitogorsk arc is low, with only minor, open folding and minor thrusting. The metamorphic grade is prenhite–pumpellyite facies. The Magnitogorsk arc–continent collision resulted in the development of an accretionary complex (Figure 3) that involved continental slope and platform sedimentary rocks (Suvanyak Complex) (Figure 2) that were detached from the margin of Baltica and thrust westward over the continental margin. The Suvanyak Complex is overthrust by syncollisional volcanoclastic sediments sourced from the accretionary complex and the Magnitogorsk arc (Zilair Formation). These units are flanked to the east by the high-pressure eclogite- and blueschist-bearing gneisses of the Maksutovo Complex (Figure 2), which records a peak metamorphic pressure of 20 ± 4 kbar (0.3 kbar is roughly equivalent to burial 1 km deep in Earth's crust) and temperature of 550 ± 50°C, and a peak metamorphic age of 380 to 370 Ma (during the Middle Devonian). The highest structural level of the accretionary complex is the Sakmara Allochthon in the south (Figure 2) and the Kraka lherzolite massif (a piece of oceanic mantle) in the north. The east-dipping Main Uralian fault, along which the arcs are sutured to the continental margin, is a fault melange that contains, among other things, several kilometre-scale fragments of oceanic crust and mantle.

Subcontinental Subduction

Little is known about what happened on the margin of the Kazakhstan Plate prior to or during its collision with Baltica, because no rocks that can be unequivocally assigned to the plate have been recognized in the Uralides. However, studies suggest subcontinental subduction and the development of a continental volcanic arc did occur. In part, the evidence for this comes from Silurian- to Devonian-age mafic to felsic gneisses and volcano-sedimentary rocks in the East Uralian zone that appear to represent a volcanic arc complex. The key piece of data for assigning these rocks to a continental arc setting is the presence of subduction-related granitoids that intrude into them. These granitoids are thought to have formed in two subduction zone settings. The first subduction related magmatism occurred from about 370 to 350 Ma (during the Late Devonian to Early Carboniferous) and produced granitoids with a recognizable older continental component. These granitoids are interpreted to have been related to the development of a continental arc during subduction of oceanic crust under the continental margin of the Kazakhstan plate. A second phase of subduction magmatism occurred from about 335 to 315 Ma (during the Early to Late Carboniferous) and produced granitoids with little, if any, continental component. These granitoids are interpreted to have been related to melting of the earlier continental arc during a subsequent or continued subduction beneath the Kazakhstan Plate margin. Magmatic activity directly related to subduction ended after the Carboniferous.

The Foreland Thrust-and-Fold Belt

The western foreland thrust-and-fold belt and foreland basin of the Uralides developed from the Late Carboniferous to the Early Triassic. The foreland thrust-and-fold belt trends roughly north and south and measures ~50 to 150 km in width from the Main Uralian Fault to the deformation front. Rocks involved in the thrust belt include Proterozoic and Archaean basement of the East European Craton, Palaeozoic platform sediments of the Baltica margin, the arc–continent collision accretionary complex, and the Permian to Early Triassic foreland basin sediments. Along its eastern margin, rocks in the Kvarkush anticline (Middle Urals) and the Bashkirian anticline (South Urals) were deformed in the Neoproterozoic, and structures were reactivated and the rocks were deformed again during the Uralide Orogeny. The structural architecture of the foreland thrust-and-fold-belt is that of a west-verging thrust stack developed above a basal detachment that lies in the Proterozoic basement. Where it has been determined by balanced cross-section restoration (the procedure of restoring the layers depicted in a cross-section to the positions they had prior to deformation), the Palaeozoic shortening is approximately 20 km or less.

Late Orogenic Strike–Slip Faulting

The internal part of the Uralides was extensively affected by a late orogenic strike–slip fault system that extends north and south for more than 700 km before it disappears beneath Mesozoic and younger sedimentary cover. Throughout much of the Middle and South Urals, this strike–slip fault system coincides with the East Uralian zone (see Figure 2), although the currently defined Main Uralian fault appears to be its western limit in the Middle Urals, so it therefore includes the Tagil Arc. Estimates of displacement along some strands of this fault system range from a few tens of kilometres to more than 100 km. Isotopic dating on one segment of the fault system indicates an age of 247 to 240 Ma (during the Late Permian to Early Triassic) for the development of fault-related mylonites, and 305 to 291 Ma (during the latest Carboniferous) for associated metamorphic rocks. The late orogenic strike–slip fault system was extensively intruded by continental-type granitoids, first in the southern part, from 292 to 280 Ma (from the latest Carboniferous to earliest Permian), and then in the northern part, from 270 to 250 Ma (during the Late to Early Permian). These granitoids have an unusually primitive Sr and Nd isotopic composition that is thought to have resulted from remelting of the older continental arc.

Crustal Structure

The crustal structure of the Uralides has been determined from the integration of surface geology with a variety of geophysical data, including reflection and refraction seismic surveys, potential fields (gravity and magnetics), and the thermal regime. The Europrobe Seismic Reflection profiling in the Urals (ESRU) and the Urals Seismic Experiment and Integrated Studies (URSEIS) experiments, together with Russian reflection and refraction seismic surveys, and the Urals Wide-Angle Reflection Seismics (UWARS) experiment, provide a large dataset for interpreting the crustal structure of the Uralides. These data show that the Uralides still preserves its bivergent collisional structural architecture and confirms the existence of a crustal root along the central axis of the orogen (Figure 4). The crustal structure has been determined by physical properties along the ESRU and, in particular, the URSEIS transects (Figure 2).

Reflection and refraction seismic data show that the Uralide crust, in the East European Craton, thickens eastward, from ~40 to ~52 km to between 50 and 55 km, across the volcanic arcs, before thinning again to between 40 and 45 km in the easternmost part of the orogen. In the South (URSEIS) and Middle (ESRU) Urals, the East European Craton part of the Uralide crust is imaged by subhorizontal to east-dipping reflectivity that can be related to its Uralide and older orogenic events (Figure 5). In both datasets the East European Craton extends eastward beneath the Magnitogorsk–Tagil zone. The Magnitogorsk (URSEIS)–Tagil (ESRU) zone displays moderate to weak upper crustal reflectivity, but the middle and

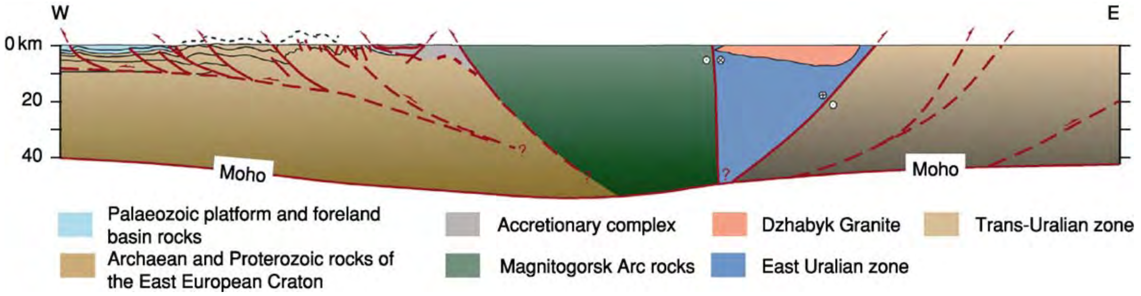


Figure 4 Schematic crustal cross section along the URSEIS profile (see Figure 2), showing the bivergence of the Uralides and the crustal root beneath the Magnitogorsk arc. The geometry of the western foreland thrust and fold belt is constrained by surface geology and reflection seismic and borehole data, and has been balanced and restored.

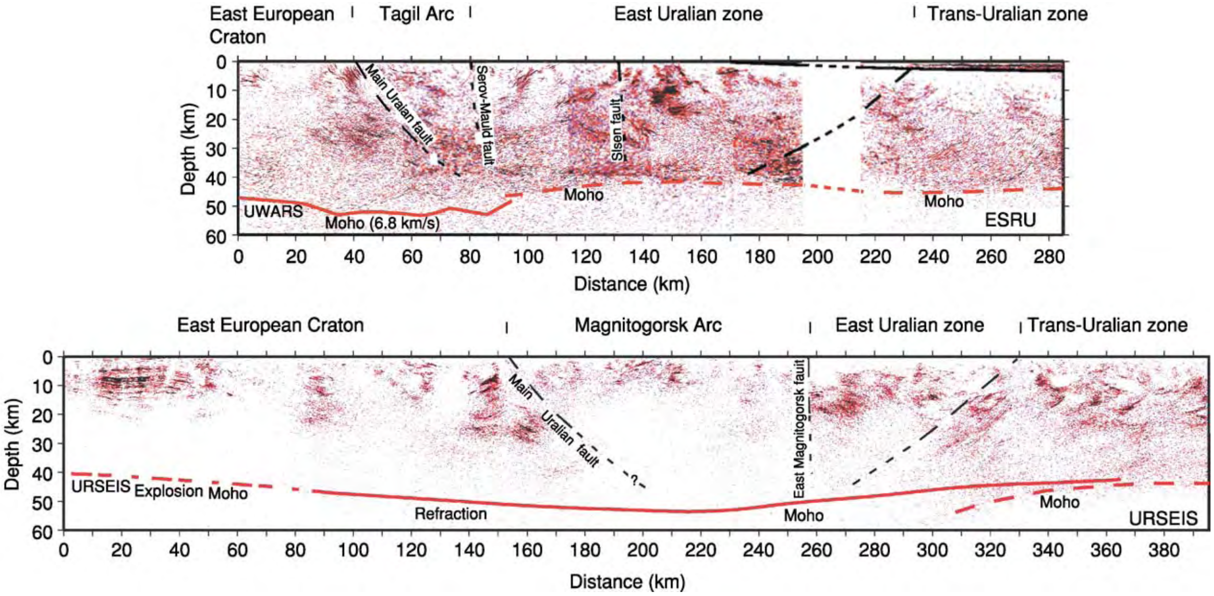


Figure 5 Automatic line drawings of the ESRU and URSEIS reflection seismic profiles, with the major crustal boundaries and the Moho indicated. These profiles clearly show the bivergent collision architecture of the Uralide orogen and the thickening of the crust towards its central axis.

lower crust reflectivity is diffuse in the case of the Magnitogorsk Arc, but quite strong in Tagil. The crust-to-mantle transition, or Moho, is not imaged beneath the Magnitogorsk zone, but is a fairly sharp transition beneath the Tagil zone. East of the arc complexes, the upper and middle crust region is imaged as clouds of diffuse reflectivity interspersed with, or cut by, sharp, predominantly west-dipping reflections that extend from the middle part of the crust into the lower crust, where it appears to merge with the Moho. In the Middle Urals (ESRU), it is characterized by abundant lower crustal reflectivity. In both the URSEIS and ESRU data, the Trans-Uralian zone dips westward beneath the East Uralian zone.

The velocity structure of the Uralide crust is best characterized along the URSEIS transect (Figure 6A and B). The upper crustal pressure wave velocities (V_p) reach up to 6.3 km s^{-1}, and the shear wave velocities (V_s) reach up to 3.9 km s^{-1}, with the higher values being in the Magnitogorsk Arc. In the middle and lower crust, V_p ranges from 6.5 to 6.8 km s^{-1}, reaching 7.1 km s^{-1} above the Moho in the central and eastern part of the transect; V_s ranges from 3.7 to 3.9 km s^{-1}, increasing to between 3.9 and 4.0 km s^{-1} at the Moho. The crust–mantle boundary is marked by an increase in V_p to $>$8.0 km s^{-1} and in V_s to $>$4.6 km s^{-1}. Using less resolved data, the Middle Urals appears to have a V_p structure similar to that of the URSEIS transect, although higher values (7.6–7.8 km s^{-1}) are reached near the Moho where the crust is thickest.

The Uralide heat flow density (HFD) is characterized by a strong minimum along the central part of the orogen (with values as low as 10 mW m^{-2}), reaching typical continental crustal values (up to

60 mW m^{-2}) on either side (Figure 7A). The short wavelength of the HFD anomaly is suggestive of a shallow origin for the minimum, although the heat production (k) data determined from surface samples are too high to allow the HFD minimum to be simulated along the URSEIS transect. This implies that more ultramafic material is present at depth beneath the Magnitogorsk zone, compared to below the East European Craton crust. Another option is that the HFD minimum is due to propagation of ground surface temperature changes to depth as a result of palaeoclimatic disturbances (such as recent climate change or glaciation). Modelling assuming the first alternative suggests that the Uralides thermal structure is characterized by relatively flat geotherms, with a Moho temperature of around 600°C (Figure 6C). This suggests that the root is not very cold and that the Magnitogorsk Arc rocks do not have a completely negligible heat production.

The Bouguer gravity anomaly in the South and Middle Urals (Figure 7B) is characterized by a low of between −60 and −45 mGal across the East European Craton, indicating upper and middle crustal densities of about 2.80 and 2.90 g cm^{-3} (with small local variations) and lower crustal densities of 2.98 and 3.02 g cm^{-3}. There is an abrupt increase in the Bouguer anomaly to between about 0 to −40 mGal in the Magnitogorsk–Tagil zone, falling to between about −70 and −40 mGal in the East Uralian zone, and about −30 to −10 mGal across the East Uralian zone. This indicates an upper crustal density of between 2.71 and 2.80 g cm^{-3} (somewhat higher in the Magnitogorsk zone), a middle crustal density of between 2.92 and 2.95 g cm^{-3}, and a lower crustal density of between 2.98 and 3.07 g cm^{-3}. The upper mantle has a density of 3.34 g cm^{-3}. The magnetic signature of the South and Middle Urals (Figure 7C) is characterized by short-wavelength features, with a long-wavelength low that reflects the magnetic character of the East European Craton. In the URSEIS transect, the magnetic crystalline basement is truncated about 50 km to the west of the Main Uralian fault, and magnetic susceptibilities of the rocks are 0 and 1.5 A m^{-1} to the east and west of this truncation, respectively. Magnetic

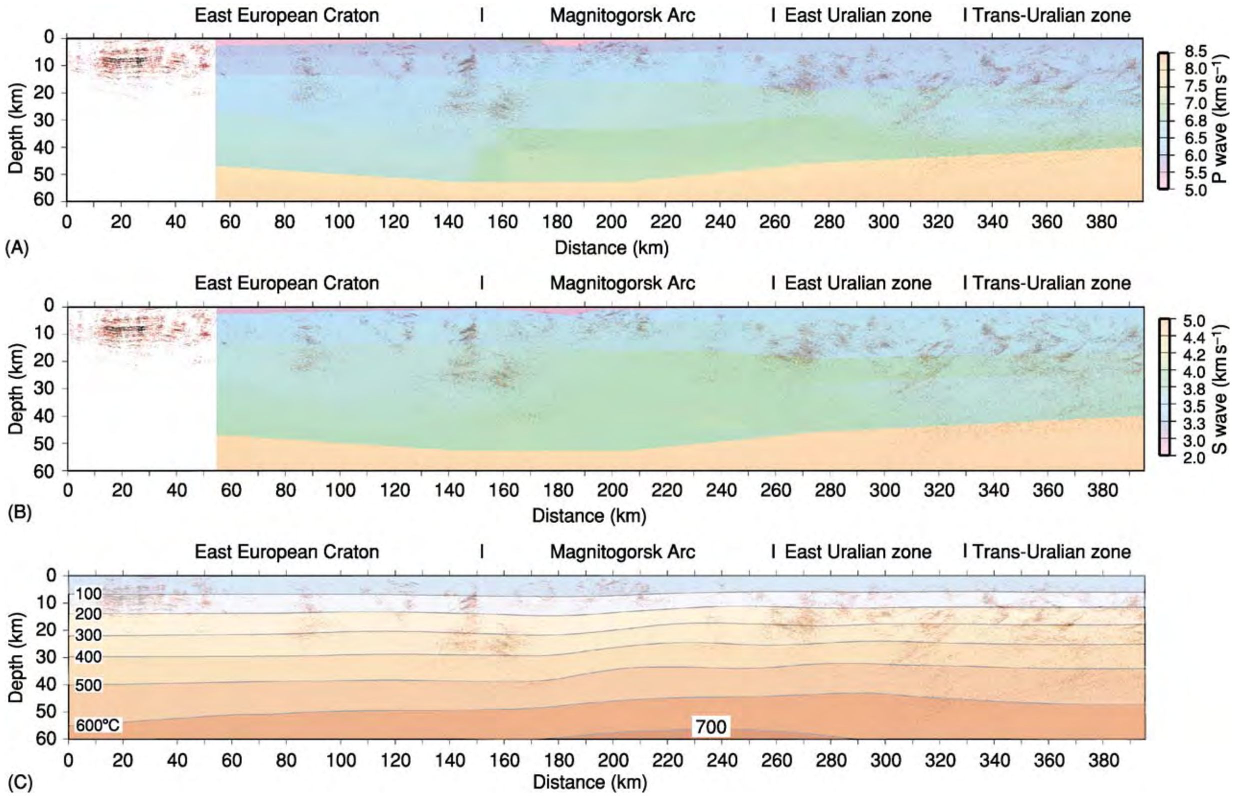


Figure 6 (A) The URSEIS reflection profile with the V_p data plotted over it; V_p is generally higher in the lower crust beneath the arc terranes. The location of the Moho is marked by the increase in V_p to greater than 8 km s^{-1}. (B) The URSEIS reflection profile with the V_s data plotted over it; V_s is generally higher in the middle and lower crust beneath the arc terranes. The location of the Moho is marked by the increase in V_p to greater than 4 km s^{-1}. (C) The URSEIS reflection profile with the present day geothermal gradient plotted over it. The temperature at the Moho is generally around 600°C, which suggests that the metamorphic grade of the Uralide crust currently does not reach granulite facies metamorphic conditions.

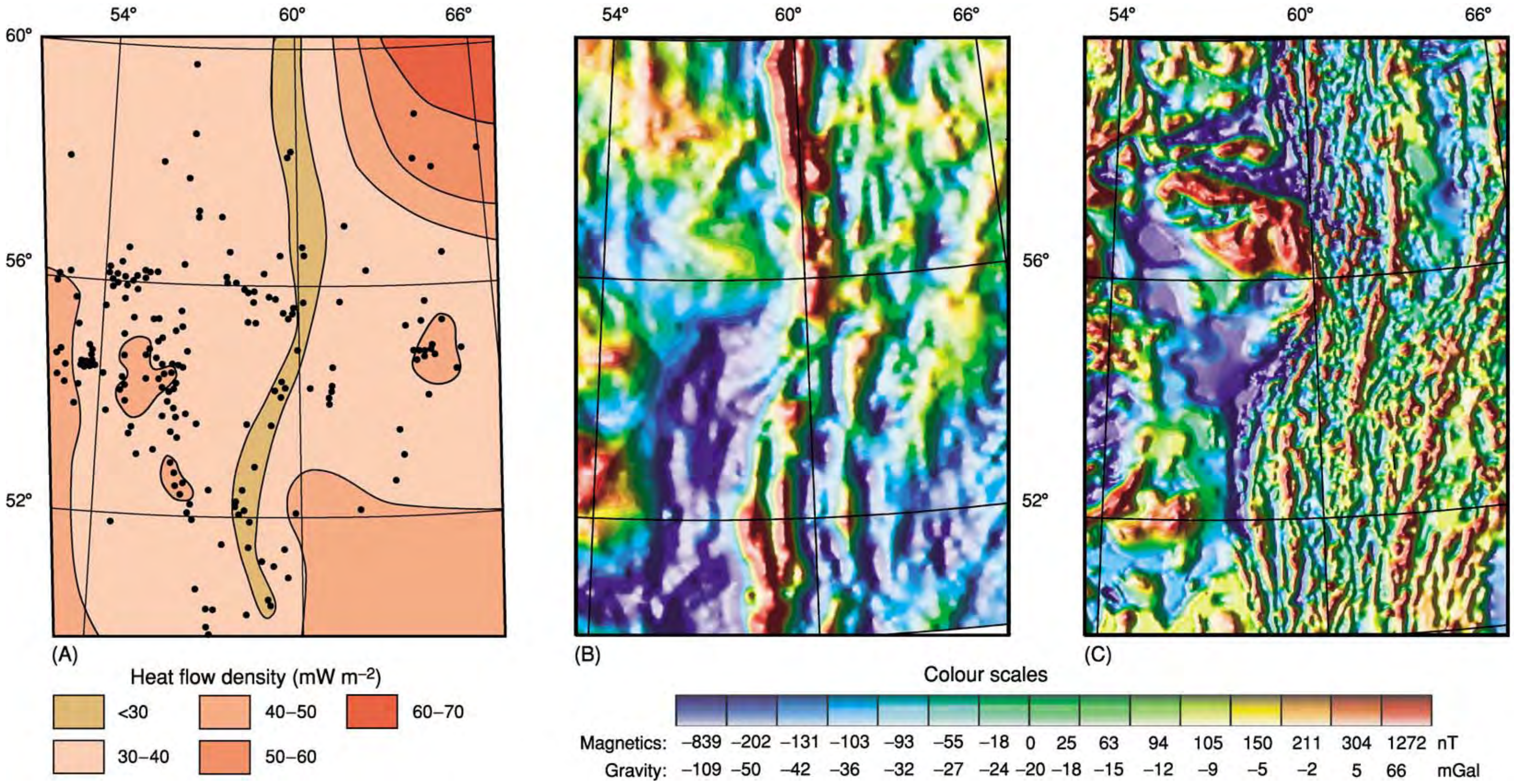


Figure 7 (A) Contoured heat flow density map of the Middle and South Urals. Dots represent data points. (B) Bouger gravity and (C) aeromagnetic maps of the Middle and South Urals. The maps are equal colour area and shaded relief, with illumination from the west. Maps courtesy of G Kimbell and C Ayala, BGS, British Geological Survey.

susceptibilities of the upper crust vary between 0.25 and 0.5 A m ^{1}, with areas reaching 2 and even 2.5 A m ^{1} locally. The middle and lower crust region has magnetic susceptibilities of 0.5 and 1.5 A m ^{1}.

Petrophysical modelling (assigning rock types based on the seismic velocity, density, and thermal data) of the Uralide crust along the URSEIS transect shows clear differences between the composition of the old continental crustal nucleus of the East European Craton and the newly added crust of the accreted arc terranes to the east. The crust of the East European Craton is more felsic than that of the Magnitogorsk and East Uralian zones, and the latter two zones have a lowermost crust with characteristics indicating a high garnet content (mafic garnet granulite) and/or the presence of hornblendite. The overall composition of the arc terranes is basaltic. The physical properties data suggest that eclogite is not present in the lower crust, or if present, it exists in such small amounts that it is below the resolution of the dataset.

Topography of the Ural Mountains

There is widespread evidence that much of the Uralides was eroded and peneplained by the Late Triassic to Jurassic. If the Uralides were peneplained by this time, when did the topography of the Ural Mountains, which locally reaches about 1900 m, form? The geomorphology of the South and Middle Ural Mountains is generally mature, being dominated by a system of smooth, north- and south-trending ridges, but with some younger features, such as deeply incised river valleys and elevated river terraces, that hint at recent uplift. The topography of the Ural Mountains is almost exclusively associated with the foreland thrust-and-fold belt and there is a strong correlation of topography with thrusts (thrusts lie in the valleys) (Figure 8). Low-temperature thermochronological studies using apatite fission track dating suggest that, with the exception of the Magnitogorsk Arc (which yields an age of 262 + 5 Ma, or Early Permian), much of the South and Middle Urals has been relatively stable since the Jurassic (ages range from 226 + 8 to 206 + 9 Ma, or Late Triassic) (Figure 8), and that since that time, very little cooling or erosion has taken place. However, there is a disturbance in the fission track age–altitude relationships across thrusts in the foreland thrust-and-fold belt, which suggests that some post-Jurassic reactivation has taken place along these faults and that the current relief of the Ural Mountains is therefore post-Uralide. (A fission track age taken from a higher altitude in the hanging wall of a thrust should be younger than one from the same stratigraphic unit at a lower level in the footwall, but this is not always the case in the Uralides.) Exactly why or when the recent relief formed is, however, not fully understood.

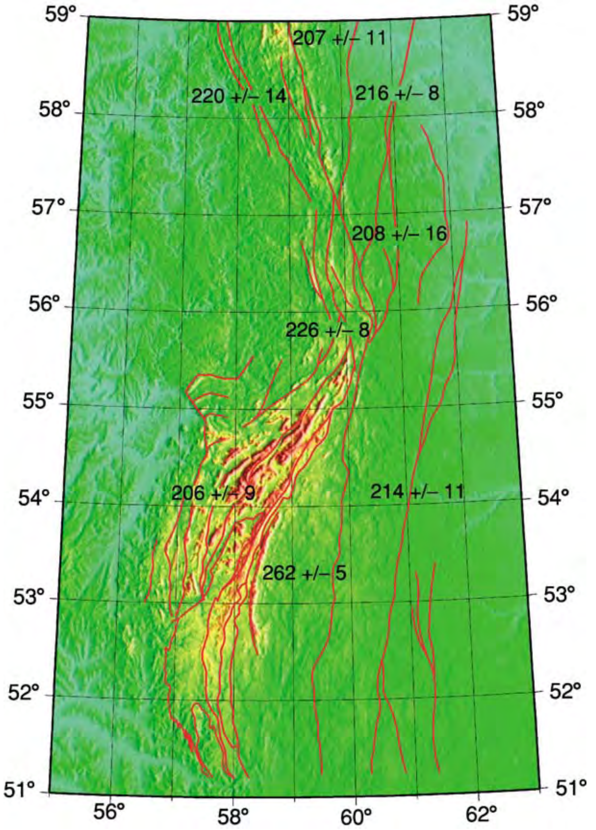


Figure 8 Topography of the South and Middle Urals, with average apatite fission track ages indicated (millions of years). Low temperature exhumation in much of the Uralides took place predominantly in the Late Triassic, although a notable exception is the Magnitogorsk Arc, where it took place in the Early Permian.

See Also

Europe: East European Craton; Timanides of Northern Russia.

Further Reading

Berzin R, Oncken O, Knapp JH, *et al.* (1996) Orogenic evolution of the Ural Mountains: results from an inte grated seismic experiment. *Science* 274: 220 221.

Brown D, Juhlin C, Alvarez Marron J, Perez Estaun A, and Oslianski A (1998) Crustal scale structure and evolution of an arc continent collision zone in the southern Urals, Russia. *Tectonics* 17: 158 170.

Brown D, Juhlin C, and Puchkov V (2002) *Mountain Building in the Uralides: Pangea to Present.* Geophysical Monograph 132. Washington, DC: American Geophysical Union.

Brown D, Carbonell R, Kukkonen I, Ayala C, and Golovonova I (2003) Composition of the Uralide crust from seismic velocities (V_p and V_s), heat flow, gravity, and magnetic data. *Earth and Planetary Science Letters* 210: 333 349.

Carbonell R, Perez Estaun A, Gallart J, *et al.* (1996) A crustal root beneath the Urals: wide angle seismic evidence. *Science* 274: 222 224.

Echtler HP, Stiller M, Steinhoff F, *et al.* (1996) Preserved collisional crustal architecture of the Southern Urals Vibroseis CMP profiling. *Science* 274: 224 226.

Glodny J, Bingen B, Austrheim H, Molina JF, and Rusin A (2002) Precise eclogitization ages deduced from Rb/Sr mineral systematics: the Maksyutov complex, Southern Urals, Russia. *Geochimica et Cosmochimica Acta* 66: 1221 1235.

Juhlin C, Friberg M, Echtler H, *et al.* (1998) Crustal structure of the Middle Urals: results from the (ESRU) Europrobe Seismic Reflection Profiling in the Urals Experiments. *Tectonics* 17: 710 725.

Knapp JH, Steer DN, Brown LD, *et al.* (1996) A lithosphere scale image of the Southern Urals from explosion source seismic reflection profiling in URSEIS '95. *Science* 274: 226 228.

Meyer FM, Kisters AFM, and Stroink L (1999) Integrated geologic studies along the URSEIS '95 transect: contributions to the understanding of the orogenic evolution of the southern Urals. *Geologische Rundschau* 87.

Perez Estuan A, Brown D, and Gee D (1997) Europrobe's Uralides Project. *Tectonophysics* 276.

Puchkov VN (1997) Structure and geodynamics of the Uralian orogen. In: Burg J P and Ford M (eds.) *Orogeny Through Time,* pp. 201 236. Special Publication 121. Oxford: Geological Society.

Savelieva GN and Nesbitt RW (1996) A synthesis of the stratigraphic and tectonic setting of the Uralian ophiolites. *Journal of the Geological Society* 153: 525 537.

Zonenshain LP, Kuzmin MI, and Natapov LM (1990) Uralian foldbelt. In: Page BM (ed.) *Geology of the USSR: A Plate Tectonic Synthesis,* pp. 27 54. American Geophysical Union Geodynamics Series 21. Washington, DC: American Geophysical Union.

Permian Basins

A Henk, Universität Freiburg, Freiburg, Germany
M J Timmerman, Universität Potsdam, Potsdam, Germany

Introduction

The Late Carboniferous to Early Permian (Stephanian to Rotliegend) evolution of western and central Europe was characterized by the formation of numerous sedimentary basins as well as widespread and voluminous magmatism in the area of the Variscan orogen (*see* **Europe:** Variscan Orogeny) and its northern foreland. This phase of intense extension and thermal perturbation of the lithosphere led to an almost complete destruction of the just-formed Variscan orogen. Within 20 My, after convergence in the external zone of the orogen had ceased, extension and erosion had already removed most of the surface topography and crustal roots, respectively, and a uniform Moho depth was restored in the area of the former orogen. Towards the end of the Permian, the crust had locally subsided even below sea-level as is indicated by the rapid transgression of the Zechstein Sea over part of the former orogen.

In this article, some general characteristics of the Permo-Carboniferous evolution in western and central Europe are first summarized. Subsequently, the fill and structure of the sedimentary basins, as well as the composition and petrogenesis of the magmatic rocks formed during this epoch, are treated in some detail. Finally, an overview of the geodynamic setting and the potential driving forces for extension – topics that have been controversially debated in past years – is given.

General Characteristics

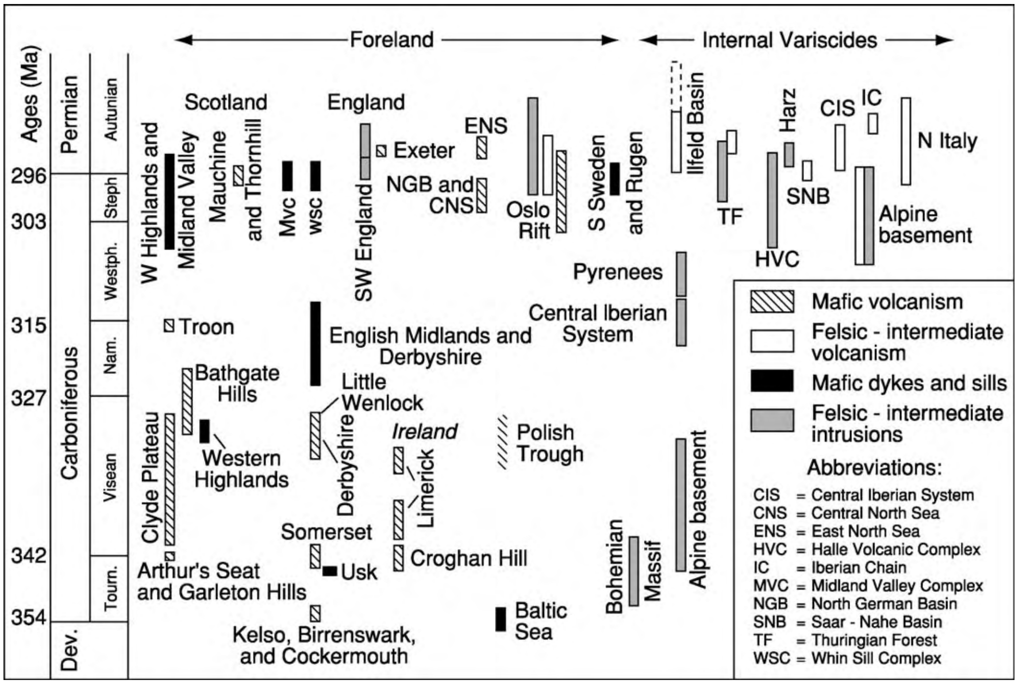
The end of convergence and the start of postconvergent evolution in the Variscan domain are documented by the youngest sediments that were still deformed by compressional movements. These occur in the most external parts of the Variscan fold belt (Ruhr basin) and are Westphalian D, about 305 Ma. However, the final stage of convergence was already accompanied by localized extension in the orogen's interior. This is documented by the rapid syn-collisional exhumation of metamorphic complexes, as well as the lateral extrusion of crustal blocks ('tectonic escape') and associated Namurian/Westphalian basin evolution next to major orogen-parallel strike–slip faults.

The subsequent Permo-Carboniferous (i.e., Stephanian to Early Permian) evolution of western and central Europe was largely controlled by a system of north-west- to south-east-trending strike–slip faults that intersected the Variscan fold belt and extended far into its foreland. Together with reactivation of older Variscan structures, movements along this fault system led to the formation of numerous transtensional half-grabens and pull-apart basins. Extension of the lithosphere and tectonic subsidence, as indicated by rapid subsidence and sediment accumulation in these fault-controlled basins, can be dated to between about 305 and 285 Ma. This correlates with a phase of intense magmatism, peaking between 300 and 285 Ma, which occurred almost synchronously over large parts of Europe – both inside and outside the area affected by Variscan orogenic processes. Subsequent thermal equilibration of the intensely disturbed European lithosphere led to thermal subsidence. During this stage, sedimentation overstepped the initially fault-controlled basin margins. The individual basins were connected and a coherent depositional area formed in large parts of Europe. At about 250 Ma, the crust had locally subsided below sea-level, as is indicated by the rapid transgression of the Zechstein Sea over part of the former orogen. Even the subsequent Mesozoic basin evolution in large parts of Europe was still controlled by thermal equilibration of the anomalies formed by Permo-Carboniferous geodynamic processes.

Basin Formation

Late Stephanian to Early Permian basin evolution started after a regional hiatus of Early Stephanian age. This hiatus is documented within the Variscan domain as well as in its northern foreland. There, Stephanian and Early Permian sediments and volcanic rocks unconformably overlie the Westphalian foreland basin sequence. In the external fold-and-thrust belt, corresponding sediments were deposited on very-low-grade to low-grade metasediments, whereas in the internal part of the Variscides, they were often directly deposited on amphibolite-facies metamorphic rocks and granite intrusions. This demonstrates that by Late Carboniferous times, the Variscides had already been deeply eroded. Only locally Stephanian sediments unconformably overlie Namurian to Westphalian precursor basins. These basins, such as the Saar-Nahe, Central Armorican Basin (i.e., Laval), and the basins east of the Elbe Line, are usually located at major fault zones that exhibit strike–slip movements and may be related to tectonic escape during the final stage of the Variscan Orogeny (see Figure 1).

The fill of the basins consists mainly of continental siliciclastic sediments (shales, sandstones, and conglomerates) and interbedded volcanic and pyroclastic deposits. During the tectonic subsidence stage, between 305 and 285 Ma, the sedimentary sequences frequently show a typical facies evolution from fluvial to lacustrine and back to fluvio-deltaic depositional environments. Deposition in fluviatile and playa-lake environments was common during the subsequent thermal subsidence stage. In some basins, the earlier deposits contain economically important coal measures (e.g., the Saar-Nahe, Saale, Autun, and Sillon Houillier basins). The gradual decrease of coal-bearing strata from the Late Carboniferous to the Early Permian reflects the northward drift of Gondwana from an equatorial setting. Due to the



▨ Mafic volcanism
☐ Felsic - intermediate volcanism
■ Mafic dykes and sills
▩ Felsic - intermediate intrusions

Abbreviations:

CIS = Central Iberian System
CNS = Central North Sea
ENS = East North Sea
HVC = Halle Volcanic Complex
IC = Iberian Chain
MVC = Midland Valley Complex
NGB = North German Basin
SNB = Saar - Nahe Basin
TF = Thuringian Forest
WSC = Whin Sill Complex

Figure 1 Age distribution of Permo Carboniferous magmatism in the Variscan domain.

resulting changes in climate, formation of substantial coal measures north of the Variscides had already ceased in the Westphalian D, whereas in the southern areas it continued until the end of the Stephanian.

The sediments and volcanic rocks accumulated during the syn-rift and subsequent thermal subsidence phases can locally reach total thicknesses of up to 8 km. However, it should be noted that the thicknesses presently observed are often only erosional remnants. At least in parts of the Variscan orogen it can be shown that a substantial part of the sediments deposited during the Stephanian to Early Permian was already eroded prior to the Zechstein transgression. In the Saar-Nahe Basin, for example, estimates based on shale compaction data and vitrinite reflectance modelling indicate removal of up to 3.7 km of sediments during this phase. Between these basins, the basement was exposed or was directly overlain by younger deposits.

The geometry and structural style of the individual basins were strongly controlled by the interplay of pre-existing structural elements and new faults that formed in conjunction with the dextral translation of Gondwana relative to Laurussia. A dextral strike–slip component is inferred from basin geometry and orientation, isopach maps of syn-rift deposits, migration direction of depocentres, the orientations and structural data from syn-sedimentary faults, simultaneous horst and basin formation, and the arrangement of volcanic centres and dykes. Thus, some of the basins trend obliquely to the structural grain of the orogen and intersect the Variscan deformation front, whereas others owe their formation to transtensional reactivation of older, compression-related faults. A prime example of the latter is the Saar-Nahe Basin in south-west Germany, which formed by oblique extensional reactivation of a crustal-scale fault structure, along which the internal zone of the Variscan orogen (Mid-German Crystalline Rise, part of the Saxothuringian Zone) was previously thrusted onto the external fold-and-thrust belt (Rhenohercynian Zone). Thus, the basins that were formed by transtensional reactivation of former thrusts frequently have half-graben structures. In contrast, pull-apart basins that opened during dextral strike–slip movements tend to be located next to the dominant north-west- to south-east-trending faults (continental-scale dextral shears parallel to the Tornquist–Teisseyre Line; e.g., the Elbe Line, Pays de Brays Fault, Bay of Biscay, Gibraltar-Minas, and Agadir fracture zones).

Several of the faults that controlled Permo-Carboniferous basin evolution were later reactivated, particularly in connection with the inversion of the Alpine foreland in the Late Cretaceous and Early Permian. These later events can sometimes obscure the Permo-Carboniferous basin geometry and fault displacement. Thus, identification of faults that were active in Permo-Carboniferous times requires careful analysis of indicators for syn-sedimentary tectonics, such as abrupt variations in thickness and facies architecture and syn-depositional deformation structures.

Magmatism

Late Carboniferous to Early Permian volcanic and plutonic activity of varying style and composition was coeval with extension and basin formation in both foreland and the internal Variscides, and occurred in crustal domains of various ages (Figure 1). Reliable U–Pb and $^{40}Ar/^{39}Ar$ mineral ages show that most of the activity occurred in the period 305 to 290 Ma. Magmatic rocks of mafic composition are common in the foreland, but are relatively rare in the internal Variscides. Layers of strongly altered air-fall tuffs, known as bentonites or tonsteins, occur in nearly every basin, but most of these originated from distal volcanic sources.

Extensional faulting took place before, during, and after volcanic activity. In the north-western part of the North German Basin, initial faulting caused the formation of horsts and grabens that strongly controlled the placement and thickness of the volcanic rocks. Faulting in the central North Sea mainly postdates the volcanic activity, whereas in the Oslo Rift, the main east and west extension was coeval with the main phase of trachyandesitic volcanism. In nearly all of the basins, lavas occur interbedded with sediments, showing that magmatic activity took place during subsidence and extension. The volcanic rocks are interbedded and alternate with Stephanian coal measures and Early to Late Permian, Rotliegend-facies alluvial and fluviatile clastic sediments, and lacustrine marls deposited in semiarid environments. Where not eroded, these are often followed by aeolian sandstones.

Foreland

In the foreland, a variety of volcanic and plutonic rocks are present in the Oslo Rift in south Norway, and large volumes of rhyolitic and andesitic volcanic rocks occur in northern Germany below younger cover and are known only from deep boreholes. More mafic magmatic activity resulted in the formation of dyke swarms and sills, such as the basaltic Whin Sill and the Midland Valley complexes in Great Britain, the basaltic dyke swarm in South Sweden, and the lamprophyre dyke swarms in the western Highlands of Scotland. Smaller volumes of mafic and felsic volcanic rock have been drilled in the central and eastern North Sea, and the Midland Valley of Scotland is known for its large number of alkaline mafic sills, vents, and subvolcanic intrusions that often contain megacrysts and xenoliths of mantle material and high-grade metamorphic crust.

Several stages of magmatism have been recognized in the Oslo Rift. During the initial stages, felsic intruded Westphalian sediments, accompanied or shortly followed by the extrusion of up to 1500 m of primitive, alkaline basalt to phonotephrite lavas in the southern parts of the rift. The subsequent main rift stage resulted in fissure eruptions of thick sequences of porphyritic trachyandesite flows that constitute the largest volume of volcanic rocks. This was accompanied by the intrusion of large amounts of syenitic magmas. Following the main stage of magmatism, intrusion of dykes and batholiths of syenitic, monzonitic to granitic composition took place, often related to the collapse of large volcanoes. The Oslo Rift extends offshore into the Skagerrak to the south, where seismic data suggest the presence of a ~1-km sequence of Late Carboniferous–Early Permian lavas.

The Whin Sill and Midland Valley complexes in northern Britain consist of a series of dolerite sills and east–north-east- to east–west-trending dykes that extend eastwards into the North Sea. The dykes and sill have subalkaline to transitional basaltic compositions. In Scania in South Sweden, a north-west-trending swarm of subalkaline dolerite dykes may be related to a few isolated, north–north-west-trending dolerite dykes on the south-west coast and Bornholm Island, and to dolerite sills that intrude Cambrian shales in Västergötland. Volcanic rocks were encountered in the drill core in the Kattegat, and seismic data suggest the presence of Early Permian volcanic edifices offshore from eastern Denmark.

The Stephanian–Autunian volcanic rocks in the north-western part of the North German Basin are mainly limited to north–north-west- to north–north-east-trending grabens, and their thicknesses vary from a few metres to ~100 m. In the north-east, they form continuous layers up to 2 km thick and may have covered an area of ~180 000 km^2. Basalt lavas, dolerite, and gabbro intrusions are subordinate, and rhyolite lavas and ignimbrites, basaltic andesite and andesite lavas, and tuffs form the largest volumes. Five eruptive stages have been recognized, and the total volume has been estimated at ~48 000 km^3, approximately 70% of which are rhyolitic lavas and ignimbrites that predominate in the eastern part of the basin.

Elsewhere in the foreland, Stephanian to Autunian magmatic activity occurred in much smaller volumes. In the central and eastern North Sea, basalt, trachyandesite, and rhyolite flows and tuffs occur interbedded with Rotliegend-facies mudstones and sandstones. The volcanic rocks reach thicknesses of up to 160 m in the central North Sea, and ~680 m in the Horn Graben. In northern Britain, up to ~100- to 240-m-thick sequences of alkaline basalt lavas and tuffs occur in the Mauchline and Thornhill basins. These unconformably overlie Westphalian sediments and are interbedded with and overlain by Early Permian aeolian red sandstones. The volcanic rocks are probably related to the approximately 60 mafic vents that occur within a 20-km radius of the Mauchline Basin and predate the aeolian sandstones.

Variscan Internides

The magmatic rocks in the internal Variscides range from granite and diorite intrusions to rhyolitic–andesitic volcanic and subvolcanic rocks. Mafic rocks are limited to diatremes, small volumes of lavas, and mafic dykes. In addition, the volcanic rocks often comprise high proportions of ignimbrites and tuffs, typical for explosive volcanism. Examples are the granite and diorite intrusions in south-west England (Cornwall), the Harz, Erzgebirge, northern Italy, and the basement of the Alps, and the volcanic rocks in basins in the Thuringian Forest, Pyrenees, Iberia, northern Italy, the Provence, and the Saar-Nahe Basin. However, not all basins contain magmatic rocks; for example, in the Stephanian to Early Permian basins in the Massif Central in France, only small volumes of volcanic rocks are present.

In south-west England, postorogenic granite emplacement was probably accompanied by the extrusion of lavas, as indicated by clasts of rhyolite and tuff within Permian sediments. The main magmatic activity occurred within the relatively short period about 4.5 Ma. Granite intrusion and subsequent uplift and erosion partly overlapped with the formation of narrow, east–west-trending, fault-bounded grabens. These contain Rotliegend-facies clastic sediments and small volumes of olivine basalt and lamprophyric lavas and agglomerates. The lamprophyre lavas are genetically related to the minette dykes that occur throughout south-west England.

The Saar-Nahe Basin contains Early Permian andesite flows and pyroclastics that are associated with subvolcanic dacite and rhyolite domes and diatremes of subalkaline basalt. In the south-westwards extension in France, Early Permian andesites, rhyolites, and rhyolitic breccias occur in boreholes and reach a thickness of over 1 km south of Nancy. The small Ilfeld Basin in the south Harz contains a ~800-m-thick sequence of sediments and thick layers of tuffs, ignimbrites, and latitic, trachytic, and rhyolitic flows. Of similar age are the undeformed or only weakly deformed granitoid and gabbro intrusions in the nearby Thuringian Forest and Harz area, and the high-level rhyolite sills near Halle. Differential uplift in the Thuringian Forest, associated with granite and diorite magmatism, led to block faulting and the formation of horsts and pull-apart basins, the latter filled by an up to 2-km-thick sequence of molasse sediments, lavas, and pyroclastic rocks of mainly trachyandesitic and rhyolitic composition.

Magmatic activity in Iberia and the Pyrenees took place in at least two stages, Late Carboniferous to Early Permian and mid-Permian to Triassic. The first phase is of predominantly calc-alkaline composition, whereas the younger magmatic rocks are more alkaline. Small basins and half-grabens contain sequences of Late Westphalian C to Autunian terrestrial sediments and volcanic rocks, the latter mainly of pyroclastic character, but also comprising volcanoclastic rocks and ash flows. Compositions range from andesite, to dacite, to rhyolite, but andesites predominate. Volcanism was accompanied by intrusion of hypabyssal sills, dykes, and domes, by high-level, often composite and hybrid, granitoid intrusions, and by several generations of granitic to dioritic dykes. The earliest volcanism in the Pyrenees may partly overlap with the intrusion of high-level calc-alkaline, often composite granitic to dioritic, plutons that have Westphalian (305–312 Ma) U–Pb zircon crystallization ages.

The pre-Mesozoic basement of the Alps, exposed in the external massifs and tectonic windows, contains Late Carboniferous to Early Permian granitoid intrusions and volcanic and sedimentary rocks. Narrow pull-apart basins contain andesite to rhyolite lavas and rhyolitic tuffs that were deposited on deformed and metamorphosed Visean rocks, suggesting several kilometres of uplift and erosion before 300 Ma. South of the Alps, near Bolzano in northern Italy, Early Permian latite, dacite, rhyodacite, and rhyolite lavas and tuffs cover an area of $\sim$4000 km^2 and locally reach a thickness of over 2 km. The Late Carboniferous to Early Permian granitoids in the basement of the Alps have a calc-alkaline, volcanic arc character, but in northern Italy, magmatic activity continued into mid-Permian times and changed to more alkaline compositions.

Petrogenesis

The varying compositions of the mafic rocks in the foreland show that they were derived from different mantle sources and evolved differently. The magmatic rocks in the Midland Valley of Scotland were derived by low-degree melting of deep mantle sources followed by rapid ascent, as indicated by their alkaline, magnesium-rich, and silica-poor composition, by the absence of low-pressure differentiates, by the presence of abundant megacrysts and of mantle and lower crustal xenoliths, and by the style of volcanism. The early alkaline mafic volcanic rocks in the Oslo Rift may have been derived from mantle sources that were metasomatically enriched by carbonatite fluids in earliest Palaeozoic times. Lithospheric extension causing decompression melting is a probable mechanism, but geochemical evidence suggests that a mantle plume component cannot be completely ruled out. In contrast, the subalkaline basalts of the Whin Sill and Midland Valley complexes indicate higher proportions of melting of shallow mantle sources. Despite being distributed over a relatively large area, their distribution need not reflect a mantle thermal anomaly of the same extent. The geometry and orientation of the dyke swarms suggest a magmatic focal region in the vicinity of the Denmark–Skagerrak region, and magma transport may have been horizontal, westwards into the North Sea and Britain. The position, trend, number, and size of the dykes may have been controlled by the regional dextral extensional stress field.

The magmatic rocks in the North German Basin and the internal Variscides have predominantly felsic to intermediate (rhyolitic to andesitic) compositions. Sr–Nd isotope data and the presence of garnet and crustal xenoliths indicate that their parent melts assimilated large amounts of crustal material, or that the rhyolitic end-members were derived by the melting of older crust. Furthermore, the scarcity of mafic magmatic rocks, in combination with the fractionated character and degree of alteration of the felsic to intermediate rocks, makes it difficult to establish the nature of the mantle sources. The granite intrusions and the felsic volcanic rocks often have calc-alkaline compositions suggesting subduction-related volcanic arc origins. This does not agree with the intracontinental extensional setting of the magmatic activity. Furthermore, most Variscan oceans had closed by Visean times. Instead, the volcanic arc signature may have been inherited from mantle sources that had been metasomatized by previous subduction events, caused by extensive assimilation of continental crust by the parent melts, and/or inherited through partial melting of older, calc-alkaline lower crust. The fact that Stephanian to Autunian mafic rocks are much rarer in the internal Variscides, compared to the foreland, suggests that the mantle-derived parent melts were unable to reach the surface directly. Instead, they stalled in lower to mid-crustal magma chambers, where they assimilated crustal material and fractionated to more felsic compositions before erupting. The large amount of crustal melts (rhyolites) in, for instance, the North German Basin may have been generated by underplating of mantle-derived melts. However, there is little geophysical evidence for lower crustal mafic intrusions in this area. In the central and eastern North Sea, Denmark, and the Skagerrak Graben, geophysical evidence such as deep crustal reflectors and Bouguer anomalies suggest that, here, volcanism may have been accompanied by crustal underplating, although the age of these features remains uncertain.

The effects of Permo-Carboniferous magmatism at various crustal levels can be studied exemplarily in the Ivrea Zone of northern Italy. This tilted crustal fragment provides one of the rare opportunities in the world to study a coherent crustal succession from the Moho to upper crustal levels. The Ivrea Zone achieved its present structure largely at the end and shortly after the Variscan Orogeny. Intrusion of mantle-derived basic magmas at or near the base of the crust caused partial melting of lower crustal rocks. This in turn generated granitic magmas that ascended to middle and upper crustal levels. At the surface, magmatism is also documented by volcanic activity in the contemporaneous sedimentary basins.

Geodynamic Setting and Driving Forces for Extension

The geodynamic setting of the Variscan domain during the Permo-Carboniferous and the driving forces for crustal re-equilibration have been matters of intense debate. A prime feature is the strong thermal perturbation of the lithosphere, which has been attributed to diverse processes such as slab detachment, delamination and thermal erosion of the mantle lithosphere, crustal extension, or ascent of mantle plumes. Some explanations even manage without substantial crustal extension and assume eclogitization of the lower crust as the main process for the disappearance of the Variscan crustal roots. In this respect, it is interesting to note that basin formation and substantial magmatism occur also in the northern foreland of the Variscides (i.e., in an area not affected by Variscan crustal and lithospheric thickening). Thus, the processes often invoked for postconvergent settings may not hold in these areas. This leads to the fundamental question: were Permo-Carboniferous basin formation and magmatism genetically linked to the preceding Variscan Orogeny, or do they document a new geodynamic regime?

Because of the close temporal and spatial relationship between crustal thickening and postconvergent extension, several authors have suggested a gravitational collapse of the Variscan orogen as the dominant process controlling the geodynamic evolution of central Europe in latest Carboniferous to Permian times. The Tibetan Plateau of China and the Basin and Range Province of the United States have been proposed as modern analogues for theLate Palaeozoic destruction of the Variscan orogen. Other authors, however, have suggested that the postconvergent destruction of the Variscides was caused primarily not by body forces, but by a change in orientation of the far-field stress regime. The onset of the Stephanian to Early Permian basin formation indeed coincides with the change in the relative movement between Gondwana and Laurussia, from head-on collision to dextral translation.

To solve this dispute, it is crucial to establish whether the Variscan orogen still had excessive crustal and lithospheric thicknesses at the beginning of the Stephanian, which would have driven orogenic collapse. Unfortunately, no direct evidence is available, neither on the palaeotopography of the Variscan orogen nor on crustal thicknesses. Combining several lines of evidence, a maximum crustal thickness of about 50 km at the onset of the Permo-Carboniferous evolution seems likely. Estimates for the total lithospheric thickness can be constrained by the widespread Visean (~330–340 Ma) high-temperature/low-pressure metamorphism and granite magmatism that are so characteristic for the Variscan orogen. The magnitude of this thermal event, which reached lower crustal temperatures in excess of 900°C, indicates the absence of thick mantle lithosphere beneath the internal zone of the Variscan orogen at this time. Thus, if delamination or convective erosion of the mantle lithosphere occurred during the Variscan Orogeny, it must have occurred prior to 330 Ma. Consequently, none of these processes can be invoked to generate surplus potential energy and gravitational collapse for the Stephanian to Early Permian evolution.

Quantitative analysis using thermomechanical finite element models suggests that the gravitational instability of the Variscan orogen was insufficient to explain the observed amount and timing of crustal extension. In order to overcome the finite strength of the crust and restore a uniform crustal thickness in the area of the former orogen and its northern foreland, tensile plate boundary stresses are required to have operated until about 285 Ma. The cause of the far-field stresses can be attributed to the dextral translation of Laurussia relative to Gondwana during the Stephanian and Early Permian. Contemporaneous crustal shortening of 300–400 km in the linked Appalachian–Mauretanide orogen also provides an estimate for the total amount of regional extension that affected western and central Europe during this stage. Local extension estimates from, for instance, subsidence analysis and section balancing are available for only a few basins in the Variscan domain (see

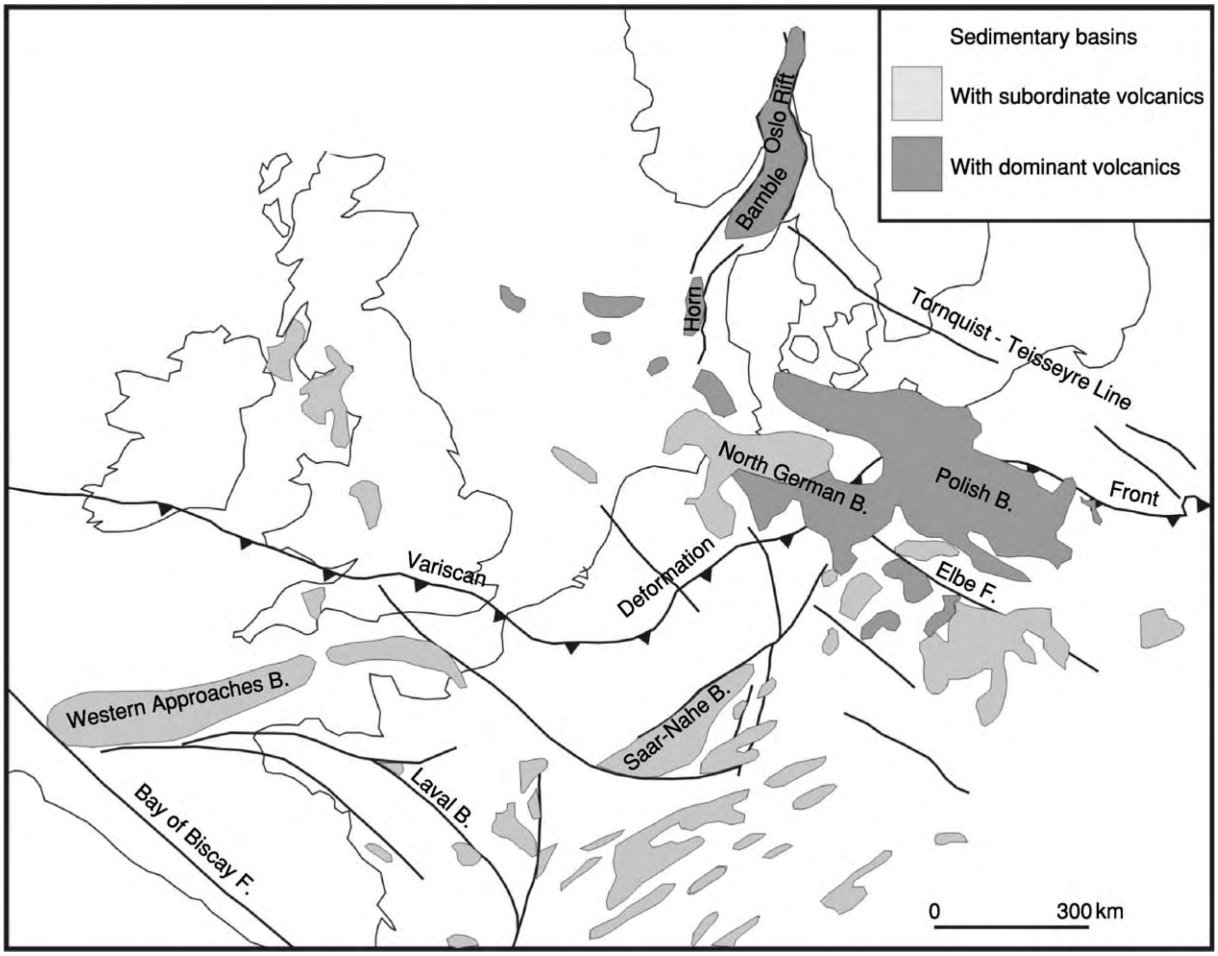


Figure 2 Permo Carboniferous sedimentary basins and volcanic centres in the Variscan realm. Modified with permission from Ziegler PA (1990) *Geological Atlas of Western and Central Europe*. The Hague: Shell Internationale Petroleum Maatschappij.

Figure 2). They indicate crustal stretching by a factor of at least 1.35. However, intrabasin highs, such as basement ridges and areas with postrift sedimentation only, do not necessarily represent unextended areas. This is because diffuse extension and ductile flow of lower crust towards the zones of maximum upper crustal extension may have contributed towards regional crustal thinning.

The decay of the thermal anomaly induced by Stephanian to Early Permian geodynamic processes, i.e., thermal contraction of the lithosphere and its re-equilibration with the asthenosphere, caused long-lasting subsidence and controlled the formation of Late Permian and Mesozoic thermal sag basins that were superimposed on the Permo-Carboniferous troughs.

See Also

Europe: Permian to Recent Evolution; Variscan Orogeny. **Lava**. **Palaeozoic:** Carboniferous; Permian. **Pyroclastics**. **Sedimentary Environments:** Alluvial Fans, Alluvial Sediments and Settings; Deserts; Lake Processes and Deposits. **Tectonics:** Rift Valleys.

Further Reading

Arthaud F and Matte P (1977) Late Paleozoic strike slip faulting in southern Europe and northern Africa: results of a right lateral shear zone between the Appalachians and the Urals. *Geological Society of America, Bulletin* 88: 1305 1320.

Benek R, Kramer W, McCann T, *et al.* (1996) Permo Carboniferous magmatism of the Northeast German Basin. *Tectonophysics* 266: 379 404.

Burg J P, van den Driessche J, and Brun J P (1994) Syn to post thickening extension in the Variscan Belt of Western Europe: modes and structural consequences. *Géologie de la France* 3: 33 51.

Cortesogno L, Cassinism G, Dallagiovannam G, Gaggerom L, Oggiano G, Ronchi A, Seno S, and Vanossi M (1998) The Variscan post collisional volcanism in Late Carboniferous Permian sequences of Ligurian Alps, Southern Alps and Sardinia (Italy): a synthesis. *Lithos* 45: 305 328.

Finger F, Roberts MP, Haunschmid B, Schermaier A, and Steyer HP (1997) Variscan granitoids of central Europe: their typology, potential sources and tectonothermal relations. *Mineralogy and Petrology* 61: 67 96.

Floyd PA, Exley CS, and Styles MT (1993) *Igneous Rocks of South West England. Geological Conservation Review*. vol. 5. London: Chapman & Hall.

Glennie KW (1999) Lower Permian Rotliegend. In: Glennie KW (ed.) *Petroleum Geology of the North Sea: Basic Concepts and Recent Advances,* pp. 137 173. Oxford: Blackwell Science.

Henk A (1999) Did the Variscides collapse or were they torn apart?: a quantitative evaluation of the driving forces for postconvergent extension in central Europe. *Tectonics* 18: 774 792.

Plein E (1995) Norddeutsches Rotliegend Becken. Rotliegend monographie, teil II. *Courier Forschungsinstitut Senckenberg* 183: 1 193.

Schaltegger U (1997) Magma pulses in the Central Variscan Belt: episodic melt generation and emplacement during lithospheric thinning. *Terra Nova* 9: 242 245.

Sundvoll B, Neumann E R, Larsen BT, and Tuen E (1990) Age relations among Oslo Rift magmatic rocks: implications for tectonic and magmatic modelling. *Tectonophysics* 178: 67 87.

Ziegler PA (1990) *Geological Atlas of Western and Central Europe*. The Hague: Shell Internationale Petroleum Maatschappij.

Permian to Recent Evolution

P A Ziegler, University of Basel, Basel, Switzerland

Introduction

Large parts of Western and Central Europe (WCE) are occupied by sedimentary basins that contain up to 8 km thick Permian to Cenozoic series (**Figure 1**). These basins are variably underlain by the Precambrian crust of the East-European-Fennoscandian Craton (*see* **Europe:** East European Craton) and its Late Precambrian to Early Palaeozoic sedimentary cover, by the Precambrian Hebridean Craton, by the Caledonian crust of the British Isles (*see* **Europe:** Caledonides of Britain and Ireland), the North Sea, Northern Germany and Poland and its Devonian and Carboniferous sedimentary cover, and by the Variscan fold belt in which orogenic activity had ceased at the end of the Westphalian. The present crustal configuration of WCE bears little relationship to the Caledonian and Variscan orogenic belts, but is closely related to the geometry of the Late Permian, Mesozoic and Cenozoic sedimentary basins and the Alpine orogen (**Figure 2**). This reflects that dynamic processes, which governed the evolution of the Late Permian and younger sedimentary basins, had a strong impact on the crustal configuration of WCE, and that the crustal roots of the Caledonian and Variscan orogens had been destroyed shortly after their consolidation.

During Permian to recent times, the megatectonic setting of WCE underwent repeated changes. Correspondingly, dynamic processes controlling the evolution and partial destruction of sedimentary basins also changed through time. Therefore, in some areas, basins of differing tectonic origin are stacked on top of one other.

The following main stage are recognized in the Late Permian to recent evolution of WCE, namely: (i) Late Permian–Early Cretaceous rifting during Pangaea breakup; (ii) Late Cretaceous–Paleocene rifting and early Alpine intraplate compression; and (iii) Eocene-recent opening of the Arctic–North Atlantic and collisional interaction of the Alpine Orogen with its foreland.

Background: Late Hercynian Wrench Tectonics and Magmatism

Following its Late Westphalian consolidation, the Variscan Orogen (*see* **Europe:** Variscan Orogeny) and its northern foreland were overprinted during the Stephanian to Early Permian by a system of continent-scale dextral shears, such as the Tornquist-Teisseyre, Bay of Biscay, Gibraltar-Minas and Agadir fracture zones which were linked by secondary sinistral and dextral shear systems. This deformation reflects a change in the Gondwana-Laurussia convergence from oblique collision to a dextral translation that was kinematically linked to continued crustal shortening in the Appalachian (Alleghanian Orogeny) and the Scythian orogens. Significantly, wrench tectonics and associated magmatic activity abated in the Variscan domain and its foreland at the transition to the Late Permian in tandem with the consolidation of the Appalachian Orogen.

Stephanian to Early Permian wrench-induced disruption of the Variscan Orogen and its foreland was accompanied by regional uplift, wide-spread extrusive and intrusive mantle-derived magmatic activity

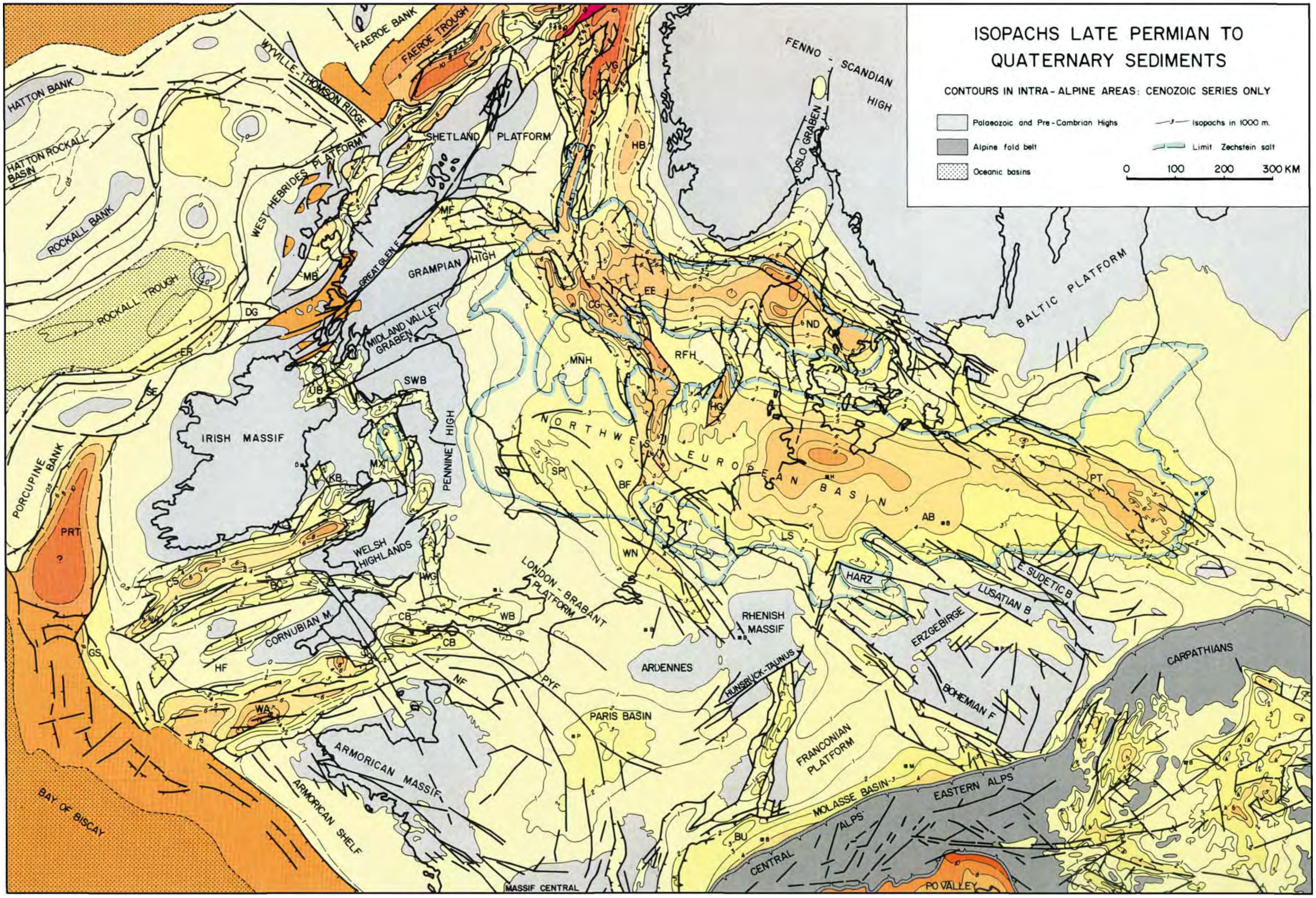


Figure 1 Total isopach of Late Permian to Cenozoic sediments.

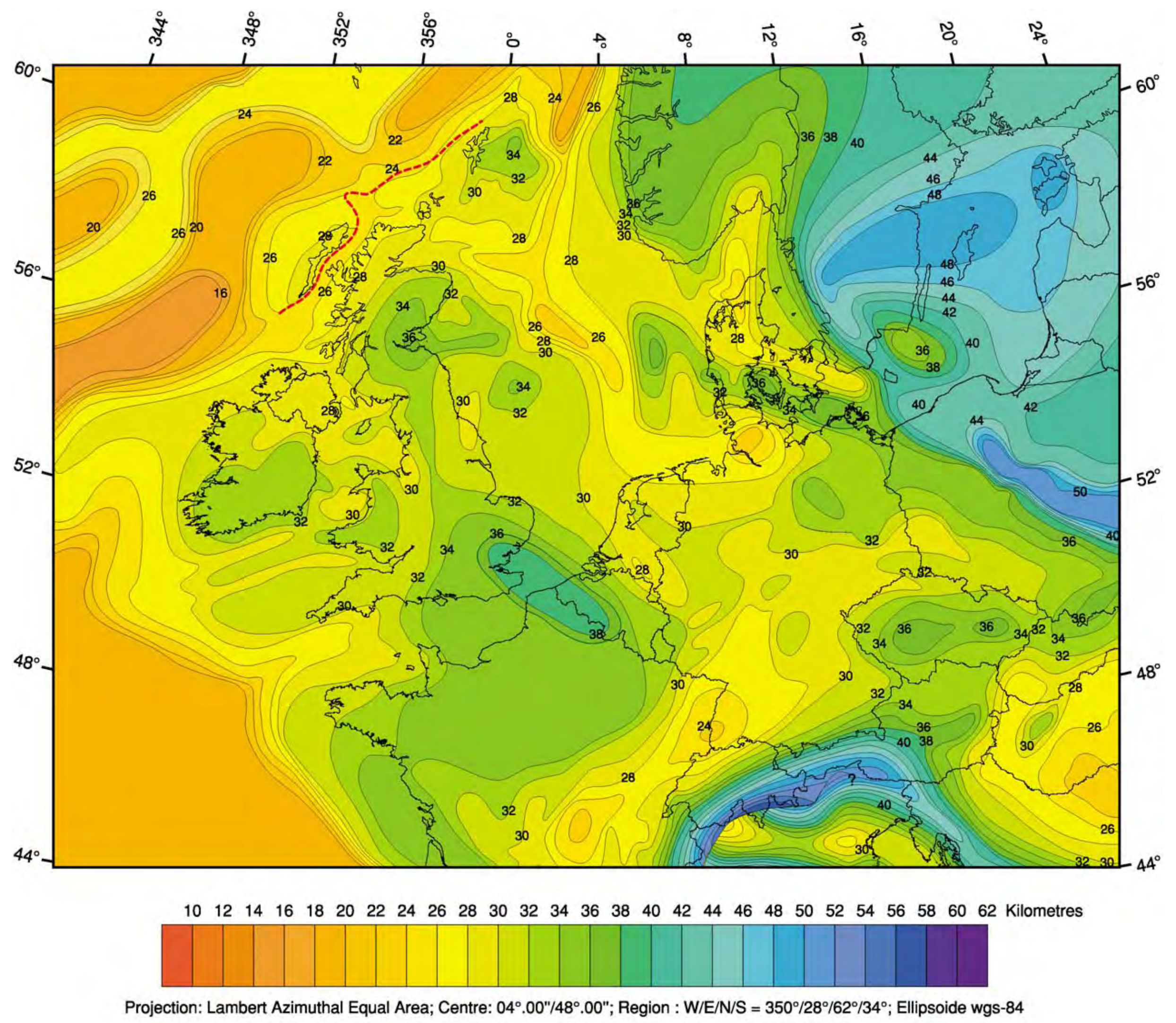


Figure 2 Moho Depth map of Western and Central Europe, contour interval 2 km (after Dèzes & Ziegler, 2004) (available in digital format).

that peaked during the Early Permian, and the subsidence of a multi-directional array of transtensional and pull-apart basins in which continental clastics accumulated. Basins that developed during this time span show a complex, polyphase structural evolution, including a late phase of transpressional deformation controlling their partial inversion. Although Stephanian to Early Permian wrench deformation locally gave rise to the uplift of extensional core complexes, crustal stretching factors were on a regional scale relatively low. Nevertheless, the model of the Cenozoic Basin-and-Range Province has been repeatedly invoked for the Stephanian-Early Permian collapse of the Variscan Orogen. Yet, kinematics controlling the development of these two provinces differed: whereas the collapse of the Variscides was wrench-dominated, extension dominated the collapse of the Cordillera.

The Stephanian to Early Permian magmatic activity can be related to wrench-induced reactivation of Variscan sutures which caused the detachment of subducted lithospheric slabs, upwelling of the asthenosphere, partial delamination and thermal thinning of the mantle-lithosphere, magmatic inflation of the remnant lithosphere and interaction of mantle-derived partial melts with the lower crust. In conjunction with slab detachment and a general reorganization of the asthenospheric flow patterns, a system of not very active mantle plumes apparently welled up to the base of the lithosphere in the area of the future Southern and Northern Permian basins, the British Isles and the Oslo Graben, causing thermal attenuation of the lithosphere and magmatic destabilisation of the crust-mantle boundary. In the domain of the Variscan Orogen, this accounted for the destruction of its orogenic roots and regional uplift. By the end of the Early Permian, its crust was thinned down on a regional scale to 30–35 km, mainly by magmatic processes and erosional unroofing and only locally by its mechanical stretching. Moreover, the thickness of the mantle-lithosphere was reduced to as little as 50 to 10 km in areas which evolved during the Late Permian and Mesozoic into intracratonic thermal sag basins, such as the Southern Permian and the Paris basins.

Late Permian–Early Cretaceous Rifting During Pangaea Breakup

Following the Stephanian–Early Permian tectonomagmatic cycle, the potential temperature of the asthenosphere returned quickly to ambient levels (1300°C). This is indicated by the late Early Permian extinction of magmatic activity and the onset of regional thermal subsidence of the lithosphere, reflecting the decay of thermal anomalies that had been introduced during the Permo-Carboniferous. In combination with erosional degradation of the remnant topography of the Variscan Orogen and cyclically rising sea-levels, progressively larger areas subsided below the erosional base level and were incorporated into a new system of intracratonic basins, comprising the Northern and Southern Permian basins, the Hessian Depression, the Paris Basin, and the Franconian Platform (Figure 3). However, in large parts of the WCE, thermal subsidence of the lithosphere was overprinted and partly interrupted by the Late Permian–Early Triassic onset of a rifting cycle which preceded and accompanied the step-wise break-up of Pangaea. Major elements of this breakup system were the southward propagating Arctic-North Atlantic and the westward propagating Neotethys rift systems. Evolution of these mega-rift systems was paralleled by the development of multi-directional grabens in the WCE, major constituents of which are the North Sea rift, the North Danish-Polish Trough, the graben systems of the Atlantic shelves and the Bay of Biscay rift (Figure 3). Development of these grabens partly involved tensional reactivation of Permo-Carboniferous fracture systems.

Late Permian

Thermal subsidence of the Northern and Southern Permian Basin commenced during the late Early Permian and persisted into Early Jurassic times, as evidenced by quantitative subsidence analyses, facies patterns and isopach maps. In these basins sedimentation commenced during the late Early Permian with the accumulation of the continental Rotliegend redbed series which attain a thickness of up to 2300 m in the axial parts of the-well defined Southern Permian Basin and of 600 m in the less well-defined Northern Permian Basin that was severely overprinted by the Mesozoic North Sea rift.

During the Late Permian, the Norwegian–Greenland Sea rift, which had come into evidence during the Late Carboniferous, propagated southward into the North-western Shelf of the British Isles, opening a seaway through which the Arctic seas transgressed via the Irish Sea and possibly the northernmost North Sea into the Northern and Southern Permian basins which, by this time, had subsided below the global sea-level (*see* **Europe:** Permian Basins). In these basins, the cyclical Zechstein carbonate, evaporite, and halite series were deposited under a tectonically quiescent regime. Following the initial transgression, the axial parts of the Northern and Southern Permian basins were characterized by deeper water conditions whereas along their margins basin-ward prograding carbonate and evaporitic shelves developed. Facies and thickness changes on these shelf series provide evidence for minor extensional faulting along the southern margin of the Southern Permian Basin. Oscillating sea-levels accounted for cyclical restriction of the Northern and Southern Permian basins in which up to 2000 m of halites, partly containing polyhalites, were deposited. During the end-Permian, global low-stand in sea-level, the Arctic seas withdrew from the WCE into the area between Norway and Greenland.

Triassic

During the Triassic, the Norwegian–Greenland Sea rift propagated southwards into the North and Central Atlantic domain, whilst the Neotethys rift systems propagated westwards through the Bay of Biscay and North-west Africa and linked up with the Atlantic rift system.

During the Early Triassic, the North Sea rift, consisting of the Horda half graben, and the Viking, Murray Firth, Central, and Horn grabens, was activated and transected the western parts of the Northern and Southern Permian basins, whereas their eastern parts were transected by the North Danish–Polish Trough (Figure 4). Simultaneously the rift systems of the Alpine domain, the Bay of Biscay, and the Western Shelves were activated. The latter included the Porcupine, Celtic Sea, and Western Approaches troughs. Crustal extension in the Celtic Sea and Western Approaches troughs involved at their eastern termination the reactivation of Permo-Carboniferous shear systems controlling the subsidence of the Channel Basin and intermittent destabilization of the Paris thermal sag basin. Significantly, Triassic–Early Jurassic rifting was

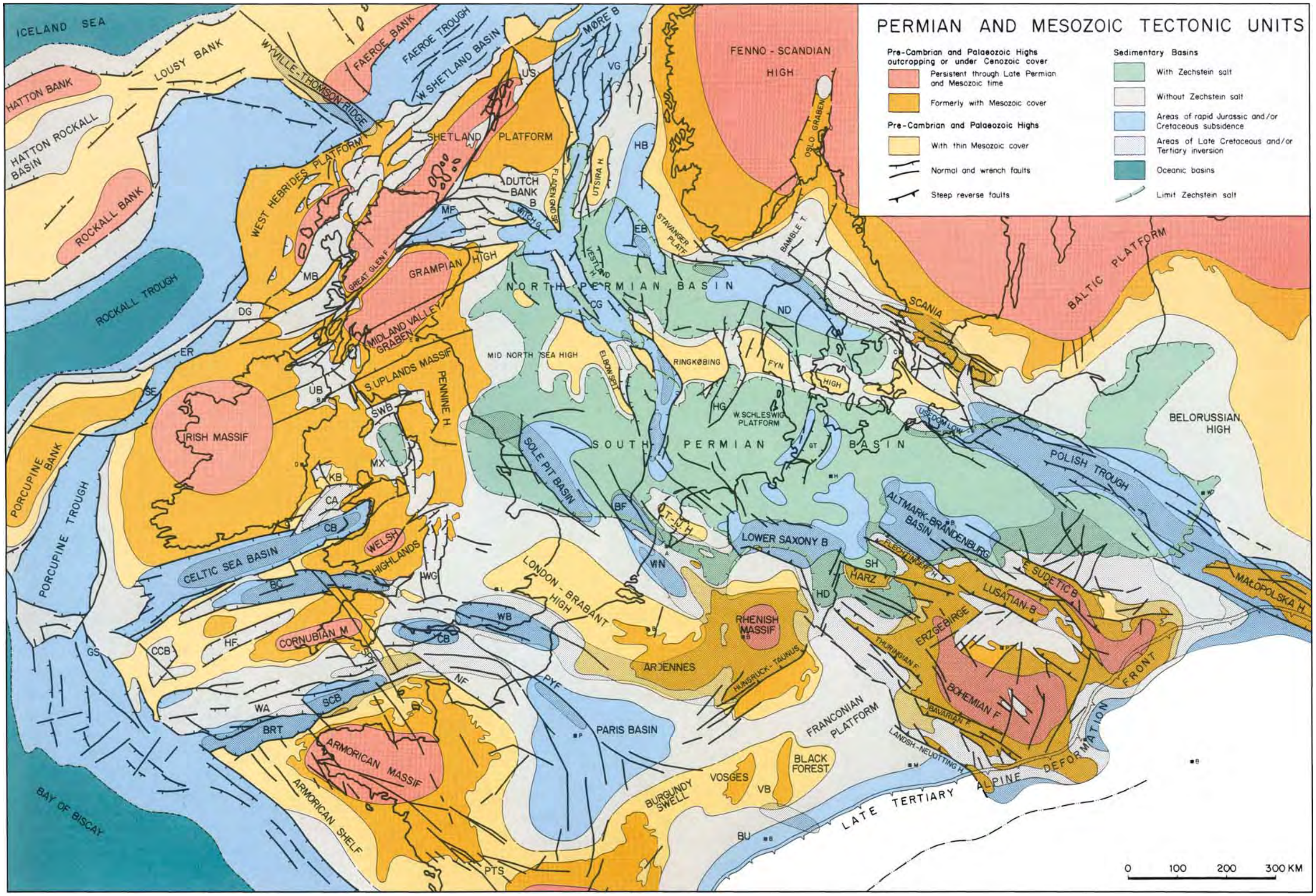


Figure 3 Permian and Mesozoic Tectonic units. For legend see **Figure 16**. Details of Enclosure from *Geological Atlas of Western and Central Europe* 2nd Edition, Peter A. Zeigler, 1990, published by Shell International Petroleum Mij. B.V., distributed by Geological Society Publishing House, Bath.

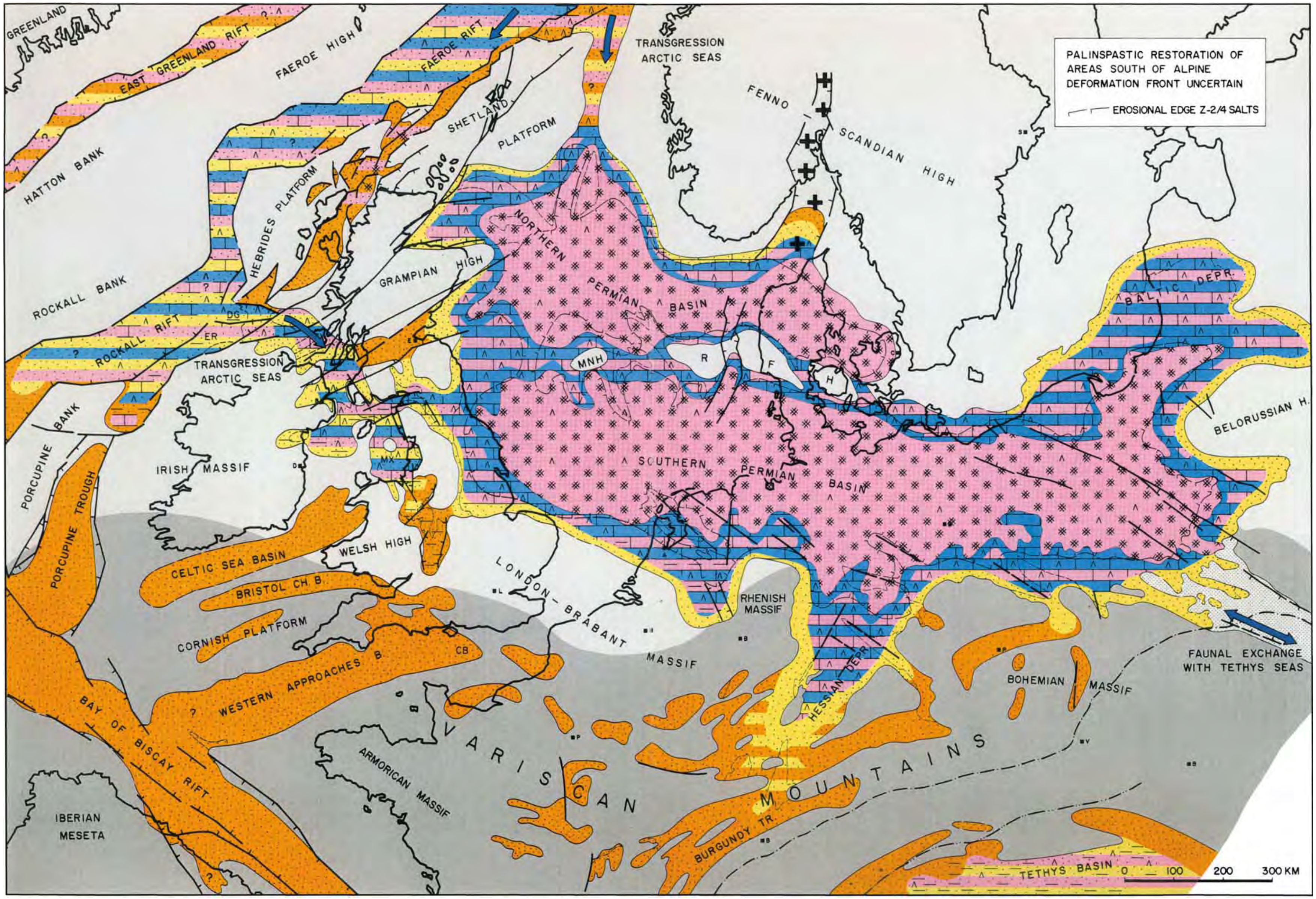


Figure 4 Late Permian, Zechstein palaeogeography. For legend see Figure 16. Details of Enclosure from *Geological Atlas of Western and Central Europe* 2nd Edition, Peter A. Zeigler, 1990, published by Shell International Petroleum Mij. B.V., distributed by Geological Society Publishing House, Bath.

accompanied by a low level of volcanic activity with the exception of the Bay of Biscay, in the evolution of which wrench faulting compensating for crustal extension in the Atlantic domain played an important role. Areas not affected by rifting continued to subside during the Triassic and Early Jurassic in response to thermal relaxation of the lithosphere, accounting in conjunction with eustatically rising sea-levels for a progressive overstepping of basin margins.

During the Early Triassic, continental to lacustrine conditions prevailed in the evolving grabens and thermal sag basins of WCE, in which the 'Bunter' red-beds were deposited. Clastics were shed into these basins from adjacent Variscan and Caledonian highs, as well as from Fennoscandia. During the late Early Triassic, the Tethys Seas ingressed the continuously subsiding Southern Permian Basin via the Polish Trough, giving rise to the deposition of carbonates in Poland and distal halites in northern Germany and the Southern North Sea (Figure 5).

As during the Middle Triassic, the trend of highs that had separated the Northern and Southern Permian basins became gradually overstepped and with this these basins coalesced, thus forming the composite North-west European Basin. The Tethys Seas advanced further into this continuously subsiding basin complex via the Polish Trough as well as via the Burgundy Trough and the Trier and Hessian depressions, establishing a broad neritic basin in which the 'Muschelkalk' carbonates, evaporites, and halites were deposited (Figure 6). Furthermore, intermittent marine transgressions advanced from the Tethyan shelves via the Bay of Biscay rift into the grabens of the Western Shelves. By contrast, continental conditions continued to prevail in the grabens of the Central and Northern North Sea and the North-western Shelf.

With the beginning of the Late Triassic (*see* **Mesozoic:** Triassic), clastic influx from Fennoscandia and eastern sources increased, causing the replacement of the carbonate-dominated Muschelkalk depositional regime by the evaporitic 'Keuper' red-beds containing halites. Whilst the Polish seaway, which had linked the Tethys and the North-west European Basin, was closed, intermittent marine transgressions advanced through the Burgundy Trough into the evolving Paris Basin and the continuously subsiding North-west European Basin, as well as through the Bay of Biscay rift into the grabens of the Western Shelves. However, continental conditions persisted in the grabens of the Central and Northern North Sea and the North-western Shelf. Only during the Rhaetian did the Arctic Seas start to advance southwards into the rifted basins of the North-western Shelf, whilst neritic conditions were established in the broad North-west European Basin (Figure 7).

The Triassic series attains thicknesses of up to 3 km in the grabens of the Western Shelves, the North Sea, and in the Polish Trough, and up to 6 km in the grabens of the North-western Shelf. In the Northern and Southern Permian Basins, the diapirism of Permian salts commenced during the Triassic, and accounted for local subsidence anomalies.

Jurassic

In conjunction with continued rifting activity and cyclically rising sea-levels, the Arctic and Tethys Seas linked up during the Rhaetian–Hettangian, via the rift systems of the North-western and Western shelves and the continuously subsiding North-west European Basin (*see* **Mesozoic:** Jurassic). In the open marine, shale-dominated North-west European Basin, which occupied much of the Southern and Central North Sea, Denmark and Germany (Figure 7), the Belemnitidae (*see* **Fossil Invertebrates:** Cephalopods (Other Than Ammonites)) developed during the Hettangian and Sinemurian. Persisting clastic influx from the East-European Platform allowed only for temporary marine incursions via the Polish Trough. Similarly, fluvio-deltaic conditions prevailed in the grabens of the Northern North Sea until the end-Hettangian to Early Sinemurian when neritic conditions were also established in these basins. By Late Simemurian times, this facilitated a broad faunal exchange between the Boreal and Tethyan realms and the dispersal of the Belemnitidae.

In response to rising sea-levels and continued crustal extension, open marine conditions were established in the Central Atlantic during the Simemurian, permitting Tethyan faunas to reach the Pacific by Pliensbachian times. In basins which were dominated by the warmer Atlantic and Tethyan waters, carbonates and shales were deposited, whilst shales prevailed in the North-west European Basin, which was dominated by the cooler Arctic waters. During the Early Jurassic, repeated stagnant water stratification gave rise to the deposition of organic-rich shales, forming important oil source-rocks, for example, in the Paris Basin and the southern parts of the North-west European Basin.

During the Late Aalenian–Early Bajocian, the Arctic seas became separated from the Tethys and the Central Atlantic in conjunction with the uplift of a large arch in the Central North Sea from which clastics were shed into the adjacent continuously subsiding basins (Figure 8). Uplift of this arch was associated with major volcanism that may be related to the impingement of a short-lived mantle plume. Open marine communications between the Arctic and the Tethys–Atlantic seas were re-opened

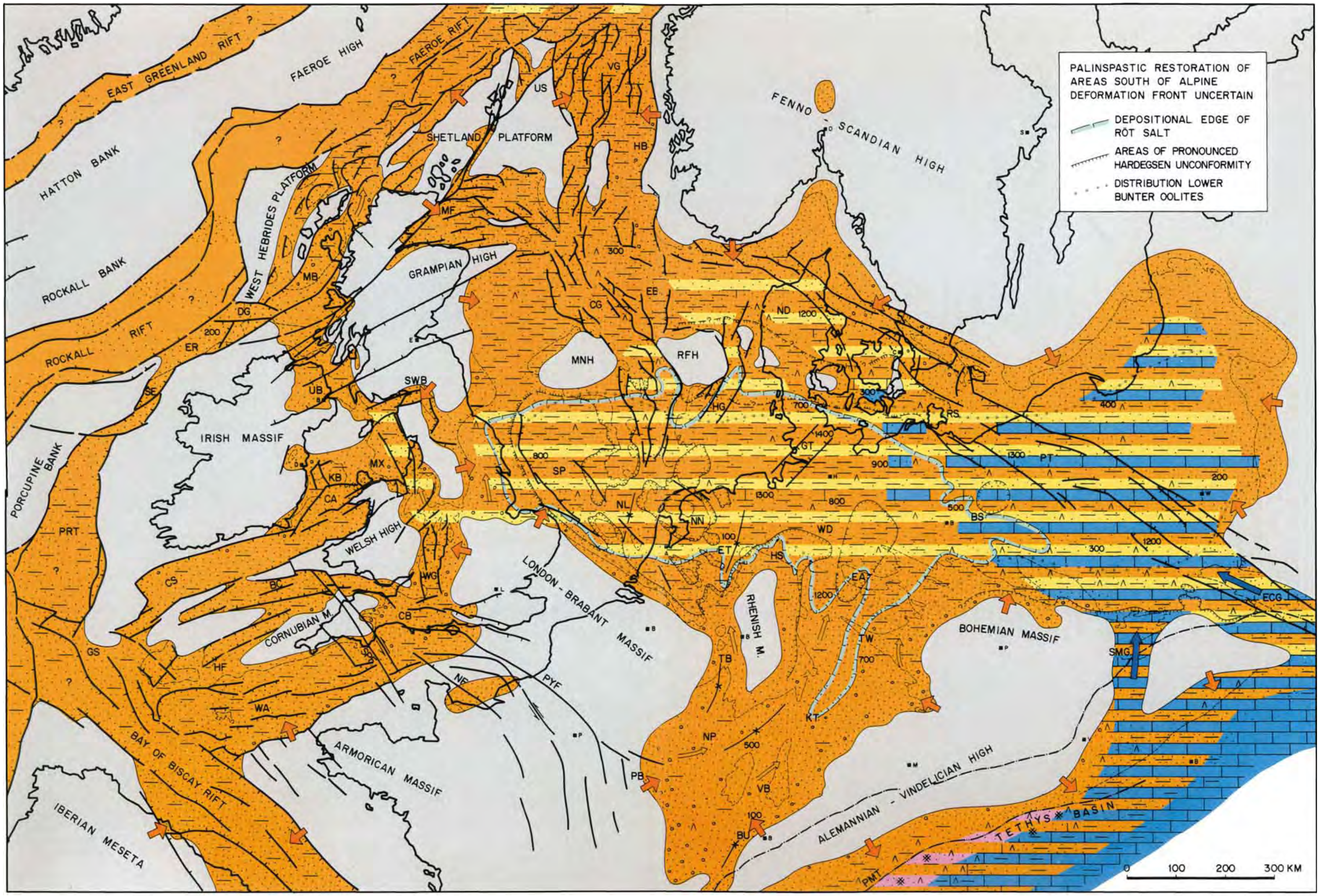


Figure 5 Scythian, Buntsandstein palaeogeography. For legend see Figure 16. Details of Enclosure from *Geological Atlas of Western and Central Europe* 2nd Edition, Peter A. Zeigler, 1990, published by Shell International Petroleum Mij. B.V., distributed by Geological Society Publishing House, Bath.

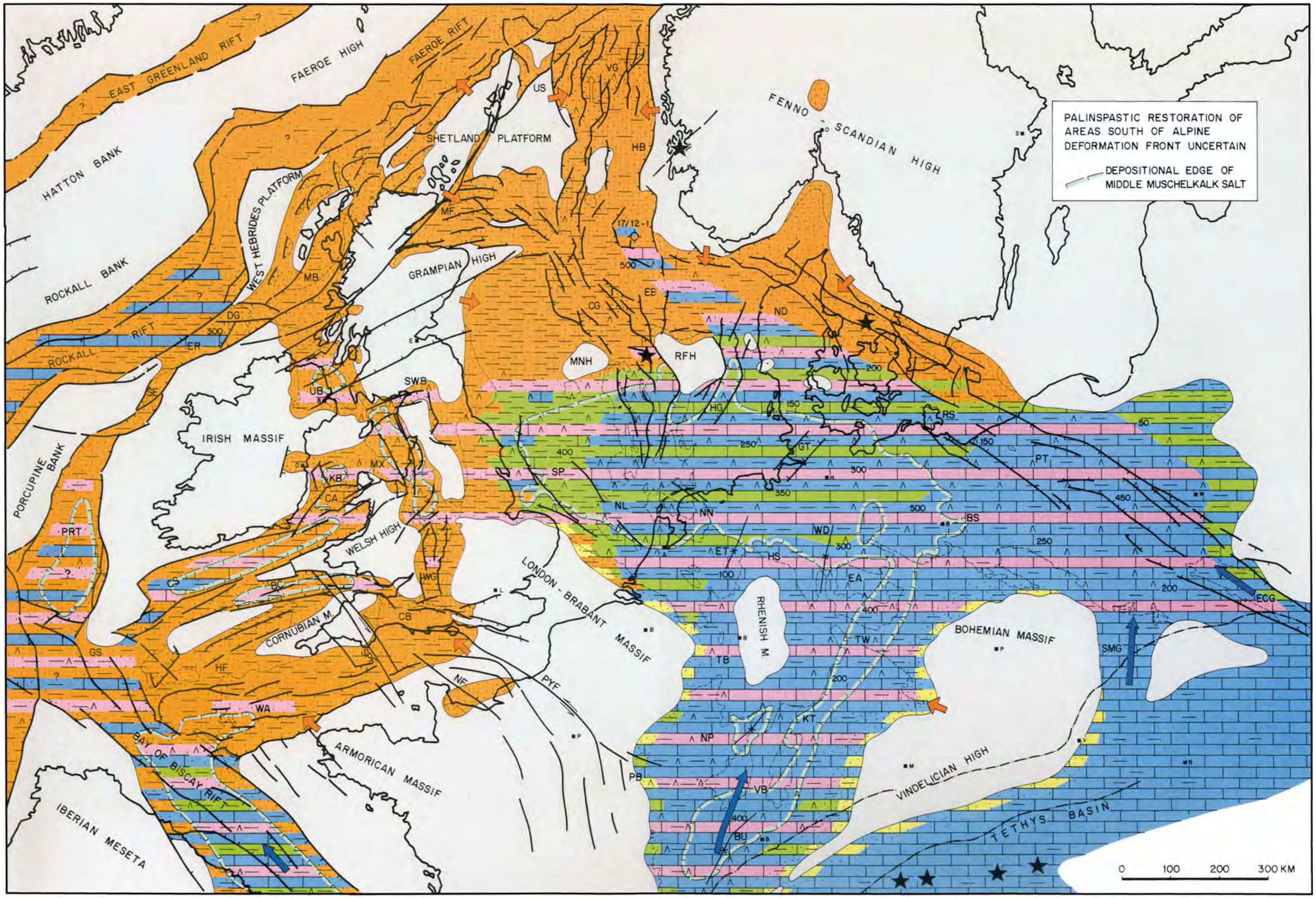


Figure 6 Anisian Ladnian, Muschelkalk palaeogeography. For legend see **Figure 16**. Details of Enclosure from *Geological Atlas of Western and Central Europe* 2nd Edition, Peter A. Zeigler, 1990, published by Shell International Petroleum Mij. B.V., distributed by Geological Society Publishing House, Bath.

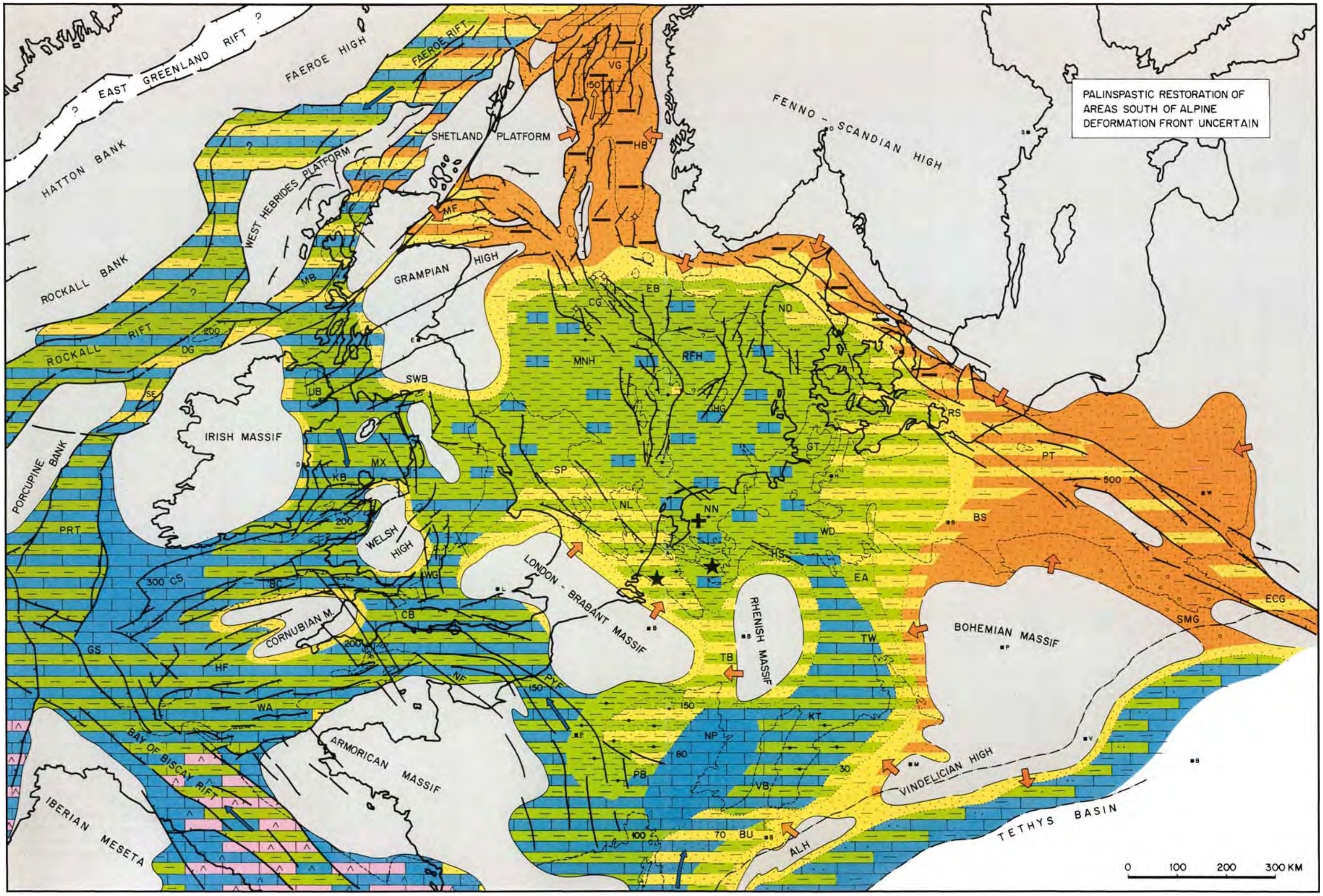


Figure 7 Rhaetian Hettangian palaeogeography. For legend see Figure 16. Details of Enclosure from *Geological Atlas of Western and Central Europe* 2nd Edition, Peter A. Zeigler, 1990, published by Shell International Petroleum Mij. B.V., distributed by Geological Society Publishing House, Bath.

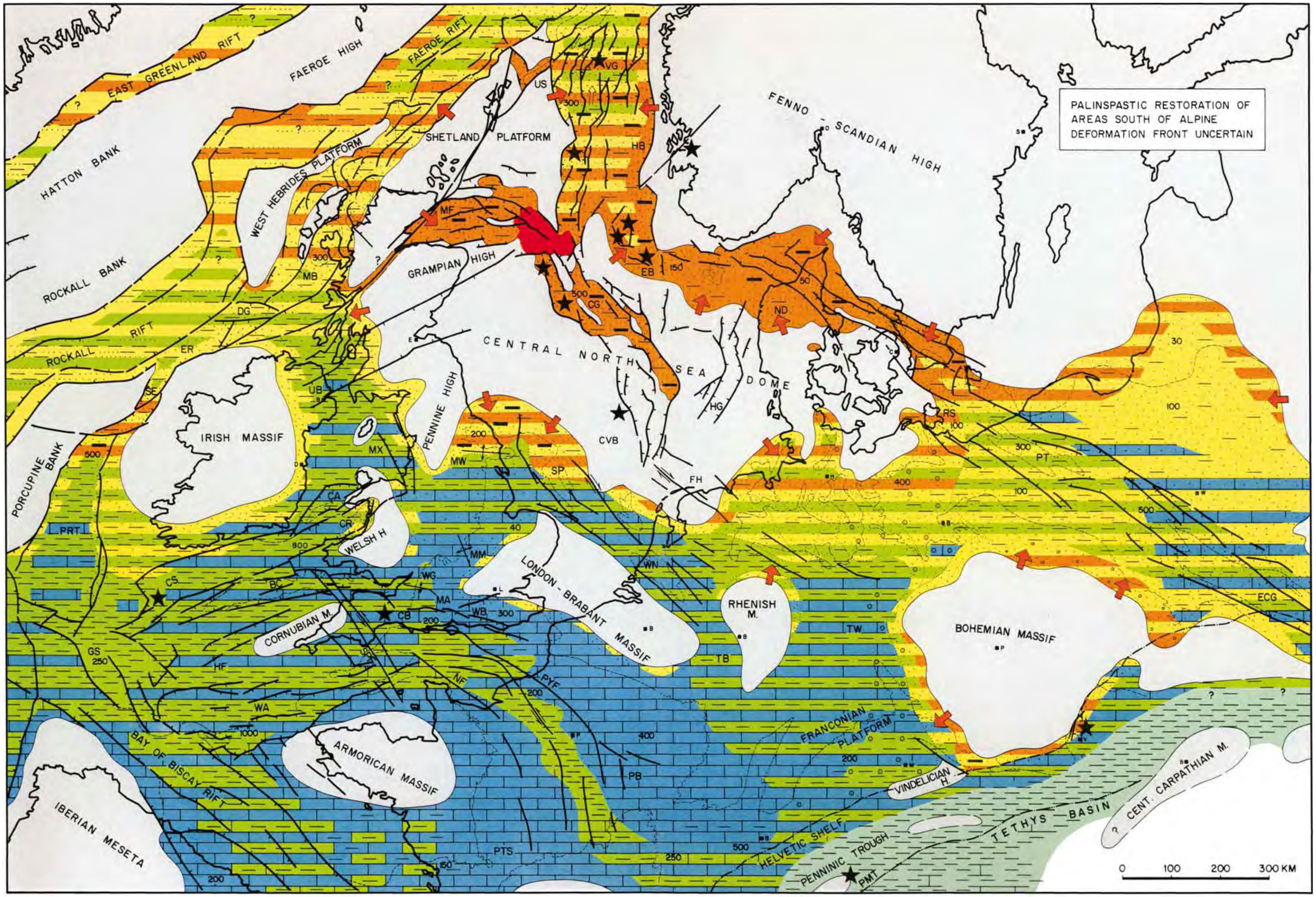


Figure 8 Bajocian Bathonian palaeogeography. For legend see Figure 16. Details of Enclosure from *Geological Atlas of Western and Central Europe* 2nd Edition, Peter A. Zeigler, 1990, published by Shell International Petroleum Mij. B.V., distributed by Geological Society Publishing House, Bath.

during the Callovian, due to the subsidence of the Central North Sea arch and continued crustal extension in the North Sea and the Faeroe-Rockall rifts. During the Middle Jurassic, carbonates prevailed in basins that were dominated by the Tethyan Seas, whereas clastics prevailed in basins dominated by Arctic waters.

Toarcian crustal separation in the Central Atlantic and Bajocian crustal separation in the Alpine Tethys caused a change in the stress systems that controlled the evolution of the WCE rifts. From then on their evolution was dominated by stresses controlling the northward propagation of the Atlantic rift system. During the Late Jurassic and earliest Cretaceous, accelerated crustal extension in the North Sea rift, now focused on its axial Viking and Central grabens, was accompanied by the development of sinistral shear systems at its southern termination which controlled the subsidence of the transtensional Sole Pit, Broad Fourteens, West Netherlands, Lower Saxony, and Sub-Hercynian basins, and the opening of a seaway cross-cutting the Bohemian Massif (Figures 3 and 9). This was accompanied by the gradual uplift of a high that extended from the London-Brabant Massif to the Bohemian Massif, partly separating the carbonate shelves of the Paris Basin and Franconian Platform from the shale and clastic dominated basins of the North Sea area. Contemporaneous accelerated rifting activity is also evident in the Western Approaches, Celtic Sea, and Porcupine troughs that opened into the Bay of Biscay and North Atlantic rifts. In the latter, crustal separation between Iberia and the Grand Banks of Newfoundland progressed gradually northwards during the Late Jurassic and Early Cretaceous. The Late Jurassic cycle of accelerated rifting and wrench tectonics was associated with only minor volcanic activity.

In the course of the Late Jurassic, deeper water conditions were established in the North Sea, Rockall-Faeroe, Porcupine, and Bay of Biscay rifts. Organic-rich Kimmeridgian–Tithonian shales deposited in these rifts and in flanking shallower water basins form the principal oil source rocks of the Central and Northern North Sea, Faeroe rift, and Channel Basin hydrocarbon provinces.

Early Cretaceous

During the earliest Cretaceous, large parts of the WCE were uplifted and subjected to erosion in response to wrench faulting and stress-induced buckling of the lithosphere, causing a relative low-stand in sea-level. Deeper marine conditions persisted; however, in the continuously subsiding Bay of Biscay, Porcupine, Rockall-Faeroe, and the North Sea rifts. On the other hand, deltaic to continental clastics were deposited in the rapidly subsiding transtensional Channel, Weald, Sole Pit, West Netherlands, Lower Saxony, and Altmark-Brandenburg basins, as well as in the Celtic Sea and Western Approaches grabens. However, the Tethys shelves, of which the Paris Basin formed an embayment, and the slowly subsiding North Danish-Polish trough were affected to a lesser extent by this tectonically-induced relative sea-level change (Figure 10).

From Neocomian to Albian times, basin margins were progressively overstepped in response to eustatically rising sea-levels and gradually abating tectonic activity in the rift- and wrench-induced basins of the WCE, whilst crustal extension progressively focused on the zone of future crustal separation between Europe and Greenland. With the onset of post-rift thermal subsidence, the North Sea Basin gradually expanded (Figure 11). Nevertheless, minor tectonic activity was still evident until Aptian times in the North Sea, the Sole Pit, Broad Fourteens, West Netherlands, Lower Saxony, and Sub-Hercynian basins, as well as along the Fennoscandian Border Zone and the border faults of the Bohemian Massif. Following Aptian crustal separation in the Bay of Biscay, the grabens on the Western Shelves became inactive and began to subside thermally. During the Early Cretaceous volcanic activity in the WCE was rather limited.

During the Neocomian, the North Sea Basin was partly separated from the Tethys by the wrench-induced Brabant-Rhenish-Bohemian high, and from the Atlantic by a high that extended from Ireland via Scotland to the Shetland Isles. However, in the course of the Aptian and Albian, shallow marine communications were reopened between the North Sea Basin and the Tethys and Western shelves via southern England and the Paris Basin. Throughout the Early Cretaceous, the WCE basin system was clastic-dominated except for the Armorican margin on which carbonates were deposited.

Late Cretaceous–Paleocene Rifting and Early Alpine Intraplate Compression

During the Late Cretaceous and Paleocene, rifting activity was centred on the Faeroe-Rockall Trough and the area between the Rockall-Hatton Bank and Greenland. Limited sea-floor spreading may have occurred in the southern parts of the Rockall Trough during the Cenomanian–Santonian. During the Campanian–Maastrichtian, the long-lived Iceland plume impinged on the Arctic-North Atlantic rift system,

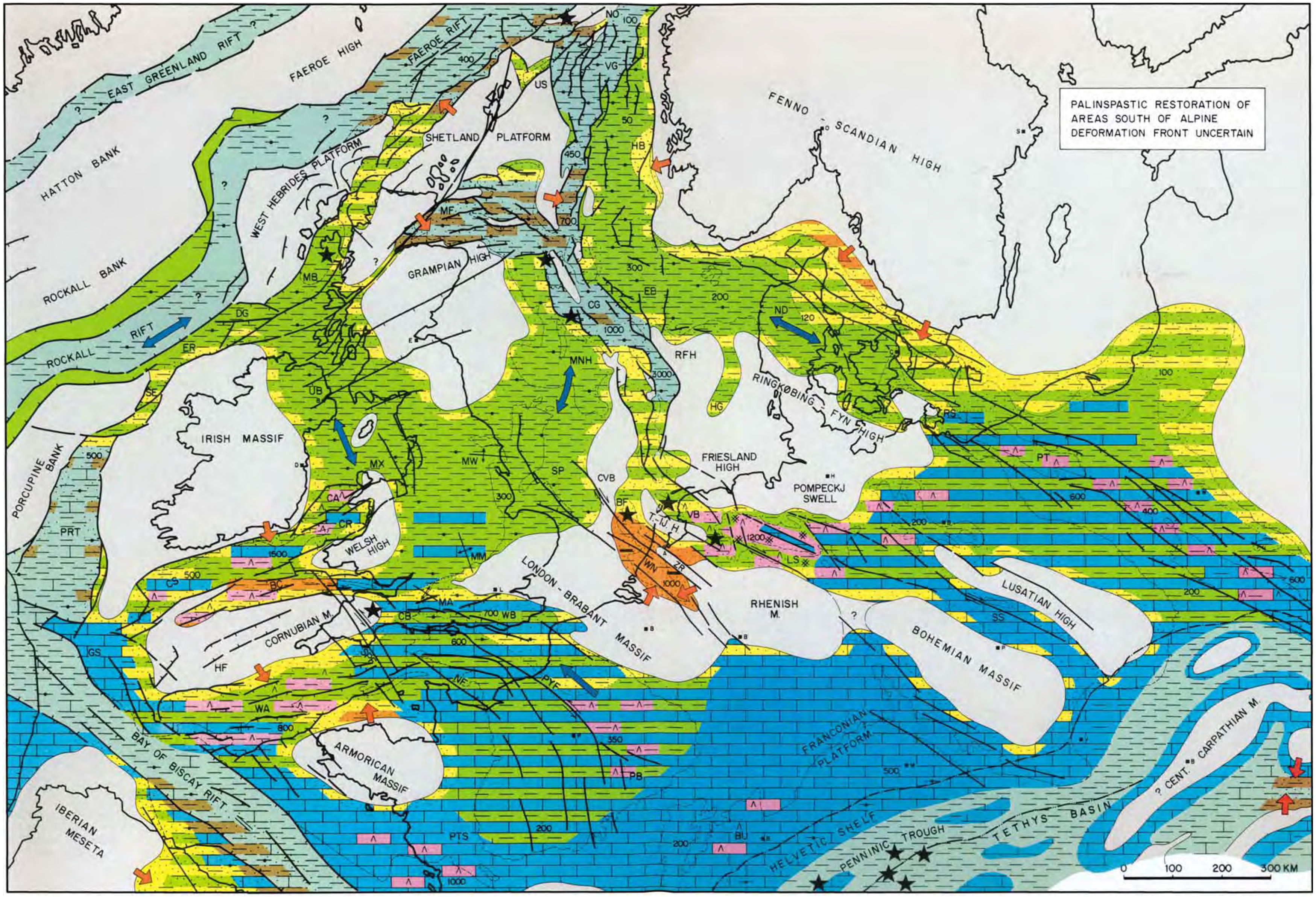


Figure 9 Kimmeridgian/Tithonian palaeogeography. For legend see Figure 16. Details of Enclosure from *Geological Atlas of Western and Central Europe* 2nd Edition, Peter A. Zeigler, 1990, published by Shell International Petroleum Mij. B.V., distributed by Geological Society Publishing House, Bath.

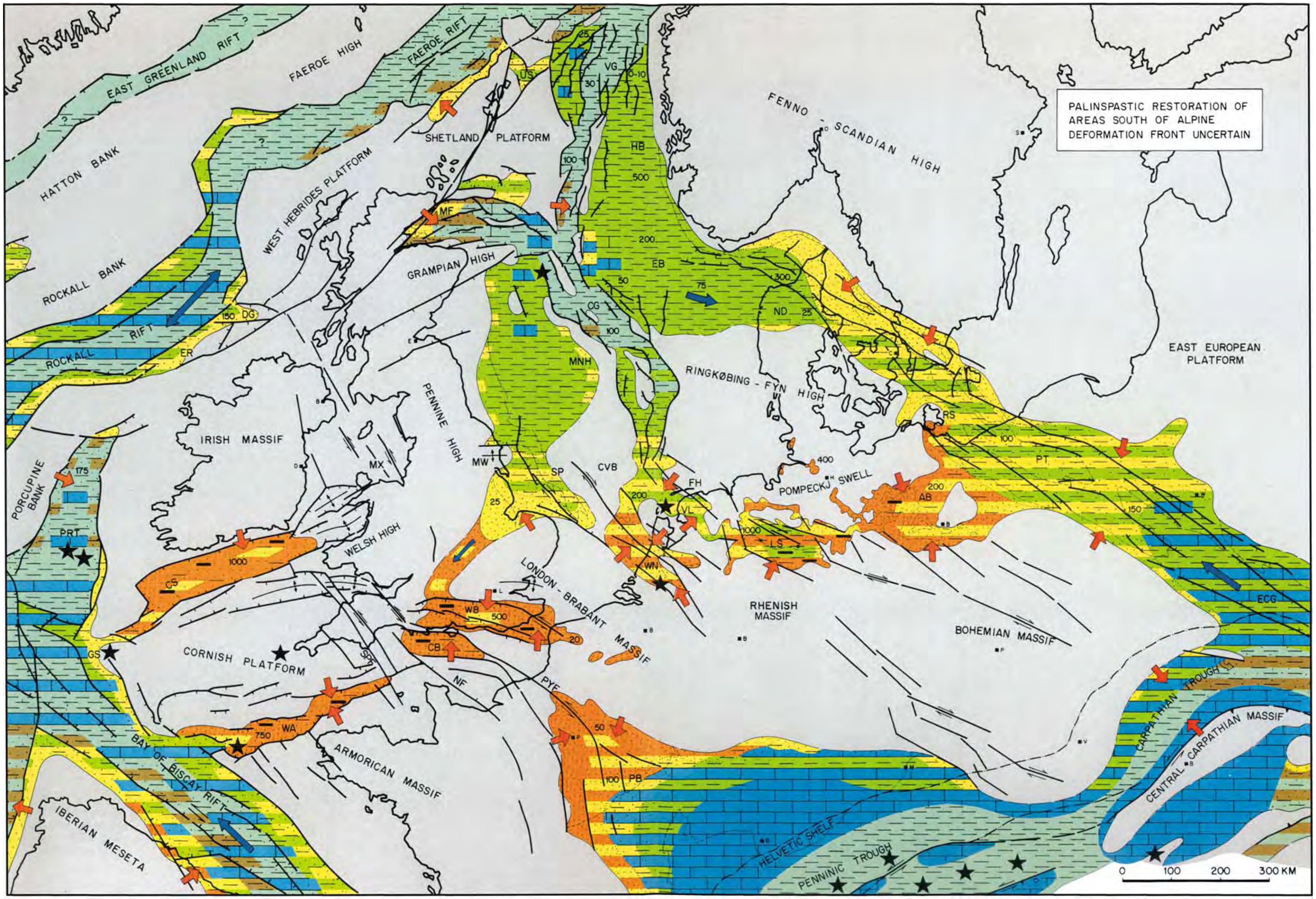


Figure 10 Berriasian Valanginian palaeogeography. For legend see Figure 16. Details of Enclosure from *Geological Atlas of Western and Central Europe* 2nd Edition, Peter A. Zeigler, 1990, published by Shell International Petroleum Mij. B.V., distributed by Geological Society Publishing House, Bath.

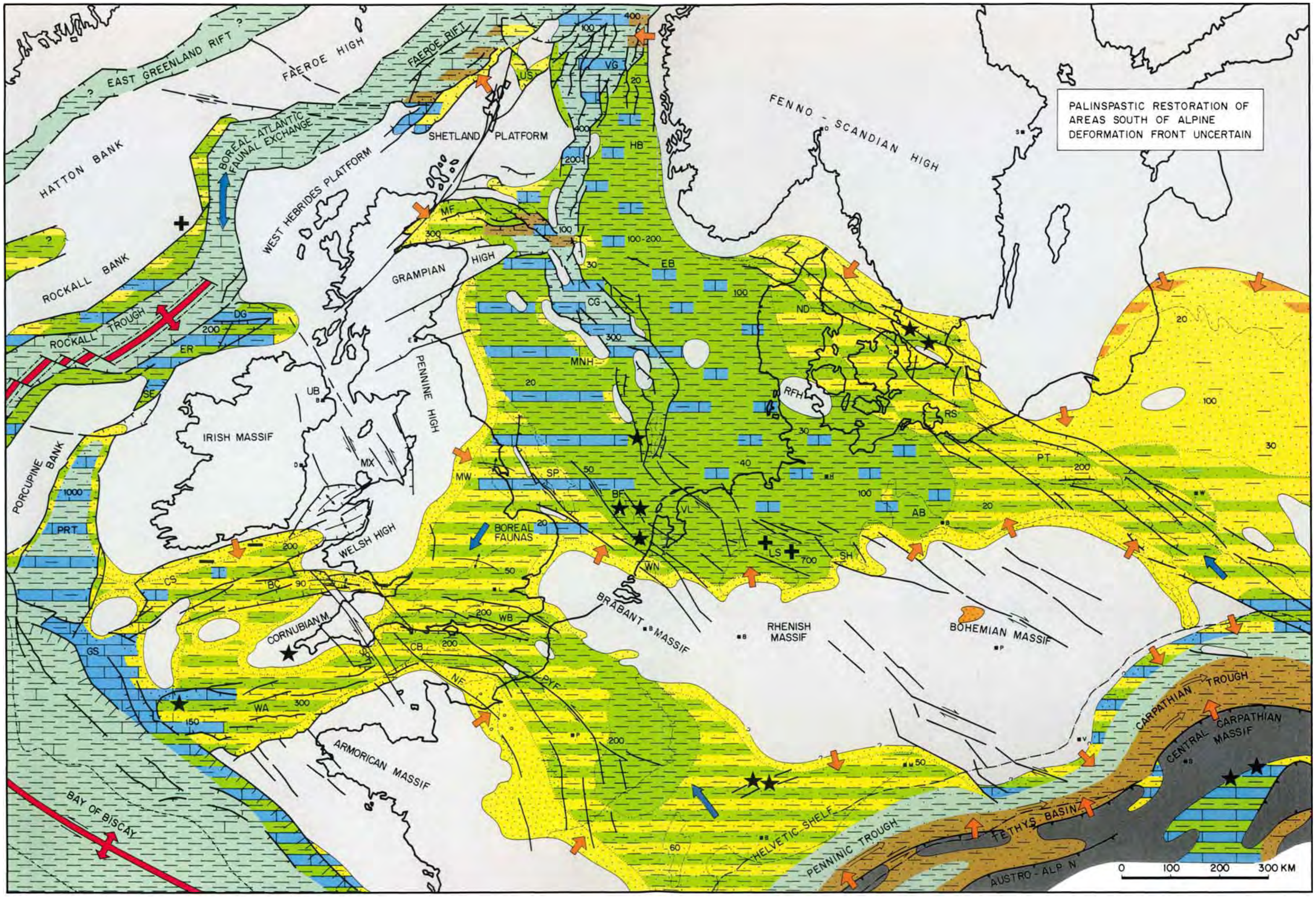


Figure 11 Aptian Albian palaeogeography. For legend see Figure 16. Details of Enclosure from *Geological Atlas of Western and Central Europe* 2nd Edition, Peter A. Zeigler, 1990, published by Shell International Petroleum Mij. B.V., distributed by Geological Society Publishing House, Bath.

giving rise to the development of the large Thulean flood basalt province that had, during the Paleocene, a radius of more than 1000 km. At the Paleocene–Eocene transition, crustal separation was achieved between Greenland and Europe to the west of the Rockall-Hatton Bank and in the Norwegian-Greenland Sea, paving the way for opening of the Arctic-North Atlantic Ocean.

During the Turonian–Santonian, Africa began to converge with Europe in a counter-clockwise rotational mode. Ensuing space constraints within the Tethyan belt were compensated by the activation of new subduction zones and the onset of closure of the Alpine Tethys and the Bay of Biscay.

Late Cretaceous

Late Cretaceous regional thermal subsidence of the North Sea Basin, the Polish Trough, the Paris Basin, and the Tethys shelves, as well as of the Western Shelves, combined with cyclically rising sea-levels, accounted for broad overstepping of the Mid-Cretaceous basin margins. This caused a drastic reduction of clastic supply to these basins in which sedimentation became carbonate-dominated from Cenomanian times onward. By end-Cretaceous times, much of the WCE was covered by the 'Chalk' sea, water depths of which ranged from neritic to bathyal, as evident in the axial parts of the North Sea Basin (Figure 12).

However, commencing in Late Turonian times, compressional stresses were exerted on the southern margin of the WCE, causing inversion of Mesozoic tensional basins and upthrusting of basement blocks by reactivation of pre-existing crustal discontinuities (Figure 3). The intra-Senonian pulse of intraplate compression, which affected the Polish Trough, the Fennoscandian Border Zone, the Bohemian Massive, the Sub-Hercynian, Lower Saxony, West Netherlands, and the Sole Pit Basins, as well as the southern parts of the North Sea rift, can be related to compressional stresses which were projected from the Alpine-Carpathian orogenic wedge through the oceanic lithosphere of the Alpine Tethys into the lithosphere of WCE. From the uplifted basement blocks of the Bohemian Massif and the Polish inversion axis, clastics began to be shed into the surrounding Chalk seas (Figure 12).

Paleocene

The second, more intense, Paleocene phase of intraplate compression, which affected about the same areas, and to a lesser degree also the Tethys shelves of the Western and Central Alps, the Paris Basin and the Channel area, probably marks the collision of the Alpine-Carpathian orogenic wedge with its East Alpine-Carpathian foreland and with the Briançonnais Terrane in the West and Central Alpine domain (*see* **Europe:** The Alps). The most distal intraplate compressional structures occurred at a distance of some 1500 km to the north-west of the contemporary collision front. The Paleocene phase of foreland compression, during which a Rocky Mountain-type array of basement blocks was upthrusted in the Bohemian Massif and the Polish Trough was deeply inverted, involved also broad lithospheric folding, causing a regional regression, the disruption of the Late Cretaceous carbonate platforms, and accelerated subsidence of the North Sea Basin. Combined with plume-related thermal doming of the British Isles during the development of the Thulean flood basalt province, sedimentation in the North Sea Basin changed at the end of the Danian from carbonates to clastics. At the same time, this basin became isolated from the Tethys and Atlantic seas but remained open to the Norwegian-Greenland Sea (Figure 13).

On the other hand, mixed carbonate-clastic environments persisted on the Western Shelves, that were not affected by this pulse of intraplate compression. However, Paleocene clastic influx into the Paris Basin from southern sources is probably related to the uplift of the Armorican-Massif Central arch in response to compressional stresses that were exerted on the shelves of Southern France during the early phases of the Pyrenean orogeny that had commenced during the Senonian. This is compatible with the Paleocene-Eocene development of the Languedoc-Provençal fold-and-thrust belt of Southern France, that involved inversion of Mesozoic extensional basins.

Paleocene intrusion of mafic dykes in the Massif Central, Vosges, Black Forest, and the Bohemian Massif, which must be related to partial melting of the lithospheric thermal boundary layer, probably reflected a mantle plume-related increase in the potential temperature of the asthenosphere, resulting in a renewed destabilization of the lithosphere. On the other hand, Paleocene compressional inversion of the Polish Trough and upthrusting of basement blocks in the Bohemian Massif apparently caused thickening of the crust (Figure 2).

Opening of the Arctic–North Atlantic and Collisional Interaction of the Alpine Orogen with its Foreland

With the Early Eocene onset of sea-floor spreading in the Arctic-North Atlantic, volcanic activity ceased on the shelves of the British Isles and the Rockall-Hatton-Faeroe Bank. With this, their post-rift thermal subsidence commenced, whilst thermal subsidence

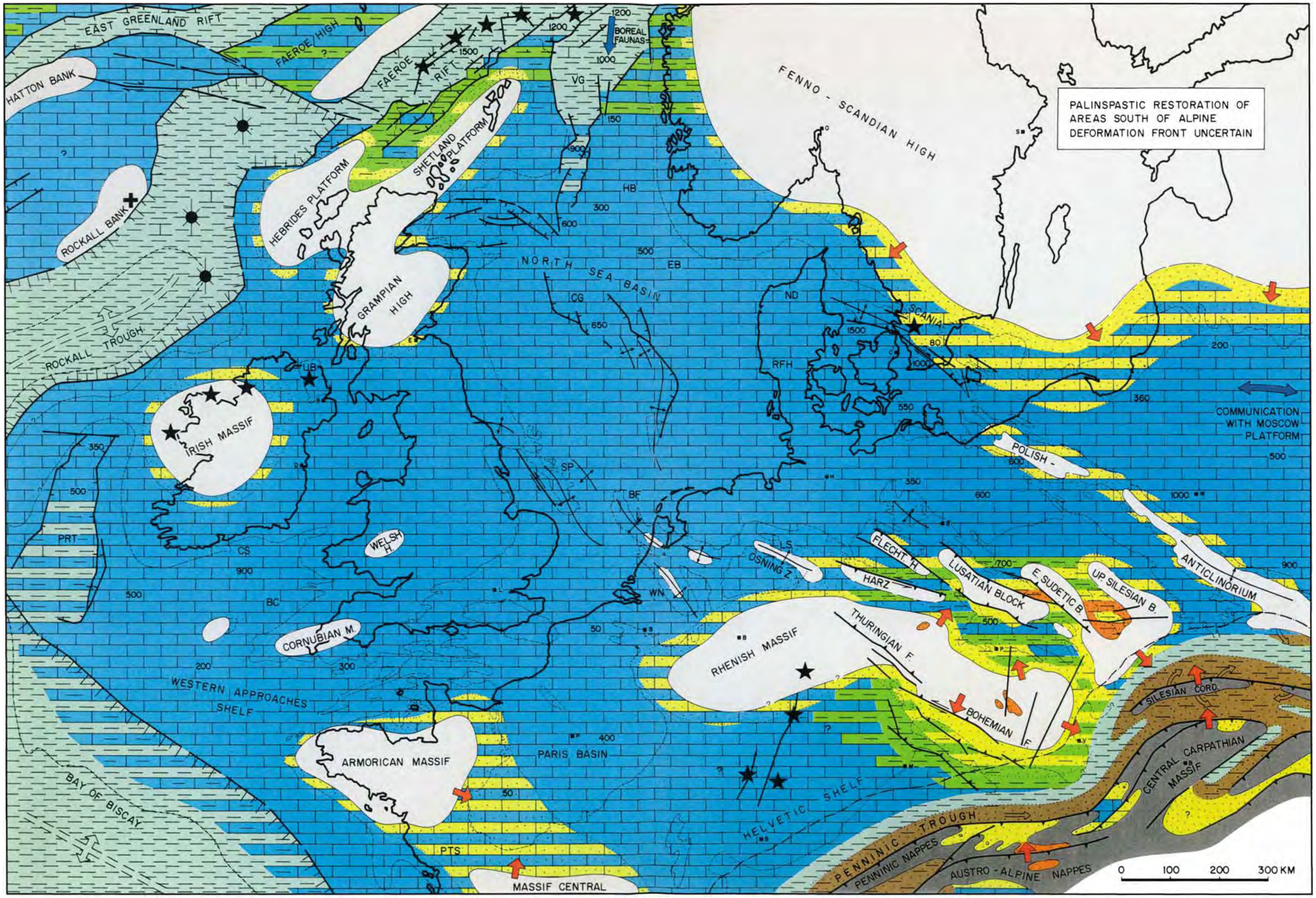


Figure 12 Senonian Danian palaeogeography. For legend see Figure 16. Details of Enclosure from *Geological Atlas of Western and Central Europe* 2nd Edition, Peter A. Zeigler, 1990, published by Shell International Petroleum Mij. B.V., distributed by Geological Society Publishing House, Bath.

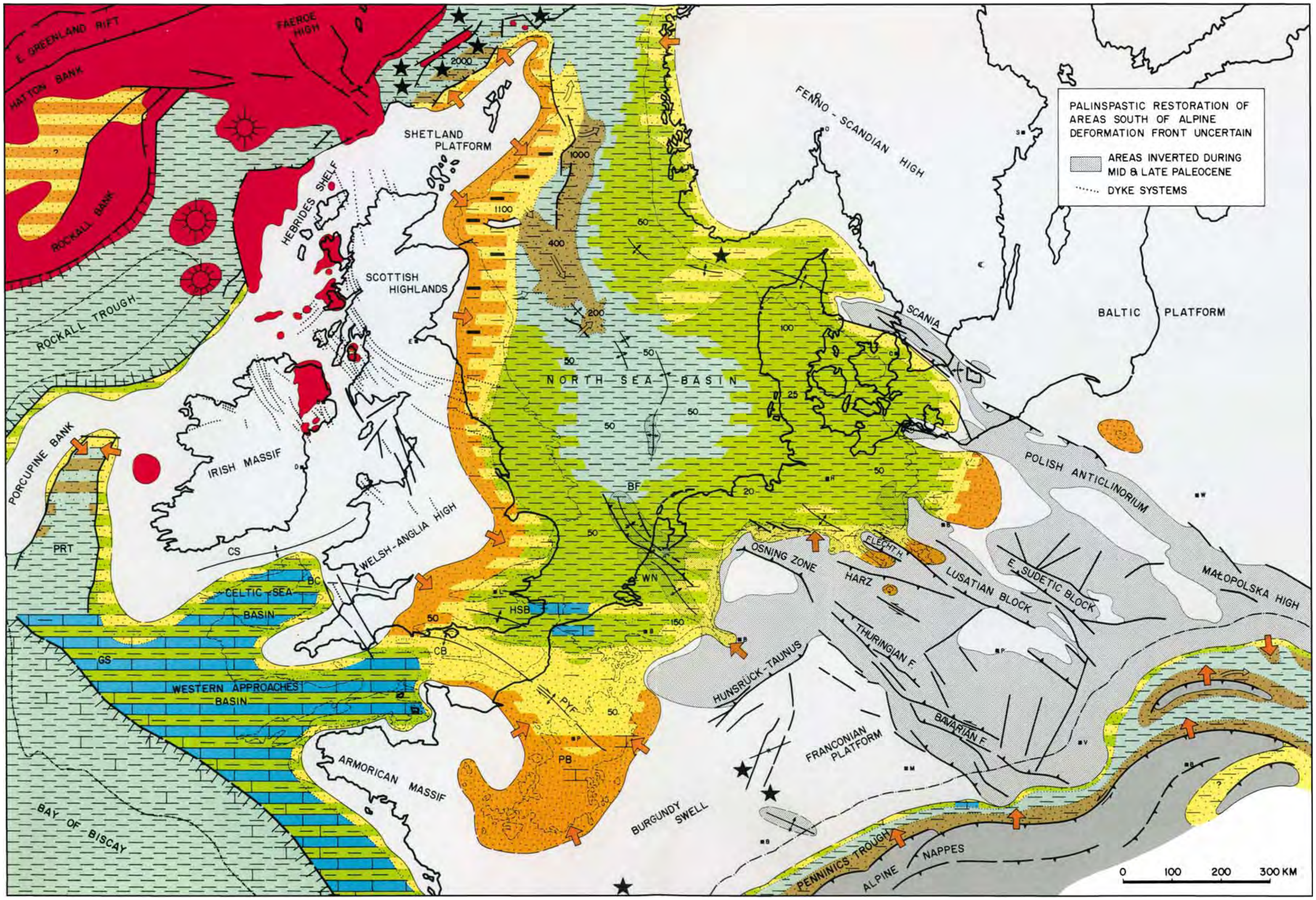


Figure 13 Late Paleocene palaeogeography. For legend see Figure 16. Details of Enclosure from *Geological Atlas of Western and Central Europe* 2nd Edition, Peter A. Zeigler, 1990, published by Shell International Petroleum Mij. B.V., distributed by Geological Society Publishing House, Bath.

of the North Sea Basin and the Western Shelves continued. During the Late Eocene and Oligocene, reorganization of sea-floor spreading axes in the Norwegian-Greenland Sea, the shelves of the British Isles were destabilized by minor wrench faulting in the prolongation of the Iceland ridge and the Charlie Gibbs fracture zones, causing the subsidence of small transtensional basins in the Irish Sea area (Figure 14). Moreover, repeated pulses of basin inversion interfered with the thermal subsidence of the Celtic Sea, Western Approaches, and Channel basins (Figure 3).

During the Eocene, thrust-loaded flexural subsidence of the foreland of the Western, Central, and Eastern Alps, and also the Carpathian foreland, commenced. Oligocene to Miocene emplacement of the East-Alpine and Carpathian nappe systems was, however, not accompanied by further intraplate compressional deformation of their forelands, thus reflecting mechanical decoupling of these orogens from their forelands. By contrast, Late Eocene–Early Oligocene and Late Oligocene–Early Miocene inversion pulses evident in the Celtic Sea, Western Approaches, Channel, Weald, Sole Pit, Broad Fourteens, and West Netherlands basins testify to intermittent and increasing mechanical coupling of the evolving West and Central Alpine Orogen with its foreland (Figures 3 and 14). Crustal shortening in the Western and Central Alps persisted during the Late Miocene and Pliocene, as evident by folding of the Jura Mountains, and may indeed still be going on, as indicated by earthquake activity and geodetic data.

In the Alpine foreland, development of the tectonically still active European Cainozoic rift system (ECRIS) commenced during the Late Eocene. Today this rift system extends over a distance of more than 1000 km from the Dutch North Sea coast to the Mediterranean. Its southern elements are the northerly-striking Limagne and the Valence and Bresse grabens, which are superimposed on and flank the Massif Central, respectively. These grabens are linked via the Burgundy transfer zone to the northerly-striking Upper Rhine Graben which bifurcates northwards into the north-west-trending Roer Graben and the north-easterly trending Hessian grabens that transect the Rhenish Massif. The north-east-striking Eger Graben, which transects the Bohemian Massif, forms an integral part of the ECRIS (Figure 14). Localization of ECRIS involved the reactivation of Permo-Carboniferous shear systems. Although characterized by relatively low crustal stretching factors, the evolution of the ECRIS was accompanied by the development of major volcanic centres on the Massif Central, the Rhenish Massif and the Bohemian Massif, particularly during Miocene and Plio-Pleistocene times. Seismic tomography indicates that mantle plumes well up beneath the Massif Central and the Rhenish Massif but not beneath the Vosges-Black Forest arch; similar data are, however, not available for the Bohemian Massif. Despite this, the evolution of the ECRIS is considered to be a clear case of passive rifting.

During the Late Eocene, the Valence, Limagne, Bresse, Upper Rhine, and Hessian grabens began to subside in response to northerly-directed compressional stresses that can be related to the collisional interaction of the Pyrenees and the Alps with their forelands. These originally-separated rifted basins coalesced during their Oligocene main extensional phase, and the Roer and Eger Grabens started. During the Late Oligocene, rifting propagated southward across the Pyrenean Orogen into the Gulf of Lions and along coastal Spain in response to back-arc extension, that was controlled by eastward roll-back of the subducted Betic-Balearic slab. By Late Burdigalian times, crustal separation was achieved, the oceanic Provençal Basin began to open, and the grabens of southern France became inactive. By contrast, the intra-continental parts of the ECRIS remained tectonically active until the present, although their subsidence has been repeatedly interrupted, possibly in conjunction with stresses controlling far-field inversion tectonics. By end-Oligocene times, magmatic activity increased on the Rhenish Shield. At the same time, the area of the triple junction between the Upper Rhine, Roer, and Hessian grabens became uplifted, presumably in response to thermal thinning of the lithosphere, interrupting the Oligocene sea-way which had linked the North Sea Basin with the Alpine foreland basin. By Middle–Late Miocene times, the Massif Central, the Vosges-Black Forest arch, and slightly later, also the Bohemian Massif, were uplifted. This was accompanied by increased mantle-derived volcanic activity. At the level of the Moho, a broad anticlinal feature extends from the Massif Central via the Burgundy Transfer zone, the Vosges-Black Forest into the Bohemian Massif (Figure 2). Uplift of these arches probably involved folding of the lithosphere in response to increased collisional coupling of the Alpine Orogen with its foreland. Uplift of the Burgundy transfer zone entailed partial erosional isolation of the Paris Basin. Under the present north-west-directed stress regime, which had developed during the Miocene and intensified during the Pliocene and reflects a combination of Alpine collisional and Arctic-North Atlantic ridge-push forces, the Upper Rhine Graben is subjected to sinistral shear, the Roer Graben is under active extension, whilst thermal uplift of the Rhenish triple junction continues. Moreover, the late phase of accelerated subsidence of the North Sea Basin, commencing in the Pliocene, as well as the

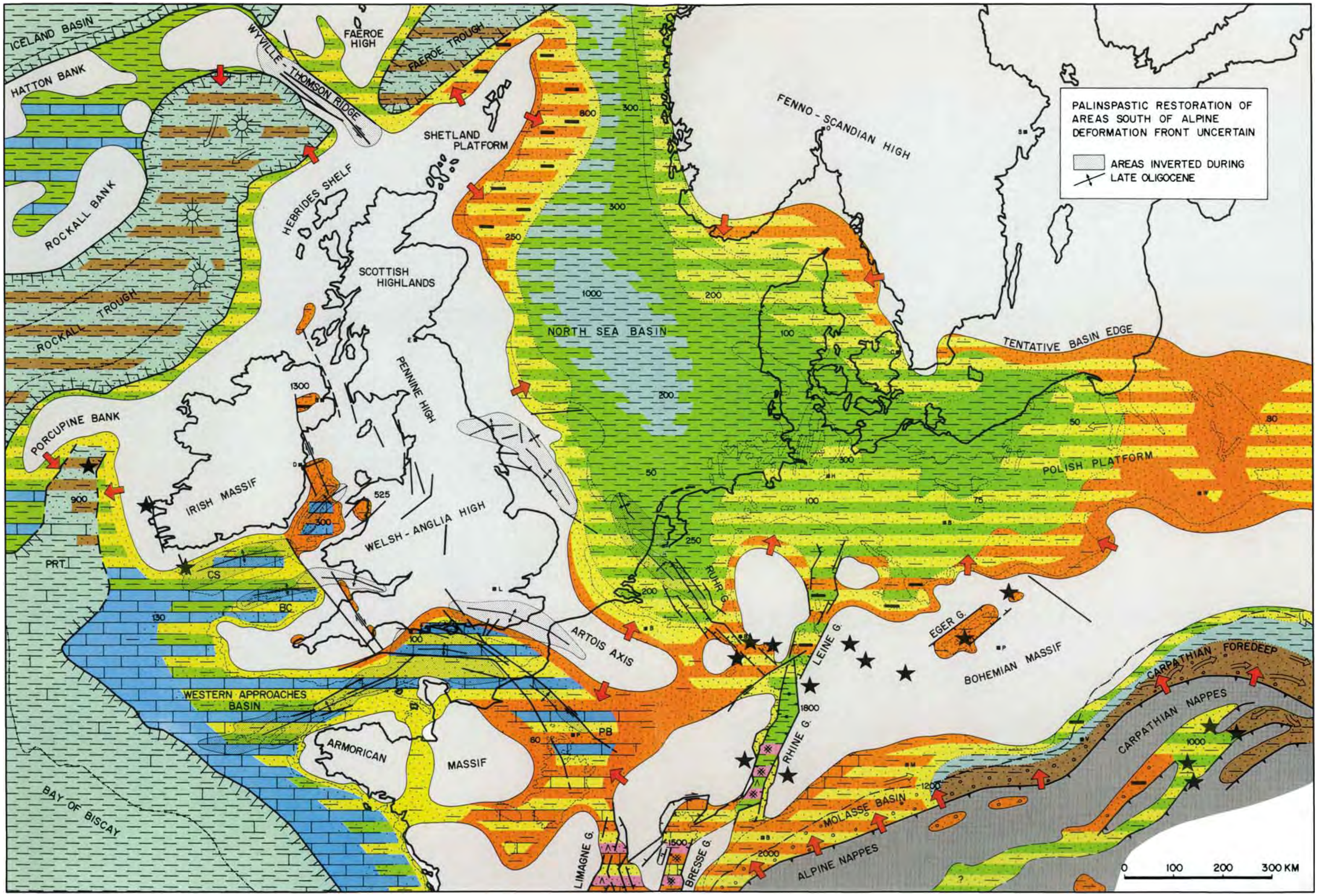


Figure 14 Oligocene palaeogeography. For legend see Figure 16. Details of Enclosure from *Geological Atlas of Western and Central Europe* 2nd Edition, Peter A. Zeigler, 1990, published by Shell International Petroleum Mij. B.V., distributed by Geological Society Publishing House, Bath.

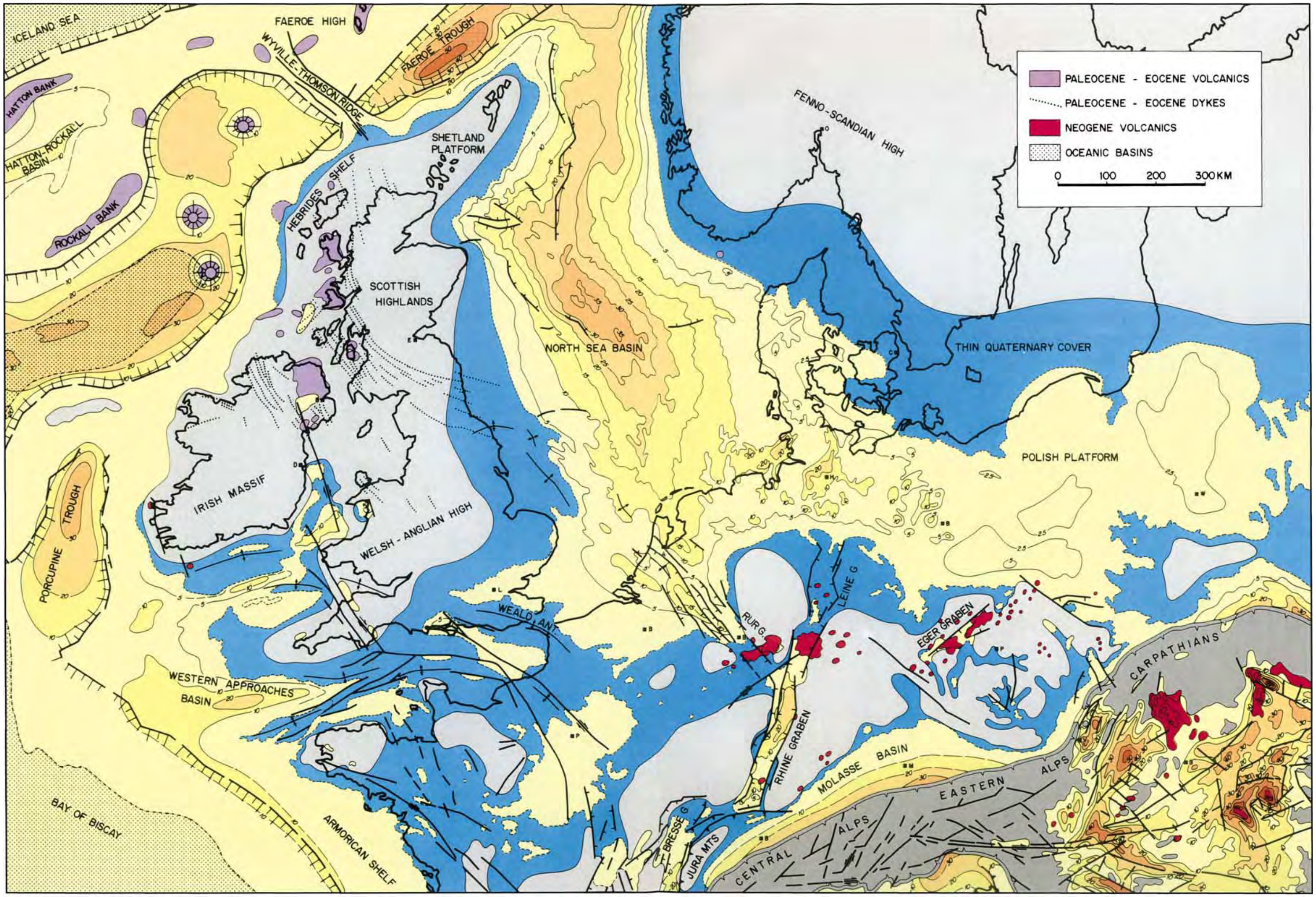


Figure 15 Isopach map of Cenozoic series. For legend see Figure 16. Details of Enclosure from *Geological Atlas of Western and Central Europe* 2nd Edition, Peter A. Zeigler, 1990, published by Shell International Petroleum Mij. B.V., distributed by Geological Society Publishing House, Bath.

Figure 16 Legend to palaeogeographic, isopach and tectonic maps and to stratigraphic correlation charts.

contemporaneous uplift of the British Isles and the Fennoscandian Shield, can be related to stress-induced deflection of the lithosphere.

Progressive uplift of the Rhenish and Bohemian massifs was coupled with the development of the modern drainage system of Central Europe and the shedding of massif clastics into the continuously subsiding North Sea Basin in which water depths gradually decreased. The Cenozoic isopach map (Figure 15) illustrates the broad saucer-shaped geometry of the North Sea thermal sag basin, the axis of which coincides with the trace of its underlying Mesozoic rift. This indicates that during Mesozoic rifting not only the crust but also the mantle-lithosphere of the North Sea Basin were significantly thinned (Figure 2). In contrast, Cenozoic sediments are generally thin on the Polish Platform, which is underlain by the deeply inverted Polish Trough that is marked by a crustal root. In northern Germany, sharp lateral Cenozoic thickness changes are related to the growth of Permian salt diapirs. A comparison of the present day erosional edges of Cenozoic basins and their depositional margins (Figure 15) illustrates the scope of Neogene to Recent uplift of the WCE in response to basin inversion and lithospheric folding.

Resources

The Permian and younger sedimentary basins of the WCE host a number of important hydrocarbon provinces, the most outstanding of which are tied to the Mesozoic North Sea rift system and the southern margin of the Southern Permian Basin. The hydrocarbon systems of the North Sea rift, as well as of the Faeroe trough, are largely tied to marine Kimmeridgian organic shales that charged reservoirs, ranging in age from Devonian to Palaeogene, with oil and gas. The gas-prone Southern Permian hydrocarbon province relies for hydrocarbon charge on Westphalian coal measures that were deposited in the Variscan foreland basin. Organic deeper water Zechstein shales and carbonates represent contributing source-rocks. Main reservoirs are Rotliegend sands, Zechstein carbonates, and Triassic sands. Early Jurassic marine organic shales control the petroleum systems of the Paris and Channel basins, as well as of the onshore parts of the North-west European basin in the Netherlands and Germany. The oil and gas province of the Aquitaine Basin relies for hydrocarbon charge mainly on marine Kimmeridgian and Berriasian shales. Hydrocarbons occurring in the Alpine foreland basin of Germany and Austria, as well as in the Upper Rhine Graben, were mainly derived from Oligocene marine shales.

Permian salts involved in diapric structures are widely exploited in the onshore parts of the WCE. Triassic salts are exploited in basins that are superimposed on Variscan crust. Polyhalites are associated with the Zechstein halites in the Southern Permian Basin and with Oligocene halites in the Upper Rhine Graben.

See Also

Africa: Rift Valley. **Europe:** Caledonides of Britain and Ireland; Permian Basins; The Alps. **Fossil Invertebrates:** Cephalopods (Other Than Ammonites). **Mesozoic:** Triassic; Jurassic. **Moho Discontinuity**.

Further Reading

BRGM, Societé Elf Aquitaine, Esso REP and SNPA (1973) *Géologie du Bassin d'Aquitaine*. Edition Bureau de Recherches Géologiques at Minières, Paris (Atlas).

Boldy SAR (ed.) (1995) Permian and Triassic Rifting in Northwest Europe. *Geological Society of London, Special Publication* 91: 263.

Cooper MA and Williams GD (eds.) (1989) Inversion Tec tonics. *Geological Society of London, Special Publication* 44: 375.

Cope JCW, Ingham JK, and Rawson PF (eds.) (1992) Atlas of Palaeogeography and Lithofacies. *Geological Society of London, Mem.* 13: 152.

Dadlez R, Marek S, and Pokorski J (eds.) (1998) *Palaeo geographical Atlas of the Epicontinental Permian and Mesozoic in Poland.* 1:2,500,000. Panstwowy Instiytuy Geologiczny, Warsaw, 75 plates.

Debrand Passard S and Courbouleix S (1984) *Synthèse géologique du sud est de la France*. Vol. 2. Atlas: Strati graphie et paléogeographie. Mém. B.R.G.M. 126.

Dèzes P, Schmid SM, and Ziegler PA (2004) Evolution of the European Cenozoic rift system: interaction of the Alpine and Pyrenean orogens with their foreland litho sphere. *Tectonophysics* (in press).

Glennie KW (ed.) (1998) *Petroleum Geology of the North Sea. Basic concepts and recent advances,* 4th Edn. Oxford: Blackwell Science.

Granet M, Wilson M, and Achauer U (1995) Imaging mantle plumes beneath the French Massif Central. *Earth Planetary Scientific Letters* 136: 199 203.

Kockel F (ed.) (1996) *Geotectonic Atlas of NW Germany, 1:300000.* Hannover, Germany: Federal Institute for Geosciences and Natural Resources.

Mégnien C (1980) Synthèse géologique du Bassin de Paris. *Mém. B.R.G.M.* 102: Vol. 2, Atlas.

Parker JR (ed.) (1993) *Petroleum Geology of Northwest Europe*. Proceedings of the 4th Conference. *Geological Society of London,* Vol. 1 & 2: 1542.

Parnell J (ed.) (1992) Basins of the Atlantic Seaboard: Petroleum geology, Sedimentology and Basin Evolu tion. *Geological Society of London, Special Publica tion* 62: 470.

Sissingh W (1998) Comparative stratigraphy of the Rhine Graben, Bresse graben and Molasse Basin: correlation of Alpine foreland events. *Tectonophysics* 300: 2249 284.

Stampfli G, Borel G, Cavazza W, Mosar J, and Ziegler PA (eds.) (2001) *The Paleotectonic Atlas of the PeriTethyan Domain.* CD ROM, European Geophysical Society.

Ziegler PA (1988) Evolution of the Arctic North Atlantic and the Western Tethys. *American Association Petroleum Geology, Mem.* 43; 198 p. and 30 plates.

Ziegler PA (1990) *Geological Atlas of Western and Central Europe,* 2nd Edn., Shell International Petroleum Minj. B.V., distrib. Geol. Soc., London, Publishing House, Bath, 238 p. and 56 encl.

Ziegler PA, Bertotti G, and Cloetingh S (2002) Dynamic processes controlling foreland development the role of mechanical (de)coupling of orogenic wedges and fore lands. *European Geophysical Society, Stephan Mueller Special Publication Series* 1: 29 91.

The Alps

O A Pfiffner, University of Bern, Bern, Switzerland

Introduction

The Alps as a Mountain Belt

The European Alps, a mountain chain with elevations reaching almost 5000 m, stretch from Nice to Vienna. The highest peak, Mont Blanc, reaches an elevation of 4807 m. Mont Blanc is part of a belt of granites that stretches from the Pelvoux massif in France to the High Tauern in Austria. The chain runs north–south from Nice, on northward, forming a 90° bend in Switzerland and then continuing eastward towards Vienna (Figure 1). It is narrowest in the transect of Switzerland. The mountain chain is dissected by numerous deeply incised valleys, some of which run parallel to the chain. To the north of the Alps, the Danube system drains into the Black Sea, the Rhine system drains into the North Sea, and the Rhone system drains into the Mediterranean. South of the Alps, the Po system drains into the Adriatic Sea.

The North-Alpine foreland basin, called the Molasse basin, stretches along the north side of the Alps. It was filled by sediments carried in by the rivers draining the Alps northward between 34 and 10 Ma. Similarly, the Po basin to the south of the Alps received the sediments from the Apennine chain and from the rivers draining the Alps southward. Both basins formed during the building of the mountain chain. The weight of the mountain chain flexed the tectonic plates on either side, creating depressions that readily filled and became shallow seas. Up to 35 km of rocks were eroded from the growing Alpine chain and accumulated in these depressions. Judging from the nature of the accumulated sediments, denudation of the growing chain kept pace with the vertical uplift. The rising mountain chain was probably never much higher than it is today.

Major Tectonic Units

The Alps formed as a result of the collision of the Eurasian and African plates, two continental tectonic plates that were initially separated by ocean basins. Starting around 100 Ma, these two plates moved closer to each other, closing the ocean basins between them and ultimately colliding. Consequently, on present-day tectonic maps of the Alps (Figure 2), it is possible to distinguish between rock suites pertaining to one or the other of these continents, or the ocean basins between them. The Helvetic zone, Jura Mountains, and the area north and west of the Alps pertain to the former European margin of the Eurasian Plate; the Austroalpine and Southalpine zones are parts of the former Adriatic margin of the African Plate. The Penninic zone is made up of sediments that accumulated in ocean basins that were located between the two continents, as well as the crustal rocks underlying these basins. During the closure of these basins and the ensuing collision of the two continental plates, the European margin was dragged down south-eastward beneath the Adriatic margin. The Eastern Alps are dominated by rocks of the upper plate, the Adriatic margin (Figure 2). Erosion has removed this upper plate almost completely in the Central and Western Alps. However, erosional remnants (termed 'klippen') of the upper plate in the Western and Central Alps, as well as pebbles carried out into the foreland by ancient rivers, prove that it once occupied much of the entire Alps.

The ocean basins between the Eurasian and Adriatic continental plates formed in response to the opening of the Atlantic Ocean. Figure 3 shows the palaeogeography at 170 and 130 Ma. Two basins, the Valais basin and the Piemont Ocean (sometimes called the Liguria–Piemont Ocean), formed between the European and Adriatic margins of the two continental plates. The two basins were separated by a microcontinent, the Briançonnais swell, which was connected to the Iberian Peninsula at the time. Plate

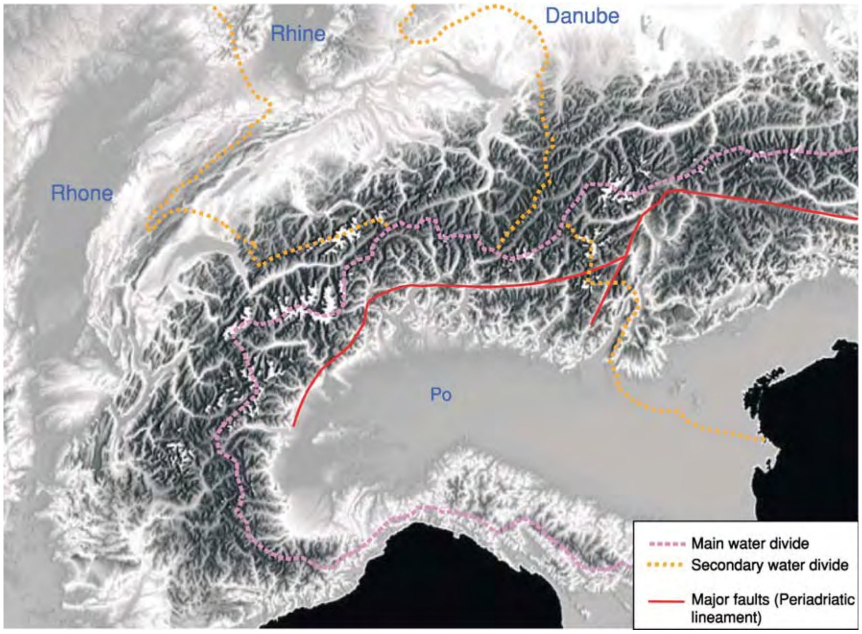


Figure 1 Digital elevation model, showing the large scale geomorphic features of the Alps and surrounding areas.

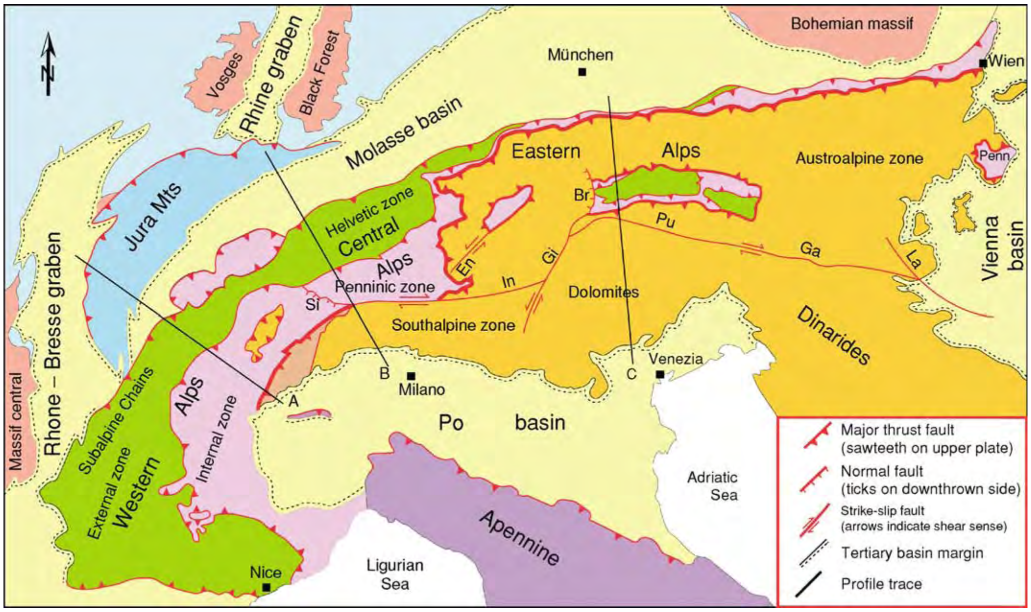


Figure 2 Tectonic map of the Alps, showing the major tectonic units. Sites A, B, and C correlate to orogen profiles A, B, and C in Figure 4. Faults: Br, Brenner; Si, Simplon; En, Engadine; In, Insubric; Gi, Giudicarie; Pu, Pustertal; Ga, Gailtal; La, Lavanttal.

movements during the opening of the Atlantic were such that the African plate moved eastward relative to the Eurasian plate, and both plates moved away from the North and South American plates. At an early stage (around 170 Ma), this sinistral plate movement occurred along a fracture zone that passed through Gibraltar. This opened the Piemont ocean in the area of the future Alps. At a later stage (around 130 Ma), when the opening of the Atlantic had proceeded further north, the sinistral movement occurred along a fault zone passing north of the Iberia–Briançonnais continental fragment along the Gulf of Biscay. This opened the Valais basin in the area of the future Alps.

The fate of the Piemont and Valais basins was controlled by convergence between the Eurasian and

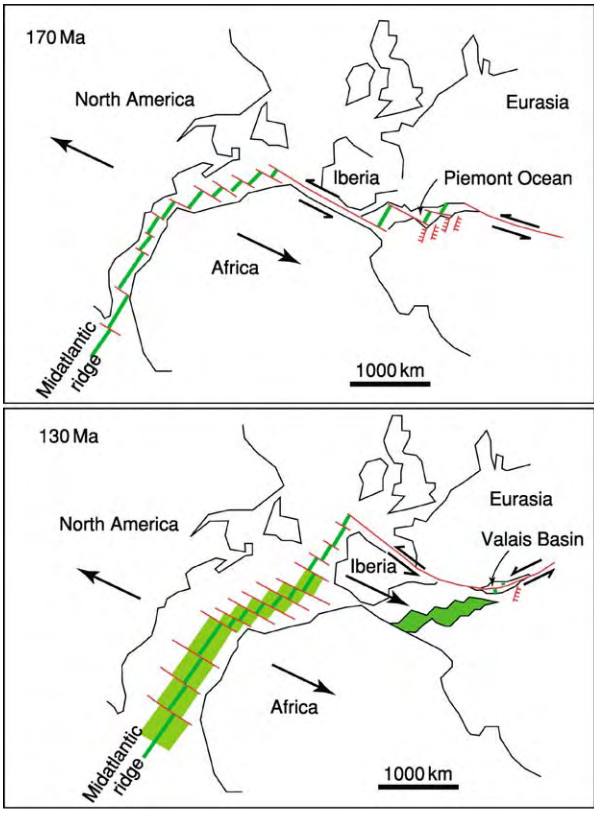


Figure 3 Palaeogeographical maps at 170 and 130 Ma, showing the future Alpine domain in the framework of the associated tectonic plates. Opening of the Piemont Ocean and the Valais Basin was closely linked to the opening of the Atlantic.

African plates. During this convergence, the basins were subducted beneath the Adriatic margin and were closed. Their lower part (the lithospheric mantle) was subducted and recycled into the mantle. Their upper parts (the crust and the sediments deposited in the two basins) were compressed and incorporated into the growing Alpine Orogen, where they now form the Penninic nappes. The European and Adriatic margins were also compressed during convergence and collision of the two plates. Consequently, the Helvetic nappes and the Jura Mountains, both of which consist essentially of Mesozoic shelf sediments, were formed on the European side. Similarly, the Austroalpine and Southalpine nappes represent the deformed Adriatic margin of the African plate.

Rock Types

When discussing the rock types that can be found in the Alps, it is useful to distinguish between rocks that formed prior to the opening of the Piemont and Valais basins and rocks that formed during and after the opening of the basins. In terms of Alpine geology, the older rocks are referred to as basement. This basement consists of two major units, crystalline rocks (granites and polymetamorphic gneisses and schists) and Palaeozoic sediments and volcanics, that pertain to mountain belts formed at 300 to 400 Ma. Granitic rocks are resistant to erosion, thus it is no surprise that they form many of the higher peaks in the Alps (including Mont Blanc).

The younger rocks are Mesozoic and Cenozoic sediments and volcanics ranging in age from 225 to 10 Ma. Large quantities of carbonates accumulated along the shelf seas of the continental margins, reaching thicknesses of more than 1 km. These carbonates now form the high cliffs that dominate the present-day morphology of the Alps. Sandstones and shales accumulated as basins on the continental slopes. In some instances, these basins were flanked by faults that formed in response to the breakup of the continents. Breccias accumulated at the foot of the steep fault scarps. The deepest part of the basins consisted of newly formed oceanic crust. Deep-sea sediments (radiolarian cherts) slowly covered the basaltic lava flows of the newly formed ocean floor.

Deep Structure of the Alps

A number of experiments have been designed to image the structure of the Alps to depths of over 50 km. Dynamite detonations and vibrator trucks located over a subsurface target generate seismic waves that travel downward and then are reflected back upward at various discontinuities in the Earth's crust. The upward-reflected waves (or echoes) are recorded by a surface array of geophones and procesed into a coherent 2-dimensional image. The resulting seismic sections can then be interpreted in terms of subsurface geological structure. Figure 4 summarizes the findings for three transects through the Alps.

Western Alps

Within the framework of ECORS-CROP (CROP = CROsta Profonda = Deep Crust; ECORS = Etude de la Croûte Continentale et Océanique par Réflexion et Réfraction Sismiques = Study of the continental and oceanic crust by reflection and refraction sesimics), a joint project between France and Italy studying the deep continental crust by reflection and refraction sesimics, researchers have profiled a transect across the Western Alps (site A in Figure 2). The profile in Figure 4A summarizes the findings of the ECORS–CROP project. The Western Alps have an asymmetric structure. On the European margin, i.e., in the western part of the Alpine orogen, the crust–mantle

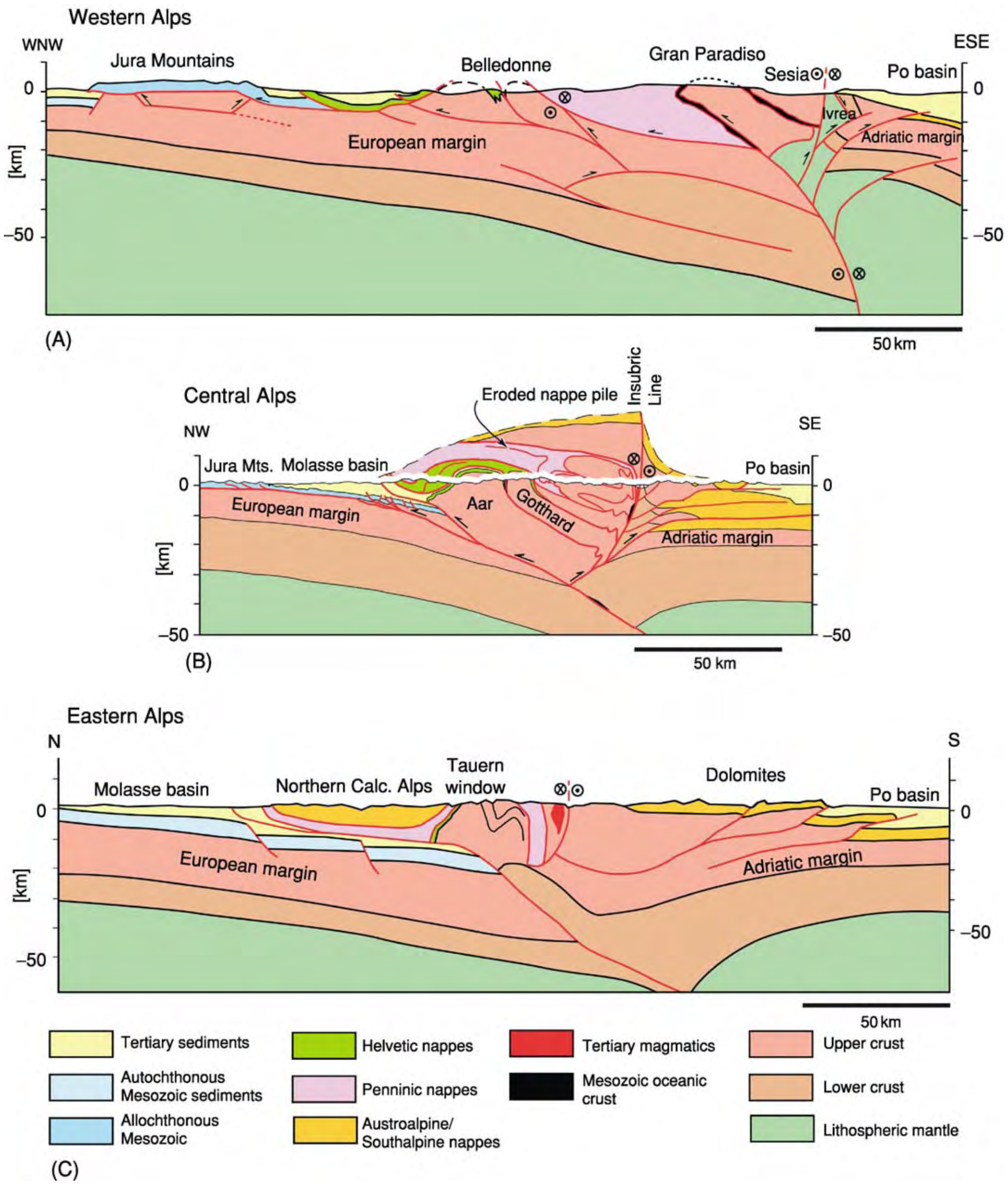


Figure 4 Three profiles through the Alps, showing the deep structure of the orogen. The profile sites are shown in Figure 2. (A) Transect through the Western Alps of France and Italy. (B) Transect through the Central Alps of Switzerland and Italy. (C) Transect through the Eastern Alps of Austria and Italy. See text for discussion.

boundary dips to the east, attaining depths exceeding 50 km beneath the orogen. Much of the crustal root is made up of lower crust, which appears to be tripled in this transect. This stacking was accomplished by thrust faults that moved entire blocks of crustal rocks upward towards the west. On the Adriatic margin, on the other hand, the crust–mantle boundary rises towards the surface, proceeding westward. This rise is accentuated by several thrust faults, which also affect the lower crust. In the case of the Ivrea zone, these lower crustal rocks and pieces of the mantle actually outcrop at the surface. The asymmetric structure of the Western Alps is a consequence of the collision of the Adriatic and European margins resulting from the convergence of the African and Eurasian plates. The ocean basins located between the two margins were highly deformed by these plate movements. The sediments, as well as their crustal substrate, were compressed, shortened, and stacked. These rocks now form the Penninic nappes shown in Figure 4A. The upper crust of the European margin was also shortened and thickened in the process of collision, forming a large-scale basement uplift (the Belledonne massif) within the Alps.

Further out towards the foreland, shelf sediments of the European margin were detached from their substrate, shortened by folding and thrusting, and transported towards the west. They now form the Jura Mountains. In addition to all of the thrusts and folds that represent WNW–ESE shortening, there were substantial movements in and out of the plane of the section shown in Figure 4A. These movements were related to strike–slip motions that displaced the Adriatic margin towards the north relative to the European margin.

Central Alps

Investigation of the Central Alps by the Swiss National Research Project (NRP) 20 has provided results complementary to those of the European GeoTraverse (EGT). The EGT study assessed the continental lithosphere that runs from the North Cape across Europe to Tunisia. Several transects were profiled across and within the Swiss Alps. Figure 4B, a profile along the central traverse of NRP 20, has been extended to include the Jura Mountains and the Po basin. As with the Western Alps, the asymmetric structure of the Central Alps evolved during convergence between Eurasia and Africa. The lower crust of the European margin extends at constant thickness beneath the Adriatic lower crustal wedge. The tip of the latter is exposed at the surface. The centre of the orogen consists of European margin upper crustal rocks that were stacked by thrust faults during plate convergence and collision. The associated heat and pressure transformed the rocks: granites became orthogneisses, sediments became paragneisses, and limestones became marbles. The mineral assemblages “frozen” in these metamorphic rocks allow determination of the temperature and pressure paths these rocks took during collision and the ensuing denudation.

Along one major fault, the Insubric Line, the nappe stack was moved upward and southward, but erosion kept pace with this uplift and removed a large section of the Austroalpine and Penninic nappes. Studies of the rocks outcropping at the surface indicate that they were once buried to depths exceeding 25 km. Strike–slip motion along the Insubric Line moved the Adriatic margin westward relative to the European margin in the later stages of the collision. In the north-western part of the orogen, the thrusting was chiefly towards the north-west. Conversely, thrusting was directed towards the south-east in the south-eastern part of the orogen. In both cases, the sequence of thrusting was from the centre of the orogen towards the forelands. This can be interpreted as the result of the collision of the European and Adriatic continental margins. Deformation of the crust occurred within the zone of contact and the deformed rocks were shoved on top of each other to form an orogenic wedge, similar to the wedge of snow forming in front of a moving snowplough. Shortening related to plate convergence and collision is particularly accentuated along this transect of the Alps. The associated uplift and erosion are indicated by the high degree of metamorphism of the exposed rocks outcropping at the surface. The focused horizontal shortening (and vertical stretching) explain why the Alpine chain is particularly narrow in this transect.

Eastern Alps

The Eastern Alps have been studied in a joint project (TRANSALP) between Germany, Austria, and Italy. A seismic transect through the Eastern Alps and the Dolomites has produced the seismic data and geological interpretations shown in Figure 4C. The crustal structure of the Eastern Alps is rather different from that of the western and central sections. The European margin shows a thin slab of lower crust dipping to the south. The Adriatic lower crust is thicker and has a piece protruding into the European upper crust. The upper crust of the European margin forms a large-scale basement uplift that is exposed in the Tauern window. The Helvetic nappes are virtually absent in this transect and the Penninic nappes make up only a proportionally small volume . The upper crust of the Adriatic margin, on the other hand, is much more voluminous. It is over 30 km thick and is shortened by thrust faults. Important strike–slip motions have displaced the Adriatic margin to the west, relative to the European margin along the Pustertal line.

Alpine Nappe Structures

Collision between the two margins of the Eurasian and African plates, like the crash of two cars, led to severe deformation (folding and fracturing) of the rocks within the zone of contact. Large-scale fractures in a zone of compression led to the formation of thrust faults that transported entire crustal blocks upward on a gently inclined fault surface. The displaced blocks were typically several hundreds of kilometres wide, 50–100 km long, and only a few kilometres thick. These thrust sheets are termed ‘nappes’, from the French for ‘tablecloth’, because of their shape. The (horizontal) shortening of the continental crust also led to the folding of layers of rock. The folds, which occur on every scale, from millimetres to kilometres, are an expression of the penetrative deformation of crustal rocks.

Thrust Faults

Geologists in the nineteenth century were puzzled when they saw older rocks lying on top of younger rocks. One of the classic locations where this is readily observed is in eastern Switzerland (Figure 5). The rugged peaks of the Tschingelhore are formed by Permian clastic sediments (roughly 260 Ma), but the rock immediately beneath the peaks consists of Jurassic limestones (formed about 150 Ma). The famous Glarus thrust is the sharp contact between the close proximity of the Permian and Jurassic units. The geographical extension of the Glarus thrust can be seen in a profile across eastern Switzerland (Figure 6). The Helvetic nappes were displaced northward along the Glarus thrust over a distance of 50 km. In order to displace a thrust sheet, a weak, basal lubricating layer must be present. In the case of the Glarus thrust, this lubricating layer is a highly sheared limestone, which was scraped off the footwall and dragged and drawn out along the thrust surface. It now forms a layer roughly 1 m thick and can be traced northward over a distance of some 30 km (in Figure 5 it is visible as a thin, light line along the horizontal contact).

As is evident from Figure 6, substantial internal deformation affected the rocks below and above the Glarus thrust. Highly sheared folds and thrust faults, repeating the various sedimentary layers many times, can be observed in the Mesozoic sediments beneath the Glarus thrust. Similarly, the rocks of the displaced block above the Glarus thrust show thrust faults and folds. One of these thrust faults, the Säntis thrust, displaced the younger Cretaceous strata much further north, compared to the older (Triassic to Jurassic) strata. In this case, the thrust fault was lubricated by a thick layer of shales, which have relatively low shear resistance.

Nappe Internal Deformation

As a thrust sheet is compressed, detached from its substrate, and transported, it undergoes internal deformation. The type of deformation depends largely on the nature of the rocks involved. In the case of a layered sequence of sedimentary rocks, folding prevails, whereas if the mechanical contrasts are low, such as sometimes occurs in a suite of granitic and gneissic rocks, the deformation may be more homogeneous. If mechanically weak layers of rocks are present, the deformation is focused and thrust faults may develop. All of these processes – folding, faulting, and homogeneous deformation – may take place jointly.

Figure 7 shows folded Cretaceous and Eocene limestones (100 to 50 Ma) that form an asymmetric anticline. The folded limestones are overlain by older rocks. Dolomites ($\sim$220 Ma) form the yellow cliff beneath the summit and crystalline basement rocks ($>$300 Ma) make up the dark summit. These older

Figure 5 The Glarus thrust in the Tschingelhore (between Flims and Elm in eastern Switzerland). The thrust fault is visible as a sharp horizontal contact between the older rocks that form the rugged peaks and the younger rocks that form the cliffs above the snowfields.

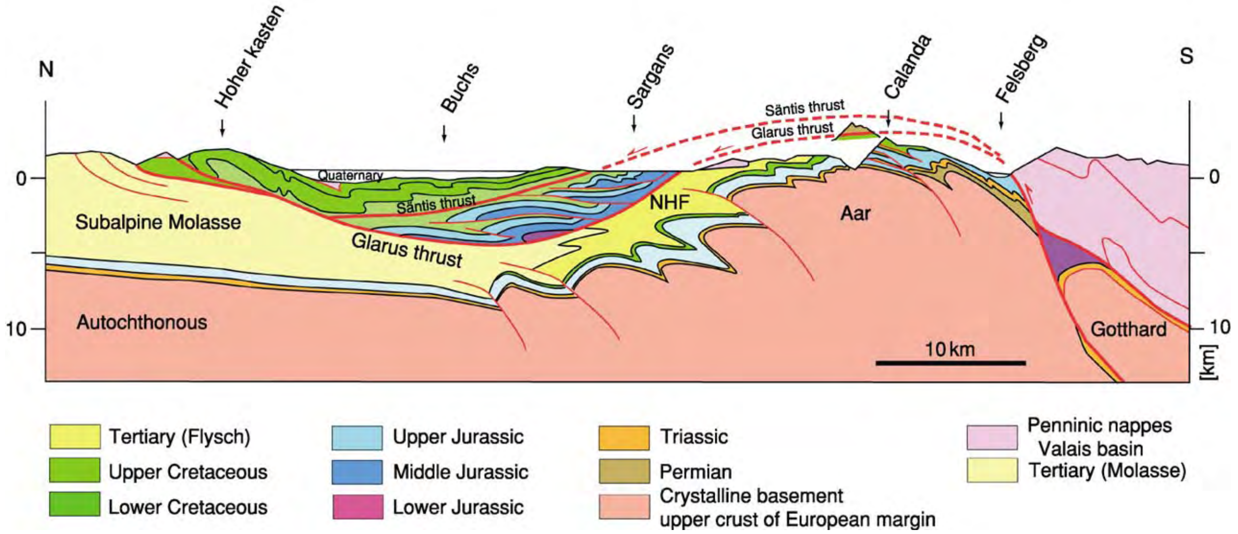


Figure 6 Profile through the Helvetic nappes of the eastern Swiss Alps. The Helvetic nappes were displaced along the Glarus thrust over a distance of around 50 km. But the rocks above and below the Glarus thrust were also intricately deformed, as is evident from the fold and thrust structures. The Santis thrust displaced the uppermost part, the Cretaceous strata, of the Helvetic nappes an additional 10 km to the north. Deeper down, the crystalline basement rocks of the Aar massif now form an anticlinal upwarp. NHF, North Helvetic Flysch.

Figure 7 Folded strata in the flank of Piz d'Artgas ('peak with arcs'), overthrust by older rocks forming the summit and the yellow cliff beneath.

rocks were emplaced along a thrust fault that is located near the base of the yellow cliff. Figure 8 is a profile across the Tauern window, where the upper crust of the European margin forms a large anticlinal fold. In the centre of the upwarp, erosion has removed the higher nappes, thus providing an insight into the formerly deeper parts of the orogen. The crystalline basement rocks in the core of the upwarp were compressed and internally shortened. From the deformed mineral grains of the rocks it is possible to determine how much horizontal shortening and vertical stretching actually occurred and to reconstruct the shape of the upwarp prior to this homogeneous deformation. The present-day shape of the Tauern upwarp (Figure 6), as well as its reconstructed geometry prior to homogeneous shortening, provide a reminder of the ductile behavior that granitic rocks can exhibit in the course of plate collision.

The Klippen nappe, a Penninic nappe in the French–Swiss Alps, is a classic example of a style of internal deformation characterized by fold-and-thrust structures (Figure 9). The lubricating layer (evaporites) at the base of the nappe consists of a thick layer of anhydrite. This rock type, which has a particularly low shear strength, forms when very shallow areas of seawater evaporate. The great thickness of the weak evaporite layer in the northern part of the section shown in Figure 9 facilitated the formation of large-scale folds, and the anhydrite was able to flow into and fill the fold cores. In the southern part of the nappe, the anhydrite layer is thinner and the deformation style is characterized by imbricate thrusting. Each thrust fault is parallel to the strata and followed the weak anhydrite layer.

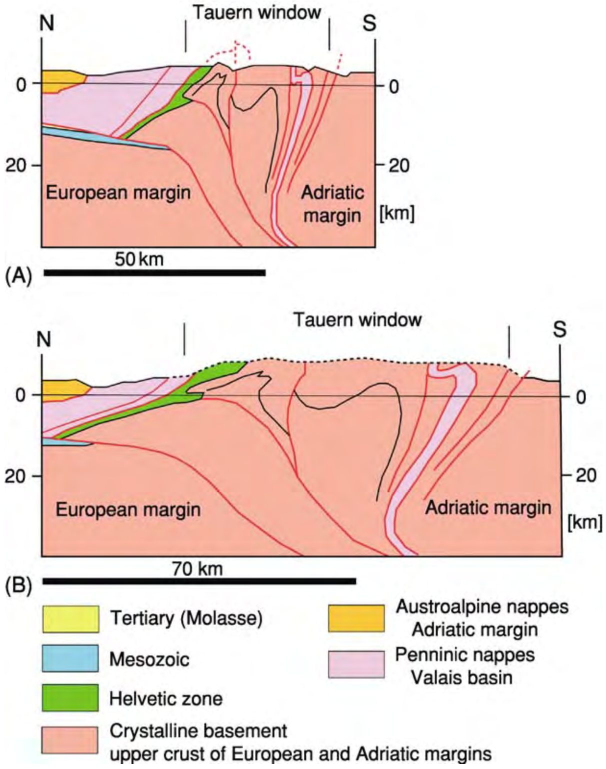


Figure 8 Profile across the Tauern window (Eastern Alps). (A) Present day geometry; (B) retrodeformed to the configuration that existed prior to homogeneous horizontal shortening and vertical stretching.

The Making of the Alps

Geologists working in the Alps had recognized early on that oceanic sediments occurred within the mountain range and were juxtaposed with rock units typical for continents. The pyramid of the Matterhorn (Figure 10), for example, is composed of crystalline basement rocks that were formed more than 300 Ma and which originated in the former (Adriatic) margin of the continental African Plate. In contrast, the base of the pyramid consists of volcanic and sedimentary rocks that formed in an ocean basin (the Piemont Ocean) 170 to 100 Ma ago. The Piemont Ocean formed in response to divergent motion between the Eurasian and African plates (see Figure 3). The Alpine Orogen evolved in a number of steps associated with relative movements between the Eurasian and African plates. The ocean basins between the two continental plates were closed in the process. The

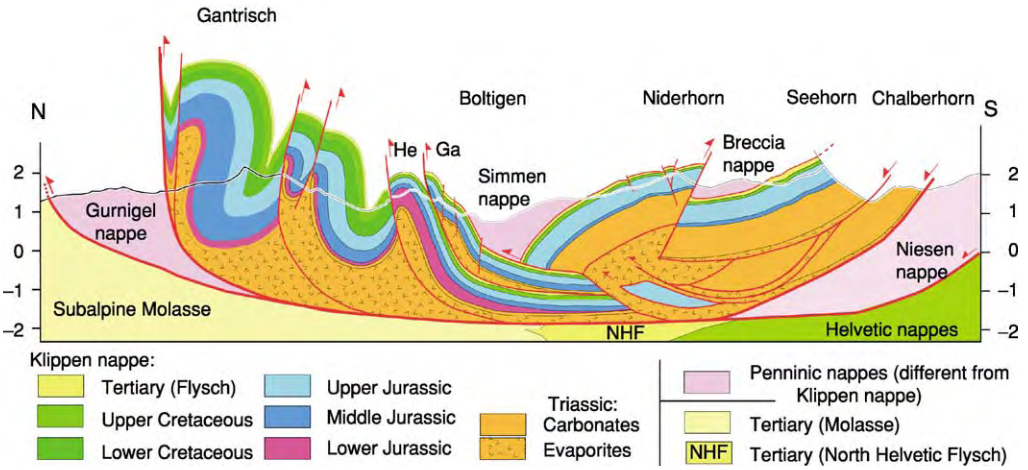


Figure 9 Profile across the Penninic Klippen nappe of the western Swiss Alps. The Klippen nappe consists of sediments of the former Briançonnais swell that have been overthrust onto sediments scraped off of the Valais basin and the Piemont ocean (the Niesen and Gurnigel nappes, respectively). The nappe internal structure of the frontal north west part of the Klippen nappe is dominated by folding, whereas in the internal south east part, imbrications stemming from thrust faulting prevail. Ga: Gastlosen thrust, He: Heiti thrust.

Figure 10 Crystalline basement rocks pertaining to the margin of the African continent build up the Matterhorn peak and overlie the younger volcanic and sedimentary rocks that formed in the Piemont Ocean.

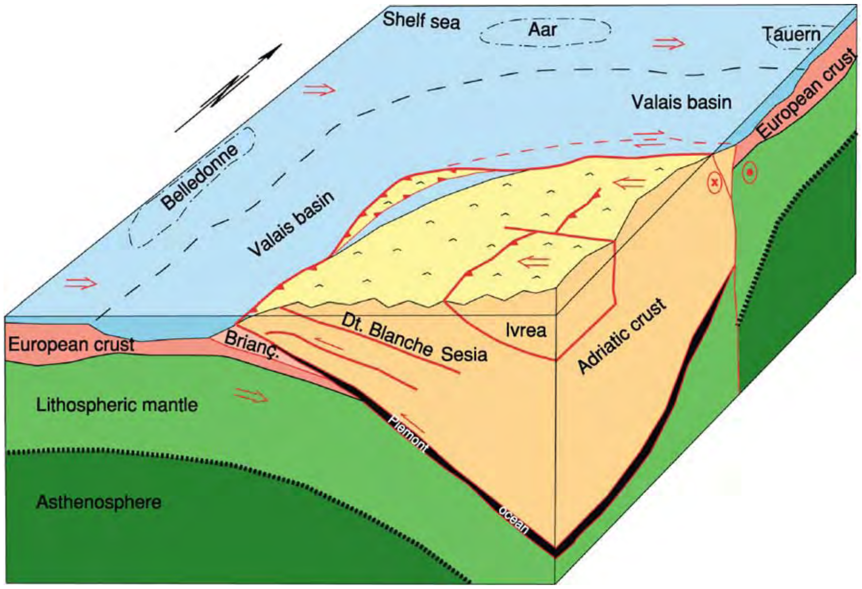


Figure 11 Block diagram showing the three dimensional geometry of the ancestral Alps at 90 Ma. An east dipping subduction zone in the Western Alps had consumed the Piemont Ocean. The Briançonnais continental fragment was entering this subduction zone. The Valais basin and the shelf seas of the European margin were the site of ongoing sedimentation.

first basin, the Piemont Ocean, closed in Cretaceous times (~100 Ma). The second basin, the Valais, closed in Tertiary times (~35 Ma). Closure of these basins resulted not only from head-on collision, but also involved strike–slip movements between the European and Adriatic margins.

During Cretaceous times, convergence between the Eurasian and African plates was directed east and west. The European margin (Figure 11) was approaching the Adriatic margin, which had already formed an ancestral mountain range. The Piemont Ocean had already been subducted along an east-dipping subduction zone. Small fragments of this ocean were scraped off of the descending plate and were attached to the upper plate, a process called ‘underplating’. The Briançonnais microcontinent

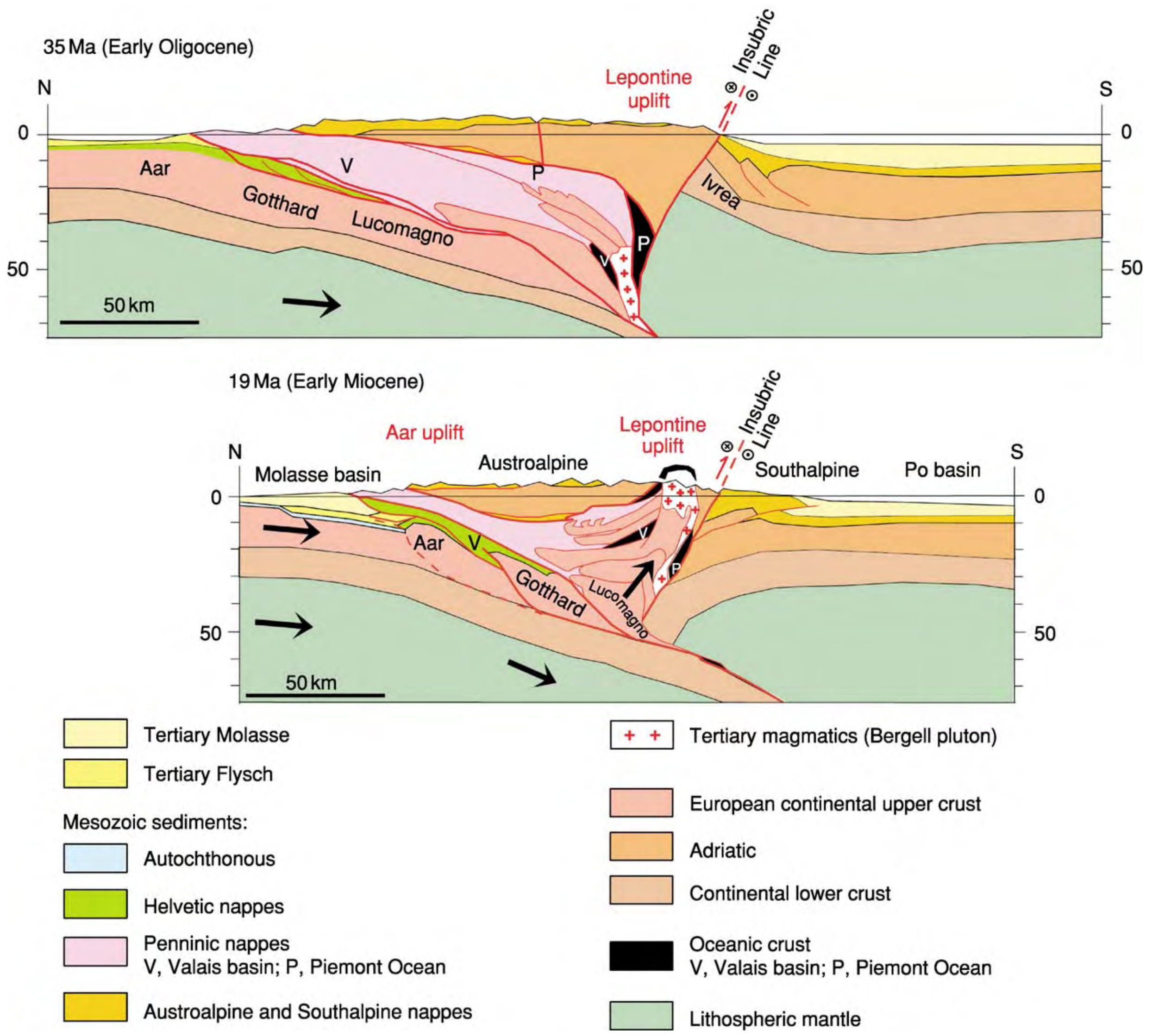


Figure 12 North south profile through the Central Alps of eastern Switzerland, reconstructed to the geometry at 32 and 19 Ma. Comparison of the two profiles reveals that the orogen grew outward with time on both sides, and that the units in the central part of the orogen were raised to a higher level by the combined action of folding and erosional denudation.

was just in the process of being subducted, but parts of it were also attached to the upper plate. The Valais basin was still the site of sedimentation, as was the shelf of the European margin. In the region of the future Central and Eastern Alps, the east–west convergence was expressed as east–west dextral strike–slip movements. At about 40 Ma, the convergence between the Eurasian and African plates changed to a north–south orientation. As a consequence, a south-dipping subduction zone evolved, into which the Valais basin gradually disappeared. Again, a number of fragments were scraped off of the descending plate and were accreted to the upper plate. About 35 Ma, the two margins started to collide. During this north–south convergence, strike–slip movements took place in the ancestral Western Alps. In the Central and Eastern Alps, the collision phase compressed the two margins and led to the stacking of crustal pieces, horizontal shortening, and vertical stretching. Figure 12 shows two stages of this collision phase in a cross-section trough the Central Alps, reconstructed for 35 and 19 Ma. The deformation of the two continental margins pushed crustal fragments up inclined thrust faults and uplifted parts of the orogen by large-scale folding and vertical stretching. As a consequence, the land surface of the ancestral Alps was uplifted. The ensuing high elevations caused precipitation and triggered enhanced erosion. Rivers built large fan deltas in the foreland of the Alps. As far as known, denudation kept pace with uplift during mountain building. Nevertheless, deep crustal fragments were exposed in the process, bringing to the surface samples of rock that had been at depths of several tens of kilometres during the early stages of the formation of the Alps.

See Also

Europe: Mediterranean Tectonics; Variscan Orogeny; Permian to Recent Evolution. **Moho Discontinuity**.

Further Reading

Debelmas J (1974) *Géologie de la France*. Paris: Doin.

Debelmas J (1979) *Découverte Géologique des Alpes du Nord*. Orleans: BRGM.

Debelmas J (1982) *Découverte Géologique des Alpes du Sud*. Orleans: BRGM.

Krenmayr HG (ed.) (2000) *Rocky Austria: A Brief Earth History of Austria*. Wien: Geological Survey of Austria.

Labhart TP (2001) *Geologie der Schweiz,* 5th edn. Thun: Ott Verlag.

Labhart TP and Decrouez D (1997) *Géologie de la Suisse*. Lausanne: Delachaux et Niestlé.

Lemoine M, deGraciansky P C, and Tricart P (2000) *De l'Océan à la Chaîne de Montagnes: Tectonique des Plaques dans les Alpes*. New York: Gordon & Breach.

Marthaler M (2001) *Le Cervin est il Africain?* Lausanne: Loisir et Pédagogie.

Marthaler M (2002) *Das Matterhorn aus Afrika: Die Entstehung der Alpen in der Erdgeschichte*. Thun: Ott Verlag.

Neubauer F and Höck V (eds.) (2000) *Aspects of Geology in Austria. Reports of the Austrian Geological Society, Special Issue 92(1999)*. Wien: Austrian Geological Society.

Nicolas A, Polino R, Hirn A, Nicolich R, and ECORS CROP Working Group (1990) ECORS CROP traverse and deep structure of the western Alps: a synthesis. In: Roure F, Heitzmann P, and Polino R (eds.) *Deep Structure of the Alps,* vol. 156, *Mémoires de la Societe Géologique de France,* pp. 15 28. Paris: Geological Society of France.

Pfiffner OA, Lehner P, Heitzmann P, Mueller St, and Steck A (eds.) (1997) *Deep Structure of the Swiss Alps. Results of NRP 20*. Basel: Birkhäuser.

Roure F, Bergerat F, Damotte B, Mugnier J L, and Polino R (1996) The ECORS CROP Alpine seismic traverse. *Mémoires de la Societe Géologique de France* 170.

TRANSALP Working Group (2002) First deep seismic reflection images of the Eastern Alps reveal giant crustal wedges. *Geophysical Research Letters* 29(10): 92 1 92 4.

Mediterranean Tectonics

E Carminati and C Doglioni, Università La Sapienza, Rome, Italy

Introduction

It is commonly accepted that Mediterranean geology has been shaped by the interplay between two plates, the African and European plates, and possibly also smaller intervening microplates. The Mediterranean was mainly affected by rifting after the Variscan Orogeny (*see* **Europe:** Variscan Orogeny): during the Mesozoic, oceanic Tethys areas and passive continental margins developed, where widespread carbonate platforms were formed. During the Late Mesozoic, the Mediterranean area was dominated by subduction zones (from east to west, the Cimmerian, Dinarides, and Alps–Betics), which inverted the extensional regime, consuming the previously formed Tethyan oceanic lithosphere and the adjacent continental margins. The composition (oceanic or continental), density, and thickness of the lithosphere inherited from the Mesozoic rift controlled the location, distribution, and evolution of the later subduction zones. The shorter wavelength of the Mediterranean orogens relative to other belts (for example, the Cordillera and the Himalayas) is due to the smaller wavelength of the lithospheric anisotropies inherited from the Tethyan rift.

The Mediterranean basin was, and still is, a collector of sediments derived from the erosion of the surrounding continents and orogens: the best examples are the Nile and Rhone deltas. In the past, other deltas deposited sediments in the bottom of the Mediterranean, and their rivers were later disconnected or abandoned: an example is the Upper Oligocene–Lower Miocene Numidian Sandstone, which was derived from Africa, deposited in the central Mediterranean basin, and partly uplifted by the Apennines accretionary prism. It is well known that, during the Messinian eustatic lowstand, the Mediterranean dried up several times, generating a salinity crisis during which thick sequences of evaporites were deposited in the basin. This generated a pulsating loading oscillation in the Mediterranean, because the repetitive removal of the water led to significant isostatic rebound across most of the basin, particularly where it was deeper, as in the Ionian, the Provençal, and the central Tyrrhenian seas.

The direction of the relative motion between Africa and Europe since the Neogene is still under debate.

Most reconstructions show directions of relative motion between north-west and north-east. Recent space geodesy data confirm this overall trend, in which Africa has a north–south component of convergence relative to Europe of about 5 mm year^{-1}, but they also show that the absolute plate-motion directions of both Europe and Africa are north-east and not north or north-west as is usually assumed (see the NASA database on present global plate motions, http://sideshow.jpl.nasa.gov:80/mbh/series.html).

The main Cenozoic subduction zones in the Mediterranean are the Alps–Betics, the Apennines–Maghrebides, and the Dinarides–Hellenides–Taurides. Closely related to the Mediterranean geodynamics are the Carpathian subduction and the Pyrenees (**Figure 1**). The Mediterranean orogens show two distinct signatures, which are similar to those occurring on opposite sides of the Pacific Ocean. High morphological and structural elevations, double vergence, thick crust, involvement of deep crustal rocks, and shallow foredeeps characterize eastwards- or north-eastwards-directed subduction zones (Alps–Betics and Dinarides–Hellenides–Taurides). Conversely, low morphological and structural elevations, single vergence, thin crust, involvement of shallow rocks, deep foredeeps, and a widely developed back-arc basin characterize the westwards-directed subduction zones of the Apennines and Carpathians. This asymmetry can be ascribed to the 'westward' drift of the lithosphere relative to the mantle, at rates of about 49 mm year^{-1} as computed from the hotspots reference frame. All Mediterranean orogens show typical thrust-belt geometries with imbricate-fan and antiformal-stack associations of thrusts. The main factor that varies between orogens and within single belts is the depth of the basal décollement. The deeper it is, the higher is the structural and morphological elevation of the related orogen.

Extensional basins are superimposed on these orogenic belts: on the western side are the Valencia, Provençal, Alboran, Algerian, and Tyrrhenian basins, on the eastern side is the Aegean Basin, and to the north is the Pannonian Basin (**Figures 2** and **3**).

The Mediterranean can be divided into western, central, and eastern basins. The western Mediterranean is younger (mainly less than 30 Ma) than the central Mediterranean and eastern Mediterranean, which are mainly relics of the Mesozoic to possibly Cenozoic Tethys Ocean.

Positive gravity anomalies occur in the deep basins (the Provençal, Tyrrhenian, and Ionian seas), where the mantle has been uplifted by rifting processes. In contrast, negative gravity anomalies occur along the subduction zones.

Western Mediterranean

A characteristic feature of the western Mediterranean is the large variation in lithospheric and crustal thickness (**Figure 5**). The lithosphere has been thinned to less than 60 km in the basins (50–60 km in the Valencia trough, 40 km in the eastern Alboran Sea, and 20–25 km in the Tyrrhenian Sea), while it is 65–80 km thick below the continental swells (Corsica–Sardinia and the Balearic promontory). The crust mimics these differences, with a thickness of 8–15 km in the basins (Valencia trough, Alboran Sea, Ligurian Sea, and Tyrrhenian Sea) and 20–30 km underneath the swells (Balearic promontory and Corsica–Sardinia), as inferred by seismic and gravity data. These lateral variations in thickness and composition are related to the rifting process that affected the western Mediterranean, which is a coherent system of interrelated irregular troughs, mainly V-shaped, that began to develop in the Late Oligocene–Early Miocene in the westernmost parts (Alboran, Valencia, Provençal basins), becoming progressively younger eastwards (eastern Balearic and Algerian basins), culminating in the presently active east–west extension in the Tyrrhenian Sea (**Figures 1, 2, 3,** and **4**). Heat flow data and thermal modelling show that the maximum heat flows are encountered in the basins: 120 mW m^{-2} in the eastern Alboran Sea, 90–100 mW m^{-2} in the Valencia trough, and more than 200 mW m^{-2} in the Tyrrhenian Sea. All these sub-basins appear to be genetically linked to the back-arc opening related to the coeval 'eastwards' rollback of the westward-directed Apennines–Maghrebides subduction zone. Extreme stretching generated oceanic crust in the Provençal (20–15 Ma), Algerian (17–10 Ma), Vavilov and Marsili (7–0 Ma) basins. Between 25 Ma and 10 Ma, the Corsica–Sardinia block rotated 60° counterclockwise (**Figures 1, 2, 3,** and **5**).

In the southern Apennines, the choking of the subduction zone with the thicker continental lithosphere of the Apulia Platform slowed the eastwards migration of the subduction hinge (**Figure 6**), whereas in the central and northern Apennines and in Calabria subduction is still active owing to the presence in the foreland of the thin continental lithosphere of the Adriatic Sea and the Mesozoic oceanic lithosphere of the Ionian Sea, allowing rollback of the subduction hinge.

The western Mediterranean basins tend to close both morphologically and structurally towards the south-west (Alboran Sea) and north-east (Ligurian Sea; **Figures 1** and **6**). The eastwards migration of the arc associated with the westwards-directed subduction generated right-lateral transpression along the entire east–west-trending northern African belt

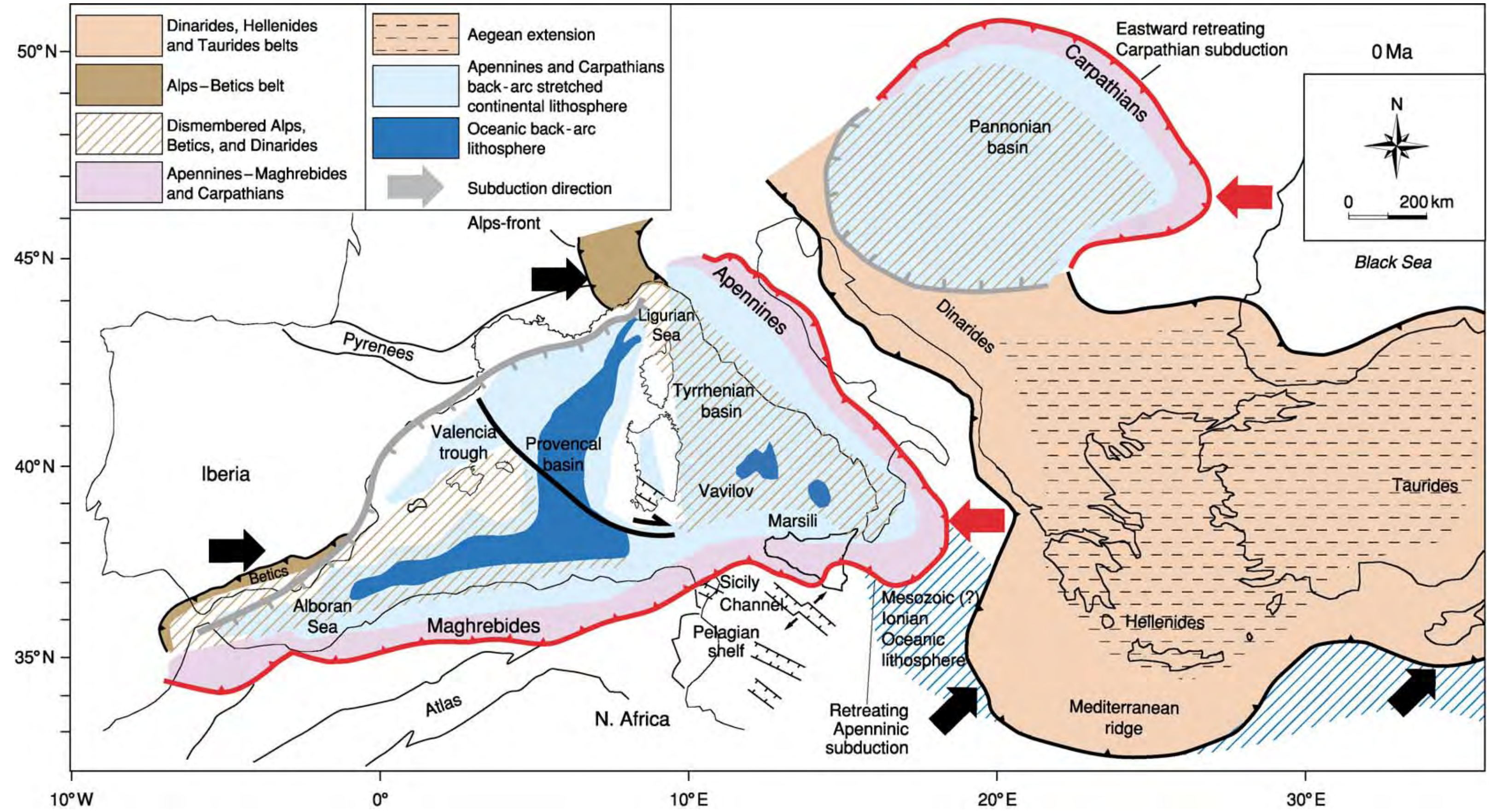


Figure 1 Present geodynamic framework. There are four subduction zones with variable active rates in the Mediterranean realm: the westwards directed Apennines–Maghrebides; the westwards directed Carpathians; the north eastwards directed Dinarides–Hellenides–Taurides; and the south eastwards directed Alps. The Apennines–Maghrebides subduction related back arc basin of the western Mediterranean stretched and scattered into segmented basins most of the products of the Alps–Betics orogen.

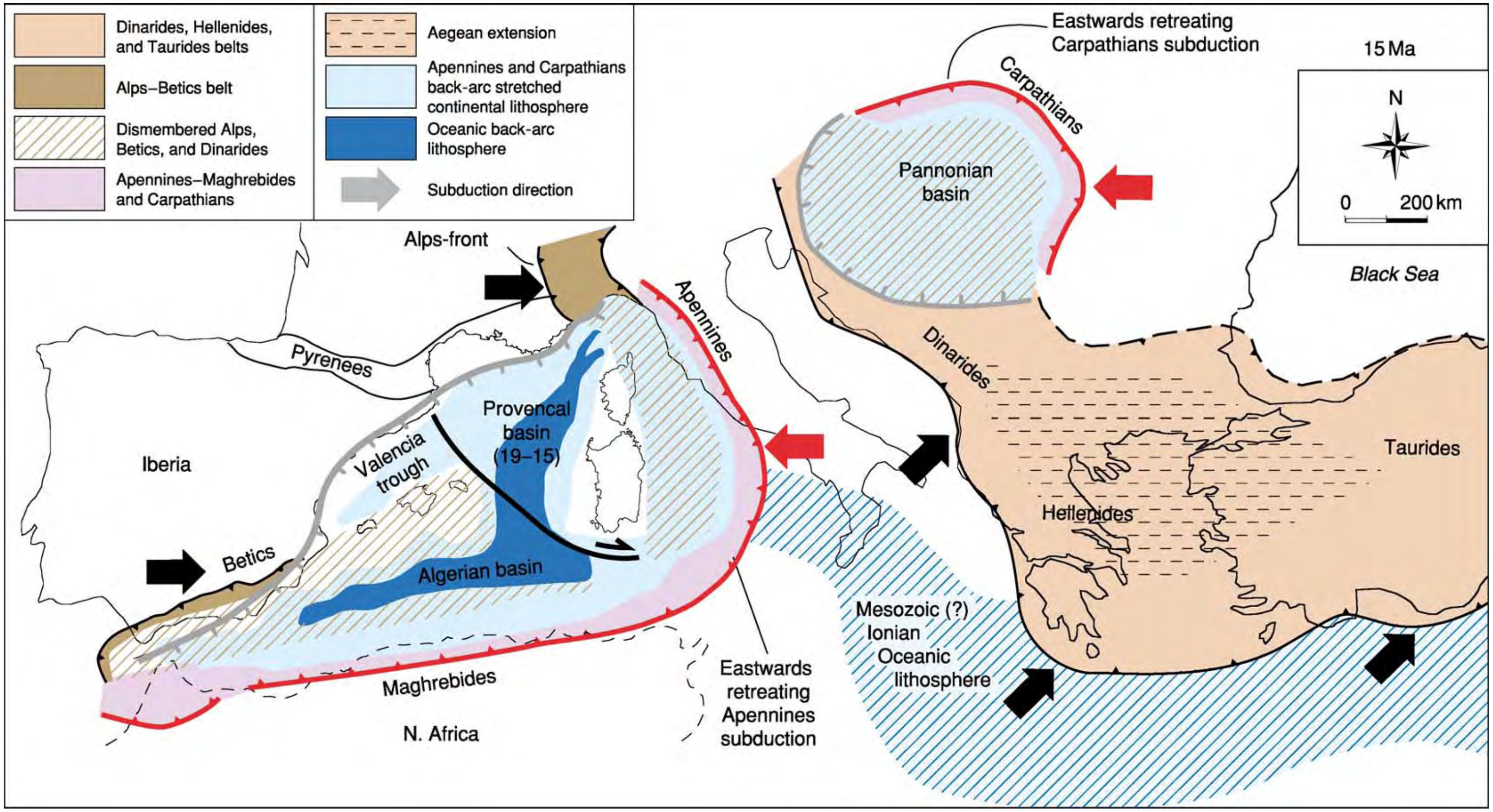


Figure 2 Palaeogeodynamics at about 15 Ma. Note the 'eastward' vergence of both the Apennines Maghrebides trench and the back arc extensional wave. The Liguro Provençal basin, the Valencia trough, and the North Algerian basin were almost completely opened at 10 Ma. The Dinarides subduction slowed down, owing to the presence of the thick Adriatic continental lithosphere to the west, whereas to the south the Hellenic subduction was very lively owing to the presence in the footwall plate of the Ionian oceanic lithosphere. The Carpathians migrated eastwards, generating the Pannonian back arc basin, with kinematics similar to those of the Apennines. Provencal basin (19 15) Age of the oceanic crust.

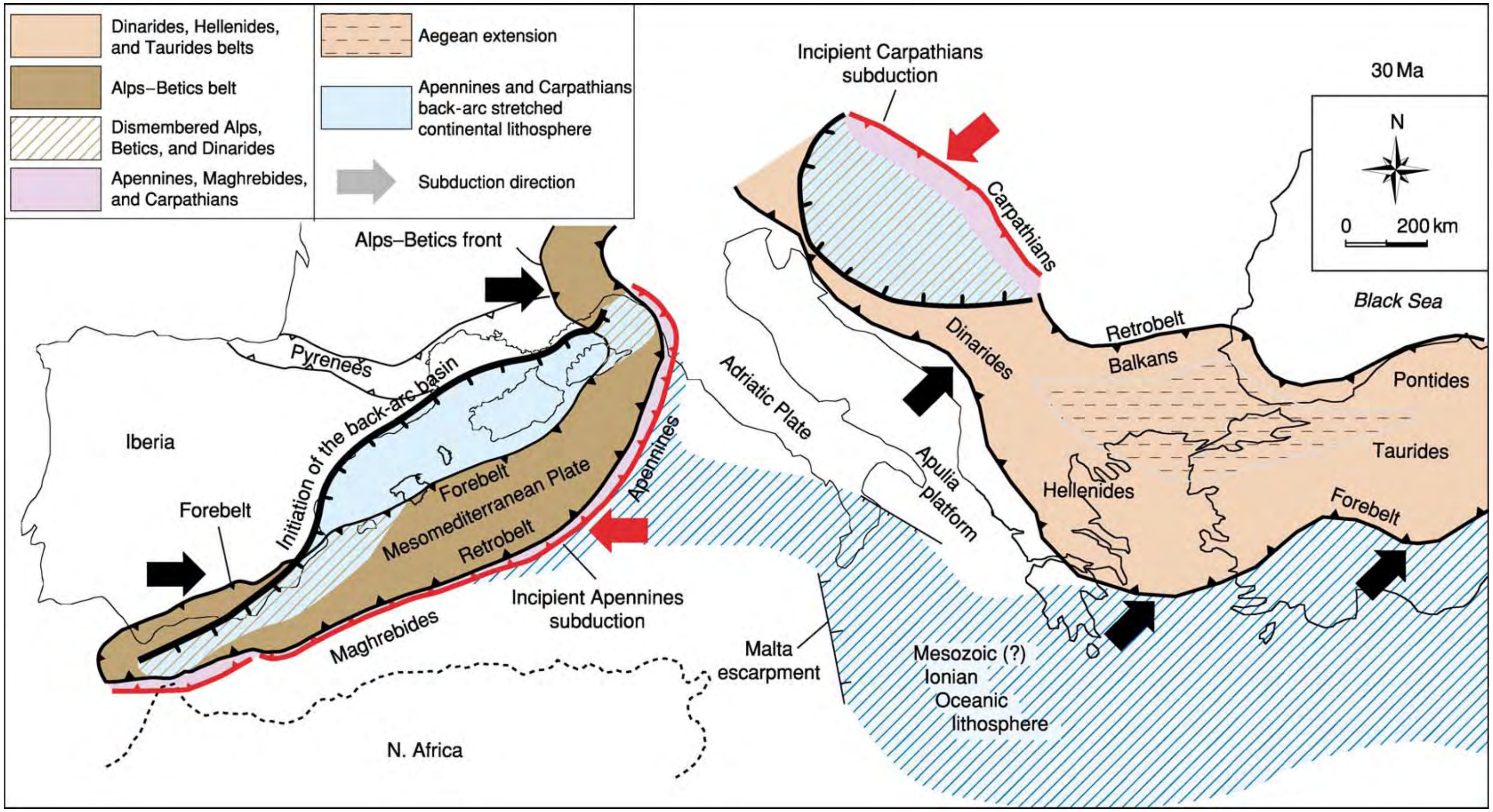


Figure 3 Palaeogeodynamics at about 30 Ma. The locations of the subduction zones were controlled by the Mesozoic palaeogeography. The Alps Betics formed along the south eastwards dipping subduction of Europe and Iberia underneath the Adriatic and Mesomediterranean plates. The Apennines developed along the Alps Betics retrobelt to the east, in which oceanic or thinned pre existing continental lithosphere was present. Similarly, the Carpathians started to develop along the Dinarides retrobelt (i.e. the Balkans). The fronts of the Alps Betics orogen were cross cut by the Apennines related subduction back arc extension.

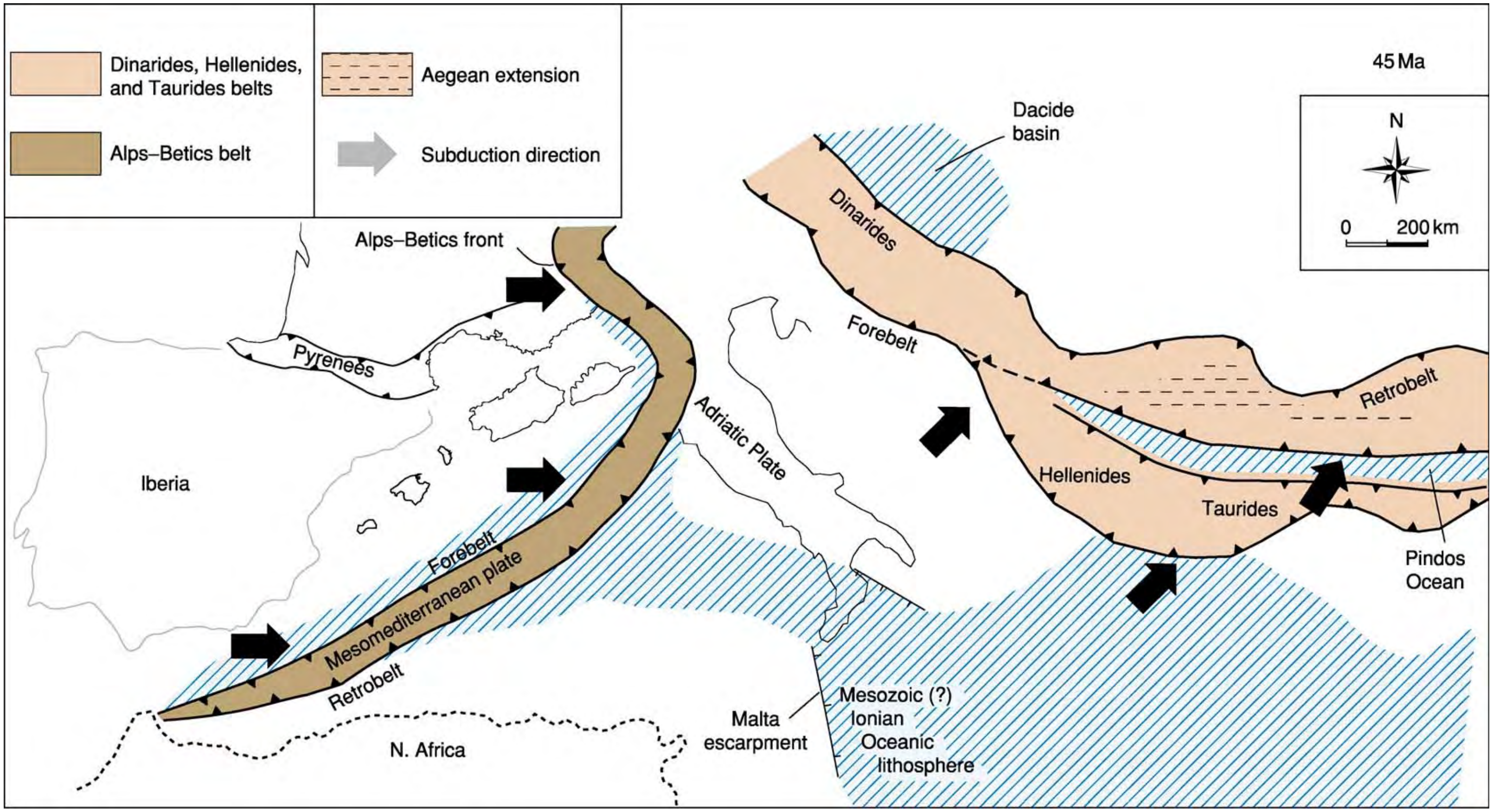


Figure 4 Palaeogeodynamics at about 45 Ma. The Alps were continuous with the Betics to Gibraltar, consuming an ocean located to the west.

(Maghrebides) and its Sicilian continuation, whereas left-lateral transtension occurs along the same trend in the back-arc setting just to the north of the African margin. The opposite tectonic setting is found in the northern margin of the arc.

Subduction retreat generated calc-alkaline and shoshonitic magmatic episodes – particularly in the western margins of the lithospheric boudins – which were followed by alkaline-tholeiitic magmatism in the back-arc to the west.

Extension partly originated in areas previously occupied by the Alps–Betics Orogen, which formed in the Cretaceous due to the 'eastwards'-directed subduction of Europe and Iberia underneath the Adriatic Plate and a hypothetical Mesomediterranean Plate (Figure 4). If Sardinia is restored to its position prior to rotation, it can be seen that during the Early Cenozoic the Alps were probably joined with the Betics in a double-vergent single belt. The western Alps, which are the forebelt of the Alps, were connected to the

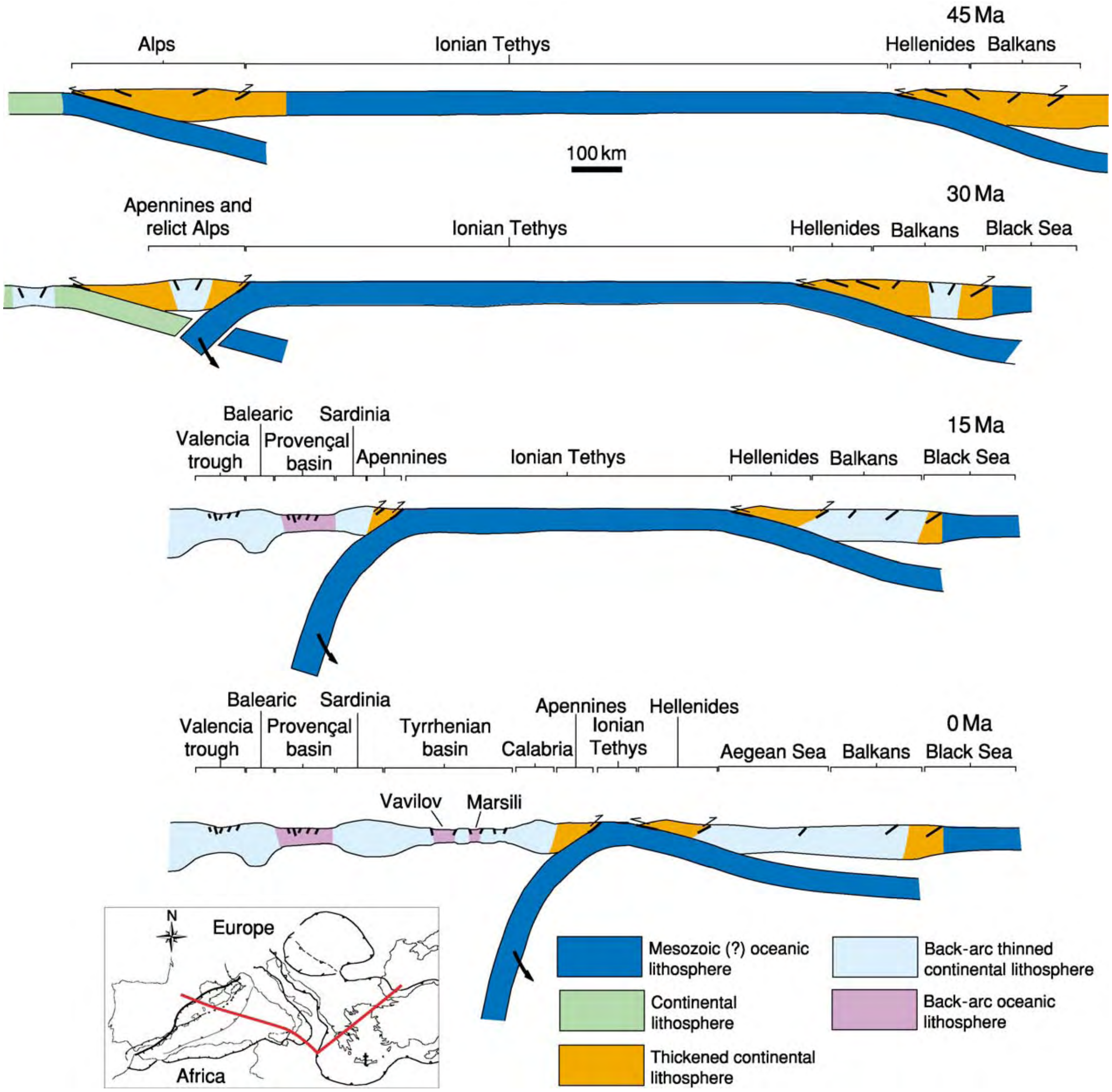


Figure 5 During the last 45 Ma, the evolution of the Mediterranean along the trace shown on the map (inset) is the result of three main subduction zones: the early eastwards directed Alpine subduction; the Apennines subduction switch along the Alps retrobelt; and the Dinarides Hellenides subduction. The last two slabs retreated at the expense of the inherited Tethyan Mesozoic oceanic or thinned continental lithosphere. In their hanging walls, a few rifts formed as back arc basins, which are progressively younger towards the subduction hinges. The slab is steeper underneath the Apennines, possibly owing to the westwards drift of the lithosphere relative to the mantle.

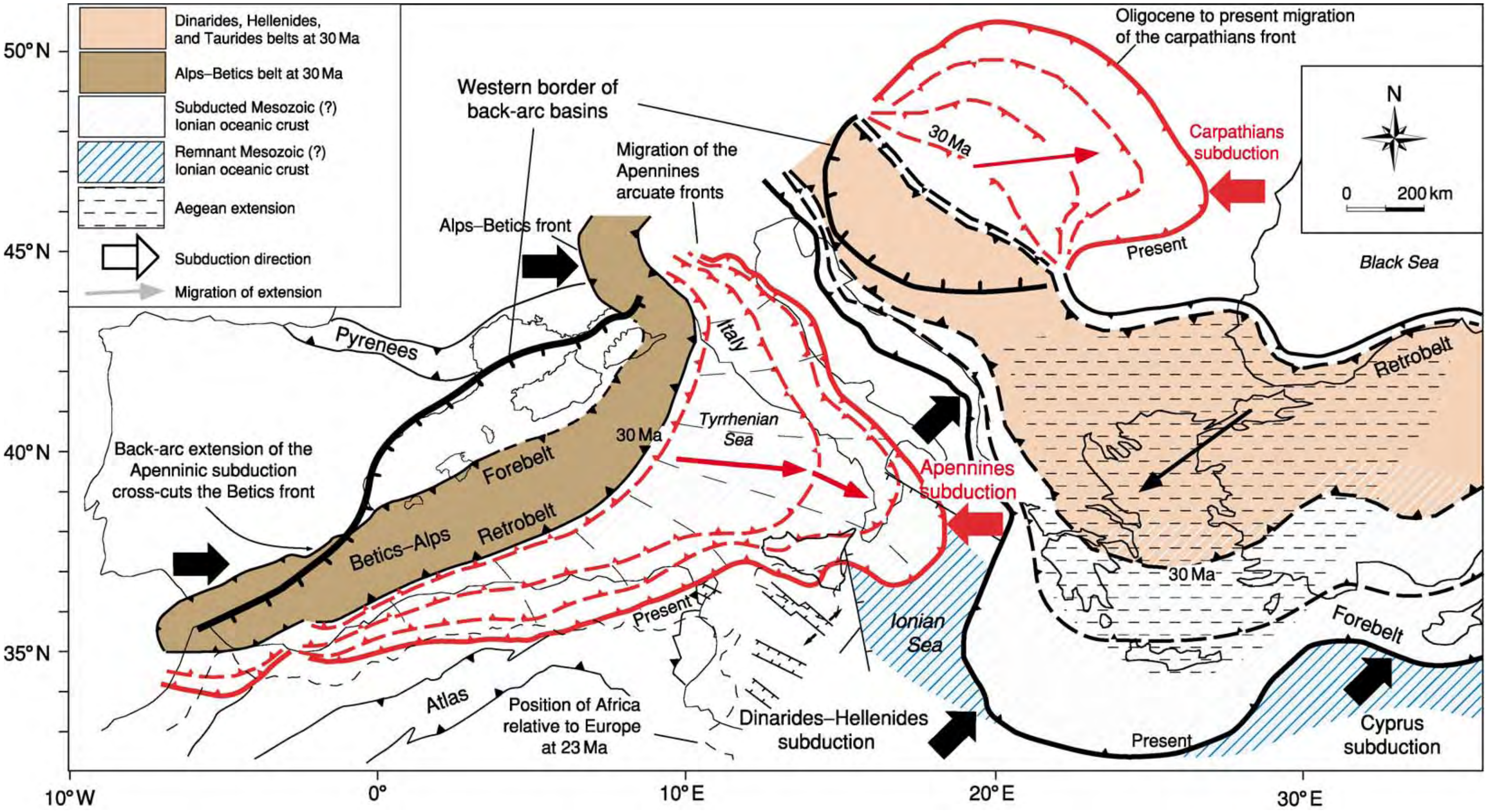


Figure 6 Main tectonic features of the Mediterranean realm, which has been shaped during the last 45 Ma by a number of subduction zones and related belts: the double vergent Alps Betics; the single eastwards vergent Apennines Maghrebides and the related western Mediterranean back arc basin; the double vergent Dinarides Hellenides Taurides and related Aegean extension; the single eastwards vergent Carpathians and the related Pannonian back arc basin; and the double vergent Pyrenees.

Alpine Corsica; the Alps continued south-westwards into the Balearic promontory and the Betics. The retrobelt of the Alps, the southern Alps, also continued from northern Italy towards the south-west. In a double-vergent orogen, the forebelt is the frontal part, which is synthetic to the subduction and verges towards the subducting plate; the retrobelt is the internal part, which is antithetic to the subduction and verges towards the interior of the overriding plate.

The westwards-directed Apennines–Maghrebides subduction started along the Alps–Betics retrobelt (Figures 3 and 5), where oceanic and thinned continental lithosphere occurred in the foreland to the east. Subduction underneath the Apennines–Maghrebides consumed inherited Tethyan domains (Figures 5 and 6). The subduction zone and the related arc migrated 'eastwards' at a speed of 25–30 mm year ^{1}.

The western Late Oligocene–Early Miocene basins of the Mediterranean nucleated both within the Betics orogen (e.g. the Alboran Sea) and in its foreland (e.g. the Valencia and Provençal troughs; Figure 3). At that time the direction of the grabens (40°–70°) was oblique to the trend of the coexisting Betics orogen (60°–80°), indicating its structural independence from the Betics Orogeny. Thus, as the extension cross-cut the orogen and also developed well outside the thrust-belt front, the westernmost basins of the Mediterranean developed independently of the Alps–Betics orogen, being related instead to the innermost early phases of back-arc extension in the hanging wall of the Apennines–Maghrebides subduction zone. In contrast to the 'eastwards'-migrating extensional basins and following the 'eastwards' retreat of the Apennines subduction zone, the Betics–Balearic thrust front was migrating 'westwards', producing interference or inversion structures.

The part of the Alps–Betics orogen that was located in the area of the Apennines–Maghrebides back-arc basin (Figure 1) has been disarticulated and spread out into the western Mediterranean (forming the metamorphic slices of Kabylie in northern Algeria and Calabria in southern Italy). Alpine type basement rocks have been dragged up in the Tyrrhenian Sea.

Similarly, boudinage of the pre-existing Alps and Dinarides orogens occurred in the Pannonian Basin, which is the Oligocene to Recent back-arc basin related to the eastwards-retreating westwards-directed Carpathian subduction zone (Figures 1, 3, and 6). In the Pannonian basin, the extension isolated boudins of continental lithosphere that had been thickened by the earlier Dinarides orogen, such as the Apuseni Mountains, which separate the Pannonian basin from the Transylvanian basin to the east. The western Mediterranean back-arc setting is comparable with Atlantic and western Pacific back-arc basins that show similar large-scale lithospheric boudinage, in which parts of earlier orogens have been scattered in the back-arc area, like the Central America Cordillera relicts that are dispersed in the Caribbean domain.

The Apennines accretionary prism formed in sequence at the front of the pre-existing Alpine retrobelt, and, therefore, the central western Apennines also contain the inherited Alpine orogen of Cretaceous to Miocene age. There was probably a temporary coexistence of opposite subductions during the Late Oligocene to Early Miocene (Figure 5). Structural and geophysical data support the presence of an eastwards-migrating asthenospheric wedge at the subduction hinge of the retreating Adriatic Plate. The subduction flip, from the Alpine eastwards-directed subduction to the Apennines westwards-directed subduction, could be reflected in the drastic increase in subsidence rates in the Apennines foredeep during the Late Oligocene to Early Miocene. Westwards-directed subduction zones, such as the Apennines, show foredeep subsidence rates that are up to 10 times higher (more than 1 mm year ^{1}) than those of the Alpine foredeeps. The subduction flip (Figure 5) could also be reflected in the larger involvement of the crust during the earlier Alpine stages than in the Apennines décollements, which mainly deformed the sedimentary cover and the phyllitic basement. It has been demonstrated that the load of the Apennine and Carpathian orogens is not sufficient to generate the 4–8 km deep Pliocene–Pleistocene foredeep basins, and a mantle origin has been proposed for the mechanism (slab pull and/or eastwards mantle flow).

Paradoxically, the extension that determined most of the western Mediterranean developed in the context of relative convergence between Africa and Europe. However, it appears that the north–south relative motion between Africa and Europe at the longitude of Tunisia has been about 135 km in the last 23 Ma, more than five times slower than the migration of the Apennines arc, which has moved more than 700 km eastwards during the last 23 Ma (Figures 1 and 6). Therefore, the eastwards migration of the Apennines–Maghrebides arc is not a consequence of the north–south relative convergence between Africa and Europe but is instead a consequence of the Apennines–Maghrebides subduction rollback, which was generated either by slab pull or by the 'eastwards' flow of the mantle relative to the lithosphere deduced from the hotspot reference frame.

The western Mediterranean developed mainly after the terminal convergence in the Pyrenees at about 20 Ma, which resulted from the Late Cretaceous to Early Tertiary counterclockwise rotation of Iberia, which was contemporaneous with the opening of the Biscay Basin.

In northern Africa, south of the Maghrebides (and the related Algerian Tell and Moroccan Riff), the Atlas Mountains represent an intraplate inversion structure, in which extensional (north-north-east-trending) and left-lateral (about east–west-trending) transtensional Mesozoic intercontinental rifts were later buckled and squeezed by Cenozoic compression and right-lateral transpression in the foreland of the Apennines–Maghrebides subduction zone. This is also indicated by the Mesozoic sequences in the Atlas ranges, which are thicker than the adjacent undeformed *mesetas*.

Central Mediterranean

The Malta escarpment (Figures 3 and 4), along the eastern coast of Sicily, is a physiographic feature that has been tectonically controlled since Triassic times. Rocks dredged from the Malta escarpment range from Mesozoic to Tertiary in age. The escarpment represents a Mesozoic continental margin that has been reactivated as a transtensional feature since the Pliocene. In spite of the Apennines and Hellenides Neogene subduction zones, two conjugate passive continental margins are preserved at the margins of the Ionian Sea, along the Malta escarpment to the south-west and the Apulian escarpment to the north-east. Based on the low heat flows (18–40 mW m ^{2}) and the 4–8 km of sedimentary cover, the Ionian Sea is probably a remnant of the Mesozoic Tethys Ocean, confined by the two conjugate passive continental margins. The transition from continental crust to oceanic crust appears to be sharper to the north-east than to the south-west. The basin between south-east Sicily and south-west Puglia was about 330 km wide. The inferred oceanic ridge could have been flattened by thermal cooling and buried by later sediments.

Stratigraphic and structural constraints to the north in the Apennines belt suggest that the Ionian Ocean continued to the north-west (Figure 5). This palaeogeography is supported by the seismicity of the Apennines slab underneath the southern Tyrrhenian Sea, which implies subducted oceanic lithosphere. The adjacent absence or paucity of deep seismicity does not imply the absence of subduction but can be interpreted as a reflection of the more ductile behaviour of the subducted continental lithosphere.

The Sicily Channel and the Pelagian shelf off the coast of eastern Tunisia have been undergoing extension since at least Pliocene times; in other words Africa is moving south-westwards in relation to Sicily (Figure 1). This process is responsible for the two grabens of Pantelleria and Malta deepening the seafloor and for the generation of active alkaline magmatism (e.g. the ephemeral Ferdinandea Island). The lithospheric extension was active whilst the Apennines–Maghrebides accretionary prism advanced, generating an interplay of two tectonic settings working together, with thrusts advancing over an orthogonal extending area, generating both thrusts cutting normal faults and normal faults offsetting thrusts. The rifting of the Sicily Channel seems to be physically connected north-westwards to the rift in south-western Sardinia (Campidano graben) and south-eastwards to the Sirte Basin, off the coast of Libya. One possibility is that this rift is linked through transfer zones in Egypt to the Red Sea and the East African Rift.

Eastern Mediterranean

The Dinarides, Hellenides, and Taurides are a polyphase orogen, representing the coalescence of at least two or three subduction zones since Mesozoic times (Figures 1, 4, 5, and 6). The orogen has a part synthetic to the north-eastwards-directed subduction, i.e. the forebelt verging south-westwards. The conjugate part of the orogen is the retrobelt, which verges north-eastwards and northwards (Balkans and Pontides). The existence of three subduction zones is supported by the occurrence of two distinct oceanic sutures, preserved as the ophiolitic suites of Vardar and the Sub-Pelagonian units, which represent two separated branches of the Mesozoic Tethyan Ocean and the present oceanic subduction of the Ionian Sea. It is commonly believed that the more internal (Vardar) suture zone is the older one.

The polyphase orogen exhibits a similar architecture to the Alps, but duplicated. The Rhodope–Serbo-Macedonian and Sakarya (northern Turkey) massifs mimic the internal massifs of the Alps, which represent the continental margin of the hanging-wall plate. On the other side, to the south-west of the Vardar oceanic suture, the Pelagonian (Macedonia–Greece) and Menderes (northern Turkey) massifs correspond to the external massifs of the Alps, representing the continental lithosphere of the footwall plate. The Pelagonian basement is at the same time the hanging-wall plate for the more external north-eastwards-directed subduction of the Sub-Pelagonian and Pindos Ocean, which was eventually closed by collision with the eastern margin of the Adriatic Plate.

However, unlike the Alps, widespread extension developed in the Dinarides–Hellenides–Taurides orogen (Figures 1 and 6). This extension resulted in the low topography of the orogen in comparison with belts such as the Alps and the Zagros or the Himalayas. In the Balkans, the Rhodope, and the Serbo-Macedonian massifs, structural and stratigraphic data indicate an interplay of compressional and extensional tectonics. A Cretaceous to Eocene compressive deformation was followed by the generation of Eocene grabens. A later (possibly Miocene) compression inverted and uplifted these grabens, but it was followed by extensional tectonics that have affected the Balkan peninsula since Pliocene times, determining the north-west-trending normal faults and the related east–west right-lateral and north–south left-lateral transtensive transfer faults. North-eastwards-directed subduction is continuing along the eastern side of the Adriatic, in the Ionian Sea underneath the Mediterranean Ridge (the accretionary prism), and on the northern side of the Levantine Sea, i.e. in the eastern Mediterranean beneath Cyprus (**Figure 1**). The convergence rates are faster underneath the Mediterranean ridge (up to 40–50 mm year ^{1}), decrease eastwards along the Cyprus segment, and have minimum values along the Adriatic coast. The convergence rate appears to be controlled by the composition of the foreland lithosphere: where it is oceanic and dense, such as in the Ionian Sea, the subduction is faster than in the Adriatic and Cyprus segments, where the downgoing lithosphere is continental and transitional oceanic–continental, respectively. In the orogen, calc-alkaline and shoshonitic magmatism has accompanied most of the subduction since Cretaceous times. The later extensional process in the anomalously called 'back-arc' is possibly responsible for the transition to the alkaline magmatic signature.

One of the best-known ophiolitic sequences in the world crops out in Cyprus: a complete oceanic section is exposed (from harzburgites and peridotites of the upper mantle to gabbros, sheeted dykes, lavas, and pelagic sediments of the crust). The island is an anticline involving the whole crust, and its culmination coincides with the Erathostene seamount in the subducting foreland. The Erathostene seamount is a structural high inherited from the Mesozoic–Cenozoic rift.

Since at least Miocene times, there has been an independent and presently active subduction along the northern margin of the Black Sea, generating the Caucasus.

Geodynamic reconstructions of the eastern Mediterranean explain the extensional tectonics either by westwards Anatolian extrusion or by gravitational collapse of thickened lithosphere. However, these mechanisms can be ruled out because plate-velocity vectors increase from eastern Anatolia to the Aegean and Greece. This contradicts the basic rule that the velocity field decreases away from the source of the energy, i.e. the supposed squeezing of Anatolia by the Arabia indenter, or the collapse of the Anatolian orogen. Moreover, the topographic gradient between Anatolia and the Ionian deep basin is too small (less than 1°) to provide sufficient energy to explain the present deformation. Instead, the simplistic view of the westward Anatolian escape would close the Aegean Sea.

The plates involved in the geodynamic reconstructions of the eastern Mediterranean are Africa, Greece, Anatolia, Eurasia, and Arabia. Deformation is very active in all these areas. The most prominent geodynamic factor shaping the eastern Mediterranean is the north-east-directed subduction of Africa underneath Greece and the Anatolian Plate (Eurasia). Seismic lines across the Cyprus Arc at the southern margin of the Anatolian Plate show clear active compression and deformation of the seafloor.

The Aegean Sea is generally considered to be a back-arc basin resulting from the aforementioned subduction. However, the Aegean Sea is characterized by a relatively thick crust (20–25 km) in spite of long-standing subduction, which has probably been active since at least the Cretaceous. The subduction zone migrated south-westwards to the present position of the Cyprus-Hellenic subduction zone, and the associated orogen was later replaced by extension. In the Aegean Sea, Alpine-type crustal thickening with high pressures and low temperatures was followed by non-coaxial crustal-scale extension. This is consistent with the initial emplacement of thrust-sheets of basement slices, which were later cross-cut by extensional or transtensional faults. In addition, extension and associated magmatism were and are migrating south-south-westwards, and have developed particularly since the Oligocene, while subduction began much earlier. 'Normal' back-arc basins (e.g. the Tyrrhenian Sea) associated with westwards-directed subduction zones opened very fast (10–20 Ma) and are always contemporaneous with the subduction. Moreover, they are characterized by oceanization and eastwards migration of extension and related magmatism, features directly surrounded by a frontal accretionary wedge. In contrast, the accretionary wedge of the Hellenic subduction zone is the south-eastern prolongation of the Dinarides thrust belt, where no back-arc rift comparable to the Tyrrhenian Sea occurs.

The extension in western Turkey, the Aegean Sea, Greece, and Bulgaria appears to be the result of differential convergence rates in the north-eastwards-directed subduction of Africa relative to the hanging wall of disrupted Eurasian lithosphere. Relative to Africa, the faster south-eastwards motion of Greece than of Cyprus–Anatolia results in the Aegean extension. The differences in velocity can be ascribed to differential decoupling with the asthenosphere. In the back-arc basins of the western Pacific the asthenosphere replaces a subducted and retreated slab; however, the Aegean rift represents a different type of extension associated with a subduction zone, in which the hanging-wall plate overrides the slab at different velocities, implying internal deformation.

According to this geodynamic scenario, during the compressive events associated with north-eastwards-directed subduction, basement rocks (both continental and ophiolitic slices) in western Anatolia and the Aegean Sea were uplifted and eroded. Later extension caused subsidence in the area, and the basement slices were partly covered by continental and marine sediments.

During its development, the Aegean extension migrated south-westwards (Figures 5 and 6). The Aegean rift affects the Aegean Sea and all of continental Greece, and it can be followed to the east, where it is widely expressed in Turkey, and to the north-west in Bulgaria, Albania, Macedonia, Serbia, and Bosnia. At the same time, from the Oligocene to the present, to the north, the Pannonian basin developed as the back-arc of the Carpathians subduction, but migrating eastwards, and affecting mainly eastern Austria, Slovenia, Croatia, Hungary, and Romania. Therefore, in the central part of the former Yugoslavia, the Pannonian and Aegean rifts meet with opposite directions of migration.

See Also

Europe: Variscan Orogeny; Permian to Recent Evolution; The Alps; Holocene. **Plate Tectonics**. **Tectonics:** Convergent Plate Boundaries and Accretionary Wedges; Mountain Building and Orogeny.

Further Reading

Berckhemer H and Hsü KJ (eds.) (1982) *Alpine Mediterranean Geodynamics*. Geodynamics Series 7. Washington: American Geophysical Union.

Calcagnile G and Panza GF (1980) The main characteristics of the lithosphere asthenosphere system in Italy and surrounding regions. *Pure and Applied Geophysics* 119: 865 879.

Carminati E, Wortel MJR, Spakman W, and Sabadini R (1998) The role of slab detachment processes in the opening of the western central Mediterranean basins: some geological and geophysical evidence. *Earth and Planetary Science Letters* 160: 651 665.

Catalano R, Doglioni C, and Merlini S (2001) On the Mesozoic Ionian basin. *Geophysical Journal International* 144: 49 64.

Cella F, Fedi M, Florio G, and Rapolla A (1998) Gravity modeling of the litho asthenosphere system in the Central Mediterranean. *Tectonophysics* 287: 117 138.

Christova C and Nikolova SB (1993) The Aegean region: deep structures and seismological properties. *Geophysical Journal International* 115: 635 653.

Dercourt J, Gaetani M, Vrielynck B, *et al.* (2000) *Atlas Peri Tethys, Paleogeographical Maps.* Geological Map of the World. Paris: CGMW.

de Voogd B, Truffert C, Chamot Rooke N, *et al.* (1992) Two ship deep seismic soundings in the basins of the Eastern Mediterranean Sea (Pasiphae cruise). *Geophysical Journal International* 109: 536 552.

Doglioni C, Gueguen E, Harabaglia P, and Mongelli F (1999) On the origin of W directed subduction zones and applications to the western Mediterranean. In: Durand B, Jolivet J, Horváth F, and Séranne M (eds.) *The Mediterranean Basins: Tertiary Extension Within The Alpine Orogen,* pp. 541 561. Special Publication 156. London: Geological Society.

Durand B, Jolivet J, Horváth F, and Séranne M (1999) *The Mediterranean Basins: Tertiary Extension Within the Alpine Orogen.* Special Publication 156. London: Geological Society.

Frizon de Lamotte D, Saint Bezar B, Bracene R, and Mercier E (2000) The two main steps of the Atlas building and geodynamics of the western Mediterranean. *Tectonics* 19: 740 761.

Gueguen E, Doglioni C, and Fernandez M (1998) On the post 25 Ma geodynamic evolution of the western Mediterranean. *Tectonophysics* 298: 259 269.

Guerrera F, Martin Algarra A, and Perrone V (1993) Late Oligocene Miocene syn / late orogenic successions in western and central Mediterranean chain from the Betic cordillera to the southern Apennines. *Terra Nova* 5: 525 544.

Huguen C, Mascle J, Chaumillon E, *et al.* (2001) Deformational styles of the eastern Mediterranean Ridge and surroundings from combined swath mapping and seismic reflection profiling. *Tectonophysics* 343: 21 47.

Kastens K, Mascle J, Auroux C, *et al.* (1988) ODP Leg 107 in the Tyrrhenian Sea: insights into passive margin and back arc basin evolution. *Geological Society of America Bulletin* 100: 1140 1156.

Réhault JP, Mascle J, and Boillot G (1984) Evolution géodynamique de la Méditerranée depuis l'Oligocène. *Memorie Società Geologica Italiana* 27: 85 96.

Robertson AHF and Grasso M (1995) Overview of the Late Tertiary Recent tectonic and palaeo environmental development of the Mediterranean region. *Terra Nova* 7: 114 127.

Stampfli G, Borel G, Cavazza W, Mosar J, and Ziegler PA (2001) *The Paleotectonic Atlas of the PeriTethyan Domain.* CD ROM. European Geophysical Society.

Stanley DJ and Wezel FC (eds.) (1985) *Geological Evolution of the Mediterranean Basin.* New York, USA: Springer Verlag.

Vai GB and Martini P (eds.) (2001) *Anatomy of an Orogen: the Apennines and Adjacent Mediterranean Basins.* Dordrecht: Kluwer Academic Publishers.

Wilson M and Bianchini G (1999) Tertiary Quaternary magmatism within the Mediterranean and surrounding regions. In: Durand B, Jolivet J, Horváth F, and Séranne M (eds.) *The Mediterranean Basins: Tertiary Extension Within The Alpine Orogen,* pp. 141 168. Special Publication 156. London: Geological Society.

Holocene

W Lemke and J Harff, Baltic Sea Research Institute Warnemünde, Rostock, Germany

Introduction

The Quaternary period comprises the shortest time interval of all geological systems. Compared to the preceding climatically stable and warm Tertiary, it is characterised by a multiple alternation of large-scale glaciations and short warm intervals in between. The latest interglacial period, which is still ongoing, is called Holocene after the Greek words 'holos' (entire) and 'ceno' (new). According to recent understanding, it began *ca*. 11 600 calendar years before present. In contrast to other epochs of the Earth's history, it is not defined and subdivided by certain floral or faunal assemblages but by climatic features. Another basic difference from former geological periods is the increasing human impact on the geosphere. In fact, some authors claim that the properties of the 'system Earth' have changed by human influence to an extent that it cannot be called natural anymore. Reconstructing the geological past, therefore, requires consideration of natural processes, as well as the results of human activity and to separate them from each other. Thus, Holocene geology is intensely interrelated not only with other natural sciences but also with human history, archaeology, and further social sciences. This adds a wealth of additional information to the data stored within geological archives. On the other hand, geological problems during the Holocene are not only a matter of actualism in the classical sense anymore. Due to the increasingly closer connection between geological processes and the development of the human society, forecasting of geological trends becomes more and more important. In this way, Charles Lyell's (*see* **Famous Geologists:** Lyell) statement about the principle of actualism could be extended to: "The knowledge about present and past is the key to the future."

Dating

When aiming for an accurate reconstruction of the geological past, dating becomes an essential issue. Looking back from recent times to the near past, dating of geological events is simply done on a high resolution by analysing the written historical archives. Further back in time, indirect methods (by using so-called proxy data) have to be used. Proxy data with a yearly resolution are related to processes which result in persistent and regularly successive yearly structures within sediments (e.g., varve sequences) or organic material like wood (dendrochronology). The latter is based on the study of tree ring patterns which are controlled mainly by climatic factors. In Europe it was used particularly for oaks in central and western Europe and for pines in northern Europe. Regionally generalised curves for these two tree species cover nearly all of the European Holocene.

Dendrochronological dates are highly valuable for calibrating dating results produced by other methods. This refers especially to isotopic dating by radiocarbon, which is widely used as a standard method for the dating of organic material within the Holocene. By comparing dendrochronological or varve counting dates with radiocarbon dating, inconsistencies within the later ones, particularly within the early Holocene, became obvious. Therefore, when looking at dates in the literature, it is crucial to consider if calibrated (calendar) years or radiocarbon years are referred to. Within this article calendar years before present (BP) are used (except for Figure 10). Other short-lived isotopes, such as ^{210}Pb are used to date processes and events in the more recent past on time-scales of decades and centuries.

Once the environmental history of a specific region is well known, assemblages of plants or animals might also help to assess the age of the deposits they are found in.

Climate

One of the most intensively studied subjects of Holocene development is climate. Ice and marine sediment cores have been used to assess climatic changes and they provide smoothed background data to more regional, or local and mostly more dramatic, climatic variations on the European continent. The onset of the Holocene is marked by a global drastic temperature increase of about 7°C at the end of the Younger Dryas, about 11 600 calendar years BP. This climate reorganisation happened during a period of not more than a few decades. Since then, the Holocene climate has been stable by comparison with the preceding glacial period. Nevertheless, minor climate fluctuations have been reconstructed. Several periods with cooler and warmer temperatures than the last century have left their traces in the geological and biological archives and also in human history (Figure 1).

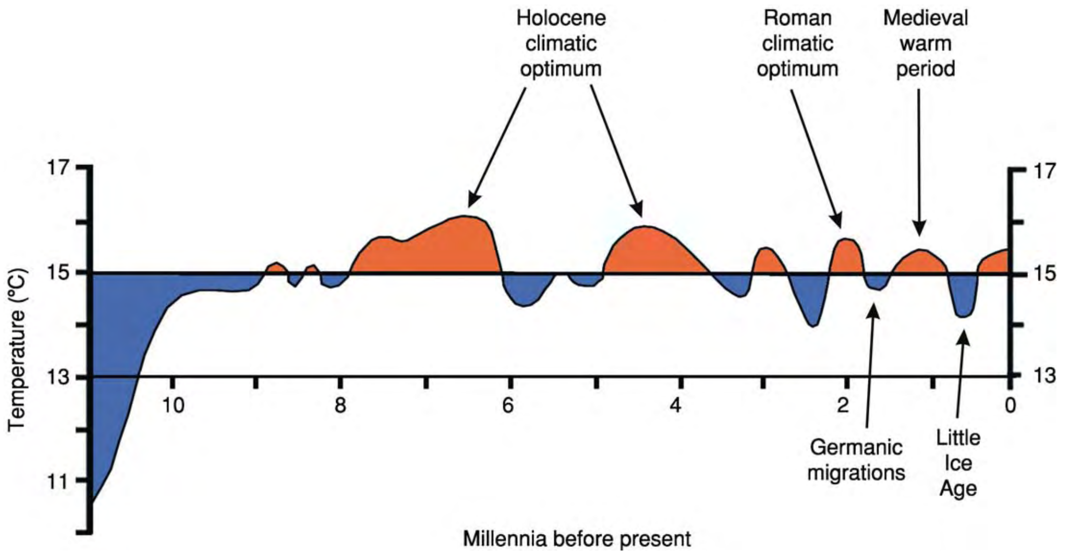


Figure 1 Mean Holocene annual temperatures. The horizontal axis is time before present (BP) in thousands of years (Ka). (Adapted from Schonwiese C (1995) *Klimaänderungen Daten, Analysen, Prognosen*. Berlin: Springer Verlag, with permission.)

During the first thousand years of the Holocene, the climate was possibly slightly cooler than today. Between 9000 and 8200 years BP, climatic conditions were slightly warmer and often moister than today. Cooler conditions throughout Europe have been interpreted from different proxy data for the period around 8200 years BP, when annual mean temperatures dropped by 2°C in central Europe and the Alpine timberline fell by about 200 m. This sudden cool phase lasted for about 200 years and wetter-than-present conditions in north-west Europe have been inferred.

Warmer temperatures prevailed in Europe during the so-called climatic optimum in the Early Holocene (8000-4500 years BP). The Early Holocene climatic optimum was characterised by warmer summers than today in Europe. For astronomical reasons, the northern hemisphere received nearly 8% more solar radiation during summers than in recent times. A northward shift of the inner tropical convergence zone (ITCZ) forced monsoonal rainfall as far north as the Mediterranean Basin. At around 5900 years BP, a short cold episode interrupted this warm phase.

Since 4500 years BP, the climatic conditions have fluctuated around a situation comparable to the recent one. Remarkable deviations occurred at about 3300 years BP, when intensified glaciation started in the Alps. About 2000 years ago BP, favourable climatic conditions promoted the development of the Roman Empire, while the Germanic migrations after its collapse went along with a cooler climate. In mediaeval times (*ca*. 1300–700 years BP), another warm period allowed the expansion of Scandinavian Vikings as far as Greenland and North America. Clear indications of warm temperatures during this time interval were also reported from northern Russia, central Europe, and the Mediterranean. The youngest climatic deterioration, known as the Little Ice Age, at about 700–150 years BP, destroyed the agricultural economic basis of the Norse settlers in Greenland, and by about 500 years BP, their population in Greenland had vanished. In the mid-seventeenth century, glaciers in the Swiss Alps advanced and rivers in England and the Netherlands often froze over during the winter. Severe cold winters have been deduced from borehole data in the Czech Republic for the time slice between 300 and 400 years BP, too. Since the mid-nineteenth century, the global temperatures have risen (Figure 1) again, a process which is still going on.

A general periodicity of 200 to 600 years for the whole Holocene climate can be inferred from various proxies. External processes including solar activity cycles and internal driving forces as volcanic eruptions are under debate as controlling factors of climate variability up until now. To what extent this climatic cyclicity is modified by human activity (e.g., extensive release of greenhouse gases) is a matter of current scientific discussion.

Naturally Changing Holocene Landscapes in Europe

The deglaciation processes, which had started within the Late Pleistocene, were accelerated in the Early Holocene. Within the first two thousand years of the

Holocene, the former prevailing tundra and steppe habitats (Figure 2) were nearly completely replaced by mixed deciduous forest (Figure 3). The first phase of tree invasion was characterised by birch and pine and later by hazel and elm. The local tree assemblages could differ considerably from each other and also from recent compositions. At around 10 000 years BP, in many parts of Europe the forest cover was still rather more open than at present, with more herbaceous glades. By 9000 years BP, the forest had become closed, but with conifers more abundant than at present in eastern Europe. Until that time, deciduous trees such as oaks or hornbeam were predominant or abundant, even in southern Europe (Figure 3). The typical recent Mediterranean vegetation with evergreen trees and shrubs started to develop after that time.

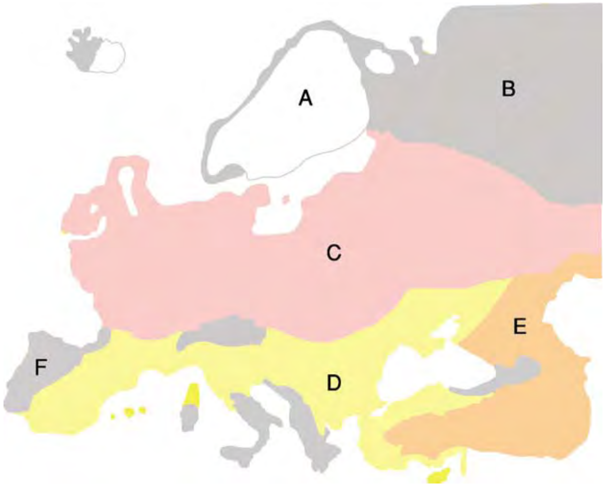


Figure 2 Vegetation zones in Europe during the Younger Dryas. (A Ice; B Polar Desert; C Steppe Tundra; D Dry Steppe; E Semi Desert; F Wooded Steppe. (Adapted from Adams 2002, with permission.)

During the climatic optimum, thermophile plants and animals extended further to the north. Possibly due to elm disease, a drastic decline in the number of elm pollen is observed in the geological records throughout Europe at about 5800 years BP.

Since 4500 years BP, there has been an increasing human influence on the faunal and floral elements of the European ecosystem. Figure 4, showing the potential present vegetation without human influence, is obviously different from the coverage conditions today.

Another result of the changing Holocene climate was a rising global sea-level by meltwater supply and thermal expansion of sea-water (eustatic sea-level rise). During the maximum Weichselian glaciation, the global sea-level was about 125 metres deeper than today. In the Early Holocene, large deglaciated areas started to uplift because of the vanishing ice load (glacio-isostasy). At the centre of the last glaciation, around the Bothnian Bay, an uplift of more than 280 m is recorded within Holocene sediments. This isostatic uplift was compensated by subsidence in more distal regions (Figure 5) within the southern Baltic Basin. By some authors, this process is assumed to be in the context of the collapse of an asthenospheric bulge in front of the retreating Weichselian ice shield (Figure 6). The combination of eustatic

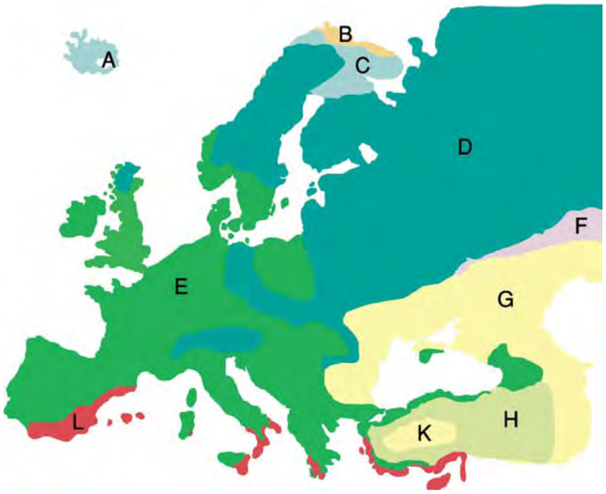


Figure 3 Vegetation zones in Europe about 9000 years BP (A Ice; B Tundra; C Open boreal woodland; D Boreal Forest; E Deciduous/Mixed Forest; F Forest Steppe; G Moist Steppe; H Woodland; K Dry Steppe; L Mediterranean Forest. (Adapted from Adams 2002, with permission.)

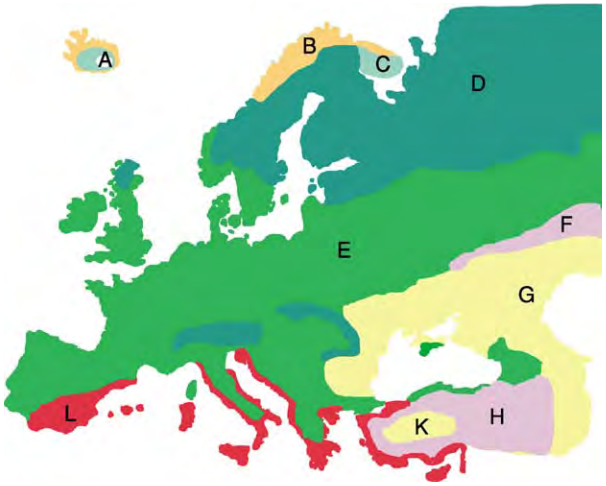


Figure 4 Potential present vegetation zones in Europe (A Ice; B Tundra; C Open boreal woodland; D Boreal Forest; E Deciduous/Mixed Forest; F Forest Steppe; G Moist Steppe; H Woodland/Wooded Steppe; K Dry Steppe; L Mediterranean Forest. (Adapted from Adams 2002, with permission.)

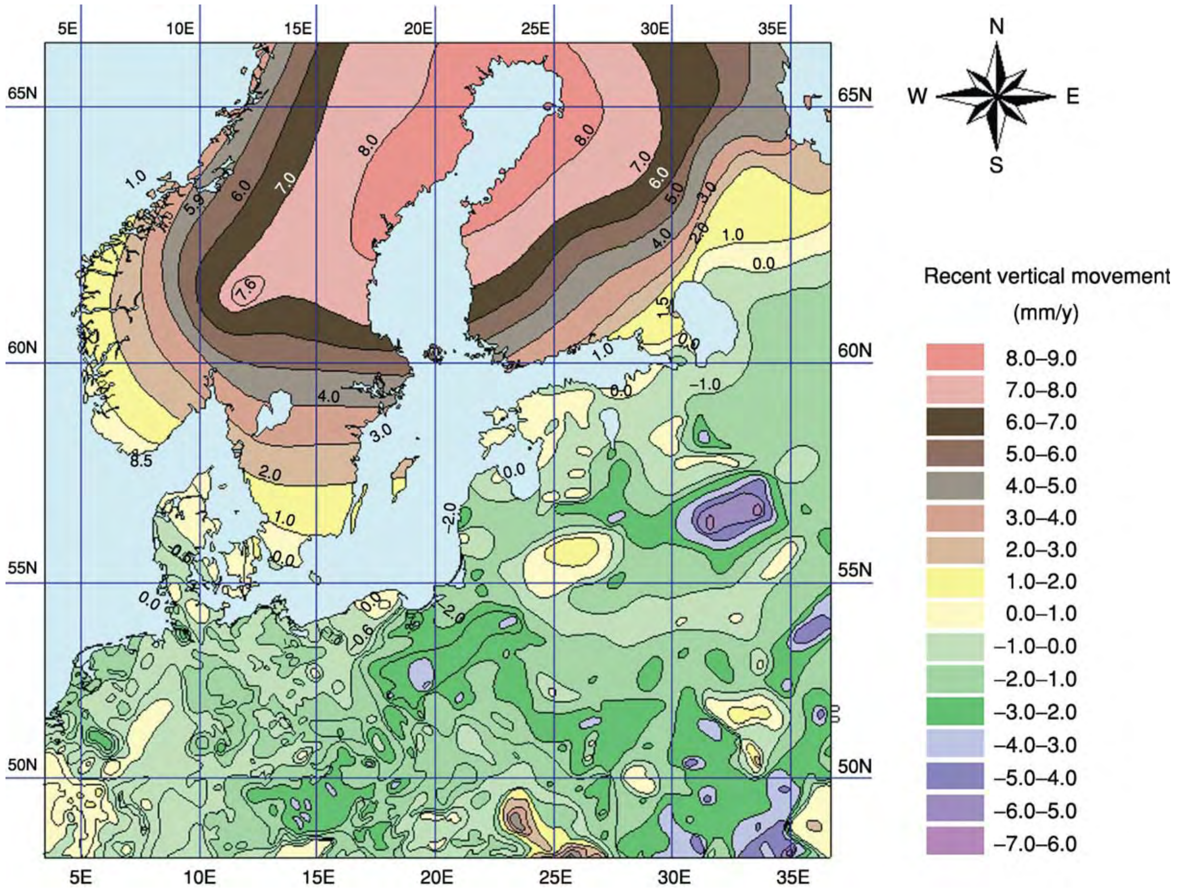


Figure 5 Recent vertical movement of the Earth's crust. In the northern part the map is dominated by the extensive, north east south west stretched uplift of Fennoscandia with maximum values of >8 mm year. The subsidence of a belt surrounding the Baltic Shield is less differentiated. (From Harff J, Frischbutter A, Lampe R, and Meyer M (2001) Sea level change in the Baltic Sea: interrelation of climatic and geological processes. In: Gerhard LC, Harrison WE, and Hanson BM (eds.) *Geological perspectives of global climate change.* Tulsa, Oklahoma, American Association of Petroleum Geologists in collaboration with the Kansas Geological Survey and the AAPG Division of Environmental Geosciences: 231 250. Reprinted by permission of the AAPG whose permission is required for further use.)

sea-level changes and isostatic uplift or subsidence, partly modified by tectonic movements, produced considerable changes in the geography of Europe during the Holocene. Depending on the geographic position of the affected area, large relative sea-level changes (positive or negative ones) have occurred (Figure 7). These changes are particularly obvious where large intracontinental basins like the recent Baltic Sea area were affected.

At the beginning of the Holocene, large parts of the Baltic Basin were filled with freshwater from the Baltic Ice Lake which was fed mainly by meltwater from a large glaciated area in North and north-eastern Europe. The only important outlet in the Öresound (between the recent Danish island Sealand and southern Sweden) was too narrow to serve as a sufficient spillway between the Baltic Ice Lake and the North Sea. The global sea-level was about 25 metres lower than in the Baltic Ice Lake. When the Scandinavian inland ice started to retreat from southern Sweden, a spillway through the central Swedish Depression was opened. As a dramatic process, half of the recent Baltic Sea's water volume drained into the Atlantic Ocean via the Kattegat and North Sea. This drainage took no longer than a few years and had an enormous impact along the former shores of the Baltic Ice Lake. Large areas previously covered by water became dry land, and southern Scandinavia became directly connected to central Europe. Saline waters of the Kattegat could enter the Baltic Basin for a time-span of a few hundred years, a stage of the Baltic Sea's development known as the *Yoldia Sea*, a phase which is dated from 11 570 to 10 700 years BP (Figure 8).

The connection between the Yoldia Sea and the Kattegat through central Sweden was located in a rapidly uplifting region. Therefore, the connection closed at about 10 700 years BP and a newly dammed-up freshwater lake was formed within the Baltic Basin. It

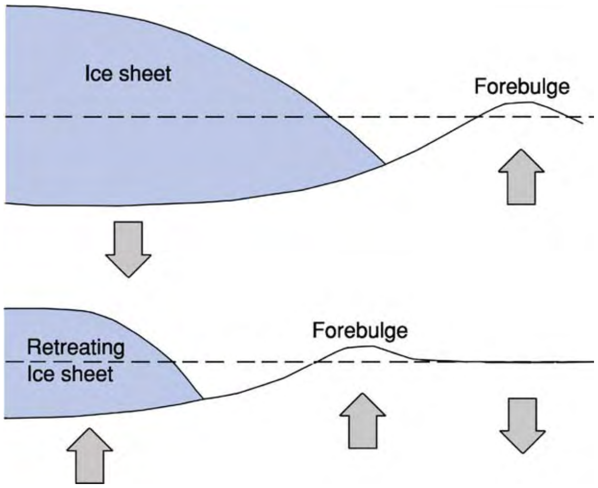


Figure 6 Principle of glacio isostatic vertical crustal move ments: (a) Ice load causes subsidence of the Earth's crust below and compensatory uplift beyond the ice margin as a fore bulge; (b) Uplift of the ice released Earth's crust and related displacement of the forebulge. (Modified after Daly 1934.)

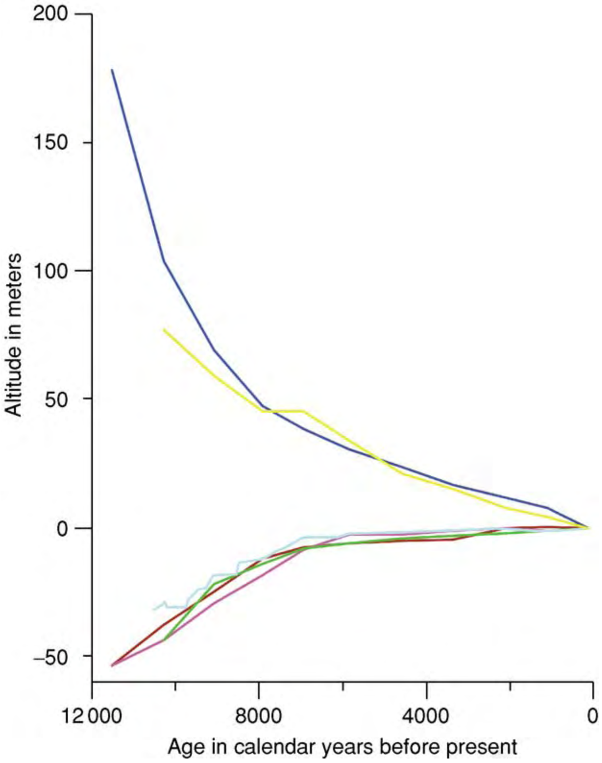


Figure 7 Schematic Holocene relative sea level curves for different European regions (blue = Oslo, Norway; yellow = south eastern Sweden; light blue = southern Baltic Sea; brown = French Atlantic coast; green = Netherlands; purple = central Mediterranean).

is called *Ancylus Lake*, with reference to a typical freshwater snail occurring in these waters (*Ancylus fluviatilis*) (Figure 9). This freshwater stage lasted more than two thousand years and was characterised by considerable short-term water level fluctuations. As differential isostatic uplift continued, the critical thresholds between the Baltic Basin and Kattegat moved from southern Sweden to the recent-formed Danish Straits. At about 8000 years BP, the eustatic sea-level rise led to the first ingressions of marine waters into the Baltic Basin. In pace with the rapidly rising global sea-level, the thresholds were flooded and a stable connection between the Kattegat and the Baltic Basin was formed. This crucial phase of the Baltic Sea's evolution is called the Littorina transgression, after a marine snail which is common in deposits of this period (*Littorina littorea*). At the onset of the Littorina transgression, the water level rose at a rate of 25 mm year within the south-western Baltic Basin (eustatic rise added to crustal subsidence, as described above), which slowed down later to about 3 mm year at about 4500 years BP (Figure 10). During a time-span of less than a thousand years, the sea-level rose by more than 20 m, implying enormous rates of coastal retreat within the southern Baltic Basin. Due to the rapidly rising sea-level at the beginning of the Littorina Stage, the glaciogenically-shaped land relief was drowned without any notable coastal erosion and longshore transport processes of sedimentary material. The resulting geographical situation is shown in Figure 11. Only after the sea-level rise slowed down at the end of the Littorina Stage (about 2000 years BP), and during the Post-Littorina Stage, was the recent spit and barrier coast formed by erosion and sediment transport. This process was mainly controlled by climatic factors, such as the wind-driven hydrographic regime which was superimposed on long-term eustatic and isostatic movements which have caused rising relative sea-levels and coastal retreat at the southern Baltic shores in recent times. Further north in Fennoscandia, isostatic uplift continuously exceeded the eustatic sea-level rise, resulting in a permanent general sea regression (Figure 7).

The changing Holocene sea level within the Mediterranean basins was and is mainly controlled by eustatic processes. In contrast to northern Europe, glacio-isostasy does not play a significant role. On the other hand, this region occupies the junction between the African-Arabian and the Eurasian plates which gives considerable tectonic activity in the different sedimentary basins of the Mediterranean Sea. Therefore, the general picture of the sea-level, development which reflects the eustatic curves, is superimposed on the regional and local tectonics.

A matter of ongoing discussion is the possibility of the reconnection of the Black Sea and the Mediterranean Sea during the period of Holocene sea-level rise. A catastrophic flood scenario at about 7500 years BP

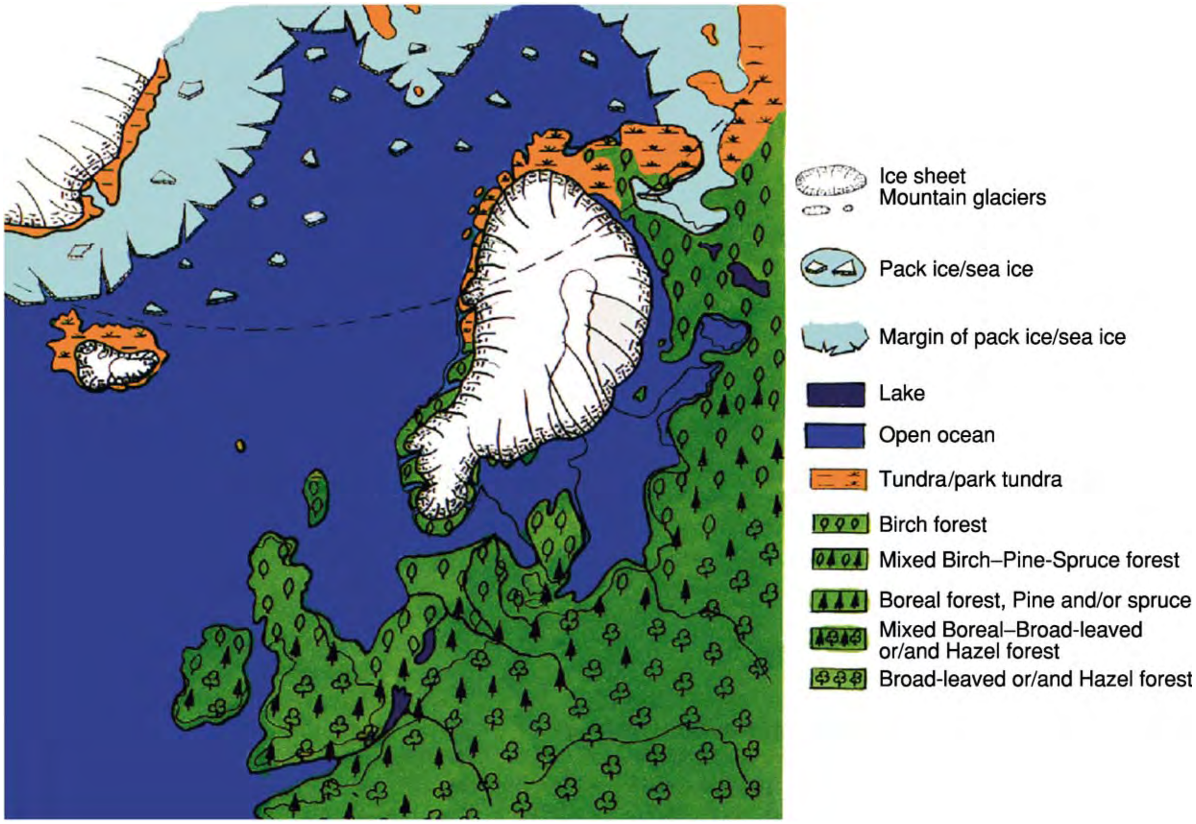


Figure 8 North western Europe during the Yoldia Sea stage of the Baltic Sea's history (about 11 000 years BP). (From Andersen BG and Borns HW (1994) *The ice age world: an introduction to Quaternary history and research with emphasis on North America and Europe during the last 2.5 million years.* Oslo: Scandinavian University Press, with permission.)

is questioned because of indications of a more complex and progressive transgression process over the past 12 000 years.

Interrelation of Human Activity and Natural Environment

In terms of human history, the onset of the Holocene is equivalent to the beginning of the Mesolithic period, which is characterised by prevailing hunter, fisher, and gatherer societies in Europe. The temperature rise at the end of the Younger Dryas caused an accelerated deglaciation. Alpine glaciers retreated to historical dimensions and the receding inland ice in Scandinavia was followed successively by tundra, steppe, and finally forests. This was accompanied by an enhanced northward migration of animals and their Mesolithic hunters. The rapidly changing natural environment, possibly in combination with human activity, resulted in the extinction of some characteristic genera of the Pleistocene megafauna such as the mammoth (*see* **Tertiary To Present:** Pleistocene and The Ice Age). Another consequence of the rapidly changing landscapes was a very variable migration pattern of the Mesolithic hunter, fisher, and gatherer groups in central and north-western Europe.

The following Neolithic period was closely connected with the introduction of farming in the various geographical regions. The first Neolithic settlements in southern Europe (Greece) are dated to about 9000 calendar years BP. In northern Germany, southern Scandinavia and the British Isles, they are more than 3000 years younger. Coming from the Near East, the new method of food-procurement spread to the Great Hungarian Plain in a first wave from 8200–7800 years BP. A second leap entered the North European Plain at around 7400 years BP. Neolithic settlers from Southeast Europe migrating along the rivers Danube and Rhine were probably responsible for the consequent social and cultural changes. In parts of Europe (e.g., western Mediterranean and northern Europe) the native Mesolithic population adopted agricultural methods to form a transitional economy.

While the hunter-fisher-gatherer societies of the Mesolithic used and manipulated the natural ecosystem without altering it considerably, the Neolithic farmers started to transform the environment according to their

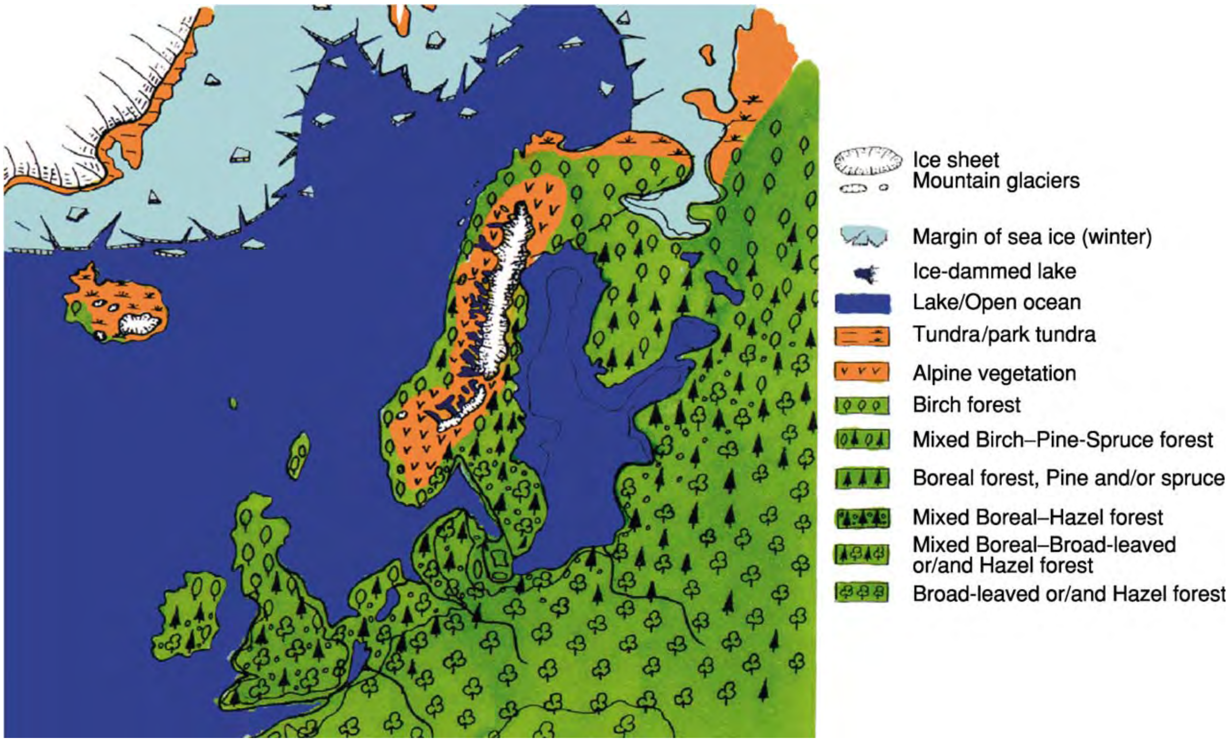


Figure 9 North western Europe during the Ancylus Lake stage of the Baltic Sea's history (about 9500 years BP). (From Andersen BG and Borns HW (1994) *The ice age world: an introduction to Quaternary history and research with emphasis on North America and Europe during the last 2.5 million years.* Oslo: Scandinavian University Press, with permission.)

needs. Cultivation started within rather small (perhaps pre-existing) open land patches in the Early Neolithic. In this phase clearances remained localised and the early Neolithic farmers were still highly dependent on intact natural habitats surrounding the cultivated land patches.

This situation changed in the later part of the European Holocene, when increasing demand for agricultural products induced a major transformation of the yet still mainly natural environment into an agricultural one. This process was certainly time-transgressive and affected some European regions more than others. But, far from being controlled solely by human action, natural feedback, combined with differing vulnerability of the existing ecosystems, amplified, shifted, or interfered with the initial direction of processes initiated by human activity. In order to adapt to the partly self-induced new conditions, human societies had to react by further cultural development which accelerated the general transformation process of the natural environment. This is exemplified by the domestication of animals which were advantaged by the anthropogenically influenced environment, while on the other hand natural competitors and predators became progressively extinct by hunting kill-off or simply by loss of habitat (extensive land use by farming). The same applies to floral assemblages, as initially, their composition was governed by natural conditions but it became increasingly influenced by agriculture. This development finally produced an increasing interdependence of cultivated plants and animals with mankind. Societies like the Irish people in nineteenth century, for example, were greatly dependent on potato growth. When the Late Potato Blight ruined all the potato crop in Ireland in the 1840s, the resulting famine lead to dramatic consequences. About 1 million people died while another 1.5 million people emigrated.

The stability of the cultural landscape which had evolved out of the interplay between natural and human influences, as well as that of human societies living there, depended on the state of their equilibrium. Changes of the natural component could cause considerable impact to the progressively complex human society and this is exemplified by climatic or geological influences. Some historical epochs with a prospering economy and politically stable conditions are connected with warmer periods.

Colder climate deviations often were characterised by political and economical instability. This refers, for example, to the time of the Germanic migrations about 450–700 AD and the Little Ice Age between 1500 and 1800 AD. Rather short-term, but possibly catastrophic, impacts derived from events like

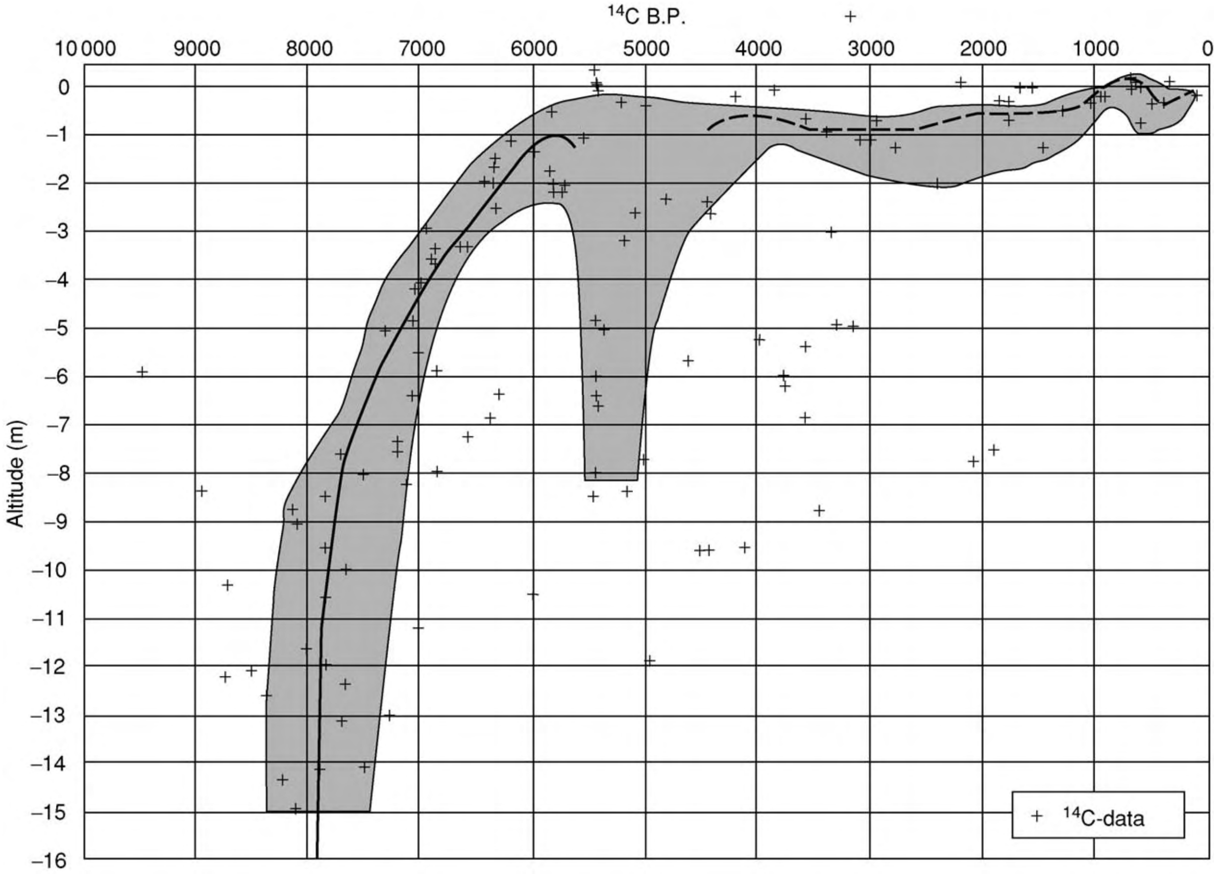


Figure 10 Preliminary curve of relative sea level change for western Pomerania, shaded area: interval of 'data confidence', solid line: local trend for western Pomerania, dashed line: estimated trend, for the time span between 6000 and 4500 years BP tectonically controlled local uplift is assumed. Note: Ages are given here in ^{14}C years BP. (From Harff J, Frischbutter A, Lampe R, and Meyer M (2001) Sea level change in the Baltic Sea: interrelation of climatic and geological processes. In: Gerhard LC, Harrison WE, and Hanson BM (eds.) *Geological perspectives of global climate change.* Tulsa, Oklahoma, American Association of Petroleum Geologists in collabor ation with the Kansas Geological Survey and the AAPG Division of Environmental Geosciences: 231 250. Reprinted by permission of the AAPG whose permission is required for further use.)

volcanism (e.g., the eruption of Vesuvius in AD 79 described by Pliny the Younger), earthquakes (e.g., the disastrous one of Lisboa in AD 1755) or floods along rivers and seashores.

On the other hand, changing social conditions also induced dramatic changes of their natural environment. Once human activity stopped or declined, the surrounding environment developed depending on its natural stability. Robust ecosystems changed back to a state similar to the original one being controlled by the natural conditions. An example for such processes is the re-forestation after depopulation during the Thirty Years' War (1618–1648 AD) in central Europe.

In other cases, human activity led to irreversible effects on the ecosystem's stability when the social structures collapsed. During the Roman period there was a land use maximum partly on metastable soils. Here, terraces were maintained, preventing soil erosion. After the invasion of eastern nomads, the land was partly abandoned and soil erosion started at a greater extent and in the worst case, barren badlands were the final result. Therefore, maximum erosion is not connected with maximum land use, but with subsequent phases in different cultural environments. Generally, when judging human impact on the natural environment, it is often regarded only as negative. However, in contrast, the creation of new metastable ecosystems more diverse than before, particularly in the early phases of the Holocene, would not have been possible without human activity.

Human Activity and Environmental Conservation

During the last 500 years, human impact on the European environment have become much more important than the natural conditions. Particularly

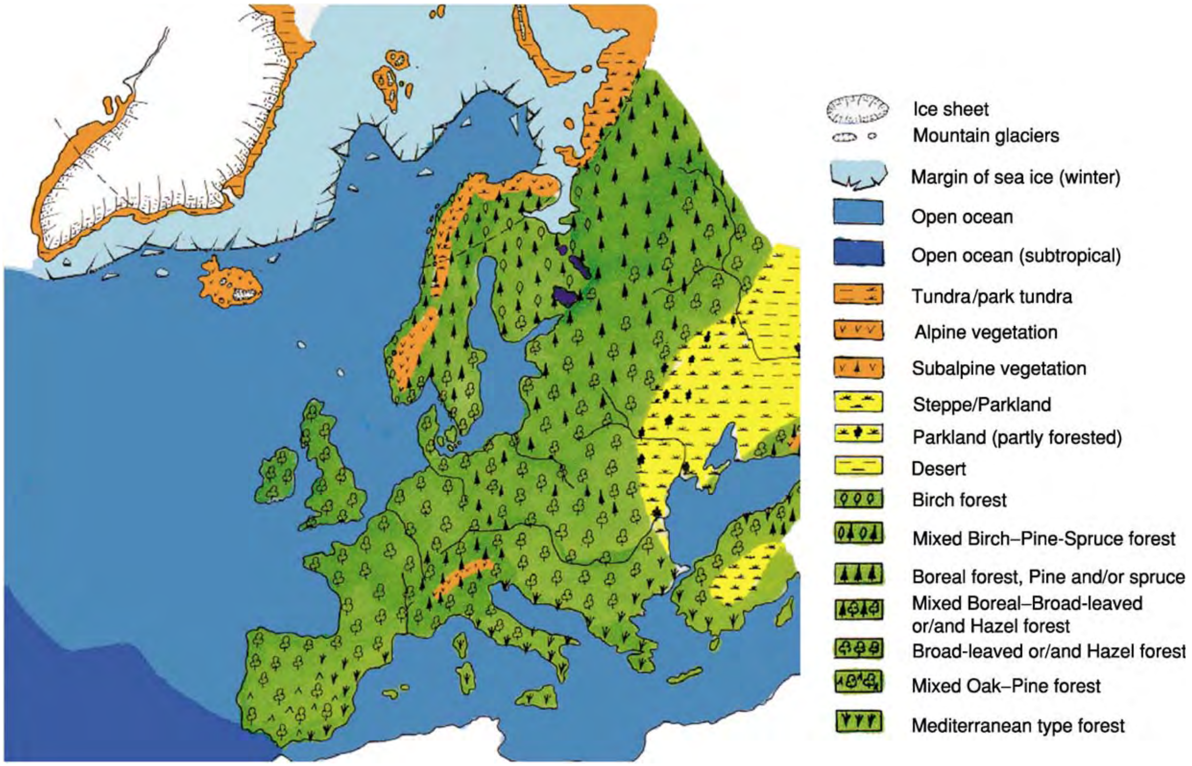


Figure 11 Europe about 4500 years BP. (From Andersen BG and Borns HW (1994) *The ice age world: an introduction to Quaternary history and research with emphasis on North America and Europe during the last 2.5 million years*. Oslo: Scandinavian University Press, with permission.)

within the last hundred years, industrial methods in agriculture have resulted in a widespread conversion of natural habitats into farming land. Accelerated human-induced soil loss has become a major problem. Measures initiated to solve the problems of modern agriculture, for example, to protect the soil against nutrient depletion, partly result in adverse effects such as eutrophication.

One of the most severe impacts on environment is connected with the beginning of industrialisation. Natural resources which had accumulated over millions of years have been exploited within decades or less. Large industrial facilities have been built on terrain of formerly less affected landscapes. Moreover, military needs have modified great parts of the terrain. Growing populations all over Europe have enhanced the conversion of natural habitats into settlement areas. At the same time, the management of industrial and municipal waste has become a major issue. The effects of industrialisation are reflected by many proxies such as the concentration of heavy metals in sub-recent deposits (Figure 12).

All these tendencies have developed ideas on nature conservation, sometimes with the idealistic approach of going back to a state where human influence is negligible. Nature conservation in this context may be regarded as a contradiction in itself because it does not recognise the vital role of human society for the natural environment. Large areas along the Netherlands's coast would have been flooded if the coast was not protected by coastal engineering. Furthermore, the dynamic character of the environmental status must be considered. For the last 8000 years of the Baltic Sea's history it can be shown, for example, that high nutrient levels already existed immediately after the Littorina Transgression. Without any remarkable human influence, organic substances and nutrients could accumulate in the sediments because of restricted vertical convection. This process has been intensified during the last centuries and decades by agriculture and the industrial release of nutrients (Figure 13). Thus, if the Helsinki Commission for the protection of the Baltic Sea aims to restore the eutrophication level of the 1950s, it must be stated that that was just one time slice of the Baltic Sea's Holocene development. One might be successful in re-establishing the concentrations of certain nutrients, but it would be in a completely different new context. As human society develops further, it cannot be expected that there will be no response by the natural environment. A certain equilibrium, including the benefit for as many species as possible including

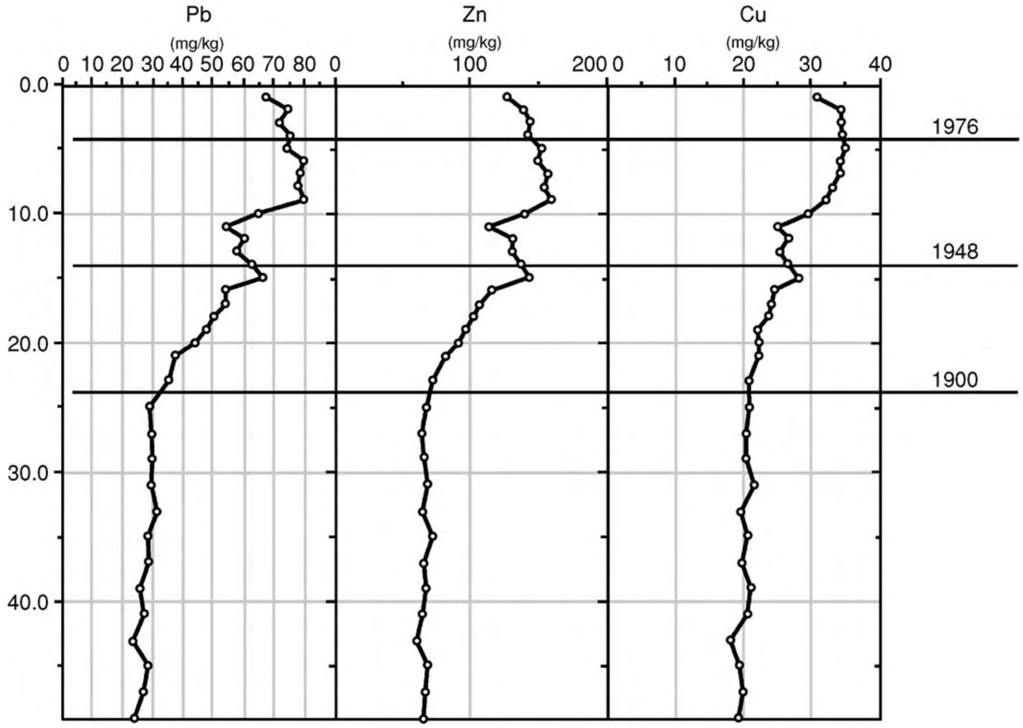


Figure 12 Enrichment of heavy metals within the last century, as recorded in a sediment core from the Arkona Basin (western Baltic Sea). (From Th. Leipe, Baltic Sea Research Institute Warnemunde, with permission.)

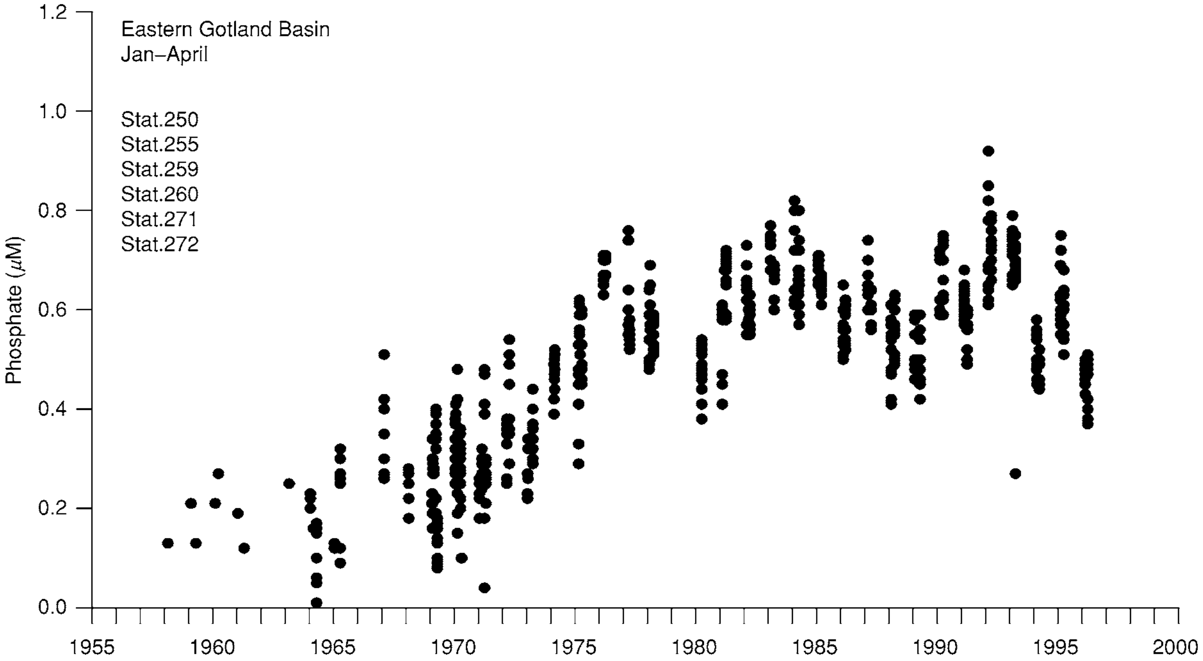


Figure 13 Trends of winter phosphate concentrations in the surface layer (0 10 m) of the eastern Gotland Basin (central Baltic Sea). (From G. Nausch, Baltic Sea Research Institute Warnemunde, with permission.)

Table 1 Summary of the climatic and historical development during the European Holocene

Calendar years BP	*Environmental Period*	*General climatic features*	*Selected historical events*	*Stages of the Baltic Sea's development*
		Modern climatic optimum, warm	World War I and II	Mya Sea
200	Subatlantic			
			French Revolution	
		Little Ice Age, mean annual temperature in Europe 1°C lower than today, cold winters, but pronounced fluctuations, glacier advances		Less brackish
400			Thirty Years' War	Mya Sea
				Lymnaea Sea
			Renaissance and reformation	Brackish
600				
			Bubonic plague kills about one third of the European population	
		Transition to colder conditions		
800				
		Medieval climatic optimum, mean annual temperature 1 1.5°C higher than today, winegrowing as far north as to the British Isles, first dry, later wet		Brackish
	Subatlantic			
1000				
			Expansion of Norman people as far as Iceland, Greenland, North America and southern Russia	
1200				
		Cold and wet period, many glacier advances		
1400				
	Subatlantic		End of the Roman Empire	Slightly brackish
1600				
			Invasion of the Huns forces the emigration of nations	Lymnaea Sea
1800				
2000		Roman climatic optimum, as warm as the Medieval optimum, mostly very wet, towards the end more dry	Foundation of the Roman Empire	
			First southward migrations of Gothic tribes	
2200				
		Pronounced cold period, mean annual temperature 1 1.5°C lower than today, very cool summers, very wet, large glacier advances	Celtic La Tène culture in big parts of Europe	
2400				
			Beginning of the Greek classic period	
			Formation of the Roman Republic	
2600			Celtic Hallstatt culture in central and western	
	Subatlantic		Europe introduces the iron age	
	Subboreal		Greek Archaic period	
2800				
3000			First Celtic tribes in eastern and central Europe	
				Brackish
3200				

(Continued)

Table 1 Continued

Calendar years BP	*Environmental Period*	*General climatic features*	*Selected historical events*	*Stages of the Baltic Sea's development*
		Predominantly warm with Distinct fluctuations, less precipitation than during Subatlantic		
			Urn field culture in central and South eastern Europe	
3400				
			First bloom of the Mycene culture in Greece	
3600				
			Beginning of the Bronze Age in northern Europe	
3800	Subboreal			
		Cold period with glacier advances, initially dry, later more wet		
4000				
4200			In Europe exists an extensive trade route network	
4400				
			Early Minoic culture at Crete Island launches the European Bronze Age	
4600				
4800				
5000			Increasing influence of Indo European people all over Europe	
5200			Introduction of the wheel in Europe	
5400			Megalithic monuments in many regions of central and western Europe	Brackish
5600	Subboreal			
5800	Atlantic			
6000			Narva culture in North eastern Europe	
6200		Rapid increase of humidity in eastern central Europe	First stone buildings at the Orkney Islands	
6400			First Neolithic settlements at the British Isles	
		Warm period, mean annual temperatures 2 3°C warmer than today, especially warm winters, very moist, former predominant pines are replaced in the forests by oaks, lime and hazel		
6600				
			Late Mesolithic Ertebølle culture in northern central Europe	
6800				
7000			Neolithic Karanovo culture in South eastern Europe	
7200			Late Mesolithic Ertebolle culture in northern Europe	
7400			Neolithic 'Bandkeramik' culture in the Loess areas of central Europe	
7600				Strongly brackish
7800				
			Mesolithic Kongemose culture in northern Europe	Littorina Sea
8000				

(Continued)

Table 1 Continued

Calendar years BP	*Environmental Period*	*General climatic features*	*Selected historical events*	*Stages of the Baltic Sea's development*
		Short lived cold interval, drop of the mean annual temperature by 2°C		Ancylus Lake
8200			First Neolithic agricultural societies in South eastern Europe and Greece	Freshwater
8400				
			The English Channel separates the British	
8600		Slightly warmer and moister than today	Isles from the continent	
8800				
9000	Atlantic		Hunter and gatherer societies	
	Boreal	Cold phase in the Carpathian Basin		
9200				
		During summers generally warmer than today, mainly open winters	Maglemose culture in northern Europe	
9400				
			Azilian culture in western Europe	
9600				
			Microliths become common	
9800				
			Use of log boats is proven	
10000				
10200	Boreal			
	Preboreal		Nomadic hunters arrive in England	
10400				
			Ahrensburg culture in northern Germany	
		Freshwater		
10600				
		Summers as warm as today, but very cold winters		
10800				Ancylus Lake
				Yoldia Sea
11000				Freshwater
11200				Regionally brackish
11400		Quick warming	Bromme culture in Denmark	
				Freshwater
11600	Preboreal	Beginning of the Holocene	Mesolithic	Yoldia Sea
	Younger Dryas	End of the Pleistocene	Palaeolithic	Baltic Ice Lake
11800				
12000		Cold period, mean annual temperatures by 5 9°C lower than today		Freshwater

mankind is desirable. Closed production cycles might be one of the important targets to achieve this.

Actualism in a New Context

The special character of the Holocene as a period which is not only part of the geological past but also an interface with the future, gives the principle of actualism a new dimension. In addition to regarding the present as being the key to the past, past and present times might be regarded as a key to the future. Predictions of future developments become more and more important to cope with possible changes in the natural environment. For this purpose, detailed knowledge about similar processes in the past is indispensable. In order to calibrate proxies from the geological record, it is necessary to analyse recent proxies by comparison with older ones, and also the written record in order to separate historical trends. Improvements in dating methods, and additional information from geological, archaeological, biological, historical and other sources will help to develop scenarios which might help the recognition and response to future challenges.

See Also

Engineering Geology: Natural and Anthropogenic Geohazards. **Famous Geologists:** Lyell. **Fossil Vertebrates:** Hominids. **Tertiary To Present:** Pleistocene and The Ice Age.

Further Reading

Adams, J *Europe during the last 150 000 years* [online at http://www.esd.ornl.gov/projects/qen/nercEurope.html]

Andersen BG and Borns HW (1994) *The ice age world: an introduction to Quaternary history and research with emphasis on North America and Europe during the last 2.5 million years.* Oslo: Scandinavian University Press.

Björck S (1995) A review of the history of the Baltic Sea, 13.0 8.0 ka BP. *Quaternary International* 27: 19 40.

Cunlifffe B (ed.) (1994) *The Oxford Illustrated Prehistory of Europe*. Oxford New York: Oxford University Press.

Donner J (1995) *The Quaternary history of Scandinavia.* Cambridge: Cambridge University Press.

Emeis K C and Dawson AG (2003) Holocene palaeoclimate records over Europe and the North Atlantic: modelling and field studies. *The Holocene* 13: 305 464.

Grove JM (1988) *The Little Ice Age.* London, New York: Routledge.

Harff J, Frischbutter A, Lampe R, and Meyer M (2001) Sea level change in the Baltic Sea: interrelation of climatic and geological processes. In: Gerhard LC, Harrison WE, and Hanson BM (eds.) *Geological perspectives of global climate change.* Tulsa, Oklahoma, American Association of Petroleum Geologists in collaboration with the Kansas Geological Survey and the AAPG Division of Environmental Geosciences: 231 250.

Litt T, *et al.* (2003) Environmental response to climate and human impact in central Europe during the last 15 000 years a German contribution to PAGES PEPIII. *Quaternary Science Reviews* 22: 1 124.

Pirazzoli PA (1991) *World atlas of Holocene sea level changes.* Elsevier Oceanography Series 58, Amsterdam, London, New York, Tokyo: Elsevier Science Publishers B.V.

Roberts N (1998) *The Holocene An environmental history.* Oxford: Blackwell Publishers Ltd.

Schönwiese C (1995) *Klimaänderungen Daten, Analysen, Prognosen.* Berlin: Springer Verlag.

EVOLUTION

S Rigby, University of Edinburgh, Edinburgh, UK
E M Harper, University of Cambridge, Cambridge, UK

Introduction

The theory of evolution by natural selection, put forward by Darwin in 1859 (*see* **Famous Geologists:** Darwin), is the greatest unifying theory of biology and palaeontology. In this context, evolution is the change that occurs between successive populations of organisms, due to their modification in response to selection pressures. The potential to change is provided by genetic variability within populations and by genetic change through time (mutation). The pressure for change to occur exists outside an organism and is provided by interactions within the environment. These interactions may be predominantly physical or biological effects. Small-scale changes in populations, giving rise to new species, are defined as microevolution. Larger-scale changes, such as the origin of new higher taxa – which may have new body plans or new organs – are defined as macroevolution. The study of evolution also includes the study of patterns of diversification and extinction. Macroevolution may be the end result of microevolution working over a long time-scale or it may be a suite of emergent properties that require unique interpretations.

Historical Background

The presence of large numbers of species on the Earth and the means by which they appeared were discussed throughout the Enlightenment, though the use of the word evolution did not become common until the twentieth century. The possibility that one species might change into another was of interest to Charles Darwin's grandfather, Erasmus Darwin, for example. Studies that focused on the ways in which species transform were begun by Jean-Baptiste Lamarck and published in his *Philosophie Zoologique* in 1809.

Lamarck argued that an 'internal force' caused offspring to differ slightly from their parents and also that acquired characters could be passed on to the next generation. One of his examples was that of giraffes, whose long necks were assumed to be a product of successive generations reaching for higher and higher leaves. He suggested that each giraffe lengthened its neck slightly by this activity and, in turn, passed on to its descendents the capacity to grow longer necks. He visualized species as forming a chain of being, from simplest to most complicated, with each species being capable of transforming into the next in line, and all existing indefinitely. In Britain this work was disseminated by both Richard Owen, who was generally supportive of the theory, and Charles Lyell (*see* **Famous Geologists:** Lyell), who was critical of it.

Charles Darwin encountered work by both of these scholars and also explored huge tracts of the natural world during his 5 years study on the *Beagle* (1831–1836). His work on a number of organisms, notably finches collected from the Galapagos Islands in the Pacific, persuaded him that organisms were adapted to their particular niche and that species were capable of change. The process by which this change could occur was a preoccupation of Darwin's in the succeeding years. As early as 1838, he had read the seminal work of Malthus on populations, but he was still working on the scope and implications of his theory when he was forced to publish by correspondence from Alfred Russel Wallace. A joint paper presented to the Linnaean Society in 1858 was followed the next year by his classic work *On the Origin of Species*.

Darwin's theory of species originating through natural selection can be set out in a small number of propositions. First, organisms produce more offspring than are able to survive and reproduce. Second, successful organisms – those that survive long enough to breed themselves – are usually those that are best adapted to the environment in which they live. Third, the characters of these parents appear in their offspring. Fourth, the repetition of this process over a long time-scale and many generations will produce new species from older ones.

The consequences of this theory are enormous. Not least, they caused scientists at the time to reconsider their assumption of a chain of life. Evolution by natural selection is a response to the local environment and is not predetermined on a grand scale. Organisms do not necessarily evolve into more complicated species over time. Amongst the general public, the theory was seen as being in conflict with a literal reading of the Bible, a view that persists amongst a religiously conservative minority.

In the years after publication, the most significant weakness of Darwin's theory was perceived to be its failure to supply a plausible mechanism for the inheritance of characters. However, this mechanism was supplied when Gregor Mendel's (1865) work on heredity was rediscovered in the early twentieth century. Mendel observed that characters were passed from parent to child in a predictable fashion depending on the relative dominance of the traits carried by each sexual partner. Characters did not 'blend' in the offspring, which is what Darwin had suggested and which astute critics had pointed out would actually have prevented evolution from occurring. These observations opened the door to the modern study of genetics. After some decades of debate, a modern consensus was reached in the 1940s, which is the basis for our current understanding of Darwin's ideas.

Evolution and Genetics: The Living Record

Evolution is possible because the genetic transmission of information from parent to offspring works as it does, in a Mendelian fashion. Subsequent work on genetics has elucidated the exact means by which this occurs and has shown how variation can be developed and sustained in a population.

The information that can be passed from one generation to the next in a population is contained on strands of DNA (deoxyribonucleic acid), or occasionally RNA (ribonucleic acid), within each cell. A DNA molecule forms from a series of nucleotides, which are joined up like beads on a string. Each nucleotide has, as one of its elements, a base. The four types of base DNA are adenine, thymine, guanine, and cytosine (usually abbreviated to A, T, G, and C). Two strings of nucleotides join via base pairs to make the double-helix shape of DNA. A always joins to T, and C always joins to G. Sequences of bases are the code that stores the information needed to produce an organism. This includes information about making the various parts of the cell or set of cells and also information about the rates at which different processes should occur and their relative timings. Each piece of information that the DNA holds is called a gene. Genes can be sequences of DNA or can be little pieces of DNA separated by other sets of bases. Most of the DNA appears to have no purpose and is called non-coding DNA. A human is produced from about 30 000 genes that use about 5% of the nucleotides of our DNA (Figure 1).

When sexual reproduction occurs, one copy of the DNA (carried on chromosomes) of each parent is passed to the children. The offspring therefore have two sets of instructions within their DNA. The pair of genes that share a common function are called alleles, and the combination of alleles controls the effect on the bearer. However, this effect will not be passed to

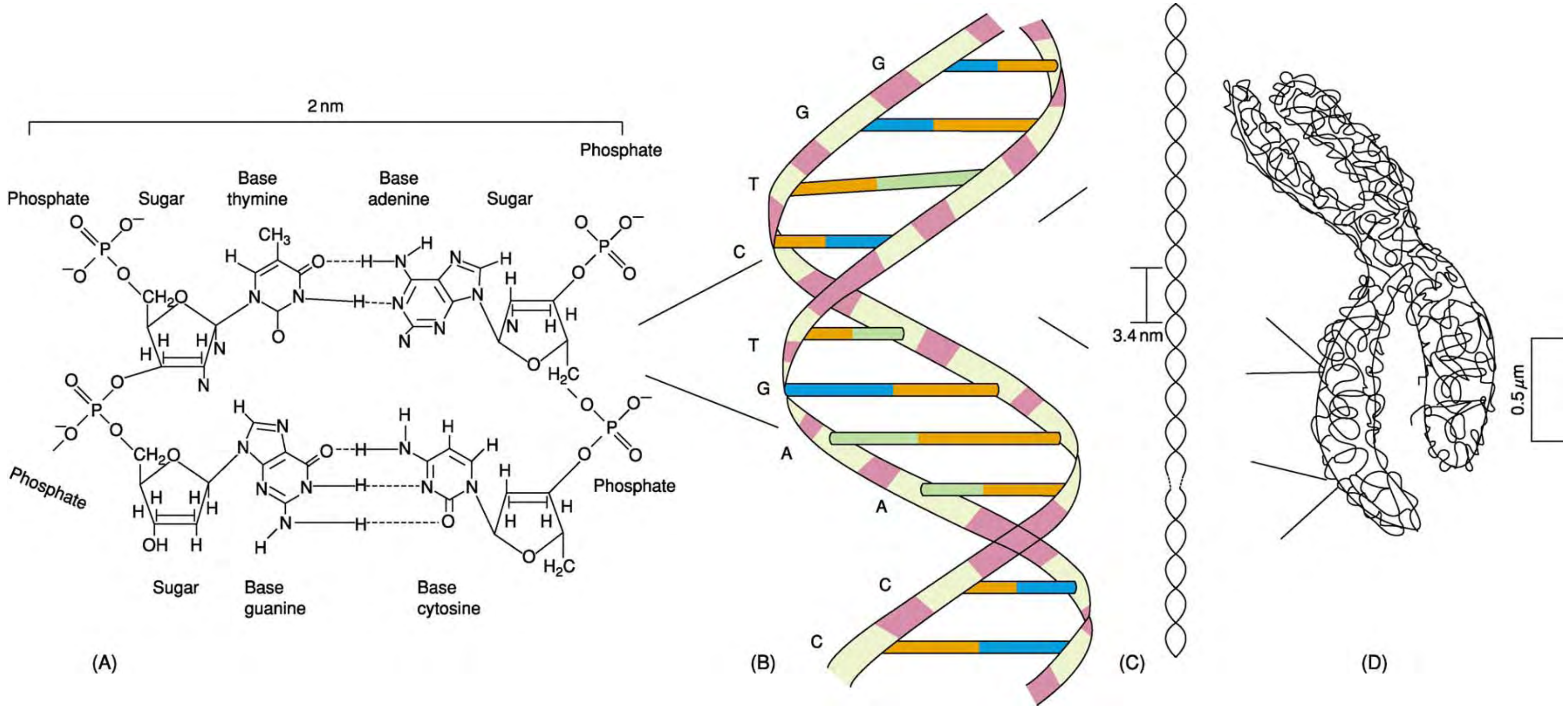


Figure 1 A diagrammatic representation of DNA, genes, and chromosomes. (A) The molecular structure of a double strand of DNA. Each strand is made up of a chain of sugars (yellow) and phosphates (purple), linked together by a set of four bases: thymine (orange), adenine (green), guanine (blue), and cytosine (red). The shapes of these bases cause adenine to bond to thymine and guanine to bond to cytosine. This makes each strand of DNA a mirror image of the other. (B) A piece of DNA carries information in the form of sets of bases, in this example GGTCTGAAC. (C) A gene is a set of useful bits of DNA, which code for a particular protein or carry out a particular instruction. Genes may be formed in pieces separated by long intervals. (D) A chromosome is a folded up cluster of DNA found in the nuclei of eukaryotic cells.

the next generation, but rather one of the alleles will. This will then combine with another allele to generate another product. Although the results of allele combination can have a complicated range of expressions in a cell or a body, the alleles don't mix, so variation is maintained.

A wide range of errors can occur when the DNA strand is replicated during reproduction. These can affect the non-coding DNA or the genes and can produce mutations of varying effect depending on whether it is genes that control the production of the body or the timing or duration of elements of this production process that are affected.

Time and Narrative: The Fossil Record

Biologists have explored theories of evolution in tandem with palaeontologists, who can retrieve narratives of evolutionary change from the fossil record. The ability to study change over millions of years is a great advantage of using fossils. Theoretically, it should be possible to study aspects of the morphologies of fossils collected bed-by-bed throughout a rock sequence in order to elucidate patterns of evolution. However, the preservation of individual fossils is often poor; most depositional events produce significant time-averaging, and the fidelity of long records of sedimentary sequences is often questionable. At some scale, all deposition is intermittent, and this means that there are gaps of some scale in all narratives retrieved from the fossil record. Fossils preserved in lakes or deep-sea cores may be less affected by this problem than fossils from more dynamic environments, and research has generally concentrated on these locations.

Microevolution

The set of potentially interbreeding members of a population forms a species, which contains a range of variation in its appearance (the phenotype) and in its genetic codes (the genotype). In practice, most living species are defined on the basis of phenotypic characteristics rather than genetic information or reproductive potential. In fossil studies of evolution, only the phenotypes are available, and the definition of a species must be based on clusters of phenotypic characters, which are taken as proxies for the potential to interbreed.

Natural selection acts on a set of individuals, so that the physical characteristics of the group and the underlying genotypes change over time. This process eventually gives rise to new species and is known as microevolution. Biologists class only gene shifts within populations as microevolution and define anything larger, including the appearance of new species, as macroevolution. To palaeontologists, the distinction is usually between speciation and anything higher, such as the emergence of new genera or of new organs.

Sometimes a species gradually changes through time until the point comes where the fossil representatives of successive populations are recognized as a different species. However, a parent species often splits into more than one offspring species or evolves into an offspring species that coexists with the parent species for some time. In this case, the original population must split into two or more subsets that cannot interbreed with one another. The two best-known methods of achieving this are called allopatric speciation and sympatric speciation.

In allopatric-speciation events a single original population is split into two geographically isolated elements. This is a common phenomenon over geological time as continents fragment, mountains rise, or sea-levels change. Each geographically separated fragment of the initial population contains only a fraction of the original genetic variation, so it may tend towards difference from the original population without any active selection, although this is now regarded as a minor component in the formation of new species. More importantly, different geographical regions will tend to produce different environmental stresses from those that were experienced before separation, leading to the selection of different successful characters in the separated populations. This eventually leads to significant changes of form in the isolated populations, which may finally produce new species.

An example of allopatric speciation has been recovered from the fossil record of Plio-Pleistocene (3–0.4 Ma) radiolarians, which are siliceous planktonic protists, collected in the North Pacific. A divergence in the forms of two sister species of the genus *Eucyrtidium* was found to have occurred at around 1.9 Ma, following a short period when the populations had been separated from one another.

During sympatric speciation the emerging species share a geographical range but may become separated over time by differences in behaviour or in resource exploitation. Adaptive pressures act differently on these populations, and different characteristics will be favoured, leading to a progressive change of form and eventually to reproductive isolation. At this point a new species will have appeared. There is some doubt about the mechanism by which species first begin to diverge without becoming geographically isolated, although the generation of new species in this way has been demonstrated for a number of types of animal. Studies on cichlid fishes in African lakes show that the most closely related species of fish often live in the same lake, rather than in adjacent lakes, as might be expected if allopatric speciation had occurred (Figure 2). Although it seems intuitively obvious that populations that become physically dissimilar will eventually be unable to produce offspring, the genetic basis for this change can be demonstrated in the laboratory but not yet fully explained.

The fossil record can be used as a tool to help in the understanding of evolution and the formation of new species. It may be that evolution progresses gradually for most of the time, an idea known as phyletic gradualism. The classic fossil example of this slow continuous process of morphological change is the study by Peter Sheldon of Ordovician trilobites recovered from deep-water shales in central Wales. Eight different genera of trilobite, including well-known forms such as *Ogygiocarella*, were found to exhibit incremental changes in rib number through the duration of one graptolite zone, which probably represents significantly less than 1 Ma. Gradual change is generally difficult to observe in the imperfect fossil record. It could be argued that in a less continuous sedimentary record (or one sampled less finely) this sequence of events would appear as a series of abrupt changes. Commonly, what is preserved is a long period where little or no change is observed followed by the abrupt appearance of a new form.

The theory of punctuated equilibrium attempts to explain this phenomenon not as the product of an imperfect fossil record but as a common pattern of evolutionary change. This is done by applying the concept of allopatric speciation to the problem. Eldridge and Gould, who developed the idea of punctuated equilibrium, argue that most species probably arise in small, geographically isolated areas and that they arise rapidly as they encounter new selection pressures. At some later time the evolved offspring species may move back into areas where it encounters its parent species and may out-compete this form. In most areas where this happens, the geological record will show one species – the parent – abruptly replaced by another – the offspring – with no intermediate steps. The chance of the isolated population being represented in the fossil record during the short period of its evolution into a new species is very slim (Figure 3). It may be that Williamson, in a study of molluscs in Plio-Pleistocene sediments from Lake Turkana, found one such rare fossil example of punctuated equilibrium. Species of gastropod and bivalve both appeared to remain static in shape for long periods of time, punctuated by brief periods when their shape changed abruptly.

Although some studies seem to show a punctuated-equilibrium style of evolution, others appear to show that evolution has progressed via phyletic gradualism, and a consensus has yet to emerge regarding these theories. In practice, most evolution is probably the result of a mixture of punctuated and gradual periods of change, partly depending on the scale of

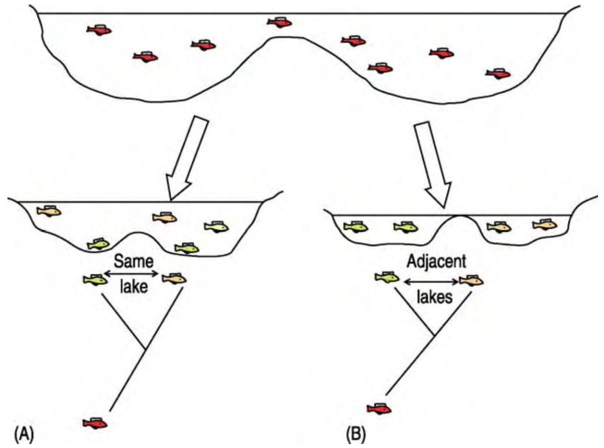


Figure 2 The difference between allopatric speciation and sympatric speciation, using the example of fishes living in lakes. (A) Sympatric speciation occurs due to changes in behaviour or mode of life, in this case by a partitioning of the original population into limnetic and benthic groups. Here, descendent species are most closely related to species living in the same lake. (B) Allopatric speciation occurs following geographical separation of the populations, in this case caused by a fall in lake level. Descendent species are most closely related to fishes living in adjacent lakes.

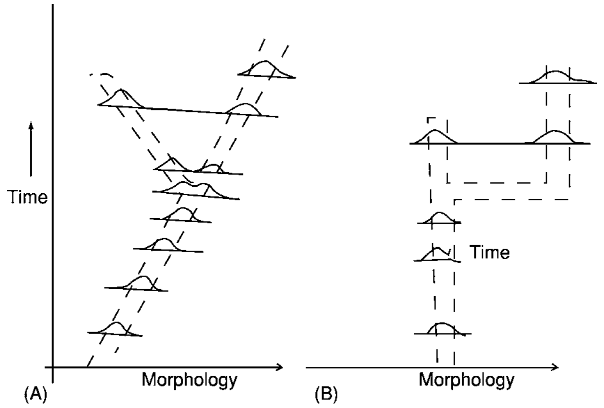


Figure 3 The differences between phyletic gradualism and punctuated equilibrium models of speciation. (A) In phyletic gradualism the shape change is gradual and populations are seen to move across morphospace continuously. Periods of speciation are relatively long and can be recorded in the fossil record. (B) In punctuated equilibrium the shape change is intermittant, rapid, and related to geographical separation of a part of the population. For most of the time the form of the population is static. In most areas no speciation event is seen, and the fossil record shows abrupt changes of morphology with no intermediate stages.

observation. Work by Johnson on Jurassic oysters (*Gryphaea*) from across western Europe provides a good example of this aggregate pattern. Change over approximately 6 Ma was generally slow, but rapid periods of change in isolated populations were also observed. One unfortunate result has been the suggestion that punctuated equilibrium is antithetical to Darwinian evolution. In this usage it is not, as even the rapid bursts of evolution implied by the theory would take place via a series of gradual (i.e. small-scale) changes in the form of the organism concerned.

Macroevolution

Macroevolution is the study of all evolutionary events or effects larger than the appearance of a new species. This includes studies of long-term change in the geological record and of the emergence of new higher taxa, for example new phyla. Linked to both of these topics is the difficult issue of how significant new structures or organs can evolve. Palaeontology is central to this study, as it provides a measure of time and can identify the most likely dates of appearance of new characters or taxa.

The single biggest and most important argument about macroevolution is whether it is a scaled-up version of microevolution or something different. If it is different, then those differences may be a reflection of the emergent properties of this complicated system and hence still reliant on microevolutionary processes occurring. More controversially, it has been argued that macroevolution includes rapid and large-scale changes of form that necessitate steps that might initially produce organisms that are less successful than their ancestors. This is completely counter to Darwinian ideas of evolution. An example of these issues can be presented via a consideration of the evolution of major groups of tetrapods.

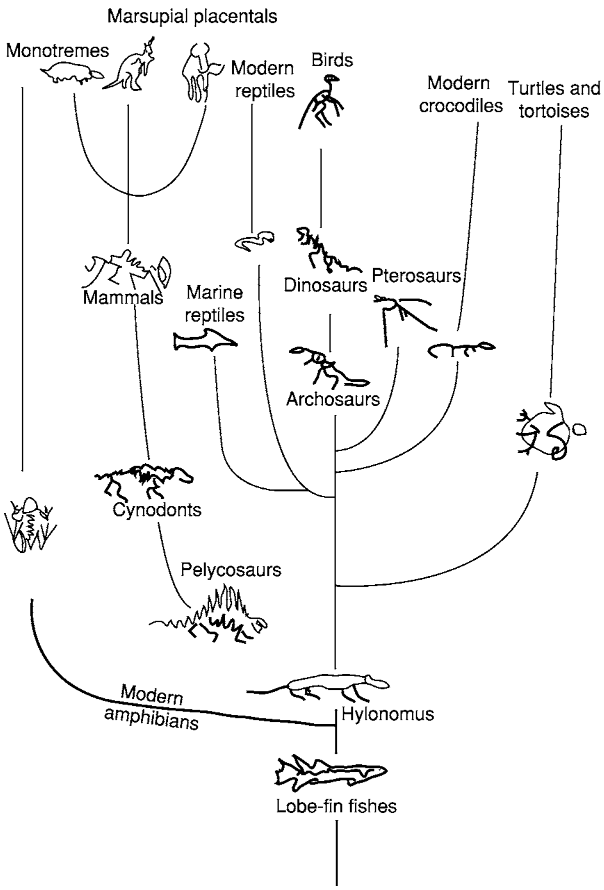


Figure 4 A simplified evolutionary tree for tetrapods, those vertebrates with pentadactyl limbs. This group of organisms evolved from lobe finned fishes during the Devonian. Whilst the evolution of mammals from cynodonts was a gradual process in which macroevolution appears to conform to microevolutionary expectations, the origin of wings in pterosaurs and birds is more difficult to explain without evoking some new process unique to evolution at this scale.

All living vertebrates with pentadactyl limbs (that is mammals, reptiles, amphibians, and birds) evolved from an ancestral fish, with the process beginning in freshwater lakes and rivers in the Devonian (Figure 4). Since then a wide variety of adaptations have appeared in these higher groups, such as feathers, fur, and wings. It can be convincingly demonstrated that some lineages, or evolving lines, acquired these characters gradually, by microevolutionary processes. The classic example of this is the origin of mammals from reptile ancestors through the Triassic and Early Jurassic. Character change occurred at a relatively constant rate throughout this 100 Ma period, and intermediate forms are well known in the fossil record. However, the process by which these new characters appeared may have controls that are not seen in microevolution and which are hinted at by the suggestion that, during this event, significant evolution tended to occur in small-carnivore groups. More controversially it has been argued that some new characters, for example wings, could not have evolved by gradual steps as they would have been useless in their early stages of development. Complicated counter arguments that invoke the possible uses of wings without flight, for example, have not cleared up the controversy.

Looking at the history of life, it is clear that there have been periods of major increases in diversity and periods of major innovation. Significant increases in diversity tend to happen after mass extinctions and are called evolutionary radiations (*see* **Biological Radiations and Speciation**). Empty niches created by the extinction event are quickly filled as organisms radiate to form new species that are able to exploit the available resources. Evolutionary radiations may also be facilitated by the appearance of major innovations, such as the evolution of hard parts by a variety of different taxa close to the Precambrian–Cambrian boundary. The two are not necessarily coupled. For example, eukaryotes, the complicated internally divided cells with which we are most familiar, evolved more than 2 Ga ago (possibly much more), but they did not become widespread or common until much later at around 1 Ga ago. They could not radiate until there was adequate oxygen present in the Earth's atmosphere and oceans, as they depended on this molecule for respiration.

It seems likely that evolutionary selection can work at the species level as well as at the level of an individual within a population. Species-level selection favours species that have lower extinction rates and higher origination rates than their 'competitors'. In the long-term, these species become more abundant at the expense of their less-successful competitors. Distinguishing between the levels of selection is extremely difficult in practice, but the theory helps to demonstrate the emergent properties of species.

Some types of extinction, or reductions in diversity, may also be explained by macroevolution and, in turn, throw light on the mechanisms of evolution. It is clear that the evolution of a new species will increase competition for resources and may force another species to become extinct if it is unable to compete successfully. The pattern of species extinction would be expected to be one of increasing chance of extinction with species age, but in some cases this does not seem to be so. Instead, species age does not appear to correlate with the likelihood of extinction. Van Valen has used this observation (which is itself somewhat contentious) to suggest a novel hypothesis for macroevolutionary patterns. He suggested that competition for resources produces a dynamic equilibrium between species, in which each will continue to evolve in order to survive. This is the core of the Red Queen hypothesis, which suggests that organisms evolve to keep their biological place or, to paraphrase the quotation from Alice in *Alice Through the Looking Glass*, they 'run to keep still'.

The characters that help organisms to survive at times of low extinction rate may be different from those that make survival of mass extinctions more likely. In other words, the criteria by which species are selected may vary with extinction rate. Specialist species tend to have greater survival potential at times when extinction rates are low and reduced survival potential when extinction rates are high. In addition, it has been suggested that small species have a greater chance of surviving mass extinctions than larger species, though the overall trend in evolution is clearly not towards smaller species.

The level of understanding of genetics is now so great that it is possible to explore macroevolution in this way. In traditional views of macroevolution, a set of ways in which different forms could be produced with small changes in the genome was known as heterochrony. The idea was that different parts of the body grew at different rates. In some examples, this might be a difference in the rate at which sexual maturity was reached relative to the rate at which the rest of the organism (the somatic portion) developed. If sexual reproduction became possible at an earlier stage in body development, this was known as pedomorphosis. The classic living example of this is the axolotl. This resembles a juvenile salamander, complete with external gills, but reproduces at this stage of development. If it is injected with extract from the thyroid gland, an axolotl will develop into an adult salamander. A genetic view of this kind of evolution is that there has been a change in the regulatory genes that switch on and off the protein-coding gene sequences within cells. If these genes start to operate at new rates, then the phenotype will change shape, in some cases dramatically.

It is now known that some genes, especially a group known as *Hox* genes, control development by instructing the different parts of the growing embryo on which part of the body should be built. It is known that these genes are more common in vertebrates than in other groups of animals and that there was a single period when these genes duplicated (or rather, duplicated twice), so that vertebrates carry four times as many of these genes as do invertebrates. This multiplication occurred between the evolution of the cephalochordates and proper vertebrates, probably during the Cambrian period. It is tempting to assume that this evolutionary event facilitated the increase in complexity needed to produce vertebrates and may have made them more 'evolvable' since. Whether or not cause and effect can be proved in this example, it points to a growing understanding of the relationship between genes and macroevolution.

See Also

Biodiversity. **Biological Radiations and Speciation**. **Famous Geologists:** Darwin; Lyell. **Fossil Invertebrates:** Trilobites. **Origin of Life**. **Palaeozoic:** Cambrian. **Precambrian:** Eukaryote Fossils.

Further Reading

Darwin C (1859) *On the Origin of Species*. Penguin Books (edited by J W Burrow).

Eldredge N and Gould SJ (1972) Punctuated equilibria: an alternative to phyletic gradualism. In: Schopf TJ (ed.)

Models in Paleobiology, pp. 82 115. San Francisco: Freeman, Cooper.

Gingerich PD (1985) Species in the fossil record: concepts, trends and transitions. *Paleobiology* 11: 27 41.

Greenwood PH (1974) *Cichlid Fishes of Lake Victoria, East Africa: The Biology and Evolution of a Fish Flock*. London: The British Museum (Natural History).

Johnson ALA and Lennon CD (1990) Evolution of gryphaeate oysters in the Mid Jurassic of Western Europe. *Palaeontology* 33: 453 485.

Ridley M (1996) *Evolution.* Oxford: Blackwell.

Sheldon PR (1987) Parallel gradualistic evolution of Ordovician trilobites. *Nature* 330: 561 563.

Skelton PW (ed.) (1993) *Evolution: A Biological and Palaeontological Approach*. Wokingham: Addison Wesley Publishing Company.

Van Valen L (1973) A new evolutionary law. *Evolutionary Theory* 1: 1 30.

Williamson PG (1981) Palaeontological documentation of speciation in Cenozoic molluscs from Turkana Basin. *Nature* 293: 437 443.

FAKE FOSSILS

D M Martill, University of Portsmouth, Portsmouth, UK

Introduction

This article considers fake fossils and the part that forgers have played in ruining scientific reputations and hindering the development of science. In some cases, financial reasons appear to have been a motive for forgery, but the impact has, nonetheless, been detrimental to science. This article is written from a certain amount of experience, the author having fallen foul of at least one fossil fabrication. Some tips are provided for those who might encounter fake fossils.

Forged fossils fall into a number of categories. Some are complete fabrications and should be considered as sculptures. They may be cast from materials that resemble rock, such as plaster or cement, or carved from real rocks. Some forgeries represent embellishments of genuine fossils, and include the addition of elements from another fossil simply to make an incomplete specimen appear more complete. Other forgeries are conversions whereby a common fossil is transformed to resemble something much rarer; others are chimeras whereby two or more fossils of different animals are united to produce quite fabulous creatures. Some composites are not manufactured deliberately to deceive; rather, many simply represent attempts to fill gaps for aesthetic purposes and to make museum displays more informative. In this latter case, no deceit is intended, but when past curators have failed to keep records of which fossils were amalgamated, taxonomic problems have arisen several years later.

In some unusual cases of forgery, remains of modern animals and plants are transmogrified into fossils by being embedded in resins or by being glued onto bedding planes (Figure 1). Deciding what constitutes a fossil forgery can be difficult. Purists might argue that any modification of a fossil represents an act of forgery, although a museum display specimen might be enhanced simply to demonstrate what a skeleton may have looked like when complete, or a damaged piece might be skilfully repaired to obscure an ugly scar or hole, perhaps caused by bad collecting practice. Certainly, Victorian museum curators thought it perfectly acceptable to construct a complete skeleton from the remains of a number of partial skeletons. One of the most famous examples includes the mounted skeleton of the giant sauropod dinosaur *Brachiosaurus brancai* that forms the centrepiece to the Humboldt Museum in Berlin. This magnificent skeleton is thought to contain the parts of at least five different individual fossils.

Cruel Hoaxes

Fake fossils represent deliberate attempts by the unscrupulous to hoodwink the unsuspecting into believing that an object is a genuine fossil. Such is human nature that as long as fossils have a financial value or can result in prestige for the discoverer or describer, then there are going to be disreputable people prepared to exploit this for their own ends, be they greed, spite, or self-betterment. This is not a new phenomenon, and has been a practice from the earliest days of palaeontology. Some faking of fossils is indeed a consequence of criminal intent to obtain money through deception, but in a number of cases, fossils have been faked in what appear to have been either jokes that have gone seriously wrong or deliberate attempts to ruin scientific reputations.

Such is the case of the now famous lying stones of Eibelstadt, near Wurzburg, Germany. This is one of the oldest, well-documented cases of fossil forgery,

Figure 1 In this crude attempt to forge a fossil, a recently dead dragonfly has been glued to the surface of a piece of limestone. Such forgeries at first can appear to be examples of excellent preservation. Be alert if a thin veneer of varnish prevents direct access to the surface of the fossil.

and concerns a certain Dr Johann Beringer. Apparently, Beringer was an extremely pompous fellow, and was despised by a number of local academics. The academics generated an enormous number of crude forgeries that they passed to Beringer via hired helpers. Despite the crude nature of the fossils, and even despite later attempts by the forgers to reveal their cruel act, Beringer published a scientific account of the fossils in 1726. The book, *Lithographiae Wirceburgensis*, figured hundreds of the faked specimens, which included sculptures of spiders in their webs, frogs, birds, and even shooting stars and moons. The flagrant act of forgery came to light just before publication, but the book includes a note dismissing the claims of forgery, such was Beringer's arrogance.

Perhaps the most famous example of deception is the case of the Piltdown Man forgeries. Many books have been written detailing this hoax and speculating on the identity of the perpetrator. Essentially, the lower jaw of an orang-utan was substantially altered and buried in a gravel pit along with some fragments of human cranium in Sussex. The remains were discovered by Charles Dawson in 1912 and then described by leading vertebrate palaeontologist Dr Arthur Smith Woodward, who was, at the time, the Keeper of the British Museum of Natural History, London. Although several observers had wondered if the fossil was a forgery, it was not definitively shown to be so until 1953, as a result of a fluorine analysis on the jawbone. This was a sad postscript to the scientific career of Sir Arthur, who had been a brilliant palaeontologist. The hoax was a scandal for British science, and it held back palaeoanthropology for several decades.

At least the Piltdown forgery concerned genuine organic remains and Arthur Smith Woodward could perhaps be forgiven for making a mistake; after all, someone had set out deliberately to deceive. In 1966, at the age of 91, noted German vertebrate palaeontologist Professor Freiderick von Huene, of Tübingen University, described a juvenile skeleton of the ichthyosaur *Leptonectes* (then called *Leptopterygius*) that had been made from cement, stained brown, and placed on a slab of rock from the Early Jurassic Posidonia Shale Formation. Not a single fossil bone was present on the specimen; it was nothing more than a sculpture, and not a very accurate one at that. Proof of this forgery came to light only when the sculpture was being cleaned by a preparator some 4 years later, and it was not revealed to the scientific world until 1976. Huene never had to face the embarrassment of this exposé because he died in 1969, and in this particular case, the published paper of Huene was not of great scientific consequence. This is in marked contrast to the paper of Arthur Smith Woodward, on the Piltdown 'fossils', which announced the presence of the oldest hominid fossils in Europe and purported to show that large human brains were an early evolutionary development. But there are some similarities in the two hoaxes. In both cases, the scientists concerned were extremely eminent and had enjoyed careers in which they had risen to the very top of their profession. It would be no surprise to learn that they had made enemies on the way up, and that some embittered rascal had sought cruel revenge. These, fortunately we hope, are rare cases.

Too Much Haste

A more recent (November 1999) case of fossil forgery resulted in considerable embarrassment for North American palaeontologist Philip J Currie, artist Stephen Czerkas, and especially for the senior assistant editor of *National Geographic Magazine*, Chris Sloan. This sorry story concerned a strange case whereby two spectacular, and quite genuine fossils, were merged together to construct a chimera comprising the back end of a small dinosaur, *Microraptor zhaoianus*, and the front end of a small fossil bird, *Yanornis martini*, both from the famous Early Cretaceous Yixian Formation of Liaoning Province, China. The two incomplete specimens were joined together to make a single, complete feathered dinosaur. Unfortunately, so much excitement was generated over the specimen that the *National Geographic Magazine* printed an article on its discovery and its perceived relevance to the 'birds are dinosaurs debate' just before the specimen was shown to be a forgery. An even more unfortunate aspect of this case occurred because, unusually for an article in the *National Geographic*, the fossil chimera was given a scientific name, *Archaeoraptor liaoningensis*, which, according to the rules of scientific nomenclature, was valid for at least part of the specimen. Paradoxically, it turned out that both halves of the chimera represented important scientific discoveries, and both were new to science. Suspicions surrounding the nature of the fossil came to light when the specimen was scanned using computed axial tomography (CAT), and it became clear that the pieces did not fit together well. A more careful examination then revealed the forgery, and although Phil Currie highlighted some problems with the fossil, these were not relayed to *National Geographic*. It was only when Chinese palaeontologist Xu Xing had met with a Chinese fossil dealer that the sorry story of the forging really emerged. But by then it was too late; the article had already appeared in the November 1999 issue and the proof of the forgery came one month later. It is to the relief of many palaeontologists that this forgery was discovered before too much damage had been done to the scientific case being made for the bird–dinosaur hypothesis, but, sadly, the furore over the forgery has distracted from the undoubted importance of the two genuine fossils.

Not all fossil forging has serious consequences for science, and indeed, some forging is undertaken in an attempt to 'improve' fossils for the fossil-buying public. Such forgeries usually involve adding embellishments to genuine fossils, or converting fossils from one type to another (Figure 2). Such embellishments do not enhance the value of the fossil; indeed, they destroy the scientific value, but they might make a fossil look attractive to the unwary purchaser. This type of forgery is common among the fossil dealers of Brazil who raid the spectacular fossil fish beds of the Santana and Crato formations of north-eastern Brazil. Here it is common to find forged fishes that have heads and tails belonging to different species. Fins may be added, and some specimens might be artificially lengthened by the insertion of several bodies into one example. Conversions are common, and it is frequent to find heads of large fossil fishes converted into frogs, and small specimens of the gonorhynchiform fish *Dastilbe* converted into lizards. To the unwary, the presence of some genuine bones is enough to encourage belief that the entire fossil is genuine.

Until recently, most of the fossils available commercially from the fossil beds of Brazil were collected to supply flea markets in the tourist centres of Brazil. The fossils were often enhanced to make them visually more attractive to tourists who probably knew very little about fossils, but who wanted to have an unusual souvenir of Brazil (Figure 3). More recently, the genuine fossils have become highly sought after by museums, and many of the rarer fossils from Brazil, such as pterosaurs and dinosaurs, command very high prices. There has thus arisen a new financial incentive for the forging of fossils. Previously, forgeries were rapidly constructed, using a sharp chisel, by the addition of a crudely engraved outline of a fish. Now, elaborate constructions are made by glueing together numerous pieces of real fossils to produce such things as pterosaur bones and crocodile heads. Much time and effort goes into these constructions, but by and large they remain crude and are easy to recognise. However, in a skull of a dinosaur that had been obtained by the Museum für Naturkunde, Stuttgart, Germany, a sagittal crest at the back of the skull was revealed to be fake only after CAT scanning. The crest, in fact, was a part of the lower jaw repositioned to make the specimen look more spectacular. There was no need for the forger to have executed this

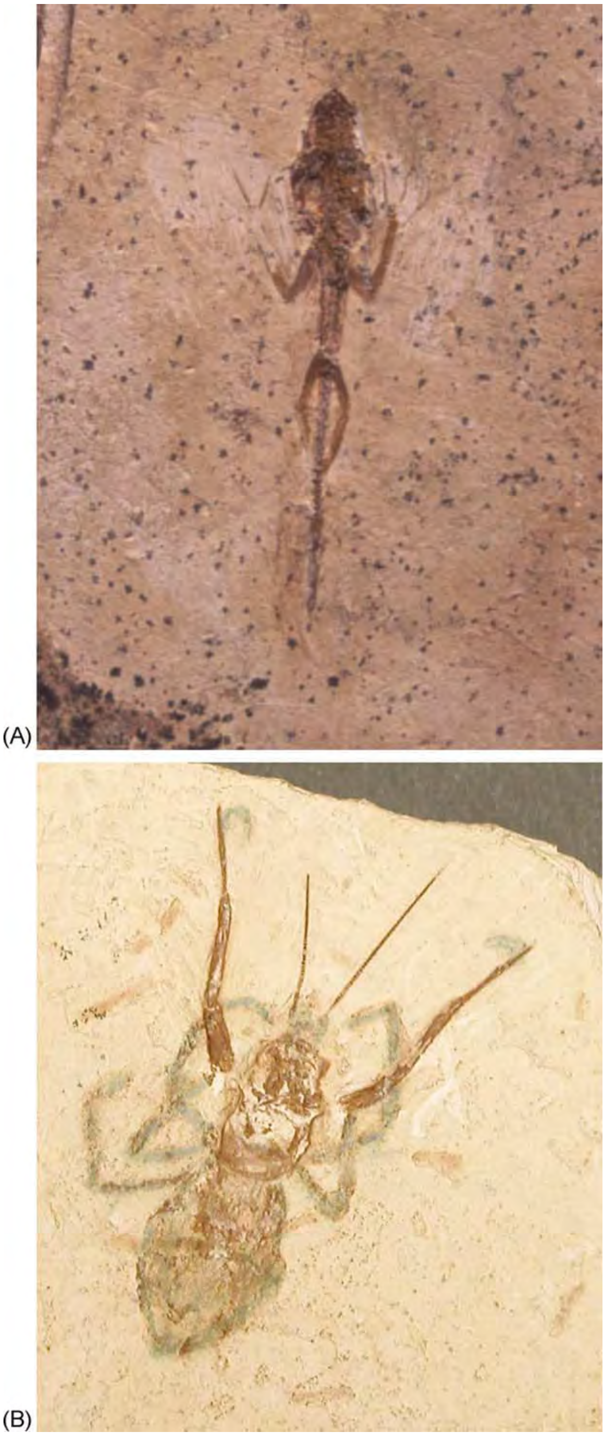


Figure 2 Examples of genuine fossils that have been altered to resemble something rare. Both are from the Nova Olinda Member of the Crato Formation, Ceará, Brazil. (A) The fossil gonorhynchiform fish *Dastilbe* has had limbs added to make it look like a lizard. (B) A fossil insect has had extra legs and claws added in ink to make it resemble something new, perhaps a spider. This fake was revealed by dropping industrial methylated spirit onto the fossil. The ink of the faked legs bled into the rock, whereas the real limbs remained intact. Both photographs by Robert Loveridge.

Figure 3 A forged composite group of fossil fishes. The fishes here are all exceptionally well preserved, but were not found in this association. They are from the famous Santana Formation, Ceará, Brazil. They have been glued together with a mix of car body filler and ground rock.

embellishment, because the skull represented a new genus and species of dinosaur and represented the most complete skull of a spinosaurid dinosaur ever found.

Amber

Amber has long been famous for its fossil inclusions, and has been used in the jewellery industry with or without fossil inclusions for thousands of years. New discoveries of amber in the Dominican Republic have resulted in a large number of forgeries. A majority of these are offered to unsuspecting tourists. Most are sold cheaply, but a number of higher priced specimens containing lizards and frogs have proved to be cleverly executed forgeries. It is not always easy to distinguish forged amber from the real thing. The hot needle test, whereby a red-hot needle is pressed into the specimen, will give off a resinous smell if the specimen is genuine amber, whereas the smell will be acrid if it is a synthetic resin; however, the test inevitably marks the specimen.

Religious Zealots

There have been several attempts by those creationists (*see* **Creationism**) who appear to feel threatened by palaeontological and geological evidence that runs counter to biblical interpretations that Earth is not terribly old and was created in a very short span of time. Rather than accept the findings of science, some supporters of a biblically based creation theory have challenged the data on which certain scientific claims are made by attempting to discredit palaeontology. Attempts to do this using logical argument have proved difficult, and so some unscrupulous individuals have attempted to undermine scientific findings by forging data. Perhaps the most notable attempt was the claim that human footprints occurred alongside those of dinosaurs at the Paluxy River site in Texas, USA. The Paluxy River site is famous for lengthy trackways of footprints of Cretaceous dinosaurs and has been made into a National Park. Reports that human footprints had been found in the same layers as the dinosaur footprints had always been treated with scepticism by the palaeontological community. Wrapped in pseudoscientific jargon, photographs of the human footprints side by side with dinosaur footprints were used as 'evidence' that humans and dinosaurs were around at the same time, and it would therefore have followed that dinosaurs could not have been millions of years old. Despite considerable protestations by scientists, it was only later that the perpetrators of the forgeries admitted that the human prints were handmade rather than footmade.

When a Fossil is Not a Forgery

One of the most important fossils, historically, is the London specimen of the small feathered bird/dinosaur *Archaeopteryx lithographica*. This fossil was widely hailed as a missing link between reptiles and birds because its exceptional preservation in the fine-grained lithographic limestones of Bavaria showed it to have a dinosaurian skeleton that included a long tail with numerous vertebrae and a beak in which the jaws possessed teeth. And yet the animal was clothed in feathers, a feature known today only in birds. This was just what devotees of Darwin needed to support the theory of evolution, and indeed *Archaeopteryx* became the archetypal missing link; it is an animal that appears to be transitional between two groups of animals, a status that is still claimed for it today. During the 1980s, two eminent scientists, but not palaeontologists or geologists, Sir Fred Hoyle and Chandra Wickramasinghe, published a claim that the London specimen of *Archaeopteryx* was a forgery. If their claim had been correct, there is no doubt it would have had important implications, but such an upheld claim would have had even more dire consequences had it been made in Victorian London. However, several independent analyses of the evidence for forgery showed quite categorically that Hoyle and Wickramasinghe were out of their depth and did not understand the nature and diversity of fossilization processes. Nevertheless, the claim, coming as it did from such noted scientists, generated considerable excitement in the media, and a number of books and papers resulted from the claim. Sadly, many palaeontologists had to devote considerable time and effort to debunk these incorrect claims.

Detecting Forgeries

It is advisable always to be suspicious of fossils bought commercially and to be very sceptical of any exceptional fossil that is provided by a ‘friendly’ noncolleague. Fossils traded commercially are quite likely to have been enhanced in order to increase their aesthetic appeal, but such improvements are usually easily detected by experienced palaeontologists. Some traders of ammonites increase the size of the ammonite by carving extra whorls into the rock.

Fabricated parts of fossils are often constructed using plastic-based fillers. These plastics will melt when probed with a hot needle, giving off an acrid smoke. Rock does not do this. Many of the spectacular trilobites from the Ordovician and Devonian of Morocco have been enhanced with fillers, and indeed some are simply casts made from moulds of genuine fossils. The casts are glued to blocks of limestone and coloured black with boot polish.Where there has been a real intent to deceive, the workmanship of the forgery is often very good and almost impossible to detect by casual inspection. Examination under a microscope may help, but when suspicions are raised, proof may come only after expensive CAT scanning or chemical analysis.

See Also

Creationism.

Further Reading

Charig AJ, Greenaway F, Milner AC, Walker CA, and Whybrow PJ (1986) *Archaeopteryx* is not a forgery. *Science* 232: 622 626.

Hoyle F and Wickramasinghe C (1986) *Archaeopteryx: The Primordial Bird.* Swansea: Christopher Davies.

Martill D (1994) Fake fossils from Brazil. *Geology Today* 1994: 36 40.

Nield T (1986) The lying stones of Eibelstadt. *Geology Today* 1986: 78 82.

Ross A (1998) *Amber: The Natural Time Capsule.* London: The Natural History Museum.

Russell M (2003) *Piltdown Man: The Secret Life of Charles Dawson.* Stroud: Tempus Publishing.

Sloan CP (1999) Feathers for *T. rex*. *National Geographic* 196(5): 98 107.

Simons LM (2000) *Archaeoraptor* fossil trail. *National Geographic* 197: 128 132.

Suess H D, Frey E, Martill D, and Scott D (2002) *Irritator challengeri*, a spinosaurid (Dinosauria: Theropoda) from the Lower Cretaceous of Brazil. *Journal of Vertebrate Paleontology* 22: 535 547.

Wild R (1976) Eine Ichthyosaurier Fälschung. *Neues Jahrbuch für Geologie und Paläontologie, Monatshefte* 1979: 382 384.

FAMOUS GEOLOGISTS

Contents

Agassiz
Cuvier
Darwin
Du Toit
Hall
Hutton
Lyell
Murchison
Sedgwick
Smith
Steno
Suess
Walther
Wegener

Agassiz

D R Oldroyd, University of New South Wales, Sydney, Australia

Son of a clergyman, Jean Louis Rodolphe Agassiz (Figure 1) was born in the village of Môtier in Canton of Fribourg, Switzerland. After schooling at Bienne and Lausanne and early acquiring an interest in natural history (particularly of fishes), he attended the universities of Zurich, Heidelberg, and Munich, intending to take a medical degree. However, at Heidelberg he began studying palaeontology under Heinrich Bronn and embryology under Friedrich Tiedemann. He also became friends with Alexander Braun and his family (later marrying his sister Cécile) and Karl Schimper, and the three young men went on botanical excursions together. At Munich, he came under the influence of Friedrich Wilhelm Schelling, Lorenz Oken and German *Naturphilosophie*. Additionally, he studied botany under Carl Friedrich von Martius, and embryology under Ignatius Döllinger. A natural philosopher should, as Agassiz represented Oken's view in his autobiography: "[construct] the universe out of his own brain, deducing from *a priori* conceptions all the relations of... living things". In practice, Agassiz certainly did not discount empirical information. Indeed, his hunt for 'facts' was one of his chief priorities. But this search was propelled by the desire to apprehend the activity of the Absolute Being in nature. Indeed, it was strongly influenced by the notion that he was examining the works of a divine Creator.

Agassiz's PhD (Erlangen/Munich) (*Selecta Genera et Species Piscium quos in Itinere per Brasiliam Annis MDCCCXVIII–MDCCCXX* [1829]) was devoted to the Brazilian fishes collected from Amazonia by the recently deceased Johann Baptist von Spix. It was tactfully dedicated to Georges Cuvier (*see* **Famous Geologists:** Cuvier). Agassiz also obtained an MD at Munich in 1830, but by then he was determined to be a naturalist not a physician. The following year Agassiz went to Paris to study comparative anatomy under Cuvier, having already examined numerous collections of fossil fish in leading museums. Cuvier was greatly impressed by the young man's work and took him under his wing, introducing him to Alexander von Humboldt, and teaching him the principles of comparative anatomy and how to reconstruct fossil fish. So Agassiz gave up the German idea of the unity of the animal kingdom and followed Cuvier's notion of there being four fundamental types in the animal kingdom. Cuvier was so impressed by Agassiz's abilities that he passed on the notes, drawings, and specimens that he had collected on fossil fish for him to study. He also ensured that other institutions made their collections available to Agassiz.

Figure 1 Louis Agassiz (1807 1873).

Cuvier died of cholera in 1832, but his influence on Agassiz was strong and permanent, particularly respecting the idea of successive geological catastrophes and the creation of new species. Agassiz's studies of fossil fish eventually yielded his great treatise *Recherches sur les Poissons Fossiles* (5 vols, 1833–1843), with the figures mostly drawn by the artist Joseph Dinkel (whom he employed over a long period); and *Monographie des Poissons du Vieux Grès Rouge* [*Old Red Sandstone*] *ou Système Dévonien des Iles Britanniques et de Russie* (1844–1845). Agassiz received the Geological Society's Wollaston Medal for his ichthyological work in 1836.

At a youthful 25 years of age, Agassiz was appointed Professor of Natural History at the small new Lyceum or Academy at Neuchâtel, back in his home region of Switzerland, and soon began to establish that institution's reputation. His early *magnum opus* made use of specimens sent to him from all over Europe, and in particular from the Old Red Sandstone of Scotland, to which country he made two visits. The later association with the amateur stonemason Hugh Miller, who arranged for Agassiz to receive specimens of Devonian fossil fish, is particularly well known through Miller's popular book *The Old Red Sandstone* (1841), and his contributions were incorporated into Agassiz's work on Devonian ichthyology.

Unfortunately, Agassiz's first marriage to Cécile Braun failed, in part because he gave so much attention to his work and partly because he came under the sway of his assistant, the geologist Edouard Desor, who pushed his way into the Agassiz household despite Cécile's objections. Moreover, Agassiz's ambitious publishing projects led to financial problems and life became difficult for him in Neuchâtel. He therefore sought the assistance of von Humboldt and Charles Lyell (*see* **Famous Geologists:** Lyell) to travel to North America, and in 1846 he went to Boston at the invitation of James Avory Lowell to give a lecture series on natural history. These were outstandingly successful, and led to his appointment as Professor of Zoology and Geology at Harvard in 1848, where he soon became one of the country's leading scientists. In 1852, he was additionally professor at the Medical School at Charleston, South Carolina, and also at Cornell University in 1868.

Declining a chair in Paris, despite the offer of most favourable terms, Agassiz committed himself to American science, pushing, with the help of endowments from Francis Calley Gray and others, for the foundation the famous Museum of Comparative Zoology at Harvard in 1858–1859 (which opened in 1860). In 1863, he helped persuade Abraham Lincoln to establish the National Academy of Sciences; the same year Agassiz was appointed a regent of the Smithsonian Institution. Agassiz had reached the top of the tree.

Subsequently, Agassiz travelled widely on both land and sea and wrote numerous scientific papers in the USA, as well as popular essays, reviews, and educational works, his writings on classification being the most influential. However, while revelling in the hospitality and opportunities that America offered, he retained a belief in the superiority of European science and culture, which later alienated some colleagues.

Left behind in Europe, Agassiz's wife had died of tuberculosis in 1848. His son joined him in America, and subsequently his two daughters. In 1850 he married Elizabeth Cary, who later founded a girls' school that later developed into Radcliffe College at Harvard. A number of Agassiz's European *epigone* followed him to America, including Desor, with whom Agassiz eventually fell out, after an unpleasant episode involving accusations of plagiarism, financial malfeasance, and worse, for which Agassiz was found to be without fault. His first wife's intuitions were more than vindicated. Desor withdrew to Europe.

Apart from collecting, naming, and describing modern and fossil fish, Agassiz also proposed a scheme for fish classification, based on their scales. This was not 'biologically' ideal, but suited the study of fossil fish, for which in many cases the scales are the best preserved remains, the bones having been cartilaginous. Thus four main orders of fish were proposed, based on their scales, rather than their crania:

1. Having plate-like scales, often with tubercles or bony points, detached from one another and irregularly arranged on a tough skin (Placoid).
2. Having large, bony, usually shiny (enameled) plate-like scales, not normally overlapping, but often interlocking in some way (Ganoid).
3. Having thin, horny, overlapping plates, each having one side with a jagged edge or comb-like projections (Ctenoid).
4. Having thin, flexible, overlapping, horny scales, smooth in outline and circular or elliptical in form (Cycloid).

For the Ganoids, Agassiz was especially interested in the modern *Lepisosteus*, which was the sole surviving modern representative of the group. So, like Cuvier (who worked on elephants, mammoths and mastodons), he specifically sought to compare living and extinct types.

Agassiz's taxonomic system was later superseded by various others, based principally on bones rather than scales, though his introduction of the Ganoidei was a substantial contribution. His taxonomy was problematic, for, while comparing fossilized and modern forms, he saw no evolutionary connection between them. On the other hand, he offered something new by the use of fossil fish for stratigraphic purposes. Moreover, in a manner that would have appealed to Cuvier, he sought to find out something about the 'conditions of existence' of his specimens as to temperature, salinity, and mode of locomotion. He supposed that prior to the Cretaceous there seemed to be less distinction between fresh-water and marine forms than at present and it might be the case that these two environments were not so marked previously as they are at present.

However, Agassiz's most important contribution to geology was his advocacy of the concept of an Ice-Age (*Eiszeit*), fundamental to Pleistocene geomorphology and stratigraphy (*see* **Tertiary To Present:** Pleistocene and The Ice Age). Curiously, it was linked to the biological ideas that he imbibed from Cuvier.

In Switzerland, the idea that the country's glaciers were formerly of greater extent had been recognized by observers back in the eighteenth century, such as the minister Bernard Friedrich Kuhn (1787). There is a report of a manuscript by a mountaineer Jean-Pierre Perraudin (1818), which described the extent of moraines and erratic boulders, and regarded striated and polished rocks as evidence of glacial action. It was perhaps Perraudin who really initiated the glacial theory in Switzerland. The highway engineer Ignaz Venetz accepted Perraudin's ideas and read a paper on the topic at Neuchâtel in 1829. The mining engineer Jean de Charpentier, director of the salt mines at Bex, also obtained information from Perraudin and in 1834 read a paper at Lucerne about the former greater extent of glaciers. (Agassiz met Charpentier when he was still at school and was partly inspired by him to become a naturalist.) However, Charpentier's paper was regarded as mistaken and was mocked, Agassiz being one of the opponents. (Historians examining Agassiz's students' lecture notes from that period have shown that he was then critical of the theory.)

But in 1836 Agassiz was in the Bex area and was shown around by Charpentier, and after calling on Venetz and examining the evidences in other parts of Switzerland he became a convert to the theory. While in Bex, Agassiz met his old student friend, the botanist and palaeontologist Karl Schimper, and the two also discussed the glacial evidence. In February 1837, Schimper gave a botanical talk at Neuchâtel, at the conclusion of which he passed round a copy of a poem that introduced the new word *Eiszeit*. By then, Agassiz had picked up the evidences and ideas in their entirety and was running with them. He presented a first outline of his views in public at the meeting of the *Société Helvétique des Sciences Naturelles* at Neuchâtel in July 1837, in what became known as the *Discours de Neuchâtel*. By 1840 Agassiz published his major study on the topic, and his most important contribution to geology: *Études sur les Glaciers*. In publishing this, he got ahead of Charpentier's *Essai sur les Glaciers* (1841), and recriminations followed, stirred up, it has been suggested, by Desor. Schimper was also annoyed with Agassiz for failing to mention him in *Études* (though he was mentioned in the *Discours*). There followed a further work on glaciers co-authored with Arnold Guyot and Desor, describing the different types of glaciers, their component parts, their motions, and a detailed account of the Aar Glacier: *Système Glaciaire: Ou Recherches sur les Glaciers, leur Mécanisme, leur Ancienne Extension et le Rôle qu'ils ont Joué dans l'Histoire de la Terre* (1847).

The *Discours* was written in haste, but provided persuasive evidence for the former extension of glaciers, at least in Switzerland, and strong arguments against the floating iceberg theory favoured by Lyell in Britain, or the common idea of glacial erratics being emplaced by catastrophic floods. On the other hand, Agassiz thought that erratic boulders might have fallen into their present positions rather than being directly transported by ice. Agassiz's *Études* was a sumptuous volume, beautifully illustrated, providing all the documentation necessary to convince readers of the former extension of glaciers. The theory could also explain the existence of the vast extent of superficial deposits ('till') over northern Europe, then

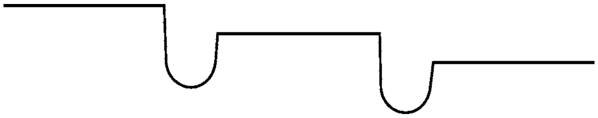


Figure 2 Sketch in Agassiz's *Discours de Neuchâtel* showing hypothetical fall of global temperature over time.

known as 'diluvium' (by association with the notion of catastrophic floods) or 'drift' (by association with the floating iceberg theory).

However, the explanatory theory advanced by Agassiz was much less persuasive than that which it was supposed to explain. It was widely held at the time (in accordance with the views of Élie de Beaumont, e.g.) that the earth was cooling. Agassiz arbitrarily assumed that it did so in a fashion indicated by his sketch (Figure 2), which appeared in both the *Discours* and the *Études*.

This graph was supposed to represent not just a cooling inorganic planet, but one inhabited by living organisms that were wiped out, however, from time to time by 'Cuvierian' catastrophes, and then replaced by different sets of organisms. The abrupt falls of temperature corresponded to the sudden disappearance of life forms, with the temperatures remaining approximately constant so long as the (supposedly heat sustaining) life forms continued in existence. Agassiz supposed that the formation of the Alps themselves was an event of recent occurrence, was preceded by a 'catastrophic' fall in temperature, and was then followed by the establishment of modern forms of life. On this view, then, the epoch preceding the present could have been of extreme cold, producing the former extended glaciation evidenced in the Alps. The onset of cold must have been sudden, from the appearance of mammoth remains in Russia. Agassiz suggested that the glaciation could have extended from the North Pole right down to the Mediterranean and Caspian seas. Thus the Great Ice-Age.

This theory was perhaps the most 'catastrophist' ever propounded by a 'respectable' geologist (other than bolide aficionados). Agassiz wrote (1838: 382):

> [T]he epoch of extreme cold which preceded the present creation... was attended by the disappearance of the animals of the diluvian epoch of geologists, as the mammoths of Siberia still attest, and preceded the uprising of the Alps, and the appearance of the animated nature of our day, as is proved by the moraines, and the presence of fish in our lakes. There was thus a complete separation between the existing creation and those which have preceded it; and, if the living species sometimes resemble in our apprehension those which are hid in the bowels of the earth, it nevertheless cannot be affirmed that they have regularly descended from them in primogeniture, or, what is the same thing, that they are identical species.

Thus Agassiz set his face against transformism or evolution and offered hyper-catastrophism and the doctrine of special creations (assuming but not then stating) that they occurred by some divine means.

Agassiz visited Britain in 1834 and 1835, chiefly in connection with his interests in fossil fish, but he also made the acquaintance of the 'diluvialist' William Buckland, who in turn visited Agassiz in Switzerland in 1838. Buckland had long been interested in the drift deposits, which he earlier has ascribed to the Noachian Flood, and introduced the distinction between 'diluvium' and 'alluvium'. He was, however, converted to Agassiz new theory during the course of his 1838 visit, realizing that features of British geology that had long puzzled him could be successfully explained in terms of the land-ice theory. In 1840, Agassiz attended meetings of the Geological Society in London and the British Association in Glasgow, and presented his glacial theory, prompting much discussion in British geological circles. However, the theory, as presented in Glasgow, tried to reconcile the new doctrine with the older idea of glacial submergence, for after the melting of the glaciers flood waters could have moved boulders and gravels (thus accounting for glaciofluvial materials).

Following the meeting, Agassiz and Buckland went on a tour of Scotland, and were satisfied that they could see most satisfactory evidence in favour of the land-ice theory, and successfully interpreted the 'Parallel Roads of Glen Roy', which Darwin had the previous year interpreted as marine shore-lines, as being due to the successive shore lines of an ice-dammed lake, an interpretation that was rapidly published in the newspaper *The Scotsman*. Following his Scottish tour, Agassiz proceeded to Ireland, where again he found ample evidence for glaciation. Returning to Scotland, he then journeyed back to London, seeing many more evidences of glaciation, and spoke at the Geological Society.

Debates about the land-ice theory rumbled on in Britain for the next quarter century. Lyell was initially converted to Agassiz's ideas, but most other influential geologists such as Roderick Murchison (*see* **Famous Geologists:** Murchison) were not. Not long after Agassiz returned to Switzerland, Lyell recanted: "he found the proposed departure from present temperature conditions too much to accept for his uniformitarianism, and he reverted to the glacial submergence theory and floating icebergs". It was not until the 1860s that more general acceptance of the land-ice theory began, with the suggestions of the surveyor Andrew Ramsay as to how glaciers might excavate the basins that are now occupied by mountain lakes; and land ice could have moved uphill to deposit marine shells on the tops of hills in North Wales. Eventually, through the theoretical work of James Croll (1875), an astronomical theory of the origin of climate change was developed, and such theory has been under discussion through to the present.

In America, Agassiz successfully applied his ideas to the interpretation of observations in the Great Lakes area, which he explored in 1848, finding new fish for his examination and ample evidences of glaciation (see his *Lake Superior; Its Physical Character, Vegetation, and Animals, Compared with those of Other and Similar Regions* [1850]), and many other regions. But most of his work in the USA was zoological rather geological. In particular, and in keeping with his long-held Cuvierian views, he was active in his criticism of Darwin's evolutionary theory, which ran counter to Agassiz's long-held beliefs about the special creation of life forms. Agassiz even resisted the idea that different varieties of animals of the same species could be produced through time, from which it followed that the different human races were essentially different species! Thus, he gave 'scientific' comfort to racial bigots. Agassiz could not comprehend how similar but different creatures of the same species could have been produced worldwide. As a special creationist, that seemed to him to be the required alternative to his 'polygenism', and as such had to be rejected. Agassiz's views in fact succeeded in driving Lyell further into the evolutionist camp. Also, because he was opposed to the idea of variation over time, Agassiz was inclined to suppose that every variety of fish he encountered represented a different species. Hence his classification became inordinately unwieldy. It is interesting that a figure, published in 1844, depicting the genealogy of his four main groups of fish, looks quite like a modern evolutionary tree, yet none of the 'branches' are shown as linking at their bases, though they 'lean towards' one another, so to speak, in a way that a later evolutionist might regard as suggestive.

The source of Agassiz's anti-evolutionism can be traced to his contacts with Oken and German *Naturphilosophie*, and associated Platonism (fused with Christian beliefs), as well as Cuvier. Species, for Agassiz, could be regarded as 'types' representing the 'thoughts' of the Creator. Because there could be no substantial natural variation over time, events such as the Ice-Age represented catastrophes of divine origin that also offered the possibility of renewed creative activity. The 'plan' of Creation was, he supposed, better understood by the natural historian than the theologian.

Agassiz was not a great geologist, despite his outstanding capacity for grasping and ordering information, and his powers as a teacher. In 1865–1866, he visited South America, funded by a wealthy Bostonian, Henry Thayer, hoping to find evidences of glaciation in the tropics. Seriously perturbed by Darwin's theory, Agassiz sought new evidence to support of his long-held ideas about catastrophes and the great Ice-Age. He wished to show that the event was of worldwide extent: so it should be possible to find evidence for it in the southern hemisphere, even in Amazonia. In Brazil, he thought he had found the evidence he sought, but he mistook boulders produced by tropical weathering for glacial erratics, and soil produced by weathering was misidentified as glacial till. His co-workers did not all agree, but Agassiz thought he had the experience and expertise to recognize glacial evidence when he saw it (though he admitted he saw no glacial striations).

Agassiz's attempt to extend his Ice-Age to equatorial regions was a failure and provided a classic example of 'theory-laden' observations. On the other hand, his recognition and advocacy of the concept of a glacial epoch and the land-ice theory (even if not original to him) was of fundamental importance, marking the beginning of glaciology and all that followed in the study of Pleistocene geology.

See Also

Creationism. **Evolution**. **Famous Geologists:** Cuvier; Darwin; Lyell; Murchison. **Fossil Vertebrates:** Fish. **History of Geology From 1835 To 1900**. **Tertiary To Present:** Pleistocene and The Ice Age.

Further Reading

Agassiz L (1887) *Geological Sketches*. New York: Houghton, Mifflin & Co.

Agassiz L (1967) *Studies on Glaciers Preceded by the Dis course of Neuchâtel Translated and Edited by Albert V. Carozzi*. New York and London: Hafner Publishing Company. (This volume contains an English translation of Agassiz's *Discours de Neuchâtel*.)

Andrews SM (1982) *The Discovery of Fossil Fishes in Scotland up to 1845 with Checklists of Agassiz's Figured Specimens*. Edinburgh: Royal Scottish Museums.

Brice WB and Figueirôa SFdeM (2001) Charles Hartt, Louis Agassiz, and the controversy over Pleistocene glaciation in Brazil. *History of Science* 39: 161 184.

Carozzi AV (1973) Agassiz's Influence on Geological Thinking in America. *Archives des Sciences Genève* 21: 5 38.

Davies GL (1969) *The Earth in Decay: A History of British Geomorphology 1758 1878*. London: Macdonald Technical and Scientific.

Gaudant J (1980) Louis Agassiz (1807 1873), fondateur de la paléoichthyologie. *Revue d'Histoire des Sciences* 33: 151 162.
Lurie E (1960) *Louis Agassiz: A Life in Science.* Chicago and London: Chicago University Press.
Marcou J (1896) *Life, Letters, and Works of Louis Agassiz,* 2 vols. New York: Macmillan (reprinted Gregg International, 1971).
North FJ (1943) Centenary of the glacial theory. *Proceedings of the Geologists' Association* 54: 1 28.

Cuvier

G Laurent, Brest, France

Introduction

Georges Cuvier was one of the grand masters of zoology in the first third of the nineteenth century. He laid the foundations of vertebrate palaeontology, and his work led to the development of the idea of stratigraphical stages through the work of Alcide d'Orbigny. Cuvier's rivalry with Lamarck, the founder of invertebrate palaeontology, manifested itself in their disputes over the subjects of catastrophism in the history of the Earth and transformism in the history of life.

Biography

Jean-Léopold-Nicolas-Frédéric (called Georges) Cuvier was born on 23 August 1769 in Montbéliard, which at that time belonged to the Duchy of Wurttemberg (Germany) but retained French as its language. At an early age he showed an interest in the study of the natural world. As he came from a Protestant family, his parents intended that he should become a pastor, but he failed his entrance examination to the seminary. Nevertheless, he obtained a scholarship to the Caroline Academy in Stuttgart, where, during the years 1784 to 1788, he received training appropriate for a future official in the service of the Duchy. In accordance with his personal interests, he also attended courses in natural history. He became friendly with Christian Heinrich Pfaff (1772–1852) and more particularly with Karl Friedrich Kielmeyer (1765–1844), who was similarly devoted to zoology and who became Professor of Zoology at the Caroline. It was he who taught Cuvier the art of dissection and gave him his 'first ideas about philosophical anatomy'.

After failing to obtain a post in the bureaucracy at the end of his period of training, Cuvier found employment as a tutor to an aristocratic Protestant family in Normandy, where he spent the years 1788 to 1795, the most disturbed period of the Revolution. He devoted his leisure time to studying botany and the anatomy of animals, particularly molluscs, which he encountered in the neighbouring coastal area. Thanks to his friends Pfaff and Kielmeyer, Cuvier maintained his links with German naturalists.

In April 1795, with the assistance of the physician and agronomist Abbé Alexandre Tessier (1742–1837), a refugee at Fécamp, Cuvier was able to establish himself in Paris. He was well received there, particularly by Etienne Geoffroy Saint-Hilaire, who was already a Professor at the Museum and with whom he became friendly. They collaborated with one another and coauthored some articles. Upon his arrival, Cuvier obtained a teaching position at the newly established college at the Panthéon. The same year he was chosen by Antoine Mertrud to fill a vacancy at the Muséum d'Histoire Naturelle. This marked the beginning of Cuvier's distinguished teaching career, both there and in the university. He was named a Member of the First Class of the Institut de France (subsequently the Académie des Sciences) when it was formed in 1795. In 1800, he was appointed to Jean Daubenton's former chair at the Collège de France. In 1802, when Mertrud died, he became titular Professor of Comparative Anatomy at the Muséum d'Histoire Naturelle. In 1803, he became Permanent Secretary of the First Class of the Institut de France.

Simultaneously, Cuvier pursued an administrative career. In 1802, he was appointed Inspector General of Public Education. In 1808, Napoleon named him Councillor of the University, which he was re-establishing, and in 1810–1811 Cuvier was one of the leading lights in the reform of higher education, first in France and subsequently in Italy, Germany, and the Netherlands. In 1813, he was a Councillor of State as 'Maître des Requêtes'. The Restoration brought him still more honours. Louis XVIII appointed him Chief Councillor of Public Education and made him a Baron in 1819. The same year he was named President of the Section of the Interior in the State Council, representing the interests of non-Catholics. In 1824, Charles X conferred on him the honour of Officer of the Légion d'Honneur, of which he had been a Chevalier since the time of the Empire. Louis-Philippe named him Pair de France in 1831. Cuvier was a member of three sections of the Institut de France: the Académie Française, the Académie des Sciences, and the Académie des Inspections et Belles-Lettres, as well as numerous foreign academies.

In 1803, Cuvier married the widow of the former fermier général Duvaucel, who had been guillotined in 1793. None of their four children survived, and their deaths caused Cuvier great distress. Cuvier died on 13 May 1832, at the height of his fame, after a short illness, the precise nature of which is unknown (although it may have been cholera or myelitis).

Cuvier's Work and Achievements

The eighteenth and nineteenth centuries were dominated by a desire to emulate the astronomical achievements of Newton in other areas of science. Cuvier aspired to be the Newton of natural history. He wanted to introduce into this field the approach that henceforth would govern all physical sciences: analyse facts, isolate them, compare them, and then try to ascertain general causes to explain the facts thus ordered according to common laws or principles. His best-known law – the one that made possible his fossil reconstructions – was the law of the correlation of organs or parts: all the parts of an organism must be suitably correlated so as to make a viable whole, capable of coping with the conditions of existence. He adopted the 'comparative' approach in the late eighteenth century when endeavouring to restore the remains of mastodons that had been sent to France from America for examination. The task was accomplished using anatomical analogies with modern elephants (for which he regarded the African and Indian types as being distinct, as were the remains of the Siberian mammoth). Applying this principle, Cuvier succeeded in reconstructing a large number of extinct forms. A single tooth, so to speak, told him everything about an organism, he triumphantly proclaimed apropos his reconstruction of the *Mosasaurus*. The immutable laws of zoology, with their wonderful constancy, which are not contradicted in any class or family, served Cuvier admirably in his arduous task of 'resurrecting' (his word) the past.

The notion of species obviously underpinned all attitudes towards, and classifications of, animated nature. It was one of the most clearly defined concepts in Cuvier's work. The most important concepts in nature were those of the individual and the species, and they were connected through the process of generation. Organized beings had two bases for natural classification: the individual, resulting from the common action of all the organs; and the species, resulting from the bonds created by the generation of individuals. From his earliest publications, and particularly in his *Tableau Èlémentaire de l'Histoire Naturelle des Animaux* (published in 1797), Cuvier gave a definition to which he remained steadfast:

> The collection of all organized bodies born one from another, or having parents in common, and all those that resemble them in the same way as they resemble each other, is called a species. [Cuvier G (1797) *Tableau Èlémentaire de l'Histoire Naturelle des Animaux*. Paris: Baudouin. p. 11]

But, in practice, in many cases – and whenever considering the past – one cannot use the descent of forms to define species. So, they must be classified by their distinctive external, and more particularly their internal, parts. Form becomes the prime consideration in the study of living bodies, and gives anatomy a role that is almost as important as that of chemistry.

Although Cuvier seems at times to have supposed that there was really nothing in nature other than the species and the individual, nevertheless the study of living forms led him to ascribe a concrete reality to another type of organization, namely that of *embranchements*. An *embranchement* was an ensemble of animal forms that had a common structural plan, which served as the basis for all external modifications. Cuvier's four *embranchements*-vertebrates, molluscs, articulata (jointed or segmented animals), and zoophytes or radiata-are still well known. If there was a 'closed system' in Cuvier's mind it would seem to have been in systematics, at the level of the *embranchements*. Each of these formed a separate whole; there was no transition or gradation from one *embranchement* to another. Other organisms would not be viable because they would not meet the conditions of existence.

The 'construction plans' of the different *embranchements* were entirely different. There is, for example, no passage from vertebrates to molluscs.

> Whatever arrangement is given to animals with back bones and those without them, one can never place one of their large classes at the end of one group, and some what similar animals at the head of the other so that the two are linked together [Cuvier G (1800) *Leçons d'Anatomie Comparée: 1*. Paris: Baudouin. p. 60]

Similarly,

> There can be no intermediary between mollusca and articulata, nor between them and the radiata, for one cannot fail to recognise the profound interval or 'salta tion' there is when one passes from one construction plan to another.

It was in this spirit that Cuvier undertook the palaeontological investigations for which he became famous.

One of his first concerns was to evaluate the significance of fossils in the reconstruction of the Earth's past. The 'documents' furnished by the successive layers – the 'charters' or 'diplomas of the history of the globe' – revealed that all organisms were not created simultaneously. There was a 'definite succession in the forms of living organisms'. If there were only unfossiliferous strata, one might claim that the various terrains were created at one and the same time. But palaeontology showed that the various classes of vertebrates do not date from the same epoch. Cuvier was certain that the oviparous quadrupeds appeared much earlier than the viviparous types, for he thought that they began with the fishes, whilst the terrestrial quadrupeds appeared long after. Moreover, there is not only an order of succession between classes but also a pronounced order of the species within the stratigraphical column.

In establishing this chronological sequence, thanks to the collaboration of his friend Alexandre Brongniart (1770–1847), who was more of a geologist than he was, Cuvier clearly affirmed the connection between fossils and geological strata – between palaeontology and stratigraphy – to the benefit of the 'true' theory of the Earth. Indeed, in 1806, he proposed to the Académie des Sciences a programme of palaeontological research that would qualify as stratigraphical. Some of the main tasks were to ascertain whether there was any regularity in the succession of fossils, to determine which species appeared first and which came later, and to discover whether these two kinds of species are never found together, or whether there are alternations in their reappearance.

In his *Recherches sur les Ossemens fossiles de Quadrupèdes* (published in 1812), Cuvier applied himself to this programme. With Brongniart, he had proposed to resolve the following questions by means of his studies. Are there animals or plants that are proper to certain strata, and which do not occur in others? Which species appear first, and which come after? Do these two sorts of species sometimes occur together? Are there alternations in their recurrence; in other words, do the first forms recur and the second ones then disappear? Have these animals and plants perished in the places where their remains are found or have they been transported there?

Cuvier noted that the fossils of the Calcaire grossier of the Paris Basin are entirely different from those of the Chalk, and he clearly enunciated the concept of 'characteristic fossils' in stratigraphy. If a formation could be characterized by its fossils, it was because it contained organisms that were entirely distinct from those of other formations. Fossils were the fundamental tools of stratigraphical determination, and Cuvier could use them to recognize a particular formation in a large number of calcareous beds. A formation previously observed in some distant canton could be recognized by the nature of the fossils in each bed. Fossils were marker signals that never failed. Indeed, there was a constant relationship between the strata and the animal and plant remains found in them. Thus an immense field of observation and research was opened up, and Cuvier never doubted that reality would correspond progressively with this programme.

Precisely determining fossil species and delimiting the places where bones were discovered would make it possible to compare not only the superposed strata but also strata that were juxtaposed, in a parallel geological situation, neither above nor below but adjacent to one another, in the same basin or at a distance in two separate basins. Cuvier was not content simply to study the Paris Basin: he extended his observations to other regions. He thought it was important to study the calcareous strata of other basins and to compare them with those of the Paris Basin. Applied successively to other cantons, this method would soon yield important generalizations, and palaeontology, too long fed by illusory conjectures, would evolve a rigorous progress similar to that of other natural sciences.

From stratigraphy, the true history of the Earth began to emerge. When fossils were studied *in situ*, or in relation to strata, they ceased to be simple curiosities and became 'historical records'. Thanks to their study according to this perspective, one could show that there had been successive epochs in the formation of the globe and that a series of different operations or processes had operated at different times.

The historical key provided by palaeontology was thus well established by Cuvier. He was enthusiastic about the grand prospect of studying the past. He wanted to be able to arrange organisms in their chronological order, know about the development of life, determine precisely which forms appeared first, and recognize the simultaneous appearance of certain species and their gradual destruction. His vision provided a research programme for geologists once they had renounced their 'just-so stories' and begun, instead, to do the work of historians. The problem of the history of the Earth was correctly posed in palaeontological and stratigraphical terms. Even though, after Cuvier and Alexandre Brongniart, there was still much to discover, the method for discovery left nothing to be desired.

Cuvier's early works made him one of the masters of comparative anatomy and also opened up a new field to him – palaeontology – that was seemingly full of promise and in which his knowledge of German science, which was then more advanced than French science, gave him a privileged position in France. His *Recherches sur les Ossemens Fossiles de Quadrupèdes* was published in 1812, but the work had been published earlier in the *Annales du Muséum*. In a communication made to the Institut de France, Cuvier had, in 1801, enunciated three hypotheses that were, even then, already known to his colleagues: former species had been entirely destroyed, or they had been modified in form, or they had been transported from one climatic zone to another. The second of these three alternative explanations had originally been suggested by Jean-Baptiste Lamarck (1744–1829), while the third was proposed by Barthélémy Faujas de St. Fond (1741–1819) and others. Cuvier preferred the first, which involved not only the disappearance but also the destruction of ancient forms. Thus, from the beginning of his scientific career, the young naturalist adopted the postulate of what came to be known as catastrophism as the basis for his palaeontological researches.

With this end in view, Cuvier applied himself to distinguishing carefully between fossil and modern forms. He was certain that none of the ancient forms had 'living analogues'. The 23 species that he had already been able to restore all appeared to have been 'destroyed' and to have become extinct. This assumption had direct consequences for geological theories: the lost species had "belonged to beings from a world anterior to our own and to beings that were destroyed by some revolutions of the globe" (Mémoire sur les espéces d'Eléphans vivantes et fossiles, Mémoire de l'institut national des Sciences et des Arts, Fructidor an VII (août-septembre, 1799), 2, 1–22: cit. p. 21). The disappearance, or, as Cuvier put it, the 'destruction', of former beings could only be explained by a 'general revolution of nature'.

The master palaeontologist laboured hard to establish the reality of the 'last catastrophe', which was related to the ideas favoured by believers of Holy Scripture. Although the picture of a total destruction of organisms did not correspond with the facts given in the Bible, Cuvier presented himself as a defender of the Noachian Deluge. He, along with Jean-André Deluc and Déodat Dolomieu (who were defenders of the idea of the Flood as a geological agent), thought that if anything was certain in geology it was that the surface of the globe had undergone a sudden revolution no more than five or six thousand years ago. But this catastrophe was, for Cuvier, only 'the last universal inundation'. That Cuvier was a supporter of general catastrophes does not, however, mean that he did not also uphold the idea of limited or local catastrophes. In the series of revolutions that he proposed, some were only partial. But just one universal catastrophe was enough to raise the problem that it posed for the continuity of life. Thus, although there had been numerous revolutions, there had not been so many creations, for migrations could play a role in some cases, as he suggested could potentially occur in New Holland (Australia). If there was an inundation of Australia that destroyed all its marsupials and the continent was subsequently colonized by animals from Asia, then the stratigraphical record in Australia would show a general catastrophic flood followed by the new creation of forms.

Cuvier never proposed a precise number of revolutions or creations. It was his follower Alcide d'Orbigny (1802–1857) who devoted himself to such calculations. He divided the stratigraphical column into 27 stages and therefore proposed a total of 28 creations, which came to serve as the basis for later stratigraphical work. Nowhere in Cuvier's *oeuvre* do we find the expression 'successive creations'. However, he originated the idea of repeated creations. From the moment when he envisaged general irruptions that "destroyed all the quadrupeds that they reached" and "caused the entire classe to perish" new creations were required to make the animals reappear. Cuvier's position on this was quite explicit, and so well known that from the beginning of the nineteenth century until his death he was considered to be the catastrophists' leader, as the eminent geologists who knew him would have agreed. Could Cuvier, who was able to speak so clearly and on occasion defend himself so well, have been misunderstood to such an extent by his contemporaries? This is inconceivable.

Cuvier had other good reasons for rejecting transformism, which were doubtless less significant for him than his catastrophism, but which had greater scientific validity. He raised a palaeontological objection that was valid even from a non-catastrophist perspective, namely the absence of intermediate or transitional forms between the former creatures and those that were more recent or extant. If the most ancient forms were the ancestors of those that followed, then one would expect to find the remains of the genealogical intermediaries. But, Cuvier objected, such transitional forms are never found. Between the *Paleotherium* and today's most similar species no such forms had at that time been found. Cephalopods do not lead on to fishes – a fact that even Lamarck did not dispute. However, Cuvier did not say how he thought new forms could have been created.

Influence: The 'Domination of Cuvier'

According to David Hull, in 1860, "on the continent, especially in France, catastrophism still reigned". But this view is mistaken.

Cuvier died in 1832 at the height of his fame. He was, however, already a controversial figure in politics, well known for his participation in the service of a succession of different regimes, and his intellectual and scientific worth were questioned by some of his contemporaries. The judgment of Goethe is well known: "No one described a fact better than he did. But he is almost devoid of philosophy. He will produce pupils well trained but with little depth" Eckernonn J.P. (1948). Gespr̈ache mir Goethe. Muncher, Deutsches Verläghaus Bong, pp. 329–330. Alexander von Humboldt, who admired Cuvier's "memorable studies of fossil bones", revealed his disagreements with him during a lecture Cuvier gave at the Collège de France, in which Cuvier criticized the ideas of Ètienne Geoffroy Saint-Hilaire.

Cuvier undoubtedly had faithful followers, especially in England; William Buckland is the best known but many others could be cited. Cuvier also had disciples in France, of whom Adolphe Brongniart is the best known and was the most devoted but there were others, for example Léonce Èlie de Beaumont and d'Orbigny (although they had some reservations). Èlie de Beaumont followed Cuvier in his rejection of fossil man, but did not do so when it came to the general destruction of life, as Darwin noted in 1859. Even d'Orbigny questioned some of Cuvier's conclusions about the Earth's past.

It is among the members of the thriving and internationally esteemed Société Géologique de France, founded in 1830 – of which, significantly, Cuvier was not a member, although Darwin (*see* **Famous Geologists:** Darwin) and Lyell (*see* **Famous Geologists:** Lyell) were – where one should look for the opponents of his catastrophist and anti-transformist opinions. Even before his death, many opposed him, including Constant Prévost (who was one of Lyell's teachers), Jean-Baptiste d'Omalius d'Halloy (the father of Belgian geology), Jules Desnoyers, André de Férussac, Marcel de Serres, and Ami Boué (one of the Society's founders). In 1830, Ami Boué, expounding Adolphe Brongniart's work on fossil plants, maintained that Cuvier's ideas were contested or rejected by 'the majority of geologists', whose names he gave. Shortly before Cuvier's death Boué declared himself to be radically opposed to "the idea put out by Mssrs Cuvier and Buckland on the universal Deluge, the universality of which was shown to be false by the most obvious facts" and also to "other opinions of M. Cuvier, namely... his hypothesis which, contrary to the natural order and to facts, admits universal cataclysms at several epochs prior to the Noachian Deluge". Gérard-Paul Deshayes, another of Lyell's tutors, recalled that in 1835 French zoologists were divided into two camps, one following Cuvier and the other following Lamarck. In 1836, Lyell himself placed Cuvier and Lamarck on the same rung. Several members of the Society, including Boué, openly declared themselves to be supporters of Lamarck.

Informed historians will not be surprised to read the evidence of Camille Dareste, who, in 1859, before Darwin became widely known, attested that the transformist theory of Lamarck had penetrated deep into the French scientific community. The domination of Cuvier after his death is a mistaken historical legend, which conscientious historians should not perpetuate.

See Also

Biblical Geology. **Creationism**. **Evolution**. **Famous Geologists:** Darwin; Lyell. **History of Geology From 1780 To 1835**. **Stratigraphical Principles**.

Further Reading

Boué A (1831) Compte rendu de la traduction allemande, par Nöggerath, des *Révolutions du Globe*, par Cuvier. *Bulletin des Sciences Naturelles et Géologie* 24: 129 130.

Coleman W (1964) *Georges Cuvier Zoologist: A Study in the History of Evolution.* Cambridge, MA: Harvard University Press.

'Collectif' (1932) *Centenaire de Cuvier, Archives du Muséum National d' Histoire Naturelle.* Paris: Masson et Cie.

Cuvier G (1797) *Tableau Èlémentaire de l' Histoire Naturelle des Animaux.* Paris: Baudouin.

Cuvier G (1800) *Leçons d' Anatomie Comparée: 1.* Paris: Baudouin.

Cuvier G (1801) Extrait d'un ouvrage sur les espèces de Quadrupèdes. *Journal de Physique* 52: 253 267.

Cuvier G (1812) *Recherches sur les Ossemens Fossiles de Quadrupèdes: 1.* Paris: Déterville.

Cuvier G (1812) *Essai sur la Géographie Minéralogique des Environs de Paris.* Paris: Baudouin.

Cuvier G (1817) *Mémoire pour servir à l'Histoire et à l' Anatomie des Mollusques, Mémoire sur la Scyllée.* Paris: Déterville.

Cuvier G (1825) *Discours sur les Révolutions de la Surface du Globe.* Paris: Dufour.

Cuvier G (1827) *Rapport Historique sur les Progrès des Sciences Naturelles depuis 1789.* Paris: Imprimerie Imperrales.

Dareste C (1859) Biographie de Lamarck. In: Hoefer JCF (ed.) *Nouvelle Biographie Générale,* 29, pp. 55 62.

Flourens P (1859) *Histoire des Travaux de Georges Cuvier,* 3rd edn. Paris: Garnier.

Hull D (1973) *Darwin and his Critics: The Reception of Darwin's Theory of Evolution by the Scientific Community.* Chicago: Chicago University Press.

Laurent G (1987) *Paléontologie et Évolution en France, 1800 1860: De Cuvier Lamarck à Darwin*. Paris: Comité des Travaux Historiques et Scientifiques.

Laurent G (2000) Paléontologie(s) et évolution au début du XIXe siècle: Cuvier et Lamarck. *Asclepio* 52: 133 212.

Outram D (1980) *Georges Cuvier, Vocation, Science and Authority in Postrevolutionary France*. Manchester: Manchester University Press.

Rudwick MJS (1997) *Georges Cuvier, Fossil Bones, and Geological Catastrophes: New Translations and Interpretations of the Primary Texts*. Chicago and London: The University of Chicago Press.

Smith JC (1993) *Georges Cuvier: An Annotated Bibliography of his Published Works*. Washington DC: Smithsonian Institution Press.

Darwin

D R Oldroyd, University of New South Wales, Sydney, Australia

Introduction

Charles Darwin (Figure 1) is chiefly remembered for his celebrated theory of the evolution of life forms and speciation, by means of natural selection. But his considerable contributions to geology should not be forgotten. Darwin, born in 1809, was brought up at Maer, Staffordshire, UK, son of a prosperous doctor. He attended school at Shrewsbury, and at age 16 proceeded to Edinburgh University to study medicine, but withdrew from the course because of his distaste for dissections and operations conducted without anaesthetics. He then went to Cambridge to take the standard arts degree, with a view to becoming a clergyman.

Darwin's interest in natural history developed while he was still at school and was furthered in Edinburgh by studies of plankton in the waters of the Firth of Forth. He attended some of Robert Jameson's mineralogical/geological lectures, which were presented according to the 'geognostic' principles of the famous eighteenth-century Freiberg teacher of 'Neptunist' theory, Abraham Werner; Darwin found the ideas taught unacceptable and he discontinued his attendance. However, he may have learnt more geology there than he later acknowledged in his autobiography. Both Darwin and his father described his time at Cambridge as wasted, which it was, so far as theological training was concerned, but Darwin continued his private studies in natural history (especially beetle collecting) and became an epigone (and later, a personal friend) of the botany professor John Henslow. Henslow imparted some geological understanding to Darwin, having earlier done a fair amount of geological work, notably in Anglesey.

Beagle Voyage

Enthused by the writings of Alexander von Humboldt, Darwin wanted to travel. In his last year at Cambridge, he planned an informal journey with friends to Tenerife, which necessitated his brushing up on geology, having earlier largely ignored the subject at Cambridge, being "so sickened with the lectures at Edinburgh". Henslow taught him to use a clinometer and gave him geological advice on his project. Possibly Darwin also attended some lectures of the geology professor Adam Sedgwick (*see* **Famous Geologists:** Sedgwick), and certainly participated in the field excursions Sedgwick ran, around Cambridge. Then, in 1831, at Henslow's suggestion, Sedgwick took Darwin along on a short field excursion in North Wales as assistant and companion, also with the idea of teaching him the rudiments of field geology. Darwin was an apt pupil, but this was essentially

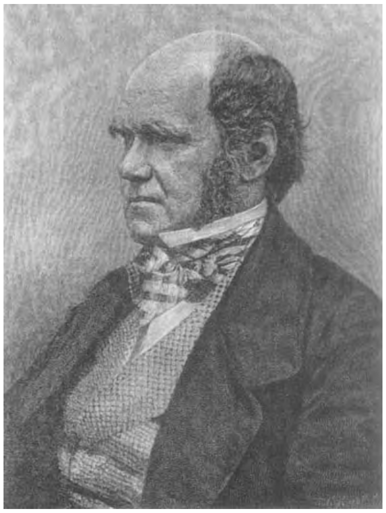


Figure 1 Charles Darwin (1809 82), from a photograph (1854?). Engraved for *Harper's Magazine*, October 1884. Reproduced from Darwin F (1887), 1, Frontispiece.

all the training he had in geology. Nevertheless, when later that year Darwin joined the *Beagle* on its epic journey and circumnavigation of the globe, he initially regarded geology as his principal scientific interest and even skill, having also done some solo field trips in south-west England.

Darwin studied Charles Lyell's *Principles of Geology* (*see* **Famous Geologists:** Lyell), the first volume of which (published in 1830) he had on board the *Beagle*; according to Darwin's later published accounts, he immediately sought to apply his understanding of Lyell's ideas to the observations made at St Iago, in the Cape Verde Islands, the *Beagle*'s first port of call. However, Darwin's field notes made at a village called St Domingo, situated in a wadi-like inland valley, reveal that he was initially thinking in catastrophist terms, such as he might have picked up from Sedgwick. Darwin thought that the valley had been created by some "great force" followed by "the agency of large bodies of water", which moved large boulders and seemed also to have deposited material similar to diluvium (which he had encountered aplenty in Wales) near the coast. Subsequently, when Darwin published his geological observations, his initial catastrophist ideas were set aside, his Cape Verde observations being presented as supportive of Lyellian geology. Thus, Darwin reported evidence for uplift of land, in the form of a fossiliferous limestone about 40 feet above sea-level, with similar material forming at sea-level today. A volcano there also seemed to have depressed strata, but the published comparison of past and present processes was unquestionably Lyellian.

Along the east coast of South America, Darwin made important collections of megafaunal remains. He made inland excursions towards the Andes and observed what appeared to be a 'stepped' structure for the Patagonian plains, indicating successive elevations. Near Tierra del Fuego, he saw snow-covered mountains, and glaciers descending close to sea-level. Travelling up the Chilean coast in 1835, he experienced a major earthquake at Concepción, with considerable uplift directly evident and in seeming accord with Lyellian geology. Darwin also travelled across Andean passes and formed the idea that the earth had been ripped apart along the line of the mountains and that igneous matter had entered the huge fissure. He even prepared a sketch map (in about 1834) of the southern part of the continent, distinguishing three main units: Tertiary strata; granites, gneiss, mica, slate, quartz rock, and clay slate; and lavas, tufas, and porphyries. A more detailed, also undated, map of the more southerly part of the continent from about the same period depicted seven units: 1. Granite, Mica slate; 2. Trappean rock and Porphyries; 3. Purple Porphyry and Tufa. Metamorphics; 4. Clay Slates; 5. Tertiary (Pliocene?); 6. d[itt]o—Recent; 7. Basaltic Lava.

Crossing the Pacific, Darwin noted that some coral islands seemed to have been elevated, as indicated by the dead coral above sea-level. Elsewhere, there seemed to have been subsidence, but the corals were growing upwards at about the same rate as the land was sinking, so that fringing reefs were formed. This evidence cohered with Lyell's idea that some parts of Earth's crust were rising while others were sinking. In Australia, Darwin crossed the Blue Mountains and observed its great cliffs and valleys, but having only a restricted view of the topography, he mistakenly supposed that the valleys were produced by marine rather than by fluvial erosion. In Tasmania, he quickly got the hang of the geology around Hobart, and possibly, on the basis of 'drop-stones', had the idea of there having been a glaciation in what he referred to as Carboniferous times (actually Permian).

On his return to England, Darwin soon became acquainted with Lyell and remained a lifelong friend; Darwin was elected a Fellow of the Geological Society and was soon on its Council, then became a Secretary and, later, Vice-President. In 1838, he published a general theoretical paper, *On the Connexion of Certain Volcanic Phenomena in South America, and on the Formation of Mountain Chains and Volcanos, as the Effects of the Same Power by which Continents are Elevated*. Published a year after he had come to accept 'transformism', this work was less 'steady state' than Lyell's geology envisaged. Darwin linked earthquakes, elevations, and volcanic eruptions. He thought that (for the Andes, at least) there were repeated uplifts and intrusions along the axis of the range, followed by cooling and consolidation. Rejecting global contraction, differential sedimentary loads, or interplanetary forces, he had no concrete suggestion as to the cause of the elevations. But he was convinced that the uplift proceeded in small stages and was ongoing, rather than occurring in one great catastrophic episode. He concluded that "the configuration of the fluid surface of the earth's nucleus is subject to some change, —its cause completely unknown, —its action slow, intermittent but irresistible". (Modern students of plume theory and the effects of actions occurring at the mantle/core boundary may find this remarkably prescient!)

Geological Publications and ideas on Glacial Phenomena

Following the *Beagle* voyage, Darwin published three major geological books: *The Structure and Distribution of Coral Reefs* (1842), *Geological Observations* *on the Volcanic Islands Visited during the Voyage of H.M.S. Beagle* (1844), and *Geological Observations on South America* (1846). In addition to matters previously discussed, he distinguished in 1846 between stratification, cleavage, and foliation, but the distinction between cleavage and bedding was probably imparted to him by Sedgwick. Also, and importantly, in 1844, Darwin initiated for petrologists the idea of gravity settling, based particularly on his observations of igneous rocks in the Galapagos Islands; this was based on the idea that crystals that first form from a cooling magma may separate out and thereby alter the chemical composition of the remaining fluid, thus producing magmatic differentiation.

Another important piece of work undertaken by Darwin post-voyage was his attempt in 1839 to explain the strange set of markings, the so-called Parallel Roads of Glen Roy, on the sides of Glen Roy in central Scotland. These controversial parallel and horizontal markings evidently marked former shorelines of some kind. In Darwin's view, they represented different marine shorelines, being formed (by analogy with ideas developed in South America) by a succession of land elevations. Erratic granite boulders were also to be found, and Darwin ascribed their deposition to floating icebergs. Indeed, the whole situation was seen and interpreted in terms of what he had seen in the Tierra del Fuego area, with Glen Roy being in some ways comparable to the Beagle Channel. So the observations were thought to accord with the notion of subsidence of the land, associated with cold and extended glaciation in Scotland. On subsequent elevation and amelioration of the climate, the supposed shorelines and erratics would be exposed. Subsequently, in 1842, Darwin thought the glacial submergence and floating iceberg theories could also be applied in North Wales.

Darwin's ideas about Glen Roy were later superseded by the idea that the marks were due to glacial lakes: i.e., water ponded in the valley by glaciers blocking its mouth at different altitudes, water having escaped over different passes at different altitudes, according to the size of the barriers. Thus, there were several distinct former lake margins at different levels. Darwin later acknowledged the superiority of this theory, and called his Glen Roy paper a "great failure" in his autobiography.

Later Years, Evolution, and the Age of the Earth

By the 1840s, Darwin's health was deteriorating and he gave up substantial fieldwork: he was beginning to focus more attention on his grand theory of evolution by natural selection, first adumbrated in 1837. His last geological paper proper (on the geology of the Falkland Islands) appeared in 1846, and that year he turned to a taxonomic study of barnacles, both modern and fossil, continuing this work at his home for eight years. His study of modern forms led to his discovery of males living as 'parasites' within the female forms, and also a gradation from hermaphrodite types, through forms with females having an 'attached' male organ and one that was parasitic but physically detached, to types whereby there was sexual dimorphism, but in which the males were 'parasitic' on the females. Thus, Darwin saw, in barnacles, evidence for the evolutionary emergence of sexual dimorphism. He then turned to the study of fossil barnacles, publishing a two-volume monograph (in 1851 and 1854). In the light of hindsight, it can be seen that these works were ordered (or the organisms classified) from an evolutionary perspective.

When Darwin eventually published his *Origin of Species* in 1859, there were two issues of principal geological interest. First, he wanted to present to the public the idea of the history of living forms as being analogous in form to a branching tree; second, he had to deal with the problem of the age of Earth. There was also the problem of the origin of life and the apparent appearance of quite new forms from time to time, particularly the appearance in the Cambrian (or Silurian as he termed it then, following Roderick Murchison (*see* **Famous Geologists:** Murchison)) of quite well-developed forms apparently without ancestors. But the actual stratigraphic record showed anything but continuity or smooth transitions. Examples of trends, with a complete presentation of the various forms in an evolving continuum, or cases of branching and speciation, were conspicuous for their absence in the fossil record. Darwin sought to answer this difficulty by appealing to the incompleteness of the stratigraphic record: many pages of the evolutionary record were missing due to weathering and erosion, or had been destroyed by metamorphism. Moreover, intermediate forms might not all have been preserved at the same locality, so it would be unreasonable to expect smooth transitions in an ascending section. Similarly, the abrupt appearance of new types in strata was a problem for Darwin (and was used as an argument against him by his contemporaries, and by critics ever since). Again he appealed to the immensity of time, the imperfection of the geological record, and geologists' incomplete knowledge of that imperfect record. Also, he pointed out that much time might be required to evolve some particularly advantageous character (such as the ability to fly), but once acquired, the increase of that character would be very rapid. So the fossil record might give the appearance of sudden changes; but the reality could actually have been one of continuous change. And sometimes anticipated ancestors might be found, as in the case of fossil whales then quite recently found in the Cretaceous. Darwin long sought the occurrence of fossil sessile barnacles and was delighted when some were eventually found in the Chalk, as he had expected would be the case one day. As to the absence of fossils older than the Cambrian (Silurian), Darwin thought that metamorphism might chiefly be responsible. He did not know that soft-bodied Precambrian fossils would one day be found; but it would have been in accordance with his expectations.

As to the age of Earth, comparison of the thickness of preserved sediments with the rate of deposition seemed to reveal the immensity of time. Assuming the erosion of the valley of the Weald in Kent as being chiefly due to the action of the sea, and thinking of erosion as proceeding at 1 inch per century, Darwin gave a figure of 306 662 400 years for the formation of the valley. This was obviously a crude estimate. Darwin halved this estimate in the second edition of the *Origin* and subsequently withdrew it all together. Modern opinion has it that Darwin's figure was much too large, but it is evident that he had a clear vision of the immensity of geological time, and thought that it could brush away many of the objections to his evolutionary theory.

In his last major work, *The Descent of Man* (published in 1871), Darwin boldly applied his theory to humans, but said little about fossil forms, and the book had little geological content. He did, however, speculate that humans had first evolved in Africa, because that was where our nearest animal relatives were found; this suggestion is still thought to be correct.

See Also

Biological Radiations and Speciation. **Evolution**. **Famous Geologists:** Lyell; Murchison; Sedgwick. **History of Geology From 1835 To 1900**.

Further Reading

Barrett PH, Gautrey PJ, Herbert S, Kohn D, and Smith S (eds.) (1987) *Charles Darwin's Notebooks, 1836 1844: Geology, Transmutation of Species, Metaphysical Enquiries*. London: British Museum (Natural History); Ithaca: Cornell University Press.

Darwin C (1842) *Geology of the Voyage of the Beagle: The Structure and Distribution of Coral Reefs*. London: Smith Elder & Co.

Darwin C (1844) *Geological Observations on the Volcanic Islands Visited during the Voyage of H.M.S. Beagle*. London: Smith Elder & Co.

Darwin C (1846) *Geological Observations on South America*. London: Smith Elder & Co. (Also numerous later editions and different publishers.)

Darwin F (ed.) (1887) *The Life and Letters of Charles Darwin, Including an Autobiographical Chapter* 3 vols. London: John Murray.

Herbert S (1986) Darwin as a geologist. *Scientific American* 254(May): 116 123.

Herbert S (1991) Charles Darwin as a prospective geological author. *British Journal for the History of Science* 24: 159 192.

Herbert S (2005) *Charles Darwin, Geologist*. Ithaca and London: Cornell University Press. In press.

Pearson PN (1996) Charles Darwin on the origin and diversity of igneous rocks. *Earth Sciences History* 15: 49 67.

Pearson PN and Nicholas CJ (2003) Charles Darwin's geological observations at Santiago (St Jago), Cape Verde Islands. *International Commission on the History of Geological Sciences 28th International Symposium. Trinity College, Dublin, Ireland . . . Programme, Abstracts & Delegates*, 41.

Rhodes FHT (1991) Darwin's search for a theory of the Earth: symmetry, simplicity and speculation. *British Journal for the History of Science* 24: 193 229.

Roberts MB (2000) I coloured a map: Darwin's attempts at geological mapping in 1831. *Archives of Natural History* 27: 69 79.

Rudwick MJS (1974) Darwin and Glen Roy: a "great failure" in scientific method? *Studies in History and Philosophy of Science* 5: 97 185.

Secord JR (1991) The discovery of a vocation: Darwin's early geology. *British Journal for the History of Science* 24: 133 157.

Stoddart DR (1976) Darwin, Lyell, and the geological significance of coral reefs. *British Journal for the History of Science* 9: 199 218.

Du Toit

J C Loock, University of the Free State, Bloemfontein, South Africa
D F Branagan, University of Sydney, Sydney, NSW, Australia

Introduction

Alexander Logie du Toit, South Africa's greatest and best-known geologist, started his career in 1903 and, over the next 20 years, mapped and investigated a large area of South Africa. He collected data simultaneously on the Dwyka Tillite, other Karoo rocks (formerly spelled Karroo), vertebrate fossils, the southern *Glossopteris* flora, Karoo dolerites and the Drakensberg basalts. During his career he expanded his interests to include other fragments of Gondwana. This led him to propagate the idea of the former existence of a southern supercontinent. He was very much aware of the pioneering work of his predecessors, especially Wegener (*see* **Famous Geologists:** Wegener), but he carried their theories further. After a short period of consolidation and publication in the early 1920s, he settled down to write his book on the 'wandering continents'. His *Our Wandering Continents* (published in 1937) cast the gauntlet at sceptics and anti-drifters. Du Toit was vindicated nearly three decades later, when geophysical studies, mainly of the ocean floors, led to the theory of plate tectonics.

Ancestry and Opportunities

Alexander du Toit (Figure 1) was fortunate in birth and ancestry. Early in the nineteenth century, Alexander Logie from Fochabers, Banffshire, Scotland, served as an officer in the 72nd Regiment in South Africa. Captain Logie married Henrietta Elizabeth Susanna du Toit, a descendant of Francois du Toit from Lille in France, who arrived at the Cape of Good Hope in 1686 as a Huguenot. As the marriage was childless, the couple adopted the infant son of Stephanus Hendrik du Toit, the brother of Mrs Logie, and his wife, Barbara Stadler. The boy was christened Alexander Logie, but he retained his du Toit surname, growing up on the family estate on the outskirts of Cape Town. Alexander married Anna Logie, daughter of Robert Clunie Logie, a brother of Captain Alexander Logie. They had four children, one of whom was Alexander Logie du Toit, born at Rondebosch near Cape Town on 14 March 1878. His relatives included some of the most prominent families then living in Cape Town. When he passed the matriculation examination in 1893 he was among the top scholars in Cape Colony.

Du Toit was sent to the University of the Cape of Good Hope, where he passed the intermediate examination with distinction in 1894 and went on to obtain the Bachelor of Arts degree in mathematics and natural science the following year, again passing with distinction.

Du Toit's descent dictated that he should study in Scotland. He qualified in mining engineering at the Royal Technical College of Glasgow in 1899 and then studied at the Royal College of Science in London. In 1901 he was appointed a lecturer at the Royal Technical College of Glasgow and also at the University of Glasgow. His studies and travels in Scotland and further afield gave him an insight into the geology and stratigraphy of the northern hemisphere and into the nature, origin, and structure of fold mountains. This newly acquired knowledge later proved crucial in his theories on continental drift. He kept his contacts with Scotland and submitted a thesis on the copper–nickeliferous layered intrusion of Insizwa in 1910, for which he was awarded a DSc degree.

In Scotland all was not study. Du Toit played the oboe, a hobby that he maintained for many years. He married Adelaide Walker in Glasgow and returned

Figure 1 Portrait of Alexander du Toit. (Photograph reproduced by courtesy of the Geological Society of South Africa.)

with her to South Africa late in 1902, where a new world and a long career stretched before him.

The Formative Years

The Geological Commission of the Cape of Good Hope, 1903–1912

The Geological Commission of the Cape of Good Hope, established in 1896 by the Cape Parliament, consisted of a veteran politician and senior civil servants. Du Toit assumed the post of geologist at the beginning of 1903, being listed on the scientific staff as Alex L du Toit BA FGS. When he left for his first field session, he was accompanied by the Commission Director, Arthur Rogers, who introduced him to the upper beds of the Cape Supergroup (Devonian) and the overlying basal units of the Karoo Supergroup (Late Carboniferous to Early Permian). This was du Toit's first contact with the Dwyka Tillite, which is exposed in the south-western corner of the Karoo outcrop area in a desert environment. We can picture him walking on the unweathered outcrops with a huge number of loose erratics lying around, which he identified according to rock type. Later, he was to trace these erratics in the Northern Cape to actual outcrops of Precambrian rocks. Little did he know at that stage that he was destined, three decades later, to be an international expert on the Dwyka Tillite and its equivalents in other fragments of Gondwana.

During his 9 years service with the Commission, du Toit mapped and studied the rocks and the strata in three areas. In the western corner of the huge area underlain by rocks of the Karoo Supergroup, he unravelled the stratigraphy of the two basal units, namely the Dwyka and Ecca Groups, paying special attention to the Dwyka glacials. Du Toit also had an interest in the Karoo Dolerite (Jurassic) and the diatremes that pierce the Karoo beds. In the north-eastern Cape, he covered an area containing rocks ranging from Early Precambrian to Quaternary. Diatremes and kimberlite intrusions again attracted his attention. Du Toit is, however, chiefly remembered as a field geologist for his detailed and accurate maps of the north-eastern Cape, where he concentrated on the Beaufort and Stormberg Groups and the plateau basalts of the Drakensberg.

An image of du Toit as a competent field geologist now emerged. He was a wiry and energetic man, who could cover long distances on foot, on a bicycle, or, on occasion, on horseback. In the more open areas to the west, his caravan-like wagon was pulled by a donkey team (Figure 2). Mapping, often in areas where large-scale maps showed farm boundaries only, was done using a small plane table and alidade. He was renowned in the geological community for being able to judge distances very accurately. When mapping dolerites, he could visualize an intrusion into Karoo beds as a three-dimensional body and hence accurately predict the locations of the dolerite outcrops. One of du Toit's ways of winning the confidence of the local inhabitants was to encourage the infirm and to dispense aspirins or coloured pills to those who feigned or claimed illness. As his fame as a 'doctor' spread, he found that he could rely on the locals for advice on the geography of a mountainous area

Figure 2 During his first field excursions, du Toit was provided with a wagon and a team of donkeys. (Photograph reproduced by courtesy of the Natural History Division of the South African Museum, Iziko Museums of Cape Town.)

or the route of a cattle track through bush and over mountains.

The results of du Toit's labours in the field are to be found in the Geological Commission's annual reports, and other observations were published in scientific journals. The 12 maps with which he was involved, either on his own or mainly in conjunction with Rogers, covered an area of 180 000 km^2 and were a source of wonder to his successors. One of his publications, on the evolution of the river systems in Griqualand West, is consulted to this day by geologists prospecting for alluvial diamonds. In the first edition (1905) of his book on the geology of the Cape Colony, Rogers referred to 'Gondwanaland', the term introduced by Suess in his *Das Antlitz der Erde* (1885) (*see* **Famous Geologists:** Suess). When Rogers prepared the second edition, he invited du Toit to be his co-author. The second edition contained expanded versions of the Gondwana hypothesis and of South African stratigraphy.

Between 1905 and 1910, du Toit worked mainly in the more arid northern parts of Cape Colony, on unfossiliferous sediments and various igneous rocks. Here, he became friends with, and was impressed by, the Dutch geologist Gustaaf Molengraaff. Du Toit's reports on the nature and petrology of the kimberlite pipes, which were often richly diamond bearing, were widely read and used by prospectors and mining interests.

The Cape Geological Commission was disbanded at the end of 1911, but du Toit had already collected in his memory and in notebooks a vast number of facts, observations, interpretations, and ideas. He was to prepare a great synthesis in the following decade.

Geological Survey, 1912–1920

Du Toit was transferred to the Geological Survey of the Union of South Africa when the four colonial surveys or commissions were amalgamated in 1912, and he suddenly found that he was free to study geological problems over the whole of South Africa. By this time, du Toit had transferred his interests to minerals and rocks. He spent more and more time on the investigation of specific sites and less time on mapping. He participated in the activities of the Geological Society of South Africa and served as President in 1918 and again in 1928. His first book, *A Physical Geography for South African Schools*, appeared in 1912. In 1914 he visited Australia to attend the meeting of the British Association, where he met the Sydney professor T W Edgeworth David and was able to examine the rock succession in eastern Australia and the evidence for Late Palaeozoic glaciation. The remarkable similarities between the records of events in two widely separated southern continents, with evidence of glaciation in Australia at the same time as that indicated by the Karoo rocks in South Africa, was striking.

As the First World War had broken out, du Toit was called up to serve as a geologist and was charged with finding water for the Union Defence Force. When the South Africans invaded German South-west Africa (now Namibia), du Toit had to find suitable sites at which to drill for water in the desert. His military service must be seen as a bonus because he found time to study the basement rocks, the Late Precambrian Nama beds, and, more importantly, the basal Karoo rocks.

Du Toit's last publication while a member of the Survey was his monumental compilation on the Karoo dolerites.

Department of Irrigation, 1920–1927

The Department of Irrigation requested du Toit's services because it needed his expertise. His relatively brief period of service saw both tragedy and triumph. Adelaide du Toit died in 1923, leaving her husband and a grown-up son. Two years later, du Toit married Evelyn Harvey.

Du Toit's many reports from this period, now mostly filed away and forgotten, dealt chiefly with dam sites and geomorphology. Nearly two decades had passed since du Toit had started his career, and his observations and the synthesis of the facts were ready for a wider audience. When the South African Association for the Advancement of Science met in Durban in 1921, du Toit was invited to deliver the popular evening lecture, for which he chose to speak on land connections between South Africa and other continents. He presented evidence in the form of vertebrate life, the migration of vertebrates, palaeoclimates, volcanism, and fold mountains to an audience that included sceptics. His main thrust was a resumé and analysis of the glacial deposits at the base of the Karoo. This was augmented by a map showing Gondwana at the close of the Carboniferous.

In the same year, the Geological Society of South Africa published du Toit's summary and analyses of Carboniferous glaciation in South Africa. The references to the pre-glacial topography, the direction of flow of continental ice-sheets, and the distribution of erratics placed southern Africa in the wider context of Gondwana. A later generation of stratigraphers provided evidence that the Dwyka glacials range in age from the Late Carboniferous to the early Permian.

Other adventures followed. In 1923, a grant from the Carnegie Institution in Washington enabled du

Toit to visit Brazil, Uruguay, Argentina, and Chile, which were previously poorly mapped and not well understood. He carried out a remarkable amount of field exploration in South America, meeting local geologists, including his old friend David Draper from South Africa, who was briefly managing the Boa Vista diamond mine in Minas Geraes. Du Toit mastered the difficult literature, which was mainly in Portuguese, Spanish, and German. His tour allowed him to study Devonian beds, fold ranges, and, as might be expected, the Karoo equivalents. Back in South Africa, he was able to show Edgeworth David, *en route* to England, the key local sites relevant to the displacement hypothesis. It was a fruitful meeting, and both geologists strengthened their support for the notion of continental drift.

The Carnegie Institution published du Toit's *Geological Comparison of South Africa with South America* in 1927. The book contained, *inter alia*, a chart showing the stratigraphic column from the Devonian to the Jurassic for selected South American countries. The boundaries of the Afro-American landmass were shown to have a bearing on the displacement hypothesis, which was becoming more widely known through the English translation of Wegener's *Origin of Continents and Oceans* (1924).

By this time, du Toit had made contact with the Dutch geologist Willem van Waterschoot van der Gracht, who was then working in the petroleum industry in the USA. Van der Gracht had become an apostle of continental drift and persuaded the American Association of Petroleum Geologists to organise a conference in New York in late 1926 to discuss the theory. Despite the contributions of several supportive American speakers, including Reginald Daly, a prejudiced group, led by Charles Schuchert, condemned the theory out of hand. Unable to attend, du Toit was dismayed by the intolerant attitudes of some people and their personal attacks on Wegener and others. He sent a paper to support Van der Gracht's publication of the proceedings, which appeared in 1928, and in the following years added further publications in journals such as the *American Journal of Science*; even as late as 1944 he made a rejoinder to G G Simpson regarding his ideas on Tertiary mammals and continental drift.

For some years, du Toit had been planning to write a textbook on the geology of South Africa. The first edition, published in 1926, was the first synthesis of its kind. A second revised edition followed in 1939. Du Toit died before he could complete his revision of the third edition, but his old friend and colleague Sidney Haughton carried on the task and the book was published in 1954.

Years of Work and Wandering

In 1927, du Toit was invited to join De Beers Consolidated Mines as a consulting geologist, specializing in diamondiferous kimberlite pipes and alluvial gravels. Once again, he could travel extensively in Africa, but now in an official capacity. These travels, to areas in which De Beers had an interest, afforded him many opportunities to study the local geology. He could also visit other countries: the USA and Canada in 1932; the USSR in 1937; and India in 1938. Little concerning the areas visited survives in published form. However, from a cache of photographs and other documents discovered recently, we know that he travelled widely in the western USSR, from the Urals down to the Ukraine and the coastal area of the Black Sea.

Du Toit's greatest contribution to geology, and also his swansong, was his book *Our Wandering Continents* (1937), in which he assembled the observations, deductions, comparisons, and syntheses of facts and theories of a lifetime. A brief description of the features of the book is in order. After a review in which he acknowledged the work of his predecessors, he referred to tectonism, volcanism, palaeoclimates, plant fossils, and geosynclines to explain and describe his grouping of the ancient continents and his theory of drifting. The distribution of glacials formed, as before, the core of his arguments. We should note that du Toit now used the term 'Gondwana' for the southern supercontinent. For the northern supercontinent he introduced the term 'Laurasia', derived from Laurentia (the eastern North American shield) and Asia. Additionally, the book contained a discussion of Arthur Holmes's suggestion of fracturing of continental blocks by subcrustal convection movements to explain spreading. Du Toit also included a figure showing the development of continental rifting that owed something to Holmes's famous figure of 1929, representing convection and continental fracture. But du Toit doubted whether subcrustal convection was 'wholly competent to account for continental drift'.

The evidence for the former linkage of the southern continents was illustrated in several convincing diagrams (Figures 3, 4, and 5), and the structural correspondences between western Europe and North America were also depicted, along with a suggested pattern for the opening of the North Atlantic (see Figures 6 and 7).

Throughout the book, du Toit challenged the geological community to accept his theories, but he did not live long enough to witness the acceptance of continental drift nearly two decades later.

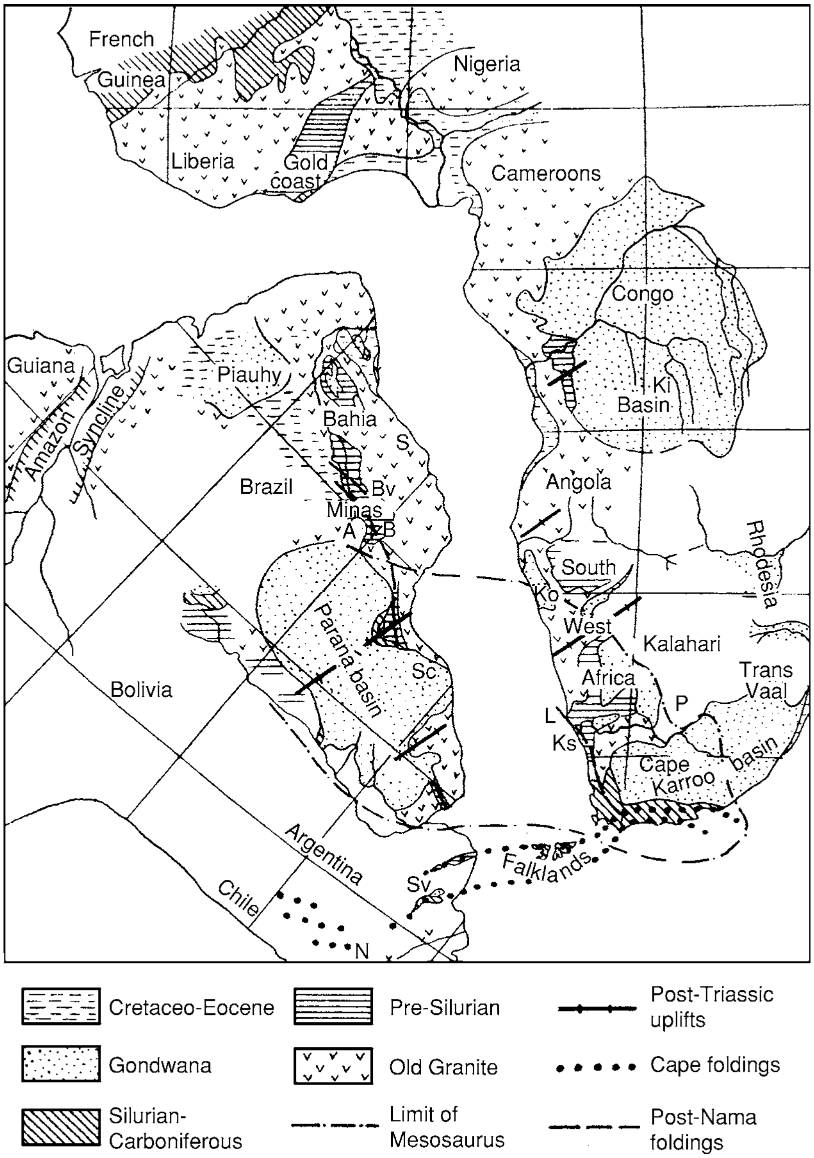


Figure 3 Analogies between South America and southern Africa, according to du Toit. (Reproduced from Du Toit AL (1937) *Our Wandering Continents: A Hypothesis of Continental Drifting.* Edinburgh: Oliver and Boyd.)

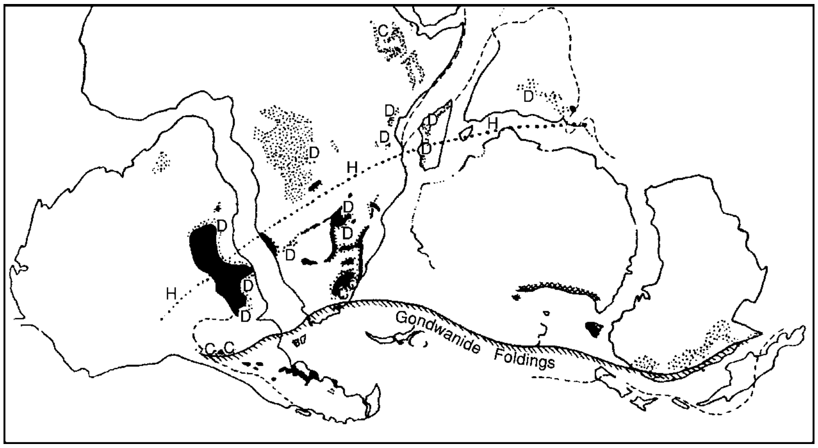


Figure 4 Structural features in 'restored' southern continents, Late Triassic and Rhaetic, according to du Toit (1937: 93).

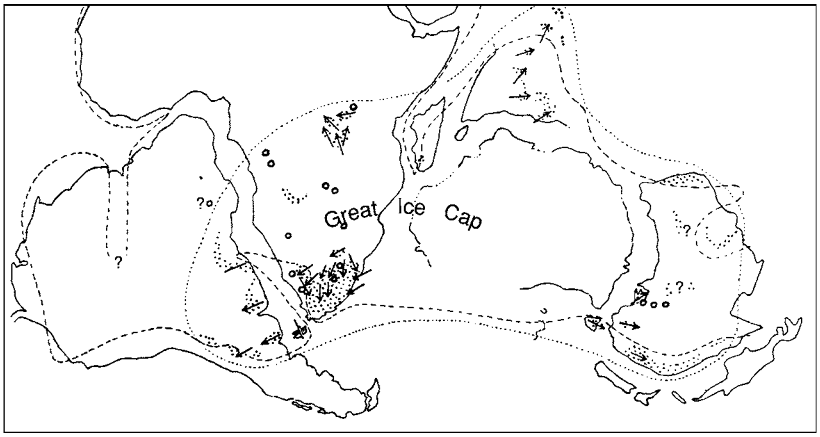


Figure 5 Arrangement of continents and areas of glaciation in the Late Carboniferous, according to du Toit (1937: 76).

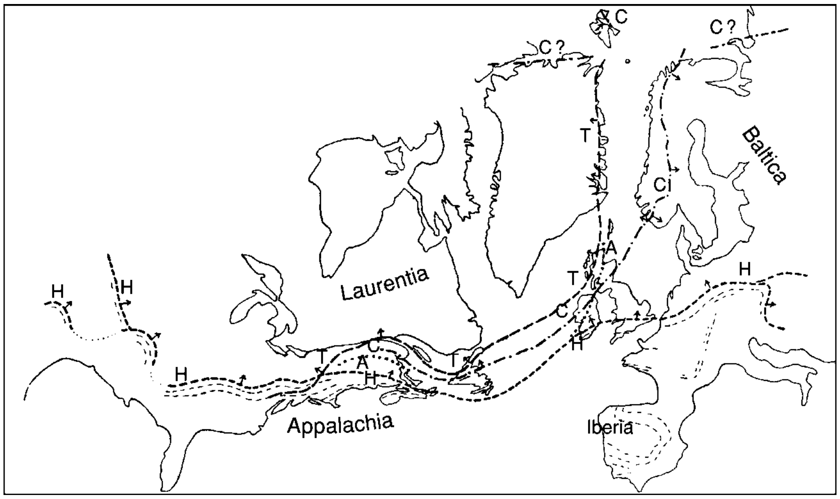


Figure 6 Structural analogies between North America and Europe, according to du Toit (1937: 145).

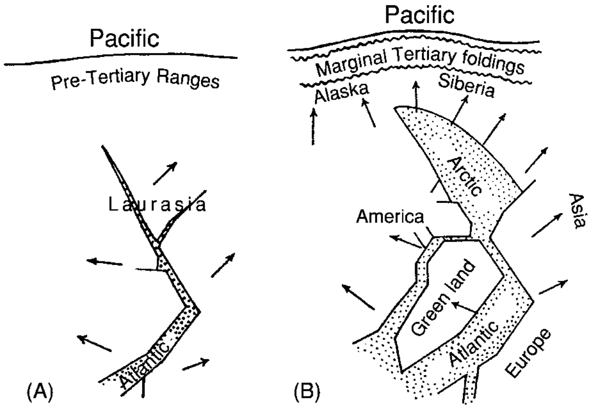


Figure 7 Evolution of the Atlantic Arctic rift, according to du Toit (1937: 222).

Retirement and Honours, 1941–1948

After retiring from De Beers, Alexander and Evelyn returned to Cape Town, where du Toit embarked on a frenzied round of activity, as his diaries for 1945 and 1946 reveal. He corresponded with friends and colleagues in South Africa and other countries, made social calls, and undertook short visits to the countryside. He continued writing scientific papers and assisted aspiring authors. His final diary entry (31 December 1946) contained a vague but ominous reference to a medical problem that was to carry him away on 25 February 1948.

During his lifetime, du Toit was awarded many medals and honorary doctorates from five South African Universities. His richly deserved Fellowship of the Royal Society of London was not awarded until 1943, possibly delayed by English prejudice against his continental drift ideas. The greatest honour of all came when the Geological Society of South Africa instituted the biennial Alex L du Toit Memorial Lecture Series in 1949. A few years later, his image appeared on a South African stamp.

A collection of his notebooks is owned by the University of Cape Town. The diaries for 1905, 1906, 1908, 1931, 1945, and 1946 are kept by the South African Museum in Cape Town. A cache of photographs and documents, together with his academic gown, was recently donated to the Museum. His awards, medals, and certificates are in the hands of a grandson.

See Also

Famous Geologists: Suess; Wegener. **Gondwanaland and Gondwana**. **History of Geology From 1835 To 1900**. **History of Geology From 1900 To 1962**. **Pangaea**. **Plate Tectonics**.

Further Reading

Branagan DF (2004) *The Knight in the Old Brown Hat: A Life of Sir T W Edgeworth David, Geologist*. Canberra: National Library of Australia.

Du Toit AL (1921) Land connections between the other continents and South Africa in the past. *South African Journal of Science* 18: 120 140.

Du Toit AL (1921) The Carboniferous glaciation of South Africa. *Transactions of the Geological Society of South Africa* 24: 188 227.

Du Toit AL (1926) *Geology of South Africa*. Edinburgh: Oliver & Boyd.

Du Toit AL (1927) *A Geological Comparison of South America with South Africa. With a Palaeontological Contribution by F R Cowper Read*. Washington: The Carnegie Institution.

Du Toit AL (1937) *Our Wandering Continents: A Hypothesis of Continental Drifting*. Edinburgh: Oliver & Boyd.

Gevers TW (1949) *The Life and Work of Alex L du Toit*. Alex L du Toit Memorial Lecture 1. Johannesburg: Geological Society of South Africa.

Haughton SH (1949 [1950]) Memorial to AL du Toit (1878 1948). *Proceedings of the Geological Society of America*: 141 149.

Haughton SH (1949) Alexander du Toit 1878 1948. *Obituary Notices of Fellows of the Royal Society of London* 6: 385 395.

Holmes A (1929) Radioactivity and Earth movements. *Transactions of the Geological Society of Glasgow* 18: 559 606.

Rogers AW and du Toit AL (1909) *An Introduction to the Geology of Cape Colony With a Chapter on the Fossil Reptiles of the Karroo Formation by Prof R Broom*, 2nd edn. London: Longmans.

Waterschoot van der Gracht WAJM (ed.) (1928) *The Theory of Continental Drift: A Symposium*. Tulsa: American Association of Petroleum Geologists.

Hall

R H Dott Jr, University of Wisconsin, Madison, WI, USA

Introduction

James Hall (1811–1898) of New York was North America's pre-eminent invertebrate palaeontologist and geologist of the nineteenth century. That he was a giant among early American geologists is evidenced by the facts that he served as President of the American Association for the Advancement of Science (1856), was a charter member of the National Academy of Sciences (1863), and was chosen to be the first President of the Geological Society of America (1889). Hall was also the best-known American geologist on the international scene in his time. As early as 1837, he was elected to membership of the Imperial Mineralogical Society of St Petersburg. Later he was the Organizing President of the International Geological Congress meetings in Buffalo, New York (1876) and in Paris (1878); he was a Vice-President of the congresses in Bologna (1881) and Berlin (1885), and he was Honorary President of the Congress in St Petersburg (1897). Hall was elected Foreign Correspondent to the Academy of Sciences of France in 1884, being its first English-speaking member. It was primarily the *Paleontology of New York*, published in 13 volumes between 1847 and 1894, that initially brought Hall his fame. However, the broader community of geologists chiefly remembers him more for the curious theory of mountains presented in his Presidential Address to the American Association for the Advancement of Science in 1857.

Early Life and Education

Hall was born near Boston in Hingham, Massachusetts on 12 September 1811. His parents had emigrated from England two years earlier, and James was their first of five children. The family was of modest means, but the young Hall was fortunate to have a gifted teacher in his public school, who stimulated an interest in nature. Through his teacher, James met several leading members of the Boston Society of Natural History. Having developed a strong interest in science, Hall was attracted to a new college in Troy, New York, which emphasized science and employed revolutionary new approaches to learning, with an active role for the student coupled with hands-on laboratory and field-trip instruction. This Rensselaer Plan was developed by Amos Eaton, with financial backing from his patron, Stephen van Rensselaer. Unable to afford commercial transportation, Hall walked the two hundred miles to Troy. At Rensselaer, he was instructed by Eaton and Ebenezer Emmons, and had for classmates such geologists-to-be as Douglas Houghton, Abram Sager, Eben Horsford, and Ezra Carr. Hall graduated with honours in 1832 and undertook a tour on foot to the Helderberg Mountains in south-eastern New York to collect Silurian and Devonian fossils. A job as a librarian allowed him to continue at Rensselaer for another year and to earn a Master of Arts degree with honours (1833). He then held an assistantship in chemistry for several more years. In 1838, he married Sarah Aikin, the daughter of a Troy lawyer; they had two daughters and two sons. Sarah died in 1895.

The New York Survey

In 1836, the New York legislature authorized a 4-year geological and natural history survey; an extension of 2 years was later authorized. Four men – William W Mather, Ebenezor Emmons, Timothy A Conrad, and Lardner Vanuxem – were in charge of four respective districts, and Lewis C Beck was mineralogist for the survey. Botanist John Torrey and zoologist James De Kay conducted the biological survey. Hall was engaged to assist his former teacher, Emmons, in the Second District in north-eastern New York, where his first assignment was to study iron deposits in the Adirondack mountains. A year later the districts were revised; Conrad was appointed State Paleontologist, and Hall was put in charge of a new Fourth District in western New York, with former Rensselaer students Horsford, Carr, and George W Boyd as his assistants. When the survey ended in 1841, only Hall and Emmons remained in New York. Hall became State Paleontologist, and Emmons became State Agriculturalist.

Lardner Vanuxem, who had studied in France, had been instrumental in introducing to America the value of fossils for subdividing strata and correlating those of similar age from place to place based upon similar fossils. Meanwhile, Timothy Conrad had gained a reputation for studies of Cenozoic fossils of the coastal plain. Thus the survey had strength in palaeontology from the start, and its staff soon developed a New York stratigraphy that set the precedent for naming stratigraphical divisions after geographical localities. Young Hall's career blossomed quickly after the monograph on the fossils and stratigraphy of the Fourth District was published in 1843. This and the other survey reports soon aroused much interest in Europe, where Palaeozoic fossils were being used to define stratigraphical subdivisions during the mid-nineteenth century. For example, Roderick Murchison's *Silurian System* appeared in 1839 (*see* **Famous Geologists:** Murchison), John Phillips's Palaeozoic Series was proposed in 1840, and Joachim Barrande's monographs on the lower Palaeozoic fossils of Bohemia began to appear in 1852. These and other authors began corresponding with Hall, and European geologists began beating a path to Albany – most notably Charles Lyell (*see* **Famous Geologists:** Lyell) during several American visits in the 1840s. During a visit in 1846, Eduard de Verneuil, a close associate of Murchison, tried to persuade Hall not to introduce the name Cambrian to the New World, but rather to use only Silurian for the lowest Palaeozoic strata – a reflection of the famous Murchison–Sedgwick feud then raging in Britain. Hall, however, was not swayed, for he was a leading exponent of the widely held 'nationalistic' view that an American stratigraphical classification was best for America.

As geological investigations in America began to mature, stratigraphical nomenclature was becoming important, especially for comparisons among different regions. Hall and others proposed that an organization be created to deal with nomenclature and other mutual problems, and so in 1838 in Albany the American Association of Geologists was created; the first formal meeting was held in Philadelphia in 1840. From this organization evolved in 1857 the American Association for the Advancement of Science, which was modelled on the British Association. Still later, the Geological Society of America was spawned in 1888 from a division of the American Association for the Advancement of Science. Hall was promptly elected President (Figure 1).

Figure 1 James Hall in 1856 at the peak of his career and only one year before his famous Presidential Address to the American Association for the Advancement of Science, in which he first proposed his theory of mountain formation. (From Clarke JM (1921) *James Hall of Albany Geologist and Paleontologist, 1811 1898.* C. Ayer Company Publishers.)

The Albany Training Ground

In 1857, Hall constructed a substantial brick laboratory building, where he worked for the rest of his life. This Albany laboratory became a veritable training school for a host of budding geologists who would distinguish themselves in the history of American science. Although universities were beginning to offer formal instruction in geology during the mid-nineteenth century, there was practically no instruction in palaeontology. As a result, apprenticeship was the principal route into that field, and Hall's laboratory was the place to apprentice. Among the many who profited from association with Hall were Charles E Beecher, Ezra S Carr, John M Clarke, Nelson H Darton, Grove K Gilbert, Ferdinand V Hayden, Eban N Horsford, Joseph Leidy, W J McGee, Fielding B Meek, Charles S Prosser, Carl Rominger, Charles Schuchert, Charles D Walcott, Charles A White, Robert P Whitfield, Josiah D Whitney, Charles Whittlesey, and Amos H Worthen.

Hall's assistants learned more from him than just palaeontology, however, for they also experienced a strong, egotistical, and irascible personality. Although his sharpest attacks were reserved for his legislative enemies, most of his assistants were also treated to his notorious outbursts. In addition to throwing vituperative verbal daggers, he sometimes brandished menacingly a stout cane or even a shotgun, kept at the ready near his desk. Perhaps the most extreme self-righteous attack was upon James T Foster, a school teacher in Greenbush, New York. Foster had the audacity to publish a popularized geological chart in 1849. Hall was so outraged that he stole aboard a boat bound for New York City and threw the entire printing of the offensive chart into the Hudson River. He had quite a time fighting the subsequent libel suit, which entangled him, Louis Agassiz (*see* **Famous Geologists:** Agassiz), James D Dana, and several other notables for several years.

Another celebrated example of Hall's irascible temper involved the prominent British geologist Charles Lyell, during his first visit to America in 1841–42. At first Hall and others were greatly flattered by the attentions of their famous visitor, but Lyell's insatiable questioning, which earned him the nickname 'Pump', and his copying of the Americans' geological maps gradually provoked resentment and a fear of being pre-empted. In March 1842, an anonymous letter signed 'Hamlet' appeared in a Boston newspaper, charging Lyell with geological piracy. It was written by Hall after some of his compatriots criticized him for being too generous in sharing the results of his research with Lyell, especially by giving him a copy of his as yet unpublished *Geologic Map of the Western and Middle United States*. Needless to say, this letter cast a chill upon the Association of American Geologists' meeting a month later, but Lyell participated as if nothing had happened. Although the charge was largely true, Hall was afterwards mortified by his rash act. For once, however, he managed to mend the damage done by his intemperate action, and he remained thereafter on good terms with Lyell.

Almost as legendary as his paranoiac outbursts was Hall's acquisitiveness for fossils. He employed every conceivable means to acquire outstanding collections. An effective technique was to flatter and invite collectors to work with him in Albany and to bring their collections. Commonly, however, when the apprentice moved on his collection did not. Hall was a workaholic who drove himself as mercilessly as he did his assistants. He could rarely say 'no' to even the most ridiculous schemes, and he ignored the entreaties of friends to ease his pace for the sake of his health.

Beyond New York

As he completed his Fourth District studies, Hall decided to see how far the New York stratigraphical classification might apply beyond his state. In 1841, he made the first of several odysseys west. With geologist David Dale Owen he made a boat trip down the Ohio River to Owen's base at New Harmony, Indiana, and, from there, he proceeded across Illinois to Missouri, Iowa, and Wisconsin. Hall was amply rewarded with evidence to support the extension of the New York stratigraphy in a broad way across the entire region. There were some significant differences, however, which he, and perhaps only he, could recognize. For example, he found that the Palaeozoic strata were much thinner to the west of New York and that there were important differences in sedimentary facies, with more clastic sediments in the east and more carbonate strata to the west. In effect, Hall had discovered the contrast between what would, much later, be termed the stable craton and the Appalachian orogenic belt. This trip also provided information to allow him to complete the *Geologic Map of the Middle and Western States*, which was incorporated into Hall's Fourth District report of 1843 – the map that Lyell had used to help prepare his own geological map of the then United States, which was published in 1845 in *Travels in North America*.

Hall's finances were always tenuous. He was easily drawn into risky ventures and also had his salary cut, or even suspended, by a frequently hostile state legislature. At least once he had to sell some of his fossil collections in order to raise money. As his reputation grew, however, opportunities for temporary outside employment helped to tide him over his New York financial droughts. These ventures also allowed him to expand his knowledge widely. One of the first such ventures took him to the Lake Superior region in 1845 to examine copper deposits for a private company. In 1847, the Federal Government authorized a geological survey to evaluate the mineral resources of northern Michigan and Wisconsin. In 1850, Hall was engaged to provide his expertise on Palaeozoic stratigraphy and palaeontology for that survey. He made two brief trips to the area (1850 and 1851), from which he gained further insights into the stratigraphy of the Great Lakes region and added to his ever-growing fossil collections. Perhaps the most important result of his work for this survey, however, was the recognition of fossil reefs in the Silurian strata of south-eastern Wisconsin. This was the first recognition of ancient reefs in North America, and perhaps in the world.

When asked to study fossils from western regions, which others had collected during various expeditions, he willingly obliged. He recognized the first known Mesozoic fossils collected by John C Fremont in the 1840s. In 1853, he agreed to let his assistants Fielding B Meek and Ferdinand V Hayden go to the White River badlands of Nebraska Territory (now in South Dakota) to collect newly discovered Cenozoic non-marine invertebrate and mammalian fossils. Meek, whose artistic as well as collecting skills were vital to Hall's enterprise, was glad to escape from his mentor for a few months. Eventually he extricated himself from Hall's empire to join the new United States Geological Survey. Meek never forgave his perceived exploitation by Hall.

When Iowa decided to undertake a geological survey in 1855 and needed a director, the Governor looked to New York, which had eclipsed all other states as well as the Federal Government in the calibre of its geological survey. Hall accepted the position with alacrity as his New York salary had been suspended in 1850 by a more than usually hostile legislature. Moreover, he welcomed the opportunity to obtain and study fossils from the new state. He soon suggested Amos Dean of Albany as the first Chancellor of the University of Iowa. Hall himself was identified as the first Professor of Geology but apparently he never lectured there. In fact, Hall mostly directed the survey from Albany and spent little time in Iowa. Four assistants did most of the actual work: Josiah D Whitney concentrated upon mineral resources, while Amos H Worthen of Illinois dealt with palaeontology, assisted by F B Meek and R P Whitfield. Hall knew that Worthen had the finest collection of crinoids in the country, so a condition of his employment was that Hall be allowed to describe them, which he did in the Iowa Survey report. Hall came to Iowa for the winter meetings of the legislature to lobby on behalf of the Survey, but payment of salaries was so erratic that he had to borrow money in Albany to keep the effort going. Finally in 1859 the survey was suspended, but two volumes had appeared in 1858.

In 1857, Illinois undertook a geological survey, and Worthen was one of three applicants to direct it. Hall wrote a glowing endorsement of him, but he also supported the other two applicants. This lapse of judgement earned him the animosity of all three applicants, and, in the end, Hall was denied access to the fossils collected by the Survey, which was a great disappointment to him.

While still working in New York and Iowa and for the Canadian Geological Survey, in 1856 Hall accepted an affiliation with Wisconsin. He joined a former Renssalaer colleague, Ezra Carr, now a professor at the University of Wisconsin, and Edward Daniels for this new effort. Hall devoted little time to the Wisconsin initiative, so Carr and Daniels were really in charge. Whitney was engaged to study the lead deposits of south-western Wisconsin, and Charles Whittlesey was engaged to study the mineral deposits of northern Wisconsin. A large volume was published in 1862, but a hostile Wisconsin legislature abruptly terminated the endeavour because it judged the results to be insufficient. It cared only about potentially economic results, so a frustrated Hall and his assistant, Robert P Whitfield, published Wisconsin's palaeontology within a New York report in 1867 (and again separately in 1871). This ingenious solution to a publication problem was typical of Hall. Much earlier he had circumvented a New York legislative edict to limit the number of expensive palaeontological monographs published simply by issuing several volumes as subdivisions of a single Part of the series, resulting ultimately in 13 separate monographs – at least twice the intended limit.

Hall became involved in several other state surveys in various capacities, ranging from advising about personnel to acting as a consultant for palaeontology or the titular head of a survey. Included were surveys of Missouri (1853 and 1871) and California (1853–1856), the transcontinental railroad survey (1853–1857), and surveys of New Jersey (1854–1857), Ohio (1854–1857), Texas (1858), Mississippi (1858), Michigan (1869–1870), and Pennsylvania (1870–1875). While this list is a testimony to his prominence, Hall's contributions to these many surveys were minor except for the identification of fossils.

Hall made his last trip to the Midwest in 1889, at the age of 77, while first President of the new Geological Society of America. His purpose was to obtain brachiopods by any and all means necessary for his latest project, namely to revise the description and classification of that great group of Palaeozoic fossils. In addition to successfully obtaining many specimens, he also met and lured to Albany a young Charles Schuchert of Cincinnati, who was destined to become his most famous protégé and a professor at Yale. The ambitious brachiopod study culminated in the last volume, Part 8, of the *Paleontology of New York*, which appeared in 1894.

During the completion of the brachiopod monograph, Hall had his last and sweetest wrangle with New York bureaucracy. The Executive Secretary of the Regents, which oversaw his programme, had become overly zealous in trying to impose strict accounting and efficiency procedures. Such a fuss developed that the legislature had to intervene. To resolve the fracas, it appointed crotchety old Hall as State Paleontologist and State Geologist for life, with complete managerial freedom. Doubtless the legislators realized that Hall's days were numbered, and, in fact, he died 3 years later. Hall must have recalled with satisfaction an earlier observation, when a particularly vicious political enemy died suddenly, that "Providence was usually on my side".

The Origin of Mountains

Hall is most widely known for his theory of mountains, which embodied the concept of the geosyncline, a term coined not by Hall but by James D Dana in 1873. In his 1857 Presidential Address to the American Association for the Advancement of Science, Hall startled his audience with a discourse on the origin of mountains rather than speaking about palaeontology and stratigraphy. In stating that "the greater the accumulation, the higher will be the mountain range", he pronounced that a great thickness of strata was a prerequisite for a mountain range composed of folded strata. Hall rejected the then popular theories of mountains of Frenchman Élie de Beaumont – that mountains formed as a result of global cooling and contraction – and the American brothers William B Rogers and Henry D Rogers, who postulated that catastrophic wrinkling of the crust resulted from wave-like movements in a fluid subcrustal zone. Instead, Hall was influenced by a suggestion by the British astronomer John F W Herschel in 1836, which anticipated the modern theory of isostasy. Herschel argued that vertical movements of the crust are caused by changes in pressure and heat at depth, which in turn are the result of erosion and deposition at the Earth's surface. The vertical adjustments towards gravitational equilibrium were accommodated by a pliable subcrust. The key element for Hall was the accumulation of thick sedimentary layers, which he imagined must depress the crust and, in the process, become wrinkled to form the structures seen in ranges such as the Appalachians. He envisioned compression of the upper layers and tension of the lower ones during subsidence – much as occurs when bending a ream of paper (Figure 2).

In 1859, Hall published the following, in the most commonly quoted source for his theory, Part 6 of the *Paleontology of New York* (*Volume 3: Descriptions and Figures of the Organic Remains of the Lower Helderberg Group and the Oriskany Sandstone. 1855–1859*, pp. 70–73. Albany: New York State Geological Survey):

> ...[t]he line of greatest depression would be along the line of greatest accumulation [that is] the course of the original transporting current. By this process of subsidence ... the diminished width of surface above caused by this curving below, will produce wrinkles and folding of the [upper] strata. That there may be rents or fractures of

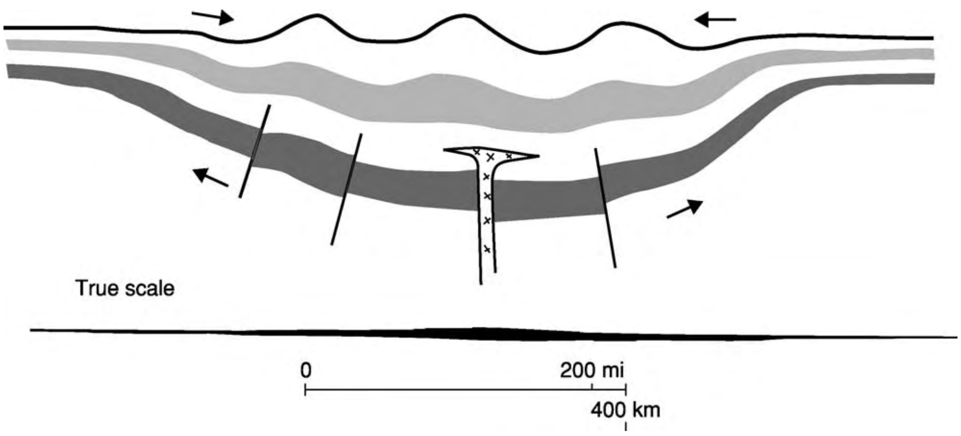


Figure 2 James Hall's theory of down warping resulting from sedimentation. The upper layers are crumpled as their circumference diminishes, whereas the lower layers are broken by tension, which allows dykes to be intruded from below. Hall never published diagrams of his theory, so this was constructed from his verbal discussion. Modified with permission from Dott RH Jr (1985) James Hall's discovery of the craton. In: Drake ET and Jordan WM (eds.) *Geologists and Ideas: A History of North American Geology*, pp. 157 167. Centennial Special, Volume 1. Boulder: Geological Society of America.

> the strata beneath is very probable, and into these may rush the fluid or semi fluid matter from below, producing trapdykes, but the folding of strata seems to be a very natural and inevitable consequence of the process of subsidence.

A year earlier, in the report of the Iowa Survey (1858), Hall had also emphasized the contrasts in thickness between the Appalachian region and the Midwest, with detailed remarks about contrasting facies as well as thicknesses in various portions of the Palaeozoic successions of the two regions. Here, too, he included a brief summary of his theory of mountains by stating that "The thickness of the entire series of sedimentary rocks, no matter how much disturbed or denuded, is not here great enough to produce mountain features" (Vol. 1, p. 42).

Hall was vague about the cause of mountain uplift. He simply ascribed it to continental-scale elevation, which he thought had no direct relation to the folding of strata within the mountains and which he did not attempt to explain. Contemporaries were quick to challenge him on this point, with Dana noting that Hall had presented "a nice theory of mountains with the mountains left out". Hall lamely denied that he ever intended to offer a complete theory of mountain building. His failure to publish the Presidential Address until 1883 may have been because of such criticisms, but his first priority was always palaeontology and he knew that the essence of his theory was to appear in both the Iowa and New York reports (as well as in an abstract in Canada) soon after his oral address.

Hall's contribution to mountain-building theory was marginal at best and was soon eclipsed by Dana's more profound and comprehensive contraction theory, which postulated that thick strata were a result of mountain-building processes rather than the cause. Nonetheless, Hall's emphasis on a cause-and-effect relationship between orogenic belts and very thick strata had a significant influence on three generations of geologists. By coining the term 'geosynclinal', which was later converted to the noun 'geosyncline', Dana formalized Hall's demonstration that Palaeozoic strata are ten times thicker in the Appalachian Mountains than in the lowlands to the west (the craton).

Even though Hall was wrong about the cause of mountain building, he was nevertheless the first person to underscore the profound stratigraphical contrasts between orogenic belts and what are now termed stable cratons. He drew attention at an early stage to large-scale stratigraphical patterns among some of the larger tectonic elements of the Earth's crust and had other shrewd stratigraphical insights that were ahead of the times. By virtue of his breadth of experience in both the cratonic and the orogenic regions of eastern North America, he was uniquely equipped to see this fundamental distinction. Coupled with his prodigious contributions to palaeontology, this assured James Hall of a prominent niche in the history of geology.

See Also

Analytical Methods: Geochronological Techniques. **Famous Geologists:** Agassiz; Lyell; Murchison; Sedgwick. **Geological Surveys**. **Stratigraphical Principles**. **Tectonics:** Mountain Building and Orogeny.

Further Reading

Clarke JM (1921) *James Hall of Albany   Geologist and Paleontologist, 1811 1898*. Albany: Privately printed.

Dana JD (1873) On some results of the Earth's contraction from cooling, including a discussion of the origin of mountains, and the nature of the Earth's interior. *American Journal of Science, 3rd series* 5: 423 495; 6: 6 14; 104 115; 161 172; 381 382.

Dott RH Jr (1979) The geosyncline   first major geological concept 'Made in America':. In: Schneer CJ (ed.) *Two Hundred Years of American Geology,* pp. 239 264. University Press of New England: Durham, New Hampshire.

Dott RH Jr (1985) James Hall's discovery of the craton. In: Drake ET and Jordan WM (eds.) *Geologists and Ideas: A History of North American Geology,* pp. 157 167. Centennial Special, Volume 1. Boulder: Geological Society of America.

Fisher DW (1978) James Hall   patriarch of American paleontology, geological organizations, and state geological surveys. *Journal of Geological Education* 26: 146 152.

Hall J (1842) Notes upon the geology of the western states. *American Journal of Science and Arts,* 1st series, 42: 51 62.

Hall J (1989) *The Natural History of New York. Part 6. Palaeontology of New York. Vol 3. Descriptions and Figures of the Organic Remains of the Lower Helderberg Group and the Oriskany Sandstone. 1855 1859.* Albany: New York State Geological Survey.

Hall J (1883) Contributions to the geological history of the North American continent. *Proceedings of the American Association for the Advancement of Science* 31: 24 69.

Hall J and Whitney JD (1858) *Report on the Geological Survey of the State of Iowa.* Des Moines: Legislature of Iowa.

Hutton

D R Oldroyd, University of New South Wales, Sydney, Australia

Introduction

The Scottish geologist, physician, farmer, philosopher, chemist, businessman, and industrialist James Hutton (1726–1797) is commonly regarded as the 'founder of modern geology', though a similar claim could be made for some others, and in some ways Hutton's thinking was not at all modern by today's standards.

Hutton's Early Career and the Beginning of His Interest in Geology

Hutton was born in Edinburgh, the son of a businessman who served for a time as City Treasurer. The young Hutton went to Edinburgh University at the age of fourteen, where he studied humanities, attended the mathematics lectures of Colin Maclaurin, and acquired a keen interest in chemistry. After his time as a student, he was briefly apprenticed to a solicitor, but eventually decided to study medicine. After taking the Edinburgh course, he went to Paris in 1747 and thence to Leiden, where he submitted a doctoral thesis in 1749. This dealt with the circulation of the blood and matters of human physiology and had a distinct chemical slant. The thesis title referred to the human body as the 'microcosm', which was traditionally regarded as having analogies with the Earth or with the whole cosmos (the 'macrocosm'). The thesis may have been the seed from which sprang Hutton's later cyclic theory of the Earth.

On returning to Britain, Hutton did not take up medicine. Instead, he went into partnership in an industrial process for extracting sal ammoniac (ammonium chloride) from soot. But Hutton sired a son, probably out of wedlock, and 'tactfully' left Edinburgh for several years to pursue a career in agriculture on two farms in Berwickshire, which he had inherited from his father. He wished to do his farming on a scientific basis, so he went to East Anglia to study the latest methods of agriculture, which he subsequently brought to Scotland. During his two years away Hutton travelled extensively and, as he became increasingly interested in the Earth, recognized the ubiquity and perpetuity of erosion and deposition and that sedimentary rocks were consolidated sediments.

Hutton worked his farms himself and experimented with agricultural techniques. Under Maclaurin, he had become acquainted with the principles of 'deism', and he had apparently lost his Christian faith at an early stage of his life. According to the deist view, God had created the Earth 'in wisdom' as a suitable place for human habitation. The existence of God was not known by courtesy of Jesus Christ, the Bible, the Church, or any other agent of revelation, but by human reason. For Hutton, divine design was manifest in Nature itself, both in the way organisms functioned and were structured and in the way the Earth was apparently well 'contrived' for human existence (with air, water, soil, animals, plants, etc. all suited to us).

But it was obvious that soil was constantly being washed into the sea, and, since it was essential for human well-being, it had somehow to be replenished. As a deist, rather than a biblical literalist, Hutton could take a grand view of time. The Earth could be millions of years old, but in that case the land would eventually be eroded to a plain and the good soil would end up as sediments in the seas. So Hutton asked himself how high ground could be regenerated to provide a source of new soil.

Hutton's Theory of Cyclic Earth Processes

In 1764, Hutton made a journey into the Highlands and began to collect geological information and specimens in a systematic manner. His farms were by then profitable and the sal ammoniac business was prospering. So he began to think of returning to Edinburgh, now as a gentleman–farmer. Probably in 1767, he rented out his farms and returned to Edinburgh (his old scandal had by then been forgotten or forgiven) to enjoy the pleasures of intellectual life in one of the great cities of the eighteenth-century Enlightenment.

Among Hutton's new friends were the economist Adam Smith, the chemist Joseph Black, who conducted experiments on heat, and the engineer and steam-engine inventor James Watt. It is likely that Watt's engines encouraged Hutton to think of heat as an agent of geological change. Perhaps the Earth had a central source of heat that might somehow drive the cycle essential for a theory of the Earth that provided for a renewal of soil? The Earth's internal heat could be analogous to the fire of Watt's engine, which drove the complicated mechanism of the engine and the machinery of a factory. But Hutton did not imagine that the Earth's internal heat was due to combustion.

Hutton's theory of the Earth was, then, developed as follows. The Earth, he thought, had a central reservoir of heat, the source or means of maintenance of which was unspecified. Rocks on the surface were broken down by weathering and erosion to form soils. Sediment was deposited in the seas by rivers, which also carved valleys. Sediments accumulated in layers on the ocean floors, and the lower layers were compressed and consolidated by the sediments deposited on top of them, assisted by the Earth's internal heat. The rock-salt deposits of Cheshire seemed to Hutton to have been melted at some time. Likewise, the grains of sand in quartzites seemed to show evidence of fusion at their edges in the process of consolidation by heat.

In time, the consolidated materials, under pressure, might become so hot that they would melt. Veins of crystalline rock, dykes or sills, could be emplaced. Moreover, Hutton supposed, great masses of molten material (which we would call magma) could be intruded into the Earth's crust, heaving it up. On cooling, this magma might crystallize to form subterranean masses of granite, which might subsequently be exposed by weathering and erosion. Thus the land would be renewed and Hutton's Earth, 'designed in wisdom', would continue indefinitely as a place suited to human habitation. The upheaval of strata was confirmed by the presence of marine fossils in strata well above sea-level. However, at the time of the first public presentation of his theory, Hutton appeared to have personal knowledge only of mineral or metallic ore veins, not granitic veins, and he did not then describe any personal examinations of large granitic bodies. He went looking for these systematically only after the preliminary presentation of his ideas.

Be that as it may, the whole process envisaged by Hutton was cyclic, for the upheaved strata would be eroded to form a new surface, on which other sediments might subsequently be deposited. So one might hope to find places where the lower layers were inclined to the horizontal and the overlying ones lie over them horizontally. Such a structure came to be known as an unconformity, and the subsequent discovery of unconformities was considered a triumph for Hutton's theory, as he apparently had the idea of such structures before he actually saw them. An unconformity could be taken to mark the end of one cycle and the commencement of the next.

Hutton's cyclic Earth processes were continuous and open ended. He did not say that the Earth was infinitely old, but as he put it in a famous sentence: 'we find no vestige of a beginning—no prospect of an end' Hutton (1788 p. 304). His cycle has been called the 'geostrophic cycle' (see Figure 1, which explicates Hutton's notion of unconformity).

Hutton's theory was formally read before the Royal Society of Edinburgh in 1785 and published in 1788. It appeared in expanded form in two volumes in his *Theory of the Earth* in 1795. Two further incomplete volumes remained unpublished in his lifetime, but the manuscripts were found in the nineteenth century and published in 1899 as Volume 3. This book described Hutton's fieldwork after the presentation of his 1785 paper.

Hutton's 1785/1788 paper did not explain the Earth's internal heat, but he tried to use field specimens to support its existence. Many of the materials that bind sediments together, such as calcareous spar, silica, etc., are not themselves water soluble.

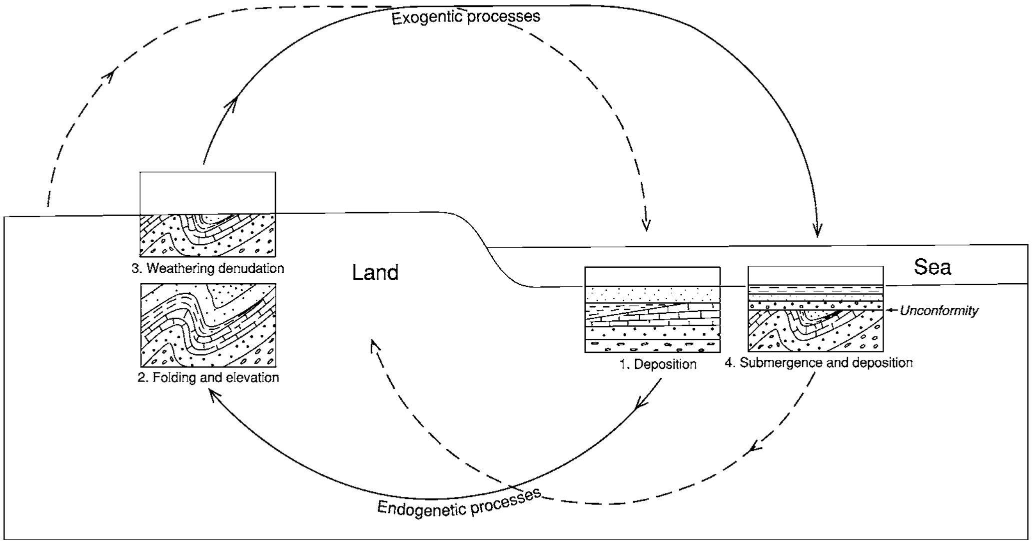


Figure 1 Representation of the 'geostrophic cycle'. Reproduced by permission of The Geologists' Association from *Proceedings of the Geologists' Association*, Tomkeieff SI, Unconformity an historical study 1962, **73**, pp. 383 417, fig. 6. © 1962 The Geologists' Association.

Therefore, Hutton argued, water could not have been the prime agent causing their consolidation. But heat could penetrate into bodies and, by fusion, could cause consolidation. So, he thought, heat must have been responsible for the penetration of sediments by flint nodules (which were injected while molten). Likewise, nodules containing crystalline spar that did not extend to their outer surfaces could not, Hutton supposed, have acquired the crystalline matter by transmission of solutions. Hutton also exhibited a specimen of 'graphic' granite, which contained crystals of quartz within feldspar within quartz. Such a texture could not, he maintained, have been produced by crystallization from aqueous solution.

There were evidently gaps in the evidence for the cyclic chain of Hutton's theory. Hutton argued that there had to be heat within the Earth and there had to be some means of elevation, even if he did not know precisely how that process worked. The hot interior was supported by the evidence of volcanoes, of course, and mines seemed to have higher temperatures at greater depths (but that was not proven by careful measurements until the nineteenth century).

Geological Evidence to Support Hutton's Theory

So Hutton's 1785/1788 paper was not in itself sufficient to persuade all his auditors or readers. At the time of its presentation, he had not recorded observations of veins of granite penetrating other rocks, nor, so far as we know, had he discovered any unconformities (although they had been reported by others without their theoretical significance being recognized). But following the public presentation of his ideas Hutton made excursions to various parts of Scotland to look for confirmatory field evidence.

In September 1785, Hutton went into the Grampians to hunt for contacts between granite and surrounding rocks into which it might have been injected while in a fused state. He was accompanied by a friend, John Clerk of Eldin, who made excellent drawings of what they saw. Hutton thought that he would find the evidence he wanted to the west of the mass of Aberdeen granite. But he may have received some hints of where to look (possibly from Clerk), for they headed directly for the valley of the River Tilt, which runs north-east from Blair Atholl. Complicated outcrops of limestones and schists were found in the valley floor, and not far up the glen they came across fine exposures of granitic veins, which sometimes cut across the country rock and elsewhere could be seen anastomosing between, or across, the laminae of the country rock. Hutton got so excited that his guides imagined that 'nothing less than ... a vein of silver or gold ... could call forth such strong marks of joy and exultation'! The granitic veins were also traced back to the large mass of granite on the north side of the glen. Hutton's joy was, of course, due to the fact that he had found what he had predicted on the basis of his theory, and the geometry of the veins was compatible only with the granite having worked its way into the country rock from below and from the granite mass.

In 1787, Hutton visited the Isle of Arran, which has a large mass of granite at its northern end, with

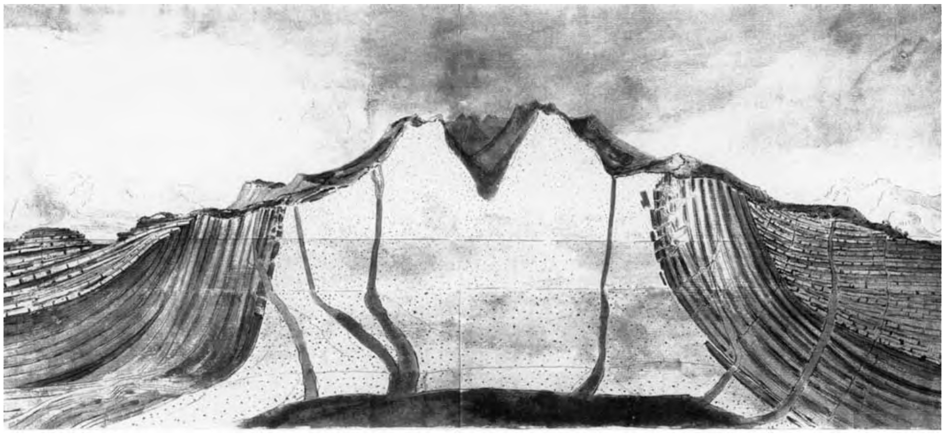


Figure 2 Cross section of the northern part of Arran, drawn by John Clerk Jr (1787). Reproduced by permission of Sir Robert Clerk of Penecuik.

surrounding tilted-up layers of schist and beds of sandstones and other sediments that are stratigraphically above the schists (but lower in terms of altitude). A large block of schist traversed by granite veins was brought back to Edinburgh to convince critics of the virtues of his theory. John Clerk's son (of the same name) (1757–1812) accompanied Hutton and produced a wonderful cross-section of the island, construed in terms of Hutton's theory (Figure 2). This section is congruent with a geological map of the area, based upon modern knowledge, which shows a domed structure of schists and sedimentary rocks, disposed around a central core of granite.

Also, at Loch Ranza on the northern tip of Arran, Hutton found his first unconformity, with the sedimentary strata (sandstones and limestones) lying over the inclined or almost vertical schists (Figure 3). Again he had found a state of affairs that he had predicted from his theory. A 'swarm' of basaltic dykes was also observed by the shore of the southern part of the island.

The Loch Ranza unconformity was not, however, wholly convincing, for the rocks were obscured by vegetation. But on their way home the travellers found a much clearer example near Jedburgh, where the road ran by the banks of the River Jed and a section revealed a splendid view of Old Red Sandstone lying horizontally over the up-ended grey gritty sandstone that is now known as Silurian greywacke (which Hutton called 'schistus' although it was not a schist).

The most famous discovery of an unconformity occurred in 1788. Hutton's upland farm was situated on 'schistus', while his main farm was on soil derived from Old Red Sandstone. The contact between the two rock types ran northwards to the coast. Hutton must have been aware of the two rock types, which were similar to those that he had seen at Jedburgh. It seemed a good plan to examine the coast, where an unconformity might be exposed. Accordingly, with his friends John Playfair (1748–1819), Professor of Mathematics at Edinburgh, and a local landowner, Sir James Hall (1761–1815), a keen amateur scientist who did some of the first experimental geology, Hutton sailed along the Berwickshire coast, past the schist terrain to that of sandstone. As anticipated, they encountered an unconformable contact, at a place called 'Siccar Point', with Old Red Sandstone overlying the schist, as at Jedburgh, but exposed in such a way that the three-dimensional structure of the contact could be examined.

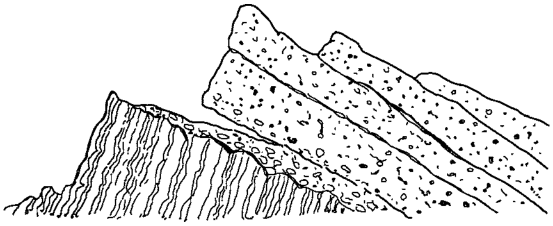


Figure 3 Unconformity at Loch Ranza, Arran, as described by Hutton and figured by Sir Archibald Geikie. Reproduced from Hutton (1899). *Theory of the Earth…Vol. III* Edited by Sir Archibald Geikie, p. 235. London: Geological Society.

The famous excursion was described in Playfair's biography of Hutton. The three men were aware that if the Siccar Point exposure was interpreted through the lens of Hutton's theory then it entailed the passage of a vast amount of time. The sediments of the greywacke were first deposited horizontally under the sea and consolidated by heat and the pressure of superincumbent material. Then the area was upheaved by forces acting from within the Earth, with hardening of the sediments by heat and pressure. The forces were such that the 'schistus' now stood almost vertically. But the strata were then subjected to weathering and erosion so that the upheaved rocks were reduced to an approximately level surface. In time, the now vertical strata subsided below the sea once more (by an inadequately explained process) and were covered by layers of sediment derived from adjacent high ground. Again there was consolidation, following which uplift occurred, exposing the strata to the elements once more, but without the overlying sandstones being folded or inclined. Thus the disposition of the rocks observed at Siccar Point could be understood – provided that time was unlimited. Playfair wrote in his biographical memoir of Hutton:

> Revolutions still more remote appeared in the distance of this extraordinary perspective. The mind seemed to grow giddy by looking so far into the abyss of time; and while we listened with earnestness and admiration to the philosopher who was now unfolding to us the order and series of these wonderful events, we became sensible how much further reason may sometimes go than imagination can venture to follow.
>
> Playfair (1805, p. 73)

Thus at Siccar Point Hutton provided evidence for (but not formal proof of) the Earth's great age and the cyclic nature of geological processes. The locality has long been recognized as one of geology's most significant field sites. Hutton acquired Playfair as a convert to his theory, and it was Playfair who popularized Hutton's ideas – Hutton's prolix style and confusing theory of heat (see below) did not gain him many adherents.

Hutton's Later Work on the Theory of Heat

In his old age, Hutton tried to give some kind of physicochemical explanation of the forces causing elevation and subsidence, but he had little success. He knew that bodies expanded when heated, and the kind of heat that produced this effect he called 'sensible heat'. He also knew that when heat was applied to a solid it increased in temperature, but on reaching the melting point it would melt without changing temperature, even though it was still being supplied with heat. In the change of state, the heat supplied to produce melting was somehow hidden. Black had called this 'latent heat'.

But the nature of heat was uncertain. Hutton thought it was a kind of weightless 'substance'. He knew that everyday objects have mass and that massive bodies are attracted to one another by gravitation. But there also seemed to be repulsive forces at work, as for example when water is boiled: steam engines exert pressure in their cylinders.

Today, we distinguish between radiant heat and heat transmitted by conduction. Hutton had no adequate concept of radiation, but he knew that heat from the sun shines on us, across space. He called it 'solar substance', and, though weightless, it somehow seemed to be absorbed by plants, though Hutton did not know how. Adding to the complications, Hutton accepted the old 'phlogiston theory' of combustion (which was collapsing at the end of the eighteenth century), according to which an inflammable material contains a weightless 'substance' or 'principle' called 'phlogiston', which is dispersed into the atmosphere during combustion. Hutton was inclined to suppose that 'solar substance' and 'phlogiston' were one and the same. (Actually, if one regards 'phlogiston' as energy, then some of the problems that Hutton was trying to understand fall into place for us.)

Hutton grappled with such problems in two books: *Dissertations on Different Subjects in Natural Philosophy* (1792); and *A Dissertation upon the Philosophy of Light, Heat, and Fire* (1794). All his arguments cannot be followed here, but he tried out the idea that objects normally attracted one another according to the inverse-square law of gravitation. Thus he spoke of 'gravitating matter'. At very close quarters, however, objects supposedly began to repel one another, according to a force law in which the distance between particles was raised to a power greater than two. The repulsive force (or 'solar substance') could supposedly take various guises: 'sensible' heat, manifested by expansion; latent heat; light; electricity; and phlogiston. So, when sediments were under extreme pressure, they might move from a compressive phase to an expansive (expanding) phase. Hence, in the geostrophic cycle, there could be alternating periods of contraction (compression or consolidation) and expansion (producing land elevation).

Hutton's theory depended on a balance of attractive (gravitational, cohesive, and concretive) and repulsive (specific, or sensible, and latent heats) forces. There could be different resultant states, arising from the forces producing elevation (expansion) and subsidence (contraction) at different times and places. But when, lacking the concept of energy, Hutton started talking about 'solar substance' in reference to solar radiation (as we would say) and thought that this 'substance' was immaterial, confusion and misunderstanding followed amongst his contemporaries; it is scarcely possible for us to make sense of his theory of heat.

Hutton's Legacy

We can see, therefore, that, for all Hutton's success in looking into the 'abyss of time' and his successful predictions of granitic veins and unconformities, his theory had significant lacunae, and the physical explanation of expansion and uplift was not really integrated into his *Theory of the Earth* and attracted little or no following. Expansion was the Achilles Heel of his theory, and the problem remained unsettled for generations. In the end, expansion and elevation were simply assumed by Huttonian cyclists.

People eventually accepted Hutton's cyclic theory, even though they could make little sense of the physical basis he envisaged. But acceptance took time, and geological theory was racked with controversy until well into the 1820s. The Professor of Natural History, Robert Jameson (1774–1854), gained control of geology teaching at Edinburgh, and even Hutton's specimens, for many years. So Huttonian theory tended to be eclipsed in Scotland for decades, despite the best efforts of Playfair and Hall, who, after Hutton's death, conducted experiments that sought to simulate the consolidation of sediments, aided by heat, and to show that limestone heated in a sealed gun barrel could be converted into something like marble without loss of carbon dioxide.

In addition to having excellent ideas about weathering and erosion, the deposition and consolidation of sediments, rates of geological change, the immense age of the Earth, and arguments in favour of geological cycles supported by evidence of unconformities, Hutton clearly appreciated the fact that many valleys have been carved by the rivers that now flow in them. Through second-hand knowledge of the Alps, he suggested that glaciers might have been much larger in the past than at present and could have deposited large blocks considerable distances from where the rock types are found *in situ*. Thus it seems that he envisaged a former colder climate than at present (due to the mountains being higher and carrying more snow) and appreciated the enormous erosive powers of glaciers.

In 1802, Playfair published his *Illustrations of the Huttonian Theory of the Earth*, which set out Hutton's doctrines in improved literary form. Hutton had referred to the ideas of the Swiss geologist Horace Bénédict de Saussure (1740–1799) about the transport of glacial debris by glaciers extended from the Alps to the Jura Mountains, and these ideas were given greater prominence by Playfair, who also wrote about the patterns of river drainage systems. But neither Hutton nor Playfair had the idea of an Ice Age. That came later, principally through the advocacy of Louis Agassiz (1807–1873) (*see* **Famous Geologists:** Agassiz).

Another Scottish geologist, the influential Charles Lyell (1797–1875) (*see* **Famous Geologists:** Lyell), accepted many of Hutton's ideas and made them almost paradigmatic, handing them on to another Scot, Archibald Geikie (1835–1924), who coined the methodological maxim: 'the present is the key to the past'. But that principle was already well established by Hutton. He used his knowledge of what he could see going on around him – on his farms and during his travels – to develop a theory about how the Earth operated as a system and how it might have been in the remote past. But Hutton's cycles were not identical. There could be local variations from one phase to the next. The Earth had a history, while operating in a law-like manner, so as to be in a steady-state when viewed on a grand scale. We can also credit Hutton with advancing the concept known today as 'deep time' – and for doing so by geological reasoning.

See Also

Famous Geologists: Agassiz; Lyell. **History of Geology Up To 1780**. **History of Geology From 1780 To 1835**. **Igneous Rocks:** Granite. **Unconformities**. **Weathering**.

Further Reading

Baxter S (2003) *Revolutions in the Earth: James Hutton and the True Age of the World.* London: Weidenfeld & Nicolson.

Dean DR (1992) *James Hutton and the History of Geology.* Ithaca: Cornell University Press.

Donovan A (1978) James Hutton, Joseph Black and the chemical theory of heat. *Ambix* 25: 176 190.

Gerstner PA (1968) James Hutton's theory of the Earth and his theory of matter. *Isis* 59: 26 31.

Gerstner PA (1971) The reaction to James Hutton's use of heat as a geological agent. *British Journal for the History of Science* 5: 353 362.

Hutton J (1788) Theory of the Earth; or an investigation of the laws observable in the composition, dissolution, and restoration of land upon the globe. *Transactions of the Royal Society of Edinburgh* 1: 209 304.

Hutton J (1795) *Theory of the Earth, with Proofs and Illus trations.* London: Edinburgh: William Creech; London: Cadell, Junior, and Davies (republished in facsimile 1972).

Hutton J (1899) *Theory of the Earth…Vol. III* Edited by Sir Archibald Geikie, p. 235. London: Geological Society.

Hutton J (1997) *James Hutton in the Field and in the Study edited by Dennis R. Dean: Being an Augmented Reprinting of Vol. III of Hutton's Theory of the Earth (I, II, 1795), as First Published by Sir Archibald Geikie (1899).* New York: Scholars' Facsimiles & Reprints, Delmar.

Jones J (1985) James Hutton's agricultural research and his life as a farmer. *Annals of Science* 42: 574 601.

McIntyre DB (1997) James Hutton's Edinburgh: the historical, social and political background. *Earth Sciences History* 16: 100 157.

McIntyre DB and McKirdy A (2001) *James Hutton: The Founder of Modern Geology.* Edinburgh: National Museums of Scotland (1st edn, 1997).

Oldroyd DR (2000) James Hutton's 'Theory of the Earth' (1788). *Episodes* 23: 196 202.

Playfair J (1805) Biographical account of the life of Dr James Hutton, F.R.S.Edin. *Transactions of the Royal Society of Edinburgh* 5: 39 99.

Şengör AMC (2001) *Is the Present the Key to the Past or the Past the Key to the Present? James Hutton and Adam Smith versus Abraham Gottlob Werner and Karl Marx in Interpreting History.* Special Paper 355. Boulder: Geological Society of America.

Tomkeieff SI (1962) Unconformity an historical study. *Proceedings of the Geologists' Association* 73: 383 417.

Lyell

D R Oldroyd, University of New South Wales, Sydney, Australia

Charles Lyell (Figure 1) was arguably the most important geologist of the nineteenth century, and his *Principles of Geology* (1st edn, 3 vols, 1830–1833; 11th edn, 1872) was a classic text that exerted much influence on the development of geology, as well as helping to shape the development of Charles Darwin's thought. Lyell's other major works were his *Elements of Geology* (titled *Manual of Elementary Geology* in some editions) (1st edn, 1838; 6th edn, 1865) and *Geological Evidences of the Antiquity of Man* (1st edn, 1863; 4th edn, 1873). He also published two books on his travels in North America.

Lyell was born into a well-to-do family at Kinnordy House, Forfarshire, Scotland, but much of his youth was spent at the family's second home in Hampshire (with a more agreeable climate). He attended a private school in Salisbury and then at Midhurst; thereafter he attended Exeter College, Oxford, where he studied mathematics and classics, but also became greatly interested in geology through the lectures of William Buckland, which students could attend as optional additions to their main curriculum. Lyell's family was considerably interested in natural history, and during his vacations they travelled extensively on the continent. Lyell also made observations on the Kinnordy estate. Even while a student, he was elected Fellow of both the Linnean and Geological Societies.

On leaving Oxford, Lyell started to train for the law at an office in London, but found the work uncongenial and complained of problems with his eyesight, and so, having (limited) independent means, he did not continue in this line of work. Rather, consorting with many of the leading geologists of the day, and travelling widely, he became virtually a full-time gentleman-geologist, being elected to the Royal Society as early as 1826. Two years later when travelling on the Continent and meeting important figures in Paris, etc., he decided to give up legal work altogether. Eventually, he acquired significant income from his geological writings. In 1832, he married Mary Horner, daughter of Leonard Horner, himself a geologist and educationist, who had learned Huttonian theory in Edinburgh. The couple, who had no children, settled in London, where Lyell became established as one of its leading scientists.

Buckland's Oxford lectures were informative and entertaining. He taught the essentials of stratigraphy, and particularly William Smith's idea that strata could be identified and correlated by their fossil contents (*see* **Famous Geologists:** Smith). But Buckland, in the religious atmosphere of Oxford, and trying to show that his science was compatible with the Bible, laid much emphasis on his studies of superficial deposits and cave remains (about which he was an authority) and sought to show that such materials

Figure 1 Charles Lyell (1797 1875).

could be explained as the result of the Noachian Flood, which in terms of biblical chronology, occurred only a few thousand years ago. Thus, Buckland's geology, which had sources in the work of Cuvier (*see* **Famous Geologists:** Cuvier), could (supposedly) lend material support to theological claims. But such a global flood would have been impossible according to the laws of nature as presently acting, and would be incompatible with geological processes seen at work today.

Lyell travelled and thought much during the 1820s; in Scotland he visited such sites as Glen Tilt and Siccar Point (*see* **Famous Geologists:** Hutton). Though greatly stimulated by Buckland, he came to reject his idea on the great role ascribed to catastrophic floods, and leaned towards the geology of Hutton, his father-in-law-to-be having attended John Playfair's 'Huttonian' lectures in Edinburgh. Abraham Werner's Neptunist theories were rejected as being incompatible with the limited solvent power of water and with Hutton's observations. In Forfarshire (which he mapped in 1824), Lyell saw marls being deposited, or already deposited, in fresh-water lakes fed by springs and associated with shells and plant remains. He knew that in the Paris area Cuvier and Brongniart had found similar fresh-water limestones, which, they thought, had no modern analogues. Thus Lyell's thinking was turned towards the idea of explaining geological phenomena in terms of presently occurring processes.

In Huttonian theory, an immense amount of time was required to make possible the great cycles of geological change that he envisaged. The question of time was one that Lyell, therefore, had to consider. Evidence for the Earth's great antiquity was produced during Lyell's journey to Sicily in 1828. He saw the huge still active volcano, Mount Etna, and it was evident that it was made up of successive lava flows. Historical information about recent flows gave an approximate idea of the rate of accumulation of the flows and the build-up of the mountain. The height of the mountain being known, one could thus form an approximate idea of its age. Further, Lyell examined shells in recent-looking strata lying below the lavas. Nearly all were still to be found today in the Mediterranean. So strata and shells of geologically recent appearance were in rocks that were very ancient in human terms, being older than flows from the volcano. (In a subsequent letter to his sister, Lyell offered that on a 'moderate computation' the shells might be 100 000 years old.) So if geologically recent rocks were ancient in human terms, rocks lower in the stratigraphic column must be exceedingly ancient. Evidently the Earth was of enormous age. In this argument, Lyell was assuming that the rate of flows at Etna occurred at approximately equal rates. He was applying the principle that nature was uniform in her operations: what was later dubbed the principle of uniformity. Lyell was a 'uniformitarian' with regard to Etna, as he had been with respect to the lake deposits in Forfarshire.

Also on his Italian journey of 1828, Lyell visited Pozzuoli on the coast near Naples. There he observed three standing columns of a Roman building, then thought to be an ancient temple. These had marks of the borings of marine organisms half way up, which suggested to Lyell that the land there had fallen below sea level since Roman times, and had subsequently risen; all this having happened without the columns toppling over. From this, Lyell inferred that the level of land was rising or falling in different places, as Hutton had previously proposed. Moreover, the processes were not sudden or catastrophic, but gradual.

Following his return to Britain, Lyell began to write his major book, which sought to establish the working methods and procedures of geology. It was to give geology its proper method and fundamental principles: hence its title, *Principles of Geology.* These may be summed up by the adage (as later stated by Archibald Geikie) that 'the present is the key to the past'. Also, for Lyell, geological processes were assumed to be 'gradual'.

Hutton's geology envisaged grand cycles of rock formation, erosion, transport, deposition, consolidation, and subsequent elevation. The rocks of each cycle were not necessarily identical in any given place, and the geologist needed to work out the history of what had happened at each locality. But overall, the earth did not have an historical direction: it did not 'progress'. Things were much the same in the past and present (humans excepted).

Lyell's views were much the same, but he placed more emphasis on fossils. He supposed that conditions were constantly changing at any given locality from one period to the next, because of the local changes of relative levels of land and sea. Climate could change too, according to whether more high land happened to be near the poles at a given time, or nearer the equator, the former state of affairs producing cooler conditions overall. So some forms would become extinct if they failed to meet the conditions of existence.

On this basis, new types of organisms also needed to come into existence from time to time. Lyell presumed that they did so, even though he did not know how this occurred. Further, he assumed that the basic animal types had always been found on the earth. On this view, there was a gradual turnover of species. His model can be represented as shown in Figure 2.

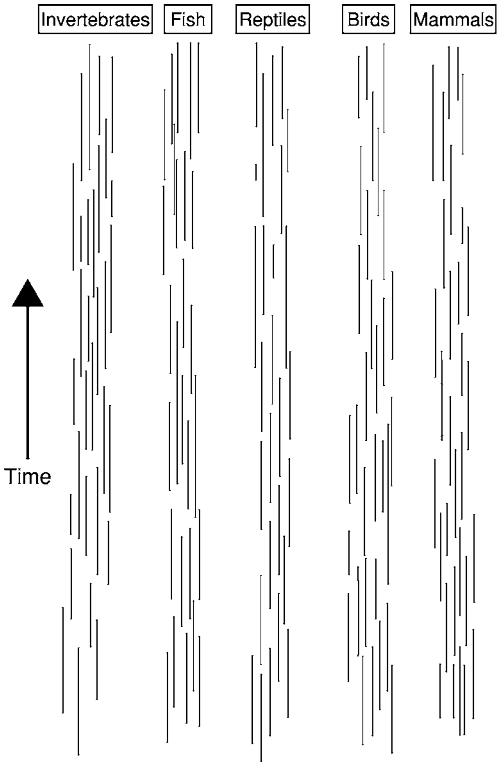


Figure 2 Diagram illustrative of Lyell's theory of species change, with 'random' creations and extinctions.

According to this model, Lyell assumed that the further back the geologist explored in time, the fewer extant species there would be. In fact, nearly all species before the beginning of the Tertiary in Europe would now be extinct. Further the Tertiary could be subdivided according to its proportions of extant fossils. The subdivisions that Lyell proposed were:

- Newer Pliocene 96% recent fossils
- Older Pliocene 52% recent fossils
- Miocene 19% recent fossils
- Eocene 3% recent fossils

Between the Secondary (later called Mesozoic) rocks and the Tertiary there was a period of non-deposition in Europe, so that Secondary fossils were now virtually all extinct. There had been a complete turnover of forms during the stratigraphic time-gap. Likewise, there was a large time-gap and turnover of forms between the Primary (Palaeozoic) forms and the Secondary types. Moreover, the further back one went through the Secondary epoch, the smaller was the percentage of forms resembling those at the top of the Secondary (i.e., in the Cretaceous); likewise, through the Palaeozoic. Lyell regarded marsupial remains found in Secondary rocks in the Stonesfield Slate near Oxford as vindication of his idea that the major animal types went back into the indefinite past. That they could not be found in the most ancient rocks was due to the fact that they had been lost by erosion or metamorphism (an important notion that Lyell first explicitly enunciated). All this was compatible with Lyell's uniformitarianism, but he had no modern empirical warrant for the supposition that new species could somehow come into being.

Lyell's ideas attracted much attention, though most geologists, looking at the stratigraphic record, found it hard to accept that there was no evidence of progress in the fossil record through time. On the other hand, his desire for geology to have its own procedures, with geological processes operating in accordance with the presently observable laws of nature, met with approval, as did his mastery of facts and grasp of palaeontological and stratigraphic detail. He wanted geology to be a science, *sui generis*, distinct from cosmology. Geologists, he thought, did not need a general 'theory of the earth' such as his 18th-century predecessors had sought to offer (though Lyell's cyclic geology was in fact fundamentally the Huttonian theory). But Lyell focused on stratigraphy and palaeontology, not 'hard rocks' such as granite.

In 1831, following the success of Volume I of his *Principles*, Lyell obtained a chair in geology at King's College, London, a newly founded Church of England establishment. By then, the putting aside of the Noachian Flood as a geological agent seemed not to concern the authorities, and Lyell's religious view were considered 'sound'. However, he only gave lectures there in 1832 and 1833. Ladies were allowed to be present for the first course, but were thought to present an undesirable distraction and their further attendance was terminated. In consequence, the attendance fell sharply, and Lyell decided that he was in part wasting his time there, so he resigned to return to his publishing and life as a gentleman geologist. In this he was eminently successful, continuing his extensive fieldwork, and involvement with the Geological Society and the British Association. Lyell served as President of the Geological Society in 1835–37, and again in 1849–51. Subsequently, he was knighted (1848), was awarded the Royal Society's Copley Medal in 1858, and served as President of the BA in 1864.

Lyell was seriously concerned with French geology. He acknowledged Cuvier's mastery of palaeontology, but rejected his 'catastrophist' theory. Lyell's *Principles* did much to counter this doctrine in contemporary Britain. In Paris in 1828, he met with the conchologist and palaeontologist Gérard Deshayes, who assisted him in the identification and stratigraphic placement of the shells he collected that year. Lyell reacted negatively to the tectonic theory of Léonce Élie de Beaumont (which envisaged mountain ranges as having been formed as a result of the

Earth's cooling and contraction) and significantly hindered its acceptance in Britain. Most importantly, Lyell gave close attention to the 'transformist' (evolutionary) theory of Jean Baptiste Lamarck in Volume II of *Principles*. Changing conditions cause new needs for organisms. To adjust to changing circumstances, organisms may alter their habits, and consequently their forms. These changes may be transmitted to subsequent generations, producing a gradual transformation of species. The first simple forms of life appeared naturally (without divine action) by spontaneous generation.

Such ideas were rejected by Lyell over many pages. His principal objection was that the stratigraphic record did not reveal smooth transitions such as Lamarck's theory would lead one to expect to find. But there were other objections, such as the inability to produce new species by breeding; and hybrids were sterile. Nevertheless, Lyell devoted much energy to thinking about what the concept of species meant, the 'laws' of distribution of species, and the extent to which they could or could not show modification due to different or changing circumstances. The problem of species and speciation was one of the main features of his book, and it set the scene for Darwin's work, and his seeing his fundamental problem to be 'the origin of species' (*see* **Famous Geologists:** Darwin).

A major problem for geologists in the first half of the nineteenth century was the large quantities of superficial deposits: gravel, tenacious clay containing unsorted rock fragments and fossil remains, and large boulders of rock distant from the nearest 'solid' outcrops of rock of that type. Such phenomena were eventually explained by the work of Louis Agassiz and his theory of an Ice Age (*see* **Famous Geologists:** Agassiz). In the early nineteenth century, these deposits were ascribed to the Noachian deluge or some like catastrophe, and William Buckland distinguished between 'diluvium' (Flood deposits) and 'alluvium' (materials deposited by rivers in the normal course of events). It was supposed that a great inundation(s) could have swept over the globe, even depositing the erratic boulders and marine shells loose at the tops of hills or mountains.

But according to Agassiz, the better explanation was that there had formerly been a colder climate with the whole of northern Europe once covered by ice, which had transported boulders, ground up the underlying rock, and deposited it, along with river gravels, over the land. The ice also could have scratched the underlying rock and transported shells to hill tops. Agassiz lectured on this to the BA in 1840, and some geologists were converted to his ideas, including Lyell. His general theory was presumed to be capable of accounting for a period of extreme cold, such as to cause widespread glaciation, if much of the high land at that time happened to be in the polar regions. But Lyell's conversion was short lived. Agassiz's theory seemed to take him too far from present analogies or present climatic conditions. So he adopted the theory that came to be called 'glacial submergence': there was a period of great cold, but not such as to produce an all-enveloping mass of land-ice. Rather, there was a general fall of land surface, causing marine submergence, accompanied by cooling, causing extension of ice-fields and the transport of boulders by drifting ice-bergs (hence the diluvial deposits are now generally called 'drift'). During his North American trip of 1845 Lyell saw floating ice in the St Lawrence River, which modern observation seemed to account for the occurrence of erratics satisfactorily in accordance with his methodology. While Lyell later accepted Agassiz's theory for the Alpine regions he never accepted the general land-ice theory, preferring the glacial submergence model.

After Darwin returned from his *Beagle* voyage in 1836, he and Lyell became close friends, but during the years before the publication of *The Origin of Species* Darwin mostly kept his emerging transformist ideas to himself. Lyell was opposed to transformism for reasons that he developed back in the 1820s, and like many he was concerned about evolution's implications for 'revealed religion' and social stability. In his Presidential Address to the Geological Society (1851), he spoke against evolutionary ideas. Man, he thought, was a very recent creation, subsequent to the mammoths. However, after Darwin revealed his ideas to Lyell about 1856, he was reluctantly converted and did his best to see the early publication of Darwin's ideas in 1858. In his *The Antiquity of Man* (1863), Lyell set forth ideas on transformism and stated his acceptance of the Darwinian theory of evolution by natural selection (though he represented it as a 'modification' of Lamarck's doctrine).

In the 1850s, Lyell had devoted a considerable amount of travel and fieldwork to the study of ancient humans, which was consistent with his general interest in the younger parts of the stratigraphic column. By that time, considerable numbers of cave deposits and flint implements had been discovered, as well as some human-like remains, notably the Neanderthal skull, found near Dusseldorf in 1857. This seemed, according to Thomas Henry Huxley's description, which Lyell quoted, to be intermediate between that of a modern human and a chimpanzee's. But Lyell cautiously (and rightly) stated that "it is at present to too exceptional, and its age too uncertain, to warrant us in relying on its abnormal and ape-like characters, as bearing on the question whether the farther back we trace Man into the past, the more we shall find him approach in bodily conformation to those species of the anthropoid quadrumana which are most akin to him in structure" (*Antiquity*, p. 375). Lyell also reported on the recently discovered *Archaeopteryx*, which might seem to be a missing link, but he also deferred to the anatomist Richard Owen's opinion that it was actually a bird. Thus, Lyell supported Darwin's evolutionism in a way that was valuable to its acceptance. But at the same time he did not push all the evidence to what we might regard as its logical conclusion.

On reading Lyell's works, one is struck by his mastery of exposition and his command of the literature, especially in stratigraphy. His influence was very great, both in his own day and subsequently. There is, however, ambiguity in the concept of 'uniformitarianism' (gradualism, steady-statism, naturalism, and 'actualism' – or the idea that modern, actually observable, processes should be used to provide geological explanations). Lyell held to all these positions. Modern geologists commonly make obeisance to uniformitarianism, without making the foregoing distinctions. Modern geology does not necessarily adhere to any of them, except in its rhetoric; for Lyell convinced people that *his* approach was the right one to adopt for geology to be regarded as a science.

See Also

Famous Geologists: Agassiz; Cuvier; Darwin; Hutton; Smith. **History of Geology From 1780 To 1835**. **History of Geology From 1835 To 1900**.

Further Reading

Gould SJ (1987) Charles Lyell, historian of time's cycle. In: Gould SJ (ed.) *Time's Arrow Time's Cycle: Myth and Metaphor in the Discovery of Geological Time*, pp. 99 179. Cambridge (Mass) and London: Harvard University Press.

Hooykaas R (1963) *Natural Law and Divine Miracle: The Principle of Uniformity in Geology, Biology and Theology*. Leiden: EJ Brill.

Lyell C (1997) *Principles of Geology* edited with an introduction by James A. Secord. London, New York, Ringwood, Toronto and Auckland: Penguin Books.

British Society for the History of Science (1976) *The British Journal for the History of Science: Lyell Centenary Issue* 9(2).

Rudwick MJS (1969) Lyell on Etna, and the antiquity of the Earth. In: Schneer CJ (ed.) *Toward a History of Geology*, pp. 288 304. Cambridge (Mass) and London: The M.I.T. Press.

Rudwick MJS (1971) Uniformity and progression: reflections on the structure of geological theory in the age of Lyell. In: Roller DHD (ed.) *Perspectives in the History of Science and Technology*, pp. 209 227. Norman: University of Oklahoma Press.

Rudwick MJS (1978) Charles Lyell's dream of a statistical palaeontology. *Palaeontology* 21: 225 244.

Rudwick MJS (1990) *"Introduction," Principles of Geology, First Edition* [in Facsimile] *Volume I Charles Lyell*, pp. vii lviii. Chicago: University of Chicago Press.

Wilson LG (1972) *Charles Lyell: The Years to 1841*. New Haven: Yale University Press.

Wilson LG (1998) *Lyell in America Transatlantic Geology, 1841 1853*. Baltimore and London: The Johns Hopkins University Press.

Murchison

D R Oldroyd, University of New South Wales, Sydney, Australia

Roderick Murchison (Figure 1) was the eldest son of a wealthy Scottish landowner at Tarradale estate, Ross-shire. Though born in Scotland, and always emphasizing his Scottish ancestry, he spent most of his career in England and spoke with an English accent. Following his father's death and mother's remarriage, Roderick was sent to school at Durham, aged 7 years, soon forming the ambition to be a soldier. At 13 years old, he attended the military college at Great Marlow where his training gave him a good 'eye for country'. He was soon involved in the 'Peninsula War' in Portugal, fighting at the Battle of Vimieira, aged only 16 years. From this victory, his unit moved into Spain where things went badly for the British army in winter conditions, with forced marches, defeat in the Battle of Coruna, and withdrawal in disarray. After a spell in Sicily, Murchison was posted to Ireland, where he led a dissolute and expensive life, and later likewise in London as a half-pay captain. With the end of the wars, he was fortunate to meet a general's daughter, Charlotte Hugonin, 3 years his senior, and they were married in 1816. They then took a leisurely tour through France, Switzerland, and Italy, and under his wife's influence his self-education began, learning French and Italian, visiting museums and galleries and some scientists and scientific institutions.

Figure 1 Roderick Murchison (1792 1871).

On their return, the Tarradale property was sold and the couple (who remained childless) resided for some years in Barnard Castle, County Durham, where Murchison devoted himself to the sporting country life. However, the two also participated in local gatherings of literary and scientific people, and met Sir Humphry Davy. Charlotte had interests in botany and mineralogy, but her husband's fox-hunting passion continued and they moved to Melton Mowbray in the hunting shires. However, urged by his wife and Davy, Murchison determined to take up geology, moving to London in 1824 and attending chemistry lectures. He joined the Geological Society, went on field trips with Buckland and others, and began serious study of the science.

Work in southern England yielded Murchison's first paper in 1825 (in which year he was elected to the Geological Society; and in 1826 to the Royal Society!), he journeyed to Scotland in 1826 and 1827, the former trip being made to Jurassic strata at Brora, following instruction on Secondary stratigraphy from William Smith (*see* **Famous Geologists:** Smith), whom he visited in Scarborough; the latter trip being with Adam Sedgwick (*see* **Famous Geologists:** Sedgwick). Murchison also visited the Continent with Lyell (*see* **Famous Geologists:** Lyell). Having gained some knowledge of how to study older rocks with Sedgwick, from 1831 Murchison and his friend and mentor determined to study the then rather little-known rocks of Wales; the so-called 'Transition Series' of Werner's Neptunist geology. He began to work from the known base of the Old Red Sandstone in the Welsh Border country, from the Wye Valley through to Cheshire.

Historians have rather detailed knowledge of Murchison's travels and scientific work as his field notebooks have been preserved, together with much correspondence. There is also a multi-volume 'journal', based on the notebooks, copied out by an amanuensis in Murchison's old age, which was intended for biographical purposes, sometimes being judiciously 'improved' by Murchison to give a favourable view of his accomplishments and ideas.

It is known, then, that in 1831 Murchison and Sedgwick planned to work out the geology of the Welsh (and Border region) Transition Series but, Sedgwick being otherwise occupied, Murchison set out alone that year, and Sedgwick arrived later, starting in North Wales, briefly with Darwin as an assistant (*see* **Famous Geologists:** Darwin). Near Ludlow, Murchison found richly fossiliferous rocks and the structure was made out successfully. He worked in his area for the next few seasons, subdividing the 'Upper Grauwacke Series' into the 'Ludlow Series', the 'Wenlock Limestone', the 'Horderley and May Hill Rocks' (later called the Caradoc Series), and the 'Builth and Llandeilo Flags', the first two being placed in the Upper Silurian and the latter two in the Lower Silurian (1835). The name Silurian was coined for a new geological system after the Silures tribe that formerly inhabited that part of Britain. The arrangement Murchison envisaged is shown in Figure 2, reproduced from his great treatise, *The Silurian System* (1839, p. 196). This *magnum opus* provided immense detail concerning the different units, figures of their characteristic fossils, and a valuable map of the geology of his Silurian 'domain'.

In 1835, Sedgwick and Murchison introduced the terms Cambrian and Silurian, though the word 'System' was not used at that time. Moreover, the boundary between the two was not then clearly defined, though following their fieldwork of 1834 Murchison stated that the upper and lower rocks had been 'dovetailed' in a manner that was satisfactory to both geologists. Unfortunately, this proved later not to have been the case and in the years that followed a serious controversy developed between the two former friends (*see* **Famous Geologists:** Sedgwick, **Palaeozoic:** Ordovician).

Murchison's approach to geology was considerably shaped by his military background. He came to regard 'his' Silurian System as personal territory, and the more parts of a map that could receive

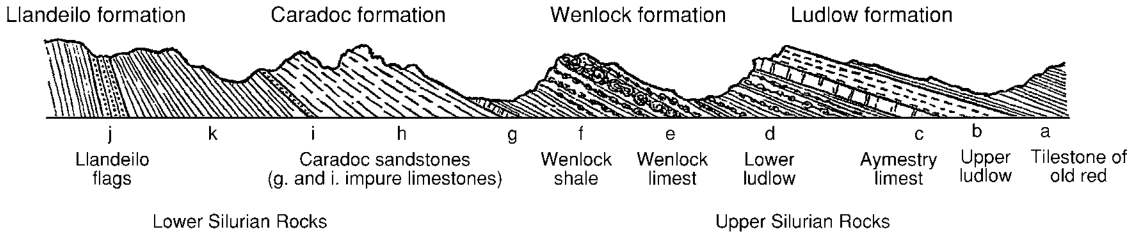


Figure 2 Arrangement of Silurian Strata, according to Murchison (1839).

Silurian colours the greater was his satisfaction! Indeed, he became known as the 'King of Siluria'. From the geological point of view it was evidently necessary to see whether the Silurian System, established in the Welsh Border area, was of general application, preferably worldwide. This grand task of spreading Silurian colours was undertaken personally by Murchison, in western and eastern Europe, the vast tracts of western Russia, and in Scotland in his old age. Other geologists were encouraged to find Silurian formations elsewhere, where Murchison's own feet had not trodden.

This expansion of Siluria also occurred in time as well as space, as Murchison sought to extend the Silurian down into the regions of Wales, where Sedgwick had established the Cambrian. So, when Murchison produced a condensed version of his stratigraphy in *Siluria* (1854), a name that suggested a kind of 'kingdom', it pushed down into Sedgwick's territory calling his rather unfossiliferous rocks of North Wales 'Lower Silurian'. The book was subtitled *The History of the Oldest Known Rocks Containing Organic Remains*. That is, Murchison claimed all the Palaeozoic rocks below the Devonian as belonging to his system.

This state of affairs arose in part because, after Sedgwick had put forward his Cambrian System, he did not there and then describe its characteristic fossils, and when this was eventually done some of his types had already been classified as Lower Silurian by Murchison. (Early on, Murchison divided his system into Lower and Upper divisions.) In *The Silurian System*, Murchison allowed the existence of fossil-containing Cambrian rocks below the Silurian, but it was supposed that they were linked by 'passage beds' to the Llandeilo. So the possibility of territorial and temporal expansion was already there. It occurred again in Russia (see below). In *Siluria* it was stated that the fossiliferous Cambrians were lateral extensions (undulations) of Silurian strata (an opinion said to have been reached in 1841), and all that was left on the map of Sedgwick's Cambrian was the apparently unfossiliferous rocks of the Harlech Dome area, some on the southern side of the Menai Strait between Anglesey and the rest of North Wales, and the unfossiliferous Longmynd rocks near Church Stretton in Shropshire. The remainder of Sedgwick's 'Cambria' was now depicted in Silurian colours. It was this encroachment, which had been going on through the 1840s, that so incensed Sedgwick (*see* **Famous Geologists:** Sedgwick). Murchison did well out of it all. He was awarded the Royal Society's Copley Medal in 1849, having been knighted in 1846.

But Murchison did not gain his honours lightly. After visiting German colleagues, he was in Russia in 1840 and 1841; Poland in 1843; Germany and Russia in 1844 and 1845; Scandinavia in 1844 and 1845; France and Germany in 1839 and 1843, and again in 1846 and 1847; and in Italy and Switzerland in 1847 and 1848. There were also journeys in Britain. In addition, Murchison served as General Secretary of the British Association, President of the Geological Society in 1841–1843 and of the Royal Geographical Society in 1843–1844. He also maintained an expensive but hospitable life style in Belgravia, London. Murchison's energy was truly remarkable.

Murchison's journeys in Russia, conducted with the French stratigrapher Philippe Edouard de Verneuil and the Russian zoologist Count Alexander Keyserling, were, for their time, of epic proportions. From St Petersburg in 1840, they travelled up to Archangel, then by indirect route to Moscow, and back to St Petersburg. Though much of the terrain is covered by drift and offers few good sections, information about Carboniferous, Devonian, and Silurian strata were obtained, partly from informants with knowledge of wells, etc. The area under St Petersburg is remarkable for having Cambrian *clays*, which by their fossils were construed by Murchison as Lower Silurian. In *The Geology of Russia* (see below) he stated that the clay contained fossils "belonging to the very oldest known Silurian or protozoic type... [and was] the true base of the Palaeozoic series, as indicated by a gradual dwindling out of animal life in the deposit of a region, where no eruptions ha[d] taken place, and where the strata are wholly unaltered". A further notable discovery, agreeable to the existence of the Devonian as a System, was the discovery of Old Red Sandstone fish in rocks that otherwise resembled the Magnesian

Limestone of England or the Zechstein of Thuringia, but were unlike the Old Red Sandstone rocks of Scotland. This confirmed the idea of the Devonian as a palaeontologically characterized system, which had different lithologies in different localities (*see* **Palaeozoic:** Devonian).

Early in his journey of 1841, Murchison met Czar Nicholas I in St Petersburg and established a good rapport with him. He was 'duchessed' by the Russian aristocracy, and formed a high opinion of it and of Russia. From Moscow, the geologists travelled east to Perm, to the west of the Urals, and thence further east to the point where they could see the plains of Siberia. Turning south on the western side of the mountains, they reached Orsk, and then headed west again, crossing the Volga and reaching the Sea of Azov, before going north again to Moscow and St Petersburg. It was on the basis of the rocks in the region of Perm that, later that year, Murchison proposed a new system, the Permian, after the ancient kingdom of Permia. Silurian, Devonian, and Carboniferous rocks were confirmed. The travellers saw considerable mining activity in the Urals. A central granitic nucleus appeared to be flanked by Silurian, Devonian, and Carboniferous strata. The Urals were eventually crossed and re-crossed on seven parallels between 60° and 54° N. As he had done for the Silurian and Devonian, Murchison named the Permian after a region with strata containing characteristic fossils.

On returning to St Petersburg, Murchison presented the Czar with a geological map, reports on the coal deposits of the Donetz Basin, and information about the alluvial gold of the Urals. It was a highly successful 'campaign' and Murchison's self-esteem rose to new heights. It was further fortified on his return to Britain by finding that the officers of the Geological Survey were obtaining results in Wales that seemed to support Murchison's views about the Silurian vis-à-vis the Cambrian.

In 1843, Murchison returned to eastern Europe, where he compared the Tatra Mountains of Poland with the Urals, and met von Humboldt in Berlin and the notable Palaeozoic palaeontologist Joachim Barrande in Prague. The same year Keyserling pursued the Russian researches in the northern Urals. In 1844, Murchison paid visits to Denmark, Norway, Sweden and Russia, and was able to find the basement rocks in Scandinavia underlying the Palaeozoics. In 1845, he was again in the north, receiving from Czar Nicholas the award of the Great Cross of St Stanislaus, for his notable contributions to Russian geology, and geologizing in the Baltic region. *The Geology of Russia in Europe and the Ural Mountains*, published in 1845 with Keyserling and Verneuil as co-authors and incorporating information from many others, was a monumental work, providing accounts of the geologists' journeys and descriptions of the geology of the half-continent. There were copious illustrations, including many sections and two large coloured geological maps; also descriptions and beautiful figures of fossils. Additionally, the book contained considerable theoretical discussion. It marked Murchison as the master stratigrapher and geological traveller of his age.

Besides becoming a leading geologist, Murchison was also active in encouraging geographical exploration, which he saw as essential to the expansion of the British Empire. As a man of influence, he interested himself in the geological appointments being made in the colonies, and his powers of patronage were considerable. For example, he assisted in the placement of Frederick McCoy in a chair at Melbourne University, which may have seemed advantageous to Murchison as it got one of Sedgwick's main allies out of the country! By 'placing' or maintaining contacts with people in different parts of the world, Murchison also gained advantage by having information channeled through his hands. He was gratified to have numerous topographical features named in his honour.

Murchison's influence became so great that he even felt qualified to offer opinions about the geology of countries that he had never visited. For example, he thought it likely that gold might be found in eastern Australia, adjacent to the hills of the 'Australian Cordillera' (mostly a rather inconspicuous topographic feature), on the basis of examination of some non-auriferous rock specimens shown him by the Polish explorer Paul Strzelecki, and on the grounds that the range contained rocks somewhat like those observed in the Urals (where alluvial gold had been found on the eastern flanks) and was similarly aligned, approximately N–S. In the event, this 'prediction' (1844) proved correct and was followed by gold rushes in the 1850s. Murchison regarded his lucky forecast as evidence that he was a "sort of authority" on Australian gold deposits. He recommended (1846) migration to Cornish tin miners, some of whom benefited from his fortuitously useful advice.

Murchison thought that the Russian gold was emplaced by quite recent tectonic activity in the Ural region, and he suggested that the range had undergone several distinct upheavals. Nevertheless, he gave credence to the theory of Élie de Beaumont that mountain ranges with different alignments were of different ages. Thus, the old Palaeozoics of Scandinavia were aligned SW–NE; the supposedly post-Carboniferous/Permian Urals ran N–S; while the post-'oolitic/chalk' of the Caucasus ran WNW–ESE.

The position of Director-General of the Geological Survey of Great Britain fell vacant in 1855, upon the decease of its founder Sir Henry De la Beche, and Murchison was appointed in his place, though already in his sixties. It proved to be an astute move, so far as the Survey's progress was concerned, for Murchison had innumerable contacts and used them to advantage to build up the organization considerably. He was indefatigably a man of organization, and competent, with his experience in the running of several societies and associations. The appointment was gratifying to Murchison as it ensured that the official maps should be constructed and coloured according to *his* interpretation and subdivisions of Palaeozoic geology, to the extent that the Cambrian was almost driven off the map for British geology.

Much of Murchison's geological work in his later years was focused on Scotland, where a separate branch of the Survey was established in 1867, though surveying had begun there back in 1854. Murchison's Scottish work involved him in the last of his three great controversies, and involved the attempted expansion of Silurian colours over the greater part of northern Scotland. There is today thought to be a great thrust-plane (the 'Moine Thrust') that runs from the north coast near Lochs Durness and Eriboll to the south-west, terminating in the southern part of Skye. To the west, one finds 'Fundamental [Lewisian] Gneiss' (so called by Murchison), overlain unconformably by the unfossiliferous Torridonian Sandstone. Lying unconformably on this there is a series of sediments, including a 'quartz rock' and the fossiliferous Durness Limestone. Over this lies the complex unit called the Moine Schists, above the thrust plane and extending eastwards until it is itself overlain unconformably by Old Red Sandstone on the eastern side of the country. But the structure near the thrust plane is complicated, with folding, inversions, and apparent duplication or repetitions of strata; in places the gneiss 'reappears', both near the thrust fault and again further east.

Murchison visited the north-west Highlands of Scotland in 1855 (with the Aberdeen University geologist James Nicol), in 1858 (with the local amateur naturalist Charles Peach), in 1859 (with the Survey officer Andrew Ramsay), and in 1860 (with the young surveyor Archibald Geikie). Fossils regarded by Murchison as Lower Silurian were found by Peach in the Durness Limestone (at a lower horizon than the thrust plane). All the strata appeared to dip gently to the southeast, with a strike approximately parallel with what is now thought to be the thrust-fault system. The outcome of all this work was that in the view of Murchison (and also Ramsay and Geikie) there was an essentially simple ascending sequence (with unconformities) from 'Fundamental Gneiss' on the west (regarded as lying at the bottom of the whole stratigraphic column for Britain) through to the Old Red Sandstone on the east, with a repetition of quartz rock into distinct upper and lower units, and also repetition of the gneiss. This meant that the Moine Schists, lying between the supposed Lower Silurian Durness Limestone and the Devonian Old Red Sandstone, though unfossiliferous, could be regarded as Silurian. So when a geological map of Scotland was published by Murchison and Geikie in 1861, large areas of northern Scotland were represented in Silurian colours. Murchison's empire was again expanding in a manner that he found most satisfactory. As to the Cambrian, Murchison allocated the unfossiliferous Torridonian Sandstone to that System, so Sedgwick saw some expansion his empire, but not by rocks with well-characterized fossils.

But Nicol's reading of the structure was very different from Murchison's. He came to the view that there was repetition of the western and eastern metamorphic rocks due to a large (high-angle) fault, and that the resultant fissure had been filled in part by some kind of igneous rock. (He was perhaps mistaking some gneiss for igneous rock.) If this interpretation were correct then placement of the Moine Schists in the Silurian would be suspect. So Nicol and Murchison fell out, and Nicol thereafter conducted his work separately from the Survey chief. The issues were debated at the British Association meeting in Aberdeen in 1859, where both geologists put forward their cases. From his stronger social position, Murchison was judged the winner by most geologists, and in fact Nicol's idea was by no means wholly correct.

The results of this encounter were most satisfactory to Murchison and Geikie, who became his mentor's advocate and eventually his sympathetic biographer. Murchison got more Silurian colour onto the geological map of Britain. In time, Geikie was appointed head of the Scottish branch of the Survey, and, when Murchison endowed a chair in geology at Edinburgh University (with Geikie's urging), it was Geikie who moved smoothly into the position, holding it concurrently with his post in the Survey. Later he became Director-General of the Survey, President of the Royal Society, and one of Britain's leading geologists.

However, the Murchison theory of the structure of the north-west Highlands was shown to be in error by Charles Callaway's and Charles Lapworth's mapwork in the early 1880s, and the reputation of the deceased Nicol was restored. Lapworth showed that the structure involved folding and thrust-faulting (a term proposed by Geikie), and suggested that the Moine Schists were in fact formed by the earth movements that gave rise to the folding and faulting, while the repetitions of rock types could be attributed to the S-shaped folding. (This suggestion was eventually taken up by Geikie's own staff, surveying in the 1880s, well after Murchison's death.) The whole episode illustrates Murchison's dominating personality and commanding social role towards the end of his career.

The reasons underlying the Cambrian–Silurian debate have been analyzed by Rudwick (1976) in the following terms. At one level, it arose because Murchison's structural interpretations were not always correct and because he confused the May Hill Sandstone (Wenlock) with the lithologically similar Caradoc Sandstone (Caradoc). Both geologists gained ideas about how to do stratigraphy from William Smith (*see* **Famous Geologists:** Smith). Smith himself started from the observation of superposed sections of rocks of characteristic structure and lithologies. Subsequently, he remarked that each rock suite had its own characteristic fossils, but he saw no reason in principle why one set of fossils should not graduate into another. Thus there could, in principle, be overlap between Cambrian and Silurian fossils. Murchison started off on a similar basis, but gradually shifted towards thinking that it was fossils that *defined* a system. Once this had happened, and he began to find 'Silurian' fossils in Sedgwick's Cambrian, then annexation of territory 'naturally' followed (given that Sedgwick was so slow in getting his 'Cambrian' fossils published). It seemed to Murchison that he was dealing with a *bona fide* system, as it preceded land plants, had few vertebrates, and was apparently distributed widely round the world. By contrast, when Sedgwick got round to palaeontological analysis about a decade after his initial fieldwork in North Wales, he thought that the break should, if anywhere, lie between Murchison's Lower and Upper Silurian; so that for Sedgwick the Cambrian should incorporate Murchison's Lower Silurian. But by then the Lower Silurian was already well established, with its fossils described.

Rudwick further points out that both geologists were opposed to Lyell's 'steady-statism' (*see* **Famous Geologists:** Lyell). They both believed that life originated at some point in the past, and Murchison wished 'his' system to be the one that contained the first evidences of life with hard-bodied remains. Hence he sought to cannibalize Murchison's Cambrian. When Barrande in Bohemia found a 'Primordial' fauna below Murchison's Lower Silurian (palaeontologically defined), it could have served as palaeontological basis for a Cambrian System. But Murchison declined to follow this path, and did not practise what he preached in the matter of the Cambrian.

Like many geologists of his day, Murchison gave considerable attention to the problem of the superficial 'drift' deposits that blanket much of Europe, and which he saw in abundance in Scandinavia, Russia, Britain, and elsewhere. In the early nineteenth century, such materials were commonly ascribed to the Noachian Flood, or later to catastrophic floods but not necessarily universal or of divine origin. In the 1840s, there were two further contending theories: that of climatic change producing an Ice Age, with land ice as the agent for the emplacement of the 'drift', as advocated by Louis Agassiz (*see* **Famous Geologists:** Agassiz); and various versions of 'glacial submergence', with cooling and changes of sea-level relative to the land such that floating icebergs could carry detritus and deposit mud and 'erratic' boulders, as envisaged by Darwin and Lyell. The 'flood theory' received some theoretical support from the Cambridge mathematician and geologist, William Hopkins, who advocate the idea of 'waves of translation'. A sudden uplift of the sea-floor might, it was suggested, produce not only waves at the ocean surface, but also wholesale lateral movement of masses of water, capable of transporting ('translating') large boulders and finer debris.

It was Hopkins' theory that Murchison favoured, in part because it was seemingly in accord with the evidences familiar to him in the Alps and elsewhere of huge earth movements, foldings, faulting, and even inversions. (Murchison had seen evidence of stratigraphic inversion in the Glarus Canton, Switzerland, when he visited the area in 1848, but subsequently disregarded it in his thoughts about the north-west Highlands of Scotland.) He accepted that retreating glaciers left moraine material in the Alpine regions, and was happy with the idea of icebergs transporting drift material. But for long he could not accept land-ice as being responsible for the huge tracts of drift on land of low relief that he saw in Russia. Besides, the evidence of striations did not seem to accord with the land-ice theory. For example, in the area of the Gulf of Bothnia he saw scratch-marks directed southeastwards, from an area of Sweden of low altitude. He did not imagine that glaciers could have come from further north, from the mountains of Arctic Sweden. Nor could he imagine that land-ice could on occasions travel uphill, transporting marine shells to hill tops. It was only in 1862 that Murchison conceded to Agassiz's land-ice theory.

Murchison was one of the heroes of the heroic age of geology. His contributions to stratigraphy, and the broadening of geological knowledge generally, were immense. He was extraordinarily energetic, and generally amiable. Other than Lyell, he was far the most influential British geologist of his day. But he exercised that influence through the hand of administrative power as much as by reasoned argument. He was a conservative in both politics and geological theory. Even his protégé Geikie, who owed him so much, concluded that Murchison "was not gifted with the philosophical spirit which evolves broad laws and principles in science. He had hardly any imaginative power. He wanted ... the genius for dealing with questions of theory ..." Possibly things would have been different if Murchison had received a university, rather than a military, education.

See Also

Famous Geologists: Agassiz; Darwin; Lyell; Sedgwick; Smith. **History of Geology From 1780 To 1835**. **History of Geology From 1835 To 1900**. **Palaeozoic:** Ordovician; Silurian; Devonian; Carboniferous.

Further Reading

Geikie A (1875) *Life of Sir Roderick I. Murchison... Based on his Journals and Letters With Notices of his Scientific Contemporaries and a Sketch of the Rise and Growth of Palaeozoic Geology.* London: John Murray (reprinted in facsimile by Gregg International Publishers Ltd, 1972).

Oldroyd DR (1990) *The Highlands Controversy: Constructing Geological Knowledge through Fieldwork in Nineteenth Century Britain.* Chicago and London: Chicago University Press.

Rudwick MJS (1972) Levels of Disagreement in the Sedgwick Murchison Controversy. *Journal of the Geological Society* 132: 373 375.

Rudwick MJS (1985) *The Great Devonian Controversy: The Shaping of Scientific Knowledge among Gentlemanly Specialists.* Chicago and London: Chicago University Press.

Secord JE (1986) *Controversy in Victorian Geology: The Cambrian Silurian Dispute.* Princeton: Princeton University Press.

Stafford RA (1989) *Scientist of Empire: Sir Roderick Murchison, Scientific Exploration and Victorian Imperialism.* Cambridge, New York, Port Chester, Melbourne and Sydney: Cambridge University Press.

Thackray JC (1976) The Murchison Sedgwick Controversy. *Journal of the Geological Society* 132: 367 372.

Sedgwick

D R Oldroyd, University of New South Wales, Sydney, Australia

Introduction

Adam Sedgwick (Figure 1) was born in 1785 in the village of Dent in the Yorkshire Dales, northern England, son of the local vicar and third of a family of seven. He died as a Fellow of Trinity College and Professor of Geology at Cambridge in 1873. He attended Sedbergh School near Dent, and with help from a notable local amateur mathematician, John Dawson of Garsdale, obtained a scholarship to Trinity, where he studied mathematics. He was named 5th Wrangler (fifth in the list of first-class honours students) in 1808. Following further examination, Sedgwick obtained a College Fellowship in 1810 and taught undergraduate mathematics. He was ordained in 1817. On becoming a permanent member of college, Sedgwick also committed himself to bachelorhood. During his life as a geologist, he proved to be extremely energetic in the field, covering large distances in a day. In Cambridge, he was quite often indisposed, but his chronic health problems apparently disappeared once he got into fieldwork.

Though a gifted mathematician, Sedgwick did not make a career in that discipline. From fragmentary autobiographical notes, he evidently had some geological interests from an early age, and he 'geologized' on the Continent in 1816. Also, he was 'introduced' to the Geological Society of London in 1818. Even so, it is surprising that his scientific accomplishments were thought sufficient to secure the Cambridge chair in geology that year. He was elected Fellow of the Royal Society in 1821, John Herschel heading the list of those who nominated him. Sedgwick was President of the Geological Society in 1829–31, and President of the British Association when it met in Cambridge in 1833.

Geological Work

On obtaining his chair, Sedgwick threw himself into geology. He started his annual fieldwork in southern England, then worked his way northwards to Northumberland, and in the years 1822–24 he made the first systematic survey of the Lake District. He obtained topographic maps of the region, 'recognized' certain rock units, and systematically covered

Figure 1 Portrait of Adam Sedgwick (1785 1873); the original is in the Sedgwick Museum, Cambridge. Copyright: The Sedgwick Museum of Earth Sciences.

the region over three seasons, colouring in his maps according to his selected lithological units. Sedgwick did not look much for fossils, but measured strata and cleavage dips and the alignments of beds, folds, faults, joints, and cleavage planes. Faults sometimes could be seen on the ground. At other times, they became apparent when the different rock types were coloured onto the maps. Sedgwick's labour and energy were immense. Armed with hammer, acid bottle, map, compass, clinometer, and notebooks, Sedgwick tried to determine the structure of that complicated region.

Trained in mathematics, the neophyte geologist was trying to ascertain whether the strata displayed any regular geometric pattern. They hardly did, but when Sedgwick published his work he sought to subsume it under the theory of Léonce Élie de Beaumont, according to which, as the Earth cooled and contracted, it supposedly formed a regular fold pattern in its crust, with mountain ranges of similar age having similar alignments. The theory never exerted much influence in Britain, and Sedgwick soon gave up the idea; but the fact that he sought to deploy the French theory suggests that he was interested in a geometrical (mathematical) theory of the earth. This was consistent with the Cambridge tradition, which found expression in the activities of the Cambridge Philosophical Society, which Sedgwick helped found in 1819. He wanted a quantitative geology, with mathematically formulated laws. However, his later Lakeland work (assisted by amateur collectors) used fossils, and by the end of his career, he had worked out a stratigraphic order for the sedimentary rocks compatible with that later developed on biostratigraphic principles. Sedgwick also recognized the rocks of central Lakeland for what they were: the relics of ancient volcanoes. He referred to water-deposited volcanic ash deposits as volcanic mud. From his Lakeland work, Sedgwick came to understand (and publish) the distinction between bedding and cleavage, but the distinction was acquired from the local amateur Jonathan Otley, who probably got it from quarrymen.

Sedgwick's Lakeland mapping revealed the existence of large tear-faults in some of the strata, and the eastern boundary of the region was marked by a huge normal fault. Hence, he suggested that the area had been affected by great earth movements. Using terminology proposed in the 1830s by his Trinity colleague, William Whewell, Sedgwick was a 'catastrophist'.

Sedgwick soon met Roderick Murchison (*see* **Famous Geologists:** Murchison), who wanted someone to show him how to make sense of ancient rocks in mountainous regions, and together they undertook a lengthy reconnaissance journey round the north coast of Scotland in 1827, unfortunately correlating the eastern and western sandstones of northern Scotland (now regarded as Devonian and Precambrian, respectively). In 1829, they made an extensive tour of the Continent, meeting European professors and travelling to Germany, Bohemia, Hungary, Austria, Switzerland, and Italy. This greatly extended Sedgwick's experience, especially in the "focus of Wernerian geology," southern Germany, which he found to be "the most decidedly volcanic secondary country I ever saw". He observed granite veins and inclined or even overturned Secondary rocks (a term used in the nineteenth century as a synonym for the German term *Floetz*; later for the strata ranging from Silurian to Cretaceous; and later restricted to the Mesozoic Era). This demolished his earlier adherence to Neptunism and he publicly repudiated the doctrine in 1831. Observations in Italy clearly suggested uplift, which was not part of the Wernerian repertoire. On the other hand, Sedgwick was inclined to ascribe the movement of (glacial) erratic boulders to the action of catastrophic floods, but some Swiss deposits could have been emplaced by the bursting of lake barriers. He rejected the idea that 'diluvium' was all deposited in the Noachian Flood, but was critical of Charles Lyell's belief that conditions on Earth were essentially similar through time (*see* **Famous Geologists:** Lyell).

In the 1830s, Sedgwick collaborated with Murchison in Wales, trying to bring order to the strata there. The strata in the mountains of Snowdonia seemed to have analogy with those of the lakes. Sedgwick tackled them, working on somewhat similar lines, and making a traverse north-west to south-east across Snowdonia in 1832. He unravelled the structure to some extent, and although the rocks seemed to have few fossils, the Cambrian System was introduced, largely on the basis of Sedgwick's work. Murchison, starting in the Welsh border region in gentler country with fossiliferous rocks, had an easier time and revealed what appeared to be a new system with its characteristic fossils, dubbed the Silurian.

However, the line of boundary between the two systems was not established at the outset of the investigations in 1834, with the result that a bitter feud subsequently broke out between Sedgwick and Murchison, with the former trying to extend 'his' system upwards, the latter endeavouring to expand 'his' downwards, eventually to the very lowest fossiliferous rocks. (The issues were very complex; it was not until after the protagonists' deaths that the issues were resolved, by Charles Lapworth, who, in 1879, proposed the Ordovician System to occupy the disputed territory between the Cambrian and Silurian. (*see* **Palaeozoic:** Ordovician) Sedgwick had complained, with reason, that materials he had sent to the Geological Society were changed so as to accord with Murchison's views, without Sedgwick's knowledge or consent. The situation grew so bad between the two that after 1853, the Geological Society declined to accept further papers by the protagonists of Siluria or Cambria. Sedgwick felt grievously ill treated and snubbed by the Society. The battle became transferred to the forum of the British Association, but after 1854, Sedgwick withdrew from that body also, so far as the Cambrian and Silurian were concerned, and continued the battle from Cambridge and in the pages of the *Philosophical Magazine*. Some of the stratigraphic formations and their classifications are shown in Figure 2. It should be noted that the rocks that Murchison allowed to be Cambrian in 1859 were the unfossiliferous Longmynd rocks, later classified as Precambrian. (Murchison also allowed Sedgwick the unfossiliferous Torridonian Sandstone in Scotland, also now regarded as Precambrian.)

Initially, Murchison won the battle, partly because he had better fossil evidence, and placed full reliance on it. Also, he had strong influence in the Geological Society and became Director of the Geological Survey in 1855, and his classifications were used by the survey officers. Sedgwick had fewer allies, mostly at Cambridge. For the rocks he was dealing with, he had to rely on structural understanding and lithologically based mapping to a greater extent than did Murchison. (Graptolites were not regularly used for stratigraphic correlation in the mid-nineteenth century.) However, Sedgwick succeeded in showing, on palaeontological grounds, that the claimed unity of Murchison's Silurian System was flawed. In 1852, Sedgwick and his assistant Frederick McCoy found that one of Murchison's Silurian formations, the Caradoc, had rocks containing two distinct faunas, as shown by the palaeontological determinations of McCoy and John Salter. There had been erroneous correlations; the same term, 'Caradoc Sandstone', had been applied to different series of rocks; and there should be an unconformity within the Caradoc, as Murchison then understood it. Sedgwick proposed the division of Murchison's Caradoc into the Caradoc Sandstone, containing fossils such as *Trinucleus*, and an upper May Hill Sandstone, containing *Pentamerus* species. This eventually turned the tide against his Silurian being regarded as a coherent system. The Survey sought to retrieve the situation by adopting the terminology 'May Hill Sandstone Llandovery' rocks, regarding them as a kind of passage or 'Intermediate Series' between the Upper and Lower Silurian. There were repercussions, too, for the interpretation of Murchison's Llandeilo.

This battle was fought with extreme vehemence. Both geologists attached their names and reputations to 'their' system. The battle seemed to exemplify the height of the colonial era, with Sedgwick and Murchison trying to extend their empires. Murchison was popularly called the 'King of Siluria'. Both men tried to rewrite history in their historical accounts of the events. In Sedgwick's case, this may have been partly due to failing memory. Earlier, there had been a bitter controversy in Devonshire, where, while still friends, Sedgwick and Murchison began to unravel the structure and stratigraphy of the area, in the process becoming involved in controversy with Henry De La Beche, the first Director of the Geological Survey. From this acrimonious debate emerged the concept of the Devonian System.

The so-called Old Red Sandstone was the unit well known in eastern Scotland as lying unconformably over Silurian strata. It was mapped by William Smith as 'Red and Dunstone'. He placed it below the Coal Measures and below a limestone that cropped out in Derbyshire and elsewhere. In Devonshire, there occurred rocks with plant remains that appeared to De la Beche to belong to the old Transition/Greywacke series (Cambrian or Silurian rocks). In Murchison's opinion, however, the plant-bearing rocks could not be so old: they must be from the Coal Measures. But he had not then been to Devon to see the rocks for himself. Murchison then combined forces with Sedgwick to combat De la Beche's interpretation. It was an issue of more than academic significance because it bore on the question of the possible extent of coal-bearing rocks.

The plant-bearing rocks overlay contorted rocks of ancient appearance, but these contained corals

<table>
<tr><td></td><td>1</td><td>2</td><td colspan="2">3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td></tr>
<tr><td>SEDGWICK
1855</td><td colspan="3">Silurian</td><td colspan="2">Upper Cambrian</td><td colspan="4">Middle Cambrian</td><td>Lower Camb.</td></tr>
<tr><td>MURCHISON
1859</td><td colspan="3">Upper Silurian</td><td colspan="6">Lower Silurian (Primordial Silurian)</td><td>Camb.</td></tr>
<tr><td>GEOLOGICAL SURVEY
1866</td><td colspan="3">Upper Silurian</td><td colspan="6">Lower Silurian</td><td>Camb.</td></tr>
<tr><td>JUKES
1857</td><td colspan="3">Upper Silurian</td><td colspan="6">Cambro-Silurian</td><td>Camb.</td></tr>
<tr><td>PHILLIPS
1855</td><td colspan="3">Upper Silurian</td><td colspan="4">Lower Silurian</td><td colspan="3">Cambrian</td></tr>
<tr><td>LYELL
1865</td><td colspan="2">Upper Silurian</td><td colspan="2">Middle Sil.</td><td colspan="3">Lower Silurian</td><td colspan="3">Cambrian</td></tr>
<tr><td>LYELL
1871</td><td colspan="4">Upper Silurian</td><td colspan="3">Lower Silurian</td><td colspan="3">Cambrian</td></tr>
<tr><td>HICKS
1874</td><td colspan="2">Upper Silurian</td><td colspan="2">Middle Sil.</td><td colspan="3">Lower Silurian</td><td colspan="3">Cambrian</td></tr>
<tr><td>LAPWORTH
1879</td><td colspan="4">Silurian</td><td colspan="3">Ordovician</td><td colspan="3">Cambrian</td></tr>
<tr><td></td><td>1</td><td>2</td><td colspan="2">3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td></tr>
<tr><td>Principal formations</td><td>Ludlow</td><td>Wenlock</td><td>Upper Llandovery = May Hill</td><td>Lower Llandovery</td><td>Bala = Caradoc Sandstone</td><td>Llandeilo</td><td>Arenig</td><td>Tremadoc</td><td>Lingula Flags</td><td>Longmynd</td></tr>
</table>

Figure 2 Classification of British Lower Palaeozoic Rocks. Reproduced from Secord JA (1986) *Controversy in Victorian Geology: The Cambrian Siurian Dispute*, p. 287. Copyright ©1986 by P.U. Press. Reprinted by permission of Princeton University Press.

different from those in the Silurian. In the opinion of the coral expert William Lonsdale, the corals were intermediate between Silurian and Coal Measure types. Thus, the Devon rocks might be situated between the Silurian and the Carboniferous. So, in 1839, Sedgwick and Murchison proposed the Devonian System, being one that had different facies in different localities. Here the palaeontological evidence of corals was taken to outweigh the uncertain stratigraphic reliability of plant fossils and the structural arguments advanced by De la Beche. The issues were debated with considerable rancour and again illustrate the character of stratigraphic controversies in the nineteenth century. But this time, Sedgwick and Murchison were on the same side and the debate never got quite so out of hand as did the Silurian/Cambrian contest.

Sedgwick as a Teacher; Other Activities, Beliefs, and Character

At Cambridge, Sedgwick gave an annual course of lectures and built up the university's geological collections, partly from his own collected specimens, but also by donations and purchases. His summer fieldwork was done at his own expense. Partly for this reason, he took a 'second job' in 1834, as a canon at Norwich Cathedral. This might have been a sinecure, but Sedgwick took his responsibilities seriously, and resided in Norwich for several months each year, also encouraging the development of a museum in the city and giving geological lectures. Sedgwick was a strong supporter of amateur science and assisted the natural history society in Kendal, near Dent.

Sedgwick was a renowned orator – or preacher and lecturer. Science lectures were not a required part of the Cambridge curriculum when he gave his first course in 1819, but he attracted many students and dons to his lectures. His course was repeated until 1859, when he was 74. He spoke extempore about geological principles and his recent fieldwork, rather than about unnecessary minutiae. His lectures, utilizing specimens and maps and diagrams to explain his ideas, were popular, and he also led groups on horseback on field excursions round Cambridge. On a famous occasion at the British Association meeting at Newcastle in 1838, he spoke in the morning at Tynemouth beach to a group attending the meeting; and by the afternoon he had attracted a crowd of thousands, expounding the relationships (as he saw them) between geology, political economy, natural theology, and patriotism, reportedly drawing tears of emotion from some auditors. As John Herschel described it, Sedgwick

> led them on from the scene around them to the wonders of the coal country below them, thence to the economy of a coal field, then to their relations to the coal owners and capitalists, then to the great principles of morality and happiness, and last to their relation to God and their own future prospects.
> (Clarke and Hughes (1890).)

In this can be seen the relationship between Sedgwick's science, social, religious, and political philosophy. Implausibly, he supposed that Britain had been specially favoured by God for its place in the world, with its deposits of coal, limestone, and iron ore.

Sedgwick was, then, devoutly religious, and a preacher as well as a teacher. From a relatively obscure Anglican background, he rose to be a Cambridge Professor and Vice-Master of Trinity, and one of the leaders of the heroic age of geology. He even met with Prince Albert (whose election to the Chancellorship of the University he promoted) to discuss reforms at Cambridge, and many of Sedgwick's suggestions were implemented. Though generally amiable, greatly liked and admired, and able to communicate with the full range of society, from quarrymen, to famous writers such as William Wordsworth or Walter Scott, to Royalty, he was uncompromising and dogmatic. He favoured Catholic emancipation, but having become an establishment figure, he did not wish to see the regular order of things upset by scientific theories that seemed to him subversive, or at odds with orthodox Anglican theology. For such reasons, he was bitterly and publicly opposed to the transmutationist ideas expressed in 1844 by Robert Chambers in his *Vestiges of the Natural History of Creation*, and was privately grieved by Charles Darwin's ideas in *The Origin of Species* (*see* **Famous Geologists:** Darwin). He also rejected the land–ice theory of Louis Agassiz (*see* **Famous Geologists:** Agassiz).

See Also

Famous Geologists: Agassiz; Darwin; Lyell; Murchison; Smith. **History of Geology From 1780 To 1835**. **History of Geology From 1835 To 1900**. **Palaeozoic:** Cambrian; Ordovician; Silurian.

Further Reading

Clark JW and Hughes TMcK (1890) *The Life and Letters of the Reverend Adam Sedgwick, LL.D., D.C.L., F.R.S., Fellow of Trinity College, Cambridge, Prebendary of Norwich, Woodwardian Professor of Geology, 1818 1873*. Cambridge: Cambridge University Press.

Oldroyd DR (2002) Adam Sedgwick: a confident mind in turmoil. In: Harman P and Mitton S (eds.) *Cambridge Scientific Minds,* pp. 64 78. Cambridge: Cambridge University Press.

Oldroyd DR (2002) *Earth, Water, Ice and Fire: Two Hundred Years of Geological Research in the English Lake District*. London: The Geological Society.

Rudwick MJS (1972) Levels of disagreement in the Sedgwick Murchison controversy. *Journal of the Geological Society* 132: 373 375.

Rudwick MJS (1985) *The Great Devonian Controversy: The Shaping of Scientific Knowledge among Gentlemanly Specialists*. Chicago and London: Chicago University Press.

Rudwick MJS (1988) A year in the life of Adam Sedgwick and company, geologists. *Archives of Natural History* 15: 243 268.

Secord JA (1986) *Controversy in Victorian Geology: The Cambrian Silurian Dispute*. Princeton and Guildford: Princeton University Press.

Sedgwick A and Murchison RI (1835/36) On the Silurian and Cambrian Systems, exhibiting the order in which the older sedimentary strata succeed each other in England and Wales. *Report of the Fifth Meeting of the British Association for the Advancement of Science held at Dublin in 1835,* pp. 59 61. London: John Murray.

Smith C (1985) Geology and mathematicians: the rise of physical geology. In: Harman PM (ed.) *Wranglers and Physicists: Studies on Cambridge Physics in the Nineteenth Century,* pp. 49 83. Manchester: Manchester University Press.

Speakman C (1969) *Adam Sedgwick Geologist and Dalesman, 1785 1873: A Biography in Twelve Themes.* Broad Oak, London, and Cambridge: The Broad Oak Press Ltd.

Smith

D R Oldroyd, University of New South Wales, Sydney, Australia

Introduction

William Smith (Figure 1) is renowned in the history of geology for differentiating and listing in order the English strata, from the Chalk down to the Coal Measures, and hence enunciating the stratigraphic principle that strata have a generally regular order of superposition and may be characterized by their fossil contents. The great geological map of much of Britain that Smith published in 1815 has also brought him much posthumous fame. As Smith put it in the memoir accompanying his map:

> [T]here is a great deal of regularity in the position and thickness of... strata; and although considerable dislocations are found in collieries and mines, and some vacancies [gaps] in the superficial courses of them, yet... the general order is preserved; and... each stratum is... possessed of properties peculiar to itself, has the same exterior characters and chemical qualities, and the same extraneous or organized fossils throughout its course. (Smith W (1815, p.2))

Background

Smith was born in 1769 in the village of Churchill, Oxfordshire, near Chipping Norton, an attractive part of Britain where Jurassic rocks (Oolitic Limestone and Lias) crop out well. His father, a blacksmith, died when he was only 7 years old, and he was then brought up by a farmer uncle at Over Norton, a few miles to the north-east, in similar type of country. Even at an early age, Smith was a keen collector of fossils. Not wishing for a life as a farm labourer, he began studying mathematics, geometry, and surveying techniques, and at age 18 became an assistant to a local surveyor.

Figure 1 William Smith (1769 1839) aged 69, engraved by TA Dean.

Smith soon became autonomous, and his work as a land surveyor evolved into the business of surveying for canal construction (initially the Somerset Coal Canal). This type of construction was then forging ahead in the days of the Industrial Revolution, and in effect, Smith became what would today be called a civil engineer. He advised on tunnel constructions, borings for coal, and mining activity in the Somerset coalfield. In 1794, he travelled to northern England on a 'fact-finding tour', in connection with canal work. He later advised on drainage projects for wealthy landowners who sought to develop their lands agriculturally, but these landowners were also interested in the mineral wealth that their estates might hold.

Development of Smith's Stratigraphic Principle

The varied experience Smith acquired, and especially that resulting from the canal cutting through different stratigraphic horizons, led, before 1796, to Smith recognizing the general aspects of his stratigraphic principle. He realized that the regular stratifications found within the coal mines could also be observed outside the mines. The Somerset Canal cuttings, which cut through two similar sections at two separate but neighbouring localities, revealed the lateral extents of strata, and Smith began to get the 'feel' of the internal structure of the earth as regards the strata of his region. In 1795, he took up residence in Bath and pondered what his work was beginning to reveal, namely, that "Nature has disposed of these singular productions [fossils] and assigned to each class its peculiar Stratum". In 1797, he wrote down a list of strata for the Bath district, listing 28 units, from Chalk down to Carboniferous Limestone, but without mentioning fossils (this is his earliest dated list that has survived). A revised version of this document (with fossils given) was dictated to two local clergymen/naturalists, the Reverends Benjamin Richardson and Joseph Townsend in 1799, who were also shown a circular map of the Bath district and one of

Somerset that Smith had geologically coloured. Not long before, Smith's employment with the Somerset Coal Canal Company had been terminated and it is possible that he was beginning to think of finding some practical and remunerated application of his new ideas. During the next few years, he was largely involved with drainage schemes, but by 1804, he was chiefly employed in 'mineral surveying'. He travelled great distances in these lines of work, and thus accumulated information that he later synthesized in the form of his celebrated geological map of 1815.

The revised stratigraphic table of 1799 had 23 stratigraphic units, from Chalk down to Coal, with thicknesses indicated, along with localities where the rocks cropped out. 'Fossils and Petrifactions' were given for each unit, but the fossil categories, as stated by him, were imprecise (corals, cochleae, ostreae, impressions of ferns, etc.). Also, some strata were characterized lithologically, or the entry was stated "No fossils known". So Smith's earliest table of strata was not based wholly on fossils, and appears to have been primarily a list of the lithological sequence of distinctly identifiable strata. At that time, he had no 'scientific' knowledge of fossils, but he was certainly collecting fossils well before 1799; his collections survive today at the British Museum. In any case, his differentiation of four blue clays and two different oolitic limestones indicates that he was making more use of fossils than is evident from the table dictated to Richardson and Townsend. They were impressed by the practical man's revelations and encouraged Smith to continue his investigations, which he did, producing a simple geological map of England and Wales in 1801. It showed quite a clear representation of the distribution of several major stratigraphic subdivisions, notably those now known as the Carboniferous, Jurassic, and Cretaceous.

With the loss of forests for shipbuilding and development of agriculture, and increasing demands for iron, there was need for coal in the years of the Napoleonic Wars, and various prospecting attempts were made in southern England, often on the illusory promise of the discovery of lignites or the occurrence of blue clays, thought by their appearance to be associated with coal measures. But Smith knew the correct order of strata by their fossils and realized that these attempts were doomed to failure. The prospectors were looking too high in the stratigraphic column. On the other hand, though Smith also gave sometimes successful advice as to where coal might be found in known coal areas, he was not always successful, due to unforeseen structural complexities, so that the coal beds present were unfortunately missed.

In fact, Smith's own entrepreneurial ventures were by no means successful. His intended book, *Accurate Delineations and Descriptions of the Natural Order of the Various Strata that are Found in the Different Parts of England and Wales, with Practical Observations Thereon*, for which a prospectus was printed in 1801, was never published because of the double bankruptcy of the intended publisher in 1801 and 1804. (The text of this prospectus was published in a 1942 article by LR Cox.) Worse, in 1798, Smith purchased a small estate at Tucking Mill near Bath, where quarrying of Bath stone was later attempted. This proved a failure because of the unexpectedly poor quality of the stone, and the financial problems that flowed from this led to Smith's subsequent financial collapse. (It is a sad irony that the experienced engineer and emerging geologist should have misjudged the stone quality and the difficulty of getting it out of the quarry.) On the other, hand, through agricultural contacts at Woburn, seat of the Duke of Bedford, Smith was introduced to the influential Sir Joseph Banks, President of the Royal Society, from whom Smith received both moral and financial support. Also, in 1801, Smith attracted, as a kind of 'pupil', the polymathic John Farey (likewise a practical man, but one of many accomplishments, including mathematics and music). They did fieldwork together and Farey became a constant advocate of Smith and his work. Smith also became acquainted with the map publisher and engraver John Cary, whose map of England and Wales and various county maps became the basis of Smith's great geological map (see later).

In 1802, Smith opened an office in Bath to conduct his affairs, and in 1804, he moved to London, where he displayed his fossils, arranging them on sloping shelves so that the fossils of each stratum were displayed in the order that they appeared in the English rocks, producing a kind of three-dimensional effect. But his efforts could only be spread thinly, and not necessarily systematically, because of the exigencies of his work. Although it is clear that Smith long intended to attempt to publish a geological map of England, Wales, and southern Scotland, this project was not in fact fulfilled until 1815, partly because of difficulties in reaching an agreement with a publisher and partly because he was continuing with his observations and collections. Smith's business took him all over the country, and he developed his topographic and geological knowledge wherever he went, to the point where he was almost overwhelmed with information and specimens.

The Geological Society of London was founded in 1807, with the wealthy George Bellas Greenough as its president. It might have seemed natural or appropriate for Smith to have joined the Society, but this did not happen. Smith was from a lower social class than were the Society's founders, and his patron, Banks, was at odds with the new group, which he saw as a rival to 'his' Royal Society. Additionally, the Society's early Fellows were chiefly interested in mineralogical matters, and they and others of influence, such as the Board of Agriculture or Professor John Kidd at Oxford, doubted the value of Smith's 'biostratigraphy'. So, to an extent, Smith was on his own, and he had to carry through his project using his own uncertain financial resources.

Eventually, however, in 1815, Smith issued his great map and its accompanying memoir – *A Delineation of the Strata of England and Wales, with Part of Scotland* and *A Memoir to the Map and Delineation of the Strata of England and Wales, with Part of Scotland*. The geological information was entered on a map specially engraved for the purpose by Cary. There followed *Strata Identified by Organized Fossils, Containing Prints on Coloured Paper of the Most Characteristic Specimens in each Stratum*, which was issued in four parts between 1816 and 1819, providing descriptions of Smith's chosen stratigraphic units and beautiful coloured illustrations of their typical fossils, and *Stratigraphical System of Organized Fossils*, which was issued in 1817. (Neither publication was fully completed.)

The main map (dedicated to Banks) was issued on 16 sheets, one being an index sheet. (The Banks copy at the British Library may be the 'master' copy, but this is not certain.) On a scale of 5 miles to the inch, the map was approximately 260 by 180 cm in size. Each copy was hand coloured (apparently using colourists employed by Cary, not always to Smith's satisfaction) and there were five issues between 1815 and 1817 (or perhaps later). Examination of surviving copies has shown that Smith added information to the new issues as it became available to him. From 1819 to 1824, he issued also various 'county maps', which depicted the geology of individual counties. These lacked 'geological rationale', in the sense that counties were not 'natural' geological regions, but they were useful commercial products. Together they formed *Smith's Geological Atlas*. A smaller country map on the scale of 15 miles to the inch, with revisions of the earlier map, was issued in 1820, and this map was also variously revised until at least 1828. Additionally, Smith produced several geological sections, including one from London to Snowdon in North Wales, in 1817, and one of the strata south of London, in 1817. A manuscript section from 1824, from Flamborough Head on the east Yorkshire coast to Whitehaven on the Cumberland coast on the west of England, is preserved at Oxford.

The 1815 map was a mighty contribution to geology, achieved largely single-handedly. Surviving pristine copies are objects of great beauty, ingeniously and impressively coloured, so as to convey almost a three-dimensional effect, by increasing the intensity of colouration towards the lower boundaries of the outcrops of the various units. The map depicted 23 stratigraphic subdivisions, and some of the colours that Smith chose (e.g., green for Chalk) survive into modern maps.

But Smith's position was financially precarious, as it had been ever since his unsuccessful quarrying venture near Bath, and eventually, in 1817, he was obliged to sell his fossil collections to the British Museum for £700, the catalogue for this being his *Stratigraphical System of Organized Fossils*. The sale only postponed Smith's financial crisis, however, and he found himself languishing in a debtors' prison for 10 weeks in 1819. He obtained release by sale of his property near Bath, but withdrew to the north of England, making a living by continuing his survey work and giving lectures in Yorkshire's major towns. He had previously been assisted by his gifted nephew John Phillips, and their association continued. Phillips was perhaps chiefly responsible for the aforementioned east–west section of the north of England, which revealed an understanding of the faulted structure of the Vale of Eden, to the east of the Lake District. Phillips also did some of the lecturing. (Phillips subsequently became one of Britain's leading geologists, a Fellow of the Royal Society and Professor at Oxford.) While in the north, Smith put both Adam Sedgwick (*see* **Famous Geologists:** Sedgwick) and Roderick Murchison (*see* **Famous Geologists:** Murchison) on the right track as to the use of fossils for stratigraphic purposes. Smith eventually settled in Scarborough and assisted in the founding of the Scarborough Philosophical Society and Museum (which was made largely to Smith's design).

Back in 1808, Smith had been visited by Greenough and other leaders of the Geological Society, but they seemed unimpressed by his work, and subsequently started compiling a collaborative Society map, based on lithological principles such as were typically used by German geologists. However, at some point during the year after Smith was imprisoned, under the influence of the publication of Smith's map, the Geological Society group changed their approach and issued their own map, using some of Smith's fossil-based data, though Greenough asserted that the utility of fossils had been "greatly over-rated". Smith, at that point in time, had been walking to the north with his nephew Phillips as companion. Smith and his supporters, such as Farey, claimed his priority rights on several occasions, but these were not fully acknowledged until there was a change of personnel in the Geological Society. Eventually, in 1831, acknowledgement was accorded Smith by making him the first recipient of the Society's Wollaston Medal (though there were objections made as to whether it was appropriate to make an award for work first announced in 1799). From the chair, the President Adam Sedgwick acknowledged his personal indebtedness to Smith's advice and dubbed him the 'Father of English Geology'. Smith's reputation was thus securely sealed, and the following year he received an annuity of £100 per year from the government. Moreover, his stratigraphic subdivisions set the pattern for work in other countries: the world followed British stratigraphy. If geology had emerged in the United States, China, or New Zealand, say, the stratigraphic column would look substantially different from that which is now used.

Influence of Smith's Work

The question of Smith's theoretical ideas in geology is important. His sections showed the strata of southern England in their correct order, where they conveniently form a 'layer-cake' stratigraphy. His main expertise was in the stratigraphy of these Mesozoic sediments (as they are now called). It is evident that strata of different lithologies were recognized first, and then Smith realized that each stratum had its characteristic fossils. Soon, he could reverse the argument and use fossils to identify the strata. Sometimes, however, he encountered problems. He thought, for example, that the poorly fossiliferous Magnesian Limestone of north Yorkshire (subsequently designated as Permian) and the Lias (now Jurassic) belonged to the same stratum, because they contained rather similar fossil fishes. Also, what is now thought of as Carboniferous Limestone and Lias were regarded on occasion by Smith as different facies of the same unit, there being no locality where the Carboniferous Limestone, Magnesian Limestone, and Lias appear in what is today regarded as the correct order, and some Carboniferous Limestone does occur in places reworked into Lias. This is not to blame Smith. He was pioneering, and mistakes were to be expected in those early days of biostratigraphy.

Throughout his career (started as a surveyor), Smith was always primarily interested in the geometrical arrangement of rocks, because this was what counted for agricultural, mining, and engineering purposes. His livelihood depended on knowing that order. Nevertheless, though it was not his primary concern, he did ask himself why the order was the way it was, how the strata came to be formed, and how long it took for them to be deposited. Smith's religious beliefs appear to have been conventional, or characteristic of his time, and involved use of the 'argument from design'. So one answer (1817) was simply that "[t]he interior of the earth…is formed upon the wisest and best principles", and that the inclinations of the strata evidenced design by making the different rock types available for human use. Fossils must "strike the admirers of nature with a degree of reverential awe and grateful admiration of the Almighty Creator". Earlier, in 1802, in a preface to a book that was never published, Smith had supported an older eighteenth-century idea that the inclinations of the strata were the result of Earth's rotation when the materials were still "in soft state or of pulpy consistence". But Smith apparently dropped this idea, which would imply that stratigraphic order did not represent chronological order. Even earlier, according to an 1844 memoir of Smith by J Phillips, Smith thought (in about 1795) that "each stratum had been successively the bed of the sea, and contained in it the mineralized monuments of the races of organic beings then in existence".

Another shred of evidence on this matter is provided by Farey, in a review he published in 1810 of Georges Cuvier and Alexandre Brongniart's 1808 memoir on the geology of the Paris area. Farey claimed that soon after Smith began his investigations, he "discovered an important law regulating all the *known alluvia*, or that which consisted of or contained the fragments and reliquia of known strata, [namely that they] were moved *from the south-east* towards the south-west" (italics in original) for material from any particular stratum seemed to have been transported beyond its western edge. This appears to have been a reference to observations of boulder clay, or 'drift', which material was ascribed by Smith (or Farey?) to "vast tidal currents which have swept over all the surface from SE. to NW., since or at the time, that the deposition of regular strata ceased". This suggests some support on Smith's part for the catastrophist doctrines espoused by Cuvier (*see* **Famous Geologists:** Cuvier). This could have accorded with Smith's religious views, but would also have involved the notion of time for the emplacement of superficial materials.

Additionally, HS Torrens has drawn attention to a Smith manuscript from about 1806; the manuscript indicates that Smith was then thinking of a vast extent of geological time: "the time required for the Perfection and Decay, and subsequent formation, into Strata which have evidently been formed in deep and quiet water". This time "would stagger the faith of Many". But Smith's lectures in Leeds in 1825 referred specifically to geological proofs of the occurrence of the deluge. He seems to have been impressed with William Buckland's recently claimed evidence for the occurrence of the Noachian Flood from cave excavations in Yorkshire. For further variety, there are 40 proof sheets at Oxford of a work to be titled *Abstract Views of Geology*, which was apparently in press at the time of Smith's death. He was speculating again about the formation of strata "from a chemical conversion of liquids and gases into the solid state, —the layering being the effect of an uncombinable excess of one of the ingredients in the layer then formed, and the vertical joints in that layer the effect of solidification...." But he was not advancing geology through such suggestions. Continuing through his late lectures and this last work, he kept reverting to his great principle of identifying and ordering strata by their fossil contents, and the utility of knowledge of this kind. His speculations about time and process had little influence on the development of geology.

Smith's strata, as given in his table of 1817, with approximate modern equivalents, following JCM Fuller (1995), were as follows:

1. London Clay – Tertiary, Lower Eocene.
2. Sand – Tertiary, Lower Eocene (Woolwich and Reading Beds).
3. Crag – Tertiary, Pliocene (Shelly Sand).
4. Sand – Tertiary, Paleocene (Thanet Sand).
5. Chalk – Upper Cretaceous, Cenomanian to Senonian).
6. Greensand – Upper Cretaceous, Albian (Upper Greensand).
7. Brickearth – Upper Cretaceous, Albian (Gault Clay).
8. Sand – Lower Cretaceous, Aptan (Lower Greensand).
9. Portland Rock – Upper Jurassic, Portlandian–Purbeck.
10. Sand – Lower Cretaceous, Wealden (Ashdown Sand).
11. Oaktree Clay – Upper Jurassic (Kimmeridge Clay) and Lower Cretaceous (Wealden).
12. Coral Rag and Pisolite – Upper Jurassic, Corallian.
13. Sand – Upper Jurassic, Corallian (Lower Calcareous Grit).
14. Clunch Clay and Shale – Upper Jurassic, Oxfordian (Oxford Clay).
15. Kelloways Stone – Upper Jurassic, Callovian.
16. Cornbrash – Middle Jurassic, Bathonian and Upper Jurassic, Callovian.
17. Sand and Sandstone – Middle Jurassic (Hinton Sand).
18. Forest Marble – Middle Jurassic, Bathonian (and Wychwood Sandstone).
19. Clay over Upper Oolite – Middle Jurassic (Bradford Clay).
20. Upper Oolite – Middle Jurassic, Bathonian (Great Oolite Limestone).
21. Fuller's Earth and Rock – Middle Jurassic, Bathonian.
22. Under Oolite – Middle Jurassic, Bajocian (Inferior Oolite).
23. Sand – Lower Jurassic, Upper Lias (Midford Sand).
24. Marlstone – Lower Jurassic, Middle Lias, Domerian.
25. Blue Marl – Lower Jurassic (Lower Lias Clay).
26. Blue Lias – Lower Jurassic (Lower Lias).
27. White Lias – Rhaetic (Lower Lias).
28. Red Marl – Triassic (Keuper Marl).
29. Redland Limestone – Permian (Magnesian Limestone).
30. Coal Measures – Pennsylvanian.
31. Mountain Limestone – Mississippian.
32. Red Rhab and Dunstone – Devonian (Old Red Sandstone).
33. Killas – Devonian and older (slates, grits).
34. Granite, Syenite, Gneiss.

It is clear, then, that in its essentials, Smith's stratigraphic order still stands to this day.

See Also

Economic Geology. **Famous Geologists:** Cuvier; Murchison; Sedgwick. **Geological Maps and Their Interpretation**. **History of Geology From 1780 To 1835**. **Palaeontology**. **Stratigraphical Principles**.

Further Reading

Cox LR (1942) New light on William Smith and his work. *Proceedings of the Yorkshire Geological Society* 25: 1 99.

Eyles JM (1969) William Smith (1769 1839): a chronology of significant dates in his life. *Proceedings of the Geological Society of London* 1657: 173 176.

Fuller JGCM (1995) *"Strata Smith" and his Stratigraphic Cross Sections, 1819: A Review of Facts Worth Knowing about the Origin of Stratigraphic Geology in the Mind of William Smith (1769 1839), an English Country Surveyor and Civil Engineer.* Tulsa: American Association of Petroleum Geologists; Bath: Geological Society Publishing House.

Knell SJ (2000) *The Culture of English Geology, 1815 1851*. Aldershot, Burlington, Singapore, and Sydney: Ashgate.

Phillips J (1844) *Memoirs of William Smith, LL.D. Author of the "Map of the Strata of England and Wales," by his Nephew and Pupil.* London: John Murray. Reprinted (1978). New York: Arno Press; and (2003) Bath: The Bath Royal Literary and Scientific Institution (with additional essays on Smith by HS Torrens).

Sheppard T (1917) William Smith: his maps and memoirs. *Proceedings of the Yorkshire Geological Society* 19: 75 253.

Smith W (1815) *A Delineation of the Strata of England and Wales, with Part of Scotland; Exhibiting the Collieries and Mines, the Marshes and Fen Lands Originally Overflowed by the Sea, and the Varieties of Soil According to the Variations in the Substrata, Illustrated by the Most Descriptive Names by W. Smith*. London: (see copy of original map at http://www.unh.edu/esci/greatmap.html).

Smith W (1815) *A Memoir to the Map and Delineation of the Strata of England and Wales, with Part of Scotland*. London: John Cary.

Smith W (1816) *Strata Identified by Organized Fossils, Containing Prints on Coloured Paper of the Most Characteristic Specimens in Each Stratum*. London: Printed by W Arding; sold by W Smith; J Sowerby; Sherwood, Neely and Jones; and Longman, Hurst, Rees, Orme and Brown. (see copy of original at http://www.unh.edu/esci/wmsmith.html).

Rudwick MJS (1996) Cuvier and Brongniart, William Smith, and the reconstruction of geohistory. *Earth Sciences History* 15: 25 36.

Torrens HS (2001) Timeless order: William Smith (1769 1839) and the search for raw materials 1800 1820. In: Lewis CLE and Knell SJ (eds.) *The Age of the Earth: From 4004 BC to AD 2002, The Geological Society, Special Publication No. 190,* pp. 61 83. London: The Geological Society.

Torrens HS (2002) *The Practice of British Geology, 1750 1850*. Aldershot and Burlington: Ashgate Variorum.

Winchester S (2001) *The Map that Changed the World*. London: Viking.

Steno

J M Hansen, Danish Research Agency, Copenhagen, Denmark

Introduction

Nicolaus Steno (Niels Stensen; Figure 1) was an anatomist, geologist, and bishop, often considered to be the founder of geology as a science. He was the first to describe the most fundamental principles of stratigraphy and crystallography, and the first to claim by rigorous arguments that fossils are the remains of former life on Earth. Steno's *principle of superposition* is simple but fundamental for all geologists and belongs to the first steps of understanding that a geology student must acquire. The principle states that geological strata were originally deposited horizontally under the action of gravity, and that the upper strata are younger than the lower are. The principle also states that inclined or folded strata have been tilted or disturbed subsequent to their deposition. The principle of superposition, theoretically, builds on Steno's statement that a crystal, sediment layer, or other kind of geological structure that takes the shape of the surface of another solid body is younger than the body from which it takes its shape. In conjunction with the principle of superposition, Steno's *principle of intersection* says that a body of rock or other geological structure is younger than those rocks or structures it is found cutting through.

Steno held forth yet another important stratigraphic principle, the *principle of reconstruction*. This states that it is possible to 'backstrip' a series of geological changes in reverse order, from the present to the past, having regard to the principles of superposition and intersection. In this way it is possible to obtain knowledge about still older situations. Thereafter, with this knowledge about past situations and their order of occurrence, the geological history of a locality, from the past to the present, can be reconstructed.

In mineralogy, Steno was the first to describe the principle of crystal growth, which leads to constant and specific angles between the sides of crystals of specific minerals (Steno's Law). In the philosophy of science and natural history, Steno founded the *principle of recognitive induction*, which made it possible to separate palaeontology and historical geology from theology.

Career, Science, and Beliefs

Steno travelled through large parts of Europe, visiting renowned scientists, academies, and universities, as if receiving scientific inspiration from the landscapes he saw. Before he was 30 years old, his anatomical studies had made him famous; in 1667, he was attached to the Medici Court in Florence, where Grand Duke Ferdinand II, impressed by Steno's anatomical and preliminary studies on fossils, made him a member of Accademia del Cimento ('Academy of Experiments'). During the next 2 years, Steno established the most important and permanent principles of what were to become the geological core disciplines: palaeontology, stratigraphy, and mineralogy.

Steno was born in Copenhagen of a Danish mother, Anne, and a Scanian father, Sten Pedersen. Niels was a fragile child and was brought up in the orthodox Protestantism of the Copenhagen of those days. Due to illness, he was kept indoors from his third to his sixth year. Isolated from other children, he listened to his parents and their friends' religious conversations

Figure 1 Steno (Danish: *Niels Stensen* or occasionally *Steensen* i.e., *Niels, the son of Sten*; Latin: *Nicolai Stenonis* or abbreviated *Steno*; French: *Nicolas Sténon*; Italian: *Niccolò Stenone*), as he most likely appeared *ca.*1667 69. Contemporaneous portrait, by the Dutch court painter Justus Sustermanns (1597 1681). The original is in the Uffizzi Gallery, Florence; a copy hangs in the Institute of Medical Anatomy, Copenhagen University. This version is from a poster made for a Steno exhibition at Tromsø University, Norway, in 1999. In other portraits of Steno can be seen his *sigilum*, a stylized asymmetrical heart from which a cross rises, a suggestion that scientific knowledge provides the highest praise to God.

and became familiar with mechanical and chemical crafts in his father's respected goldsmith shop. But shortly after Niels' recovery from illness, his father suddenly died. Because the boy was gifted, he was sent to Vor Frue Skole, a Lutheran academy, where he learned his fluent Latin from the enthusiastic Ole Borch. Borch also took Niels and the other pupils on botanical excursions around Copenhagen. Subsequently, Borch became one of Denmark's most distinguished intellectuals, holding professorships at the university in poetry, philology, chemistry, and botany. More than anyone else, Borch turned the young Steno towards becoming a scientist. He visited Steno in Holland and their friendship continued until the end of Steno's life.

In his eighteenth year, Steno enrolled at Copenhagen University, where he came under the influence of Thomas Bartholin, who, as head of the Faculty of Medicine, was famous as discoverer of the lymphatic vessels. At the time, Denmark and Sweden were at war. The city was besieged, the university was closed, and Steno assisted the students' defence of the barricades. Following the cease-fire in 1659, he managed to get out of the city by taking a ship to Rostock. Shortly afterwards, Steno turned up in Amsterdam, encountering a new world of scientific opportunities.

Steno soon moved to Leiden, where he continued to study the glands, the muscles, and the heart, and where he developed friendships with Jan Swammerdam and Baruch Spinoza. Steno went to Paris in 1665, where he presented his theories on the human brain and on muscles. According to a contemporary reviewer, Steno "turned upside down what is basic in medicine". Thus, besides Steno's geological and philosophical contributions, he is also famous for some important anatomical discoveries. The discovery of the duct from the parotid gland to the mouth is named *ductus stenonianus*, in recognition of just one of Steno's many anatomical contributions. Also important, but hardly recognized, was his description and understanding of the threefold division of the body fluids. In 1665, Steno gave the first modern description of the human brain, contradicting the interpretations of Galen, Willis, and Descartes. Moreover, the modern understanding of the anatomy and function of muscles and muscle fibres should be attributed to Steno.

After a year in Paris, Steno travelled to Italy, where his geological interest was to flourish as a member of the Medici court in Florence from 1667. On his way, he passed through southern France, where, at the University of Montpellier, he met the Englishmen Martin Lister and John Ray. Steno's geological achievements from his years in Tuscany, and his contribution to the principles of modern science, as they developed from Bacon, Galileo, and Descartes, have hardly received the reputation they deserve. Being a Dane (writing for the most part in a beautiful Latin, which was then in decline, rather than in the up-and-coming French and Italian, and only privately in English and German), he did not contribute to the national pride and fame of any large country or court. Moreover, his contemporary reputation was hindered by his criticisms of some of his most influential scientific contemporaries and by a superficial understanding of his religious conversion to Catholicism in the year (1669) that he wrote his most important geological work, *De Solido Intra Solidum Naturaliter Contento Dissertationis Prodromus* (translation: *The Prodromus to a Dissertation Concerning Solids Naturally Enclosed in Other Solids*). Eighteenth-century writers were puzzled by Steno's conversion from Protestantism to Catholicism and by his shift from an academic to a clerical career. His geological methods were, however, promptly applied in England, Germany, and Italy, but his name was rarely mentioned before Lyell, Humboldt, and Élie de Beaumont drew attention to his work in the 1830s.

In 1671, Steno was recalled to Copenhagen by Christian V, who, due to Steno's Catholic faith, could not employ him as a university professor. Instead, the king made him Royal Anatomist, a title invented for the purpose. After a long journey through large parts of south-eastern Europe, Steno finally arrived in Copenhagen. However, in 1675, he obtained the king's permission to leave, and shortly after he became a Catholic priest in Italy. Soon the Pope called him to Hannover and Steno was appointed 'Bishop of Titiopolis', a now-defunct city of the Eastern Roman empire in Asia Minor. In reality, Steno was called by the Pope to lead the Catholic mission in northern Germany, Denmark, and Norway.

In Hannover, Steno met the Duke's librarian, Leibniz, who, after having read Steno's geological work, *De Solido*, in 1669, was inspired to write his theory, *Protogea*, concerning the Earth's origin. Leibniz became an admirer of Steno's geology and used him as scientific mentor and 'censor'. However, Leibniz was unsuccessful in persuading Steno to recommence his geological work. Most likely, Steno told Leibniz that he had, while in Florence, expanded his geological 'prodromus' (*De Solido*) into a more comprehensive geological dissertation, and had given it to Holger Jacobaeus, who was later Professor of Geography at Copenhagen University. After Steno's death, Leibniz wrote to several scientists, attempting to find out what had happened to Steno's geological papers. However, the extended version of *De Solido* has never been found.

After the period in Hannover, Steno dealt with numerous theological matters. These writings are collected in his *Opera Theologica* and *Letters*, and Steno did not really write on science any more, except for a philosophical letter to Spinoza. In the letter, written a few days before he was made 'Bishop of Titiopolis', Steno criticized Spinoza for having adopted a materialistic 'religion'. Steno died in Schwerin when he was 48 years old, weakened through several years of religiously inspired poverty and self-inflicted fasting. He was beatified in 1988.

Philosophy of Fossils and Recognition

In 1667, at the request of Ferdinand II, Steno wrote a brief dissertation (*Canis carchariae dissectum caput*) on the similarity between the teeth of living sharks and so-called tongue-stones, or glossopetrae (*glossa*: tongue; *petrus*: stone), now interpreted as fossilized shark teeth. Through this work, Steno is considered to be the first scientist to have established a series of empirical and rigorous arguments in order to describe fossils and interpret them as the geologically preserved remains of former living organisms. In six 'conjectures' (*conjecturae*), Steno explained that solid 'bodies' resembling parts of marine animals are indeed the remains of the things they resemble, provided the resemblance is found on every scale and in every visible detail. Further, Steno stated that such things do not grow in the Earth but have been deposited there by natural processes in the past, and that fossils should not be understood as inexplicable imprints of God's finger, but as representatives of the things they resemble (Figure 2).

In *De Solido* (1669), Steno made a general statement of his six 'conjectures' on the origin on fossils, formulated 2 years earlier. Now Steno enunciated a general geological, as well as a basic philosophical, principle: the *principle of recognition*: "If a solid body resembles another solid body in all respects, not only in the state of its surface but also in the internal arrangement of the parts and particles, it will resemble it also in the method and place of production". Steno mentioned that the similarity may only be structural and textural, and not necessarily chemical. He noticed that some fossil shells have been petrified, or the material substituted by other 'smallest parts' (elements, minerals, and sedimentary particles), different from the materials of which they were originally formed. The shape and visible structures of the original body may still be preserved even though the original material has been substituted by sediment or some type of mineral other than that of the original crystal, bone, or shell. Giving examples and descriptions, Steno further declared that the 'principle of recognition' is valid not only for fossils but also for geological strata, crystals, and any other solid body embedded in the earth. Nature's laws are 'univocal', thus similar conditions produce similar products.

A scientist should believe in direct observations and in reasoning derived from the observations, even if such reasoning implies a historical development of Earth and dramatic changes in the distribution of land and sea through time. A scientist should not trust speculations when they are contradicted by observations. The human ability to recognize things is inherent and makes possible the basic method of empirical science. In 1673, Steno further explained his conception of the senses and human reasoning: "It is not the function of the senses to display things as they are or to judge them, but to transmit to the reason those conditions of the things to be examined, which are sufficient for acquiring a knowledge of things appropriate to man's purpose". In other words, Steno realized that recognition is the most fundamental cognitive capacity in humans. It is prior to recognitive induction, or generalization, in science. Recognition is a prerequisite for generalization. *A priori* skills are 'above' *a posteriori* skills. Thus, Steno's principle of recognition is not only a geological principle, but also a general theory about cause and effect and regularity in nature. Furthermore,

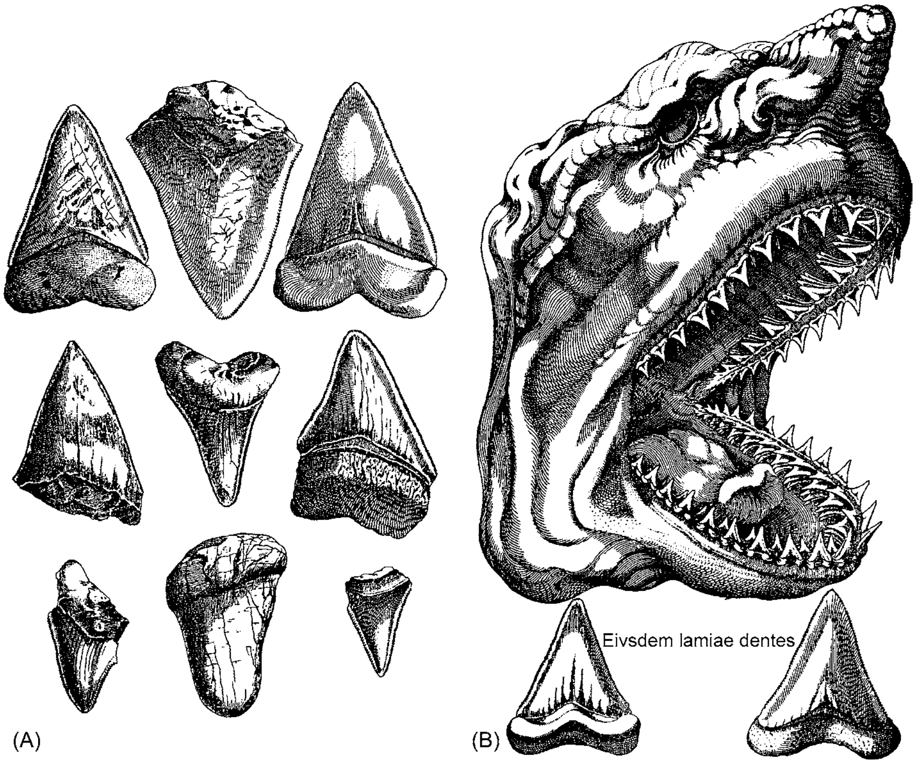


Figure 2 Steno's personal drawings of fossil shark teeth (A), compared with a contemporaneous artist's imaginative drawing of a monster shark (B) caught near Livorno and dissected by Steno. Note how the artist has been instructed to draw the anatomically correct arrangement of the teeth, which, when they are worn off, are substituted by new teeth rolling forward (another discovery made by Steno).

his principle was probably history's first theory of perception that was built on anatomical arguments relating to the capacity of the sensing organs and the brain to obtain relevant, albeit partial, knowledge from the evidence provided by nature.

Philosophy of Stratigraphy and Reconstruction

During his travels through the mountainous regions of southern Europe and during his numerous excursions in Tuscany, Steno noticed a large number of geological structures. In Steno's time, these structures had found no scientific explanation. At the time, most, if not all, scientists believed that landscapes and structures in Earth's crust derived from the time of Creation or had been formed during the turmoil of the Flood. However, through his studies of fossils, Steno became convinced that Earth had a history, accessible to scientific and human understanding over and above (though not generally conflicting with) the explanations in the Scripture. His interest was directed to all scales, not simply the small-scale structures of minerals and fossils. The structure of rocks, strata, and formations and their similarity on both sides of gorges and valleys became another philosophical and perceptual problem that Steno wanted to solve. Similarly, he wanted to find methods for exploring the history of the large-scale structures of mountain ranges.

Through his work on fossils from 1667, and 2 years later in *De Solido*, Steno dared to formulate and apply the core of his philosophy of science. This led to his definition of the fundamental stratigraphic principles of superposition and intersection, on how to find chronological and causal clues in geological bodies, in order to reconstruct their history. Moreover, it led to understanding of the general principles of crystal growth (Steno's Law). In consequence, *De Solido* is generally considered as the first scientific work on geology.

The basis of Steno's geological methods was a combined actualistic and 'forensic' procedure, proclaimed in the introduction of *De Solido*: "Given a substance endowed with a certain shape, and produced according to the laws of nature, to find in the body itself clues disclosing the place and manner of its production". Steno proclaimed that geological structures should be read according to the assumption that the present laws of nature were also in operation in the past.

Moreover, the natural structures of solid rocks and strata provided clues that could lead to 'demonstrative certain' understandings of how and in which environment (place) any given geological structure has been produced.

De Solido is structured in five parts: (1) the aforementioned introduction to the Grand Duke, (2) a philosophical part, in which the fundamental principles and methods are explained and formulated in general terms, (3) an empirical part, with numerous examples on how to apply the fundamental principles, (4) a description and interpretation of Tuscany's geological history, based on Steno's fundamental principles, and (5) acknowledgements from Steno's 'peer reviewers', the eminent scientists Vincenzio Vivianni and Francesco Redi. In the second (philosophical) part of *De Solido*, Steno summarized his geological understanding in three general 'propositions' about the way nature works and how it should be 'read'. The proposition on fossils and recognition was derived from Steno's previous work on fossils. A new proposition described the basis of stratigraphy and reconstruction: "If a solid body is enclosed on all sides by another body, the first of the two to harden [to attain a certain form] was that one which, when both touch, transferred its own surface characteristics to the surface of the other". From this general 'proposition' Steno developed a series of chronological principles, including the principles of superposition, intersection, and reconstruction. He gave a long series of examples from sediments, volcanic rocks, veins, crystals, fossils, etc., on how his principles works in practice. He realized that the principles would make it possible to reconstruct a scientifically plausible description of the historic development of Earth. He emphasized that the reconstructive method could show the succession and type of geological events, though the duration of the geological history was still unknown: "On this issue Nature is silent, only Scripture speaks".

By the help of his general 'propositions' and the associated superposition principle in Part 4 of *De Solido*, Steno showed how the geological history of Tuscany could be separated into six stages. There were two stages when the region was flooded by water and when its geological strata were deposited, two stages when it was flat and dry land formed by crustal uplifting of the strata previously deposited in water, and two stages when it was an uneven mountainous landscape eroded by rivers and deformed so that previously horizontal strata had been tilted, and again covered by younger horizontal strata.

In a cartoon-like series of didactic drawings (see Figure 3A), Steno showed how to reconstruct a region's geological history. The reconstruction must begin with the present state of affairs. Then, by the help of the superposition principle, it must be discovered what the situation was immediately prior to the present. When that is known, the situation immediately before this second-last situation must be discovered, and so on, with the third-last, fourth-last, etc., until it is impossible to identify any older situation. Then, when the different situations and their order of occurrence are known, the geological history can be reconstructed, beginning with the oldest known situation.

By good fortune, Steno's approval from the 'peer reviewers' had been easy to obtain, because he had worked in an area that could be interpreted relatively easily, and where there appeared to be no serious contradiction between Nature and Scripture. Tuscany had been flooded twice, first at the time of the Creation, before animals and plants lived on Earth (for which reason no fossils are found in the first sediments deposited by water), and again during the Flood and other marine transgressions (that is, after the creation of animals and plants, for which reason fossils are found in the sediments deposited during the Flood or later).

Philosophy of Crystals and Growth

Steno's third proposition in *De Solido* deals with the nature of growth. This included all kinds of natural growth, no matter whether it takes place in the inorganic or in the organic realm. Steno was inspired by Kepler's mathematical study of dense packing of 'atoms' and how snow crystals become a certain shape. From his own studies of sediments and crystals, however, Steno realized that crystal growth will give rise to regular external forms that cannot be produced by sedimentary processes, but also to structures more complex than those that can be produced by the packing of identical 'atoms' (see Figure 3B). Steno insisted that growth must be understood as a general problem, not only for crystals. He concluded that "if a solid body was produced according to the laws of nature, it was produced from a fluid." On reading *De Solido* and Steno's earlier anatomical papers, it is easy to understand how he reached this perceptive view. Superficially, the third proposition may seem odd. However, it reflects deep insight into change: all changes are results of motion. Motion is expressed in three basic forms: (1) as when we make a journey or an animal is running (i.e., change of location), (2) as when water runs in a river (i.e., flow), and (3) as "the first and hitherto unknown cause of motion", which Steno had already (in *De Thermis*) described as heat, and now also described as the motion of matter's smallest parts (i.e., diffusion). Thus, Steno envisaged three fundamental types of change: in modern terminology, this is change of location (or dislocation), flow, and diffusion.

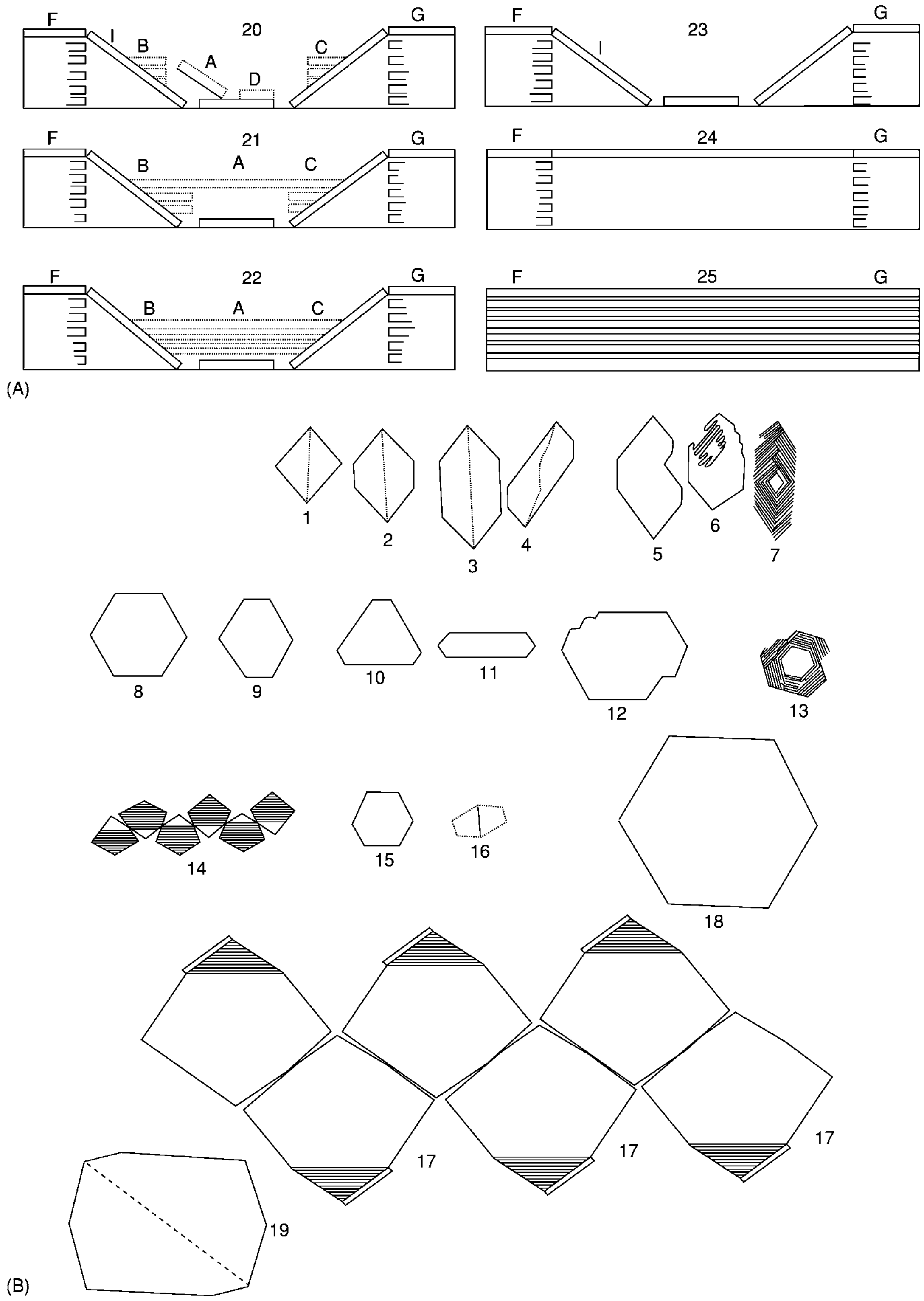


Figure 3 (A) Steno's model showing how to reconstruct the geological history of Tuscany. (B) Drawings of various crystal forms and indications of how crystals grow and dissolve.

All other kinds of change can be reduced to combinations of these three basic types!

Now, after this third proposition, Steno explained geological types of growth. Sediments increase with the help of gravity, which adds sand grains and other dispersed particles to the bottom from a slurry, or by traction of particles along the bottom until they hit obstacles. Crystals also grow by external addition, but for different reasons. The attractor is not gravity, but some unknown force, because crystals may grow even from the roof of a cavity or cavern. Steno explained geological growth as follows:

> Additions made directly to a solid from an external fluid sometimes fall to the bottom because of their own weight, as is the case with sediments; sometimes the additions are made from a penetrating fluid that directs material to the solid on all sides, as is the case of incrustations, or only to certain parts of the solid, as is the case of those bodies that show thread like forms, branches, and angular bodies.

There is no doubt that Steno held that all kinds of solid growth, whether inorganic or organic, sooner or later could be explained as being due to an external addition of dissolved or dispersed material to preexisting surfaces. The unsolved problem appeared not to be how growth takes place, but how crystals begin to crystallize and maintain a certain form as expressed by the constant angle between their crystal faces. The solution to these problems had to be found in the imperceptible smallest parts of the crystallizing matter, because neither the first 'germ' of a crystal nor its developing faces take shape from the substrate on which it grows. The growing crystal 'moulds' the substrate, but its faces are different from the latter. Kepler's paper on the dense packing of identical 'atoms' did not explain this, because it would imply that crystals could only be hexagonal, trigonal, or cubic, but not rhombic, and certainly not monoclinic or triclinic, as Steno knew some crystals to be.

Philosophy of Science and the Limits of Knowledge

The fundamentals of Steno's inductive/empirical philosophy of science were formulated in 1665 in Paris, when Steno gave history's first modern description of the human brain, including realistic drawings, completely different from those of his contemporaries Willis and Descartes. Steno opposed Descartes' conception of the brain and showed that Descartes' theory was built on pure speculation about God's impact on the human will, acting through the pineal gland. In Descartes' opinion, God controlled the soul and the human will by vibrating and rotating the pineal gland at the centre of the brain. Then, when the pineal gland, by such 'divine impacts', made the gland touch various parts of the brain, the contacts supposedly made the body react correspondingly, as if the brain were a mechanical machine and God its driver. On the contrary, said Steno, the brain is so fragile, and its structures so fine and complicated, that it "cannot even comprehend itself".

Prompted by his anatomical researches, matured through his founding of a scientific basis for the study of Earth, and made humble by his strong religious beliefs, Steno came to a clear and rigorous philosophy of science, close to that of modern scientists. During his geological studies of Tuscany and under the influence of what he had come to understand about Earth's history and changes, Steno developed his philosophy of science in order to separate geology and medicine from theology. When encountering something that is not understood, it is necessary to find something in it that is intelligible and compare it with something that is known and can be produced. The philosophical basis for *De Solido* can thus be summarized as follows: Specific recognitive induction (recognition) and specific empirical investigations (experiments, dissections, fieldwork, etc.) must regulate more general deductions and speculations (generalizations), but must not overrule deductive reasoning and perspectives, which for obvious reasons cannot be observed by the human senses or comprehended by the human mind. The past must be studied through knowledge about the present, but the past and present realities are much greater than scientific knowledge about it can ever be.

This should be understood to mean that it is not possible to observe the past *per se*, but only 'imprints' of past events. So, when seeking to interpret the past, primary emphasis must be placed on those clues that can actually be observed. This should be understood so that explanations about the inability to know anything directly are not neglected. However, such perspectives must, in contrast to Descartes' misuse of deduction, always respect what is known with the aid of the senses and by rigorous reasoning.

This way of thinking led to a general 'Kantian' (though pre-Kant) theory of human perception and interpretation of nature. In his Copenhagen lecture of 1673, Steno generalized his views on the difference between things 'as we see them' and things as they are 'in themselves'. There will always be a difference between nature as it is, and nature as humans interpret it. Modesty, caution, and scientific rigour should be key in attempts to understand things. Humans must believe in their immediate sensory capacities, and in what scientifically founded investigations reveal to the senses. Finally, it is important to believe in what the senses transmit to the reasoning. However, because human senses are imperfect and reasoning capacity is incomplete, the truth can only be approached, and complete understanding cannot be obtained.

Steno concentrated his philosophy of science in the following sentence, which has been cited more than anything else from his hand:

> *Pulchra sunt quae videntur,*
> *Pulchriora quae sciuntur,*
> *Longe pulcherrima quae ignorantur.*

In English, this reads as follows:

> Beautiful is what we see,
> More beautiful is what we know,
> Most beautiful is that about which we are insensible.

This has erroneously been interpreted to mean that Steno ranked religious belief above scientific knowledge. But he explained the aphorism in this way:

> Yes indeed, after having rejected all the errors of the senses, who would not repeat: beautiful is what appears to the senses without dissection; more beautiful what dissection draws forth from the hidden interior; yet far the most beautiful is what, escaping the senses, is revealed by reasoning helped by what the senses have already perceived.

In Steno's philosophy, humility about scientific principles and scientific understanding offered the highest praise to God. Scientific knowledge must never be ruled by clerical beliefs and powers. On the contrary, science will guide us towards the truth, which, however, will never be fully understood because of our limited sense capacity and imperfect intellectual resources.

See Also

History of Geology Up To 1780.

Further Reading

Cutler A (2003) *The Seashell on the Mountaintop. A Story of Science, Sainthood, and the Humble Genius who Discovered a New History of the Earth*. New York: EP Dutton.

Garboe A (1954) Nicolaus Steno (Niels Stensen) and Erasmus Bartholinus: two 17th century Danish scientists and the foundation of exact geology and crystallography. *Bulletin of the Geological Survey of Denmark. 4th Series* 3: 1 12.

Garboe A (1960) Niels Stensen's (Steno's) lost geological manuscript. *Bulletin of the Geological Survey of Denmark* 14: 243 246.

Gould SJ (1981) The titular bishop of Titiopolis. *Natural History* 90: 20 24.

Kardel T (1994) Steno: life, science, philosophy (with Niels Stensen's *Prooemium* or preface to a demonstration in the Copenhagen Anatomical Theater in the year 1673, and Holger Jacobaeus: Niels Stensen's Anatomical demonstration No. XVI, and other texts translated from Latin). *Acta Historica Scientiarum Naturalium et Medicinalium* 42: 1 159.

Moe H (1988) *Nicolaus Steno: An Illustrated Biography.* Copenhagen: Rhodos.

Noe Nygaard A (1986) Nicolaus Steno, paleontologist, geologist, crystallographer. In: Poulsen JE and Snorrason E (eds.) *Nicolaus Steno 1638 1686. A Re consideration by Danish Scientists,* pp. 167 190. Copenhagen: Nordisk Insulinlaboratorium.

Rodolico F (1971) Niels Stensen, founder of the geology of Tuscany. *Acta Historica Scientiarum Naturalium et Medicinalium* 23: 237 243.

Rudwick MJS (1972) *The Meaning of Fossils*. New York: MacDonald, Elsevier.

Scherz G (ed.) (1969) *Steno: Geological Papers* (translated by AJ Pollock). Odense: Odense University Press.

Steno N (1669) *De Solido Intra Solidum Naturaliter Contento Dissertationis Prodromus*. Florence: Ex Typographia Sub Signo Stellae (English translation in Scherz, 1969).

Suess

B Fritscher, Munich University, Munich, Germany

Introduction

Eduard Suess (Figure 1) was the most influential European geologist of late nineteenth and early twentieth centuries. As a professor of geology, he taught at the University of Vienna for nearly 45 years. In his major works, *Die Entstehung der Alpen* (*The Origin of the Alps*; 1875), and in the comprehensive *Das Antlitz der Erde* (*The Face of the Earth*; 1883–1909), he elaborated a 'global tectonics', based on the contracting hypothesis. By his works he created a new image of the structure and the formation of fold mountains and introduced basic terms of twentieth-century structural geology, such as the Laurentian and Angara Shields, the huge southern continent Gondwana, and the Tethys (as a former central sea, the precursor of the Mediterranean). Moreover, Suess was one of the pioneers of the doctrine of nappe folding in the Alps and he

Figure 1 Eduard Suess (1831 1914), in a portrait published in his posthumous 1916 memoir, *Erinnerungen*.

founded the concept of 'eustatic' sea-level changes. Suess is also remembered as an engineer and politician. He was a long-serving member of the Austrian national parliament and promoted and planned a new water supply for his home town, thus becoming one of the creators of modern Vienna.

Scientist, Engineer, and Politician

The founder of 'global tectonics', Suess was born into a well-connected bourgeois family. His father, a wool merchant, born in Germany, had worked and travelled in various European countries before setting up a wool business in London in 1828. Here, Eduard Suess was born on 20 August 1831. Three years later, the family moved to Prague, where Suess, as he later recalled, arrived as a 'complete English child' who understood not a single German word. In 1845, his father took over a leather factory near Vienna.

The young Eduard got a polyglot education from English, Belgian, and German tutors. In 1847, he entered the Polytechnic 'High School' (now Technical University) at Vienna, but soon left the revolutionary town of 1848, where he had participated in demonstrations, for the University of Prague. His early subjects were higher mathematics, physics, and descriptive geometry. The impressive collection of Silurian fossils at the museum in Prague roused his interest in geology, and he started to make excursions to nearby fossil-rich areas. Back at the Vienna Polytechnic in 1849, he continued his palaeontological studies in the surroundings of Vienna. In 1850, he presented a scientific paper (published in 1851) on Bohemian graptolites to Wilhelm Haidinger (1795–1871), who was then director of the newly established Austrian Geological Survey. His very first publication, however, had been a chapter on geology for a tourist guide of the Carlsbad region; Suess had written the chapter in 1850 during a visit to the region for a 'cure'.

After returning to Vienna, Suess was imprisoned for his participation in the demonstrations in 1848. Although he was released just a few weeks later, he had to leave the Polytechnic School, thus never acquiring a doctorate or any other formal university qualification. Nevertheless, in 1852, he was appointed an assistant in the Imperial Mineralogical Collection in Vienna. Assigned to classify the brachiopods of the collection, he published some pioneering studies in this field. His efforts to become a Privatdozent (private lecturer) at the university failed for his lack of a doctorate. By his palaeontological work, however, and by early travels accompanying well-known geologists such as Franz von Hauer (1822–99), Arnold Escher von der Linth (1807–72), Paul Deshayes (1797–1875), and Ernst Beyrich (1815–96), Suess already had a name among earth scientists. Thus, in 1857, on the recommendation of leading Vienna geologists, he was appointed 'professor extraordinary' of palaeontology at the university, and 5 years later was appointed the same position in geology. In 1867, Suess was appointed to a full professorship in geology, which he retained until 1901. For about three decades, Suess also travelled extensively throughout Europe.

An early engagement in school and university education marked the beginnings of Suess' political career. In 1862, he published an essay on the soils and the water supply of Vienna, showing that the numerous epidemics of that time, particularly typhoid, resulted from the city's water supply, which was at that time mainly based on wells. The following year, Suess was elected a member of the town council and was named head of a commission to study the water supply. He suggested that water should be brought by an aqueduct from mountains springs, about 70 km away; 10 years later, in 1873, the new pipeline began to operate, and the number of deaths from typhoid fever was subsequently reduced to one-fourth. Suess' second famous engineering project was the regulation of the Danube, designed to prevent the frequent flooding of the lower lying areas of Vienna. A canal was opened in 1875, and after 1876, there were no more major floods. Suess was also a member of the Diet of Lower Austria from 1869 to 1874, and he held a seat in the Austrian

Parliament from 1873 to 1896, being mainly engaged in implementing liberal reforms in the school system.

Suess was also an ordinary member of the Austrian Academy of Sciences from 1867, and served as its President between 1898 and 1911. The advancement of scientific cooperation between different scientific disciplines and different national Academies was one of his main goals. He also promoted the foundation of the Institute of Radium Research in Vienna in 1910. Suess was elected a member of various European scientific academies and the Geological Society of London awarded to him the Wollaston Medal in 1896. Suess died in Vienna on 26 April 1914 and was buried at Marz (Burgenland, Austria), where his family owned a small estate.

Palaeontology, Stratigraphy, and Eustatic Sea-Level Changes

Following his early studies of Silurian graptolites in Bohemia, Suess' main fields of study at the Imperial Mineralogical Collection were the palaeontology and stratigraphy of the Tertiary strata of the Vienna Basin. He entered scientific virgin territory by his brachiopod studies and was the first Austrian palaeontologist to be engaged in the classification of Tertiary mammals, thus providing an overview of the mammalian fauna of the Vienna Basin. Suess set the comparative method against the prevailing emphasis on merely taxonomic classification. Focusing on the anatomy of fossil species and comparing their modes of life with those of existing species, he created an early form of palaeobiology. As early as 1859, he discussed the ecology of brachiopods, and in 1875 he first used the term 'biosphere' (Biosphäre) to denote the distinct terrestrial sphere of the living organisms, which could be regarded as the surface of the lithosphere. These approaches were continued by his longstanding assistant Othenio Abel (1875–1946), who became one of the founders of modern palaeobiology.

Suess' stratigraphic work was mainly concerned with the Alps. It brought a new structural view to alpine stratigraphy, distinguishing chronological and spatial units. Suess cleared up stratigraphic problems of the European Rhaetian and Miocene, the latter in particular in the region of Eggenburg (lower Austria), north of Vienna. Basic studies related to the correlation of Alpine Triassic, Jurassic, and Cretaceous formations with their equivalents outside the Alps, and, together with Albert Oppel (1831–65), Suess correlated the development of the Alpine and Swabian Triassic strata.

The region of Eggenburg was the area where Suess first developed his concept of eustatic sea-level changes. In surveying the Tertiary beds, he remarked the regular height of the ancient shorelines of this area (Figure 2). These regularities seemed hardly explicable by an uplift of the land but, rather, by a fall of the sea-level. In 1885, after visiting Norway, he thought his ideas of the fall of the sea-level were confirmed by the stepped, horizontal terraces he had observed on the sides of the fjords and other valleys. Thus, 3 years later, he presented his theory of 'eustatic movements', i.e., of large-scale changes of the sea-level (separate from orogenic belts), which could be observed at approximately the same height over large parts of the earth.

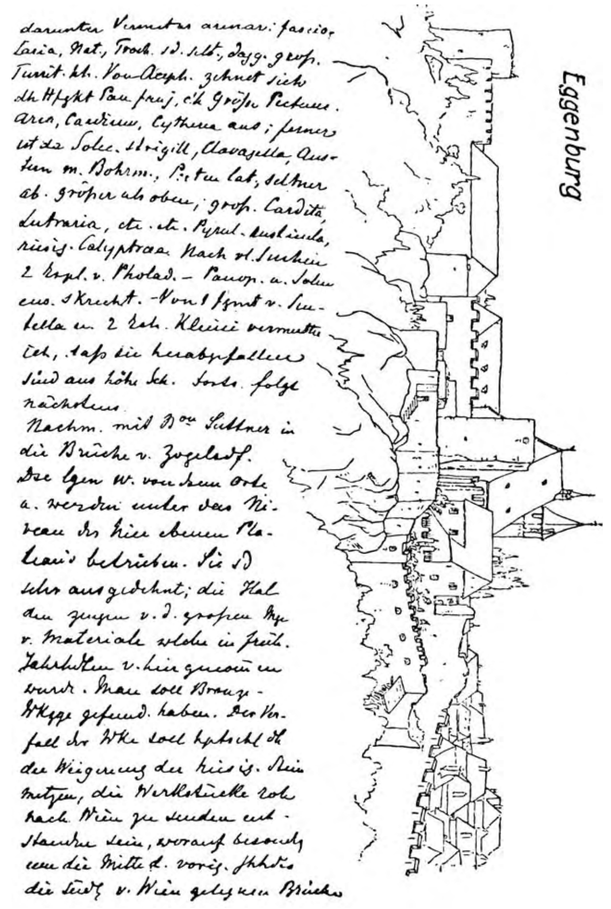
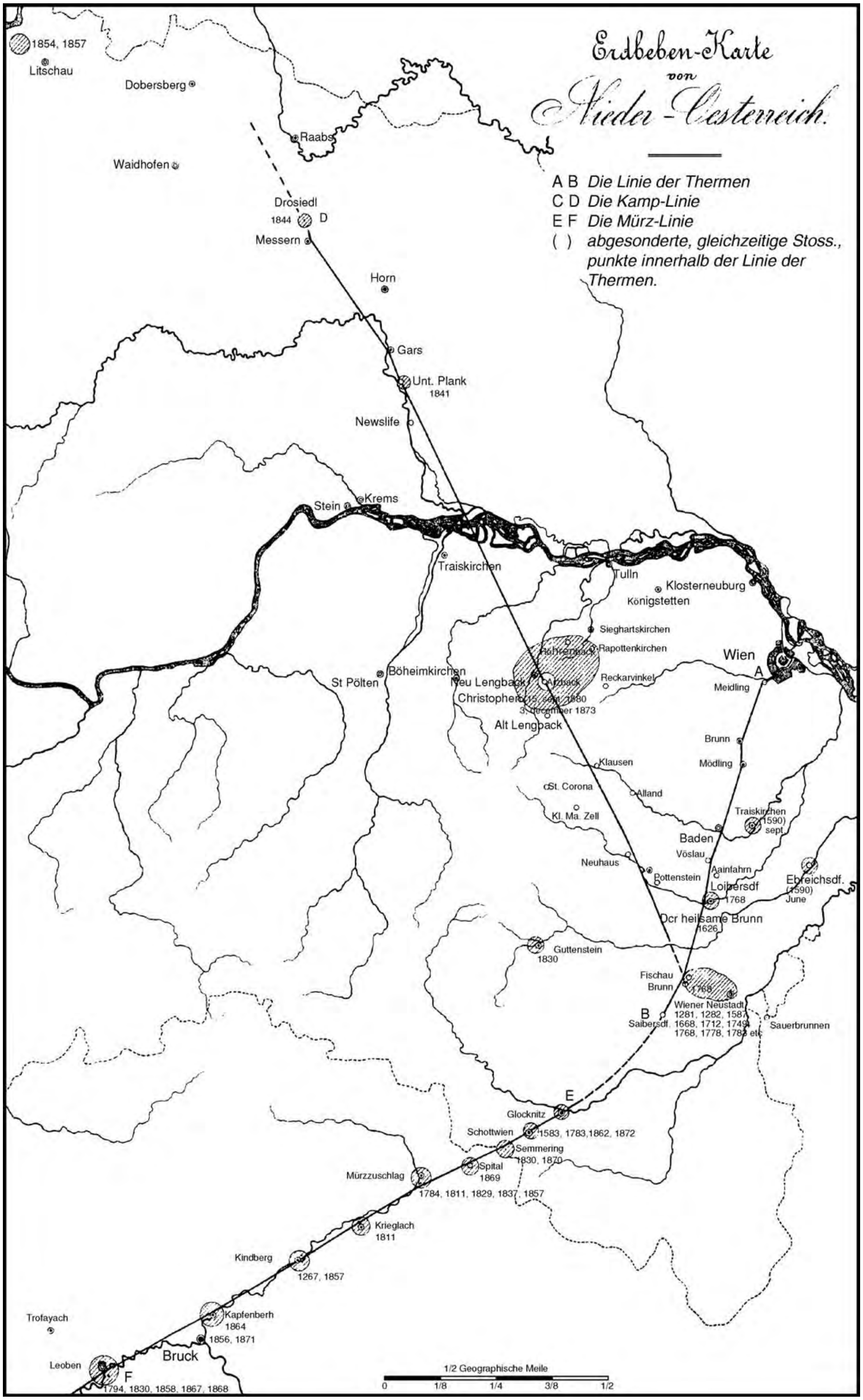


Figure 2 Sketch of Eggenburg (lower Austria) by Eduard Suess, from Suess' geological diaries, as published in *Erinnerungen* in 1916. The region of Eggenburg was crucial for Suess' ideas on large scale variations of sea level. In surveying the regularity of the ancient shorelines of this area, Suess first thought of what were later called 'eustatic' movements of the sea.

The Origin of the Alps

In 1865, Suess was commissioned to produce a survey of the geology of the Austrian Empire, comprising at that time Hungary, Czechoslovakia, and parts of Romania and Poland. In the following


Erdbeben-Karte
von
Nieder-Oesterreich.
A B Die Linie der Thermen
C D Die Kamp-Linie
E F Die Mürz-Linie
( ) abgesonderte, gleichzeitige Stoss., punkte innerhalb der Linie der Thermen.
1854, 1857
Litschau
Dobersberg
Raabs
Waidhofen
Drosiedl
1844
D
Messern
Horn
Gars
Unt. Plank
1841
Newslife
Stein
Krems
Traiskirchen
Tulln
Klosterneuburg
Königstetten
Sieghartskirchen
Rapottenkirchen
Wien
A
Röhrenback
Böheimkirchen
St Pölten
Neu Lengback
Anzback
Christopher
15, sept. 1880
3, december 1873
Alt Lengback
Reckarvinkel
Meidling
Brunn
Mödling
Klausen
St. Corona
Alland
Kl. Ma. Zell
Traiskirchen
(1590)
sept.
Baden
Vöslau
Neuhaus
Pottenstein
Aainfahrn
Loibersdf
1768
Ebreichsdf.
(1590)
June
Dcr heilsame Brunn
1626
Guttenstein
1830
Fischau
Brunn
1768
Wiener Neustadt
1281, 1282, 1587
1668, 1712, 1749
1768, 1778, 1783 etc
B
Saibersdf.
Sauerbrunnen
E
Glocknitz
Schottwien
1583, 1783, 1862, 1872
Semmering
1830, 1870
Spital
1869
Mürzzuschlag
1784, 1811, 1829, 1837, 1857
Krieglach
1811
Kindberg
1267, 1857
Trofayach
Kapfenberh
1864
1856, 1871
Bruck
Leoben
F
1794, 1830, 1858, 1867, 1868
1/2 Geographische Meile
0
1/8
1/4
3/8
1/2

years, he extended his field studies to the Carpathians, the Sudetes, and the Apennines. In applying the comparative method, which he had already used in his palaeontological work, he found that all these mountain ranges (including the Alps) had certain structures in common. As a result, in 1875, he published his first major book, *The Origin of the Alps*, a comprehensive discussion of the origin and the structure of mountain chains, anticipating most of his later ideas on tectonics.

Suess's actual entry to these ideas involved studies on earthquakes, following a visit to Calabria in 1871, where Suess witnessed the devastation of the great earthquake of 1870. This gave rise to the idea of a systematic compilation of historical accounts of earth tremors in lower Austria, i.e., in a region far away from any volcanic activity. Suess found that these Austrian earthquakes were distributed along specific lines that cut across quite different rock formations. Thus, in 1873, he published two major articles on earthquakes in lower Austria and in southern Italy, concluding that earthquakes are restricted to specific structures within Earth's crust and are thus due to the same forces as those that gave rise to the formation of mountains (Figure 3).

The revolutionary concept that Suess now set forth in *The Origin of the Alps* abandoned the idea of similarity of the structure of mountain chains, which had dominated geology for nearly a century. Contrary to the theory of mountain formation by vertical upheavals due to eruptive rocks, favoured by Leopold von Buch (1774–1853) and Leonce Élie de Beaumont (1798–1874), Suess set forth his view of horizontal movements as the essential cause of the formation of folded mountain chains, entertaining the idea of unilateral horizontal overthrustings by tangential pressures, in the case of the Alps, directed from south to north. And Suess stated a fundamental difference between the mountain chains and their older, rigid 'forelands' (Vorländer), which act, so to speak, as 'earth dams' against the mobile chains. As a further characteristic feature of developing mountain chains, he put forward the idea of 'hinterlands', i.e., their usually curved and relatively depressed 'inner' sides (Figure 4). These inner sides were the location of volcanism and earthquakes, due to the tensions caused by the movement of the newly forming mountains towards the forelands. Already during his visit to the volcanic areas of southern Italy in 1871, Suess had thought of volcanoes and earthquakes as being associated with regions of rupture within Earth's crust. Thus, earthquakes and volcanism were closely related to the process of mountain formation. Suess explicitly stated that volcanoes and intrusions were just the side effects of mountain formation.

Figure 4 Sketch of the main lines of strike for the folds of the Carpathians and the Balkans, from the first volume (1885) of Suess' *Das Antlitz der Erde*. The Carpathian Mountains, in particular, initiated Suess' thoughts about groups of curved lines, and tangential and unilateral movements, as opposed to the prevailing assumption of a symmetrical structure of folded mountain chains. The actual structure of the mountain chains (i.e., the disposition of the folds) depended on the form of the foreland and the resistance of its materials.

A Contracting Earth

In *The Origin of the Alps*, Suess also first introduced his ideas on the forces that could have brought about mountain ranges and their accompanying phenomena: namely, the contraction or shrinking of the earth due to its continuous cooling. In the 1840s, the American geologist James Dwight Dana (1813–1895) had proposed the idea of unequal radial contraction of the earth due to cooling, linking this to the origin of earthquakes. In the early 1870s, Dana and other geologists extended the contraction hypothesis as the basic assumption of the formation of mountains. Thus, Suess was neither the creator nor the chief advocate of the contraction hypothesis. However, it was Suess who – referring to Dana, to Robert Mallet's (1810–81) ideas on 'volcanic energy', and to Charles Darwin's (1809–82) theory of the origin of earthquakes due to the formation of fissures – most clearly used the contraction hypothesis for more than

Figure 3 A map from Suess' 1873 book *Die Erdbeben Nieder Österreichs*, showing the distribution of earthquakes in lower Austria. According to the position of the strongest effect of the respective earthquakes (years given near the names of the villages), Suess constructed three 'earthquake lines' (Erdbebenlinien). The Thermen Linie (A B) was named according to the thermal and sulphur springs along this line (for instance, at Brunn, Voslau, and Meidling). The Kamp Linie (C D) and the Murz Linie (E F) were both named according to rivers of the area.

30 years as a basis for explaining all the tectonic features of Earth's crust.

Arguing against the theory of mountain formation by volcanic elevations, Suess pointed to his studies in the Alps, which had offered only a single example that might confirm this theory: in the Euganean Hills (Italy), where a mass of Jurassic and Cretaceous limestones seemed to have been moved by trachyte. In the final chapter of *The Origin of the Alps*, Suess related the structures of mountain ranges to irregular earth contraction, and first used his famous phrase of 'the face of the earth'. He emphasized that, notwithstanding the irregularity of the contraction, its direction seemed to have remained the same over large regions and extremely long periods of time. In this, Suess was anticipating his later distinction of Caledonian, Armorican, and Variscan folding.

Global View: The Face of the Earth

In 1885, Suess published the first volume of his best known work, *The Face of the Earth*. The second volume followed in 1888, and the two parts of Volume 3 appeared in 1901 and 1909, respectively. An English translation in five volumes was published between 1904 and 1924. The most important edition became the French translation under the direction of Emmanuel de Margerie (1862–1953), in which thousands of new footnotes and about 500 figures were added. The monumental work was not simply an extension and more detailed discussion of Suess' previously formulated ideas. Rather, *The Face of the Earth* provided a 'global tectonics'. Embodying the results of his own travels and research all over Europe, Suess compiled the then-known materials relating to the tectonic structures of Earth's crust and created the 'language' that made possible a global view of the planet's tectonic features (Figure 5).

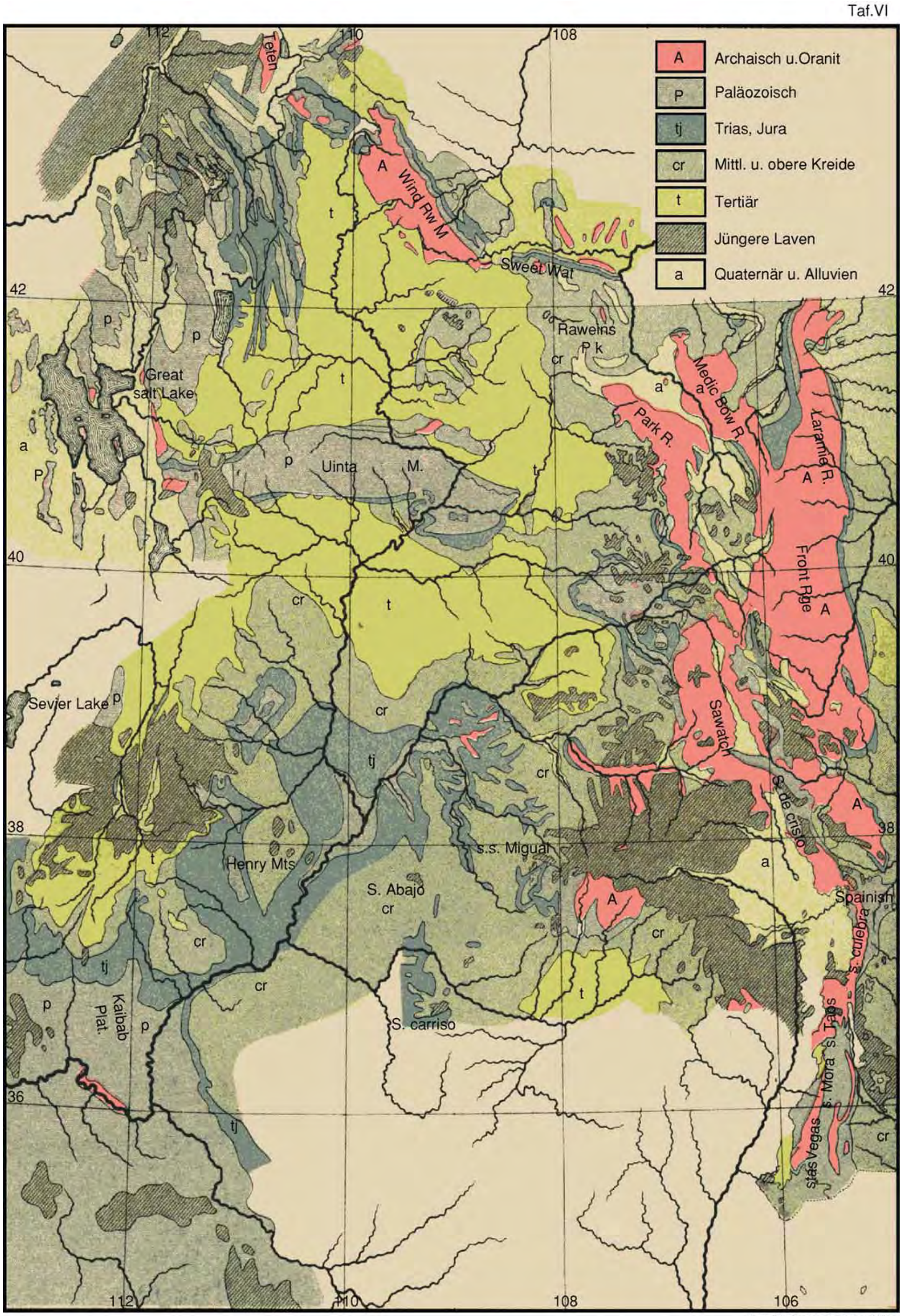
At the commencement of the tectonic development of the present landmasses, Suess distinguished four ancient continents: Laurentia, Angaraland (in what is now northern/central Siberia), Gondwana, and Antarctica (Figure 6). A further ancient block was the Baltic Shield. The later term for this, Fennoscandia, however, was actually not used by Suess. Between the Eurasian and the Indo-African blocks extended a series of younger mountain chains. Suess thought of them as having originated in the Tethys, an ancient Mediterranean sea encompassing half the globe, from Central America to the Sunda Islands (the name was coined according to the Greek goddess of the sea). A second series of young mountain chains, forming festoons and garlands, encircled the Pacific Ocean.

In addition to the youngest epoch of folding, the Alpine Orogeny, Suess distinguished two more principal stages of mountain formation in Europe: the Caledonian and the Variscan (also known as the Armorican and the Hercynian) orogenies. Their ages increased in passing from south to north. Though the most recent (Alpine) chains around the Mediterranean were already subsiding, the older ones, now eroded and covered by younger deposits, are presently at rest, but could be reactivated. According to Suess' theory of mountain formation, the subsidence and the formation of large grabens were predominant characteristics of Earth's crust, as against horizontal dislocations, such as folds and overthrusts. With reference to Suess' hypothesized periods of mountain formation, in 1887 the French geologist Marcel Bertrand (1847–1907) developed a tectonic classification of Earth's history, also correlating the different periods of folding with those of increased igneous intrusions. Thus, Suess' periods of folding became a continuous process that had built up the European continent from south to north.

In addition to these basics of a 'global tectonics', Suess introduced a series of further large-scale features of Earth's surface, such as his distinction of different types of coasts. According to the direction of the axes of fold mountains, he distinguished a 'Pacific' and an 'Atlantic' type. Around the Pacific Ocean, the fold lines were running more or less parallel to the coast, whereas in the case of the Atlantic Ocean, they ran approximately at right angles to the coast.

In the first volume of *The Face of the Earth*, Suess also further elaborated his ideas on contraction. Meanwhile, the hypothesis had received strong support by the Swiss geologist Albert Heim (1849–1937), who, in 1878, had calculated the supposed reduction of the circumference of Earth due to its shrinking or contraction. For the Jura mountains and the Alps, Heim estimated a relative compression of about four-fifths and one-half, respectively (compared to the original width, which was estimated by mentally 'smoothing out' both chains). Relative to the full circumference of Earth, the shrinking of Earth due to the folding would be almost 1%. Such a process (for which Heim thought a cooling of 200°C would be sufficient) going on throughout the whole history of Earth was quite plausible according to the contraction hypothesis.

Relying on his ideas on earthquake lines, and on further detailed discussions of the various phenomena of Earth tremors, Suess constructed a system of dislocations in rock formations, due to the reduction of the volume of the globe. The tensions produced by the process of contraction would tend to differentiate into tangential and radial tensions, thus producing both horizontal (i.e., pushing and folding) and



Kart. lith Anst.v.G.Freytag,Wien.

Figure 5 Map showing divergent branching (Virgation) of the Rocky Mountains, from the first volume (1885) of Suess' *Das Antlitz der Erde*. To convey the complexity of the structures of folded mountain chains, Suess often used the term 'Virgation', meaning the rodlike spreading out, or dispersion, of the individual branches towards their forelands. Such an order also meant that a region that was the 'hinterland' for one branch served as the 'foreland' for another. Translation of the key: A, Archaean rocks and granite; p, Palaeozoic; tj, Triassic and Jurassic; cr, Cretaceous; t, Tertiary; cross hatching, younger lava; a, Quaternary and alluvium.

Figure 6 Section of Suess' map (from the third volume, part 2, of *Das Antlitz der Erde*), showing the arrangement of the tectonic units of Earth. The rose coloured areas are Laurentia and Gondwana; brown, Caledonides and Saharides; green, Asian structures with Angaraland; pale violet, Oceaniden, Australia, and Antarctica; yellow, Cape Mountains; cross hatched, volcanic islands of Atlantic type.

vertical (or subsiding) movements. Thus, Suess distinguished two groups of dislocations, one produced by more or less horizontal movements of mountains, the other one by more or less vertical movements (i.e., by subsidence). At the end of the first volume of *The Face of the Earth*, Suess gave his famous statement of the history (and the future) of Earth: "What we are witnessing is the collapse of the terrestrial globe". He also linked Earth's contraction to the development of life. Subsidences had made possible the accumulation of water in the deep oceans, and, at the same time, the emergence of the continents, which became the home of organisms that breathe with lungs.

This idea of the linkage of the development of life on Earth to its tectonic history was elaborated at the end of the last volume of *The Face of the Earth*. Suess claimed that his ancient shields (Laurentia, Angaraland, Gondwana, and Antarctica) were the essential regions for the development of life. These areas supposedly did not participate in folding and transgressions for a long time. Consequently, the development of life in these areas should show fewer disturbances than elsewhere, and therefore Suess called them 'asylums' (Asyle). From the asylums, the distribution of the vegetation should have started again after the great tectonic changes.

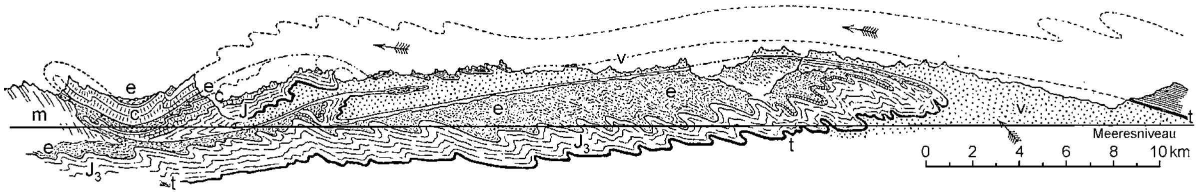


Figure 7 The Glarus double fold (according to a sketch by Albert Heim, as published in volume 3, part 2, of Suess' *Das Antlitz der Erde*), explained as a single overthrust towards the north. m, Tertiary Nagelfluh; e, Lower Tertiary flysch; c, Cretaceous; J, Jurassic; t, Helvetic Triassic; V, Verrucano. The sea level (Meeresniveau) is indicated.

Nappe Folding

The publication of *The Face of the Earth* covered a period of 24 years, and some of Suess' early concepts were changed during this time. The most striking example was the doctrine of nappe fold structures, as developed from the 1880s. Suess may have implicitly accepted large overthrusts as processes of mountain formation, at least since the 1870s. And he actually played a significant role in the development of the new doctrine, which is commonly ascribed to Marcel Bertrand. As early as 1883, Suess gave a new interpretation of Heim's 'Glarus double fold' (Glarner Doppelfalte) in the Glarus Canton of Switzerland (Figure 7). Whereas Heim had the idea of two folds from both north and south, Suess suggested a single overthrust toward the north. Nevertheless, it was not until publication of the third volume of *The Face of the Earth*, and in particular its second part, that Suess included nappe folding in his discussion of the processes of mountain formation. He discussed several of these structures, as found in the young European mountain ranges, accepting also large amounts of overthrusting, such as, for instance, in the case of the Dent Blanche massif in Switzerland, which had recently been proposed by the Swiss geologist Émile Argand (1879–1940).

Suess in the Twentieth Century

European geological thinking was deeply influenced by Suess' tectonics: a great number of his concepts and terms became standard in twentieth century earth sciences and are still current. On the other hand, Suess' basic theoretical assumption, i.e., the hypothesis of the contracting Earth, was subjected to critiques, even during the publication of *The Face of the Earth*. In 1912, only 3 years after the publication of the last volume and 2 years before Suess' death, Alfred Wegener (1880–1930) came up with quite new ideas on the origin of continents and oceans. Wegener was Suess' most serious critic. Moreover, the recognition of radiogenic heat, produced within Earth, threw grave doubt on the idea of a cooling and contracting planet. Nevertheless, it was the Viennese geologist who provided Wegener with essential information about the large-scale features of Earth, indicating former connections and movements. Thus, notwithstanding Suess' errors, his work remains among the most impressive and comprehensive scientific theories of Earth ever written.

See Also

Africa: Pan-African Orogeny; North African Phanerozoic; Rift Valley. **Andes**. **Antarctic**. **Argentina**. **Australia:** Proterozoic; Tasman Orogenic Belt. **Brazil**. **Famous Geologists:** Wegener. **History of Geology From 1835 To 1900**. **New Zealand**. **Oceania (Including Fiji, PNG and Solomons)**. **Plate Tectonics**. **Shields**. **Tectonics:** Earthquakes; Folding; Mountain Building and Orogeny. **Volcanoes**.

Further Reading

Cernajsek T, Csendes P, Mentschl C, and Seidl J (1999) "...hat durch bedeutende Leistungen...das Wohl der Gemeinde mächtig gefördert." *Eduard Suess und die Entwicklung Wiens zur modernen Großstadt*, Österreichisches Biographisches Lexikon Schriftenreihe 5. Vienna: Institut Österreichisches Biographisches Lexikon.

Greene MT (1982) *Geology in the Nineteenth Century. Changing Views of a Changing World*. Ithaca and London: Cornell University Press.

Hamann G (ed.) (1983) *Eduard Suess zum Gedenken (20.VIII.1831 26.IV.1914)*. Sitzungsberichte der Österreichischen Akademie der Wissenschaften, Philologisch historische Klasse 422. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.

Şengör AMC (1982) Eduard Suess' relations to the pre 1950 schools of thought in global tectonics. *Geologische Rundschau* 71: 381 420.

Şengör AMC (1998) Die Tethys: vor 100 Jahren und heute. *Mitteilungen der Österreichischen Geologischen Gesellschaft* 89: 5 177.

Şengör AMC (2003) *The Large Wave Deformations of the Lithosphere: Materials for a History of the Evolution of Thought from the Earliest Times to Plate Tectonics, Memoir 196*. Boulder: Geological Society of America.

Suess E (1862) *Der Boden der Stadt Wien nach seiner Bildungsweise, Beschaffenheit und seinen Beziehungen zum bürgerlichen Leben: Eine geologische Studie.* Vienna: W Braumüller.
Suess E (1873) *Die Erdbeben Nieder Österreichs.* Vienna: K Gerold's Sohn.
Suess E (1875) *Die Entstehung der Alpen.* Vienna: W Braumüller.
Suess E (1885 1909) *Das Antlitz der Erde* (three volumes; vol. 3 in two parts). Prague and Leipzig: F Tempsky and G Freytag.
Suess E (1897 1918) *La Face de la Terre* (three volumes; vol. 3 in four parts) (Traduit et annoté sous la direction de E de Margerie. Avec une préface par M Bertrand). Paris: Librairie Armand Colin.
Suess E (1904 1924) *The Face of the Earth* (five volumes). (Translated by HBC Sollas, under the direction of WJ Sollas.) Oxford: Clarendon Press.
Suess E (1916) *Erinnerungen.* Leipzig: S Hirzel.
Tollmann A (1981/1982) Die Bedeutung von Eduard Suess für die Deckenlehre. *Mitteilungen der Österreichischen Geologischen Gesellschaft,* (special volume to mark the 150th anniversary of E. Suess' birth) 74/75: 27 40.
Wegmann E (1976) Eduard Suess. In: Gillispie CC (ed.) *Dictionary of Scientific Biography* 13, pp. 143 149. New York: Charles Scribner.

Walther

I Seibold, University Library, Freiburg, Germany

Career

Johannes Walther (Figure 1) was one of the early pioneers in sedimentology, introducing a modern approach that combined both lithological and biological aspects.

Walther was born on 20 July 1860, the son of a vicar in Neustadt/Orla in the German province of Thuringia. In his boyhood he was handicapped by a nervous disease (probably caused by a serious fall) that prevented his regular attendance at school and as a result he did not receive a leaving certificate. But his intelligence and enthusiasm for science were so evident that he obtained special permission to study at the University of Jena and his health was restored during his period as a student. Walther was awarded his PhD in zoology in 1882, following which he went to Leipzig and Munich to study geology and palaeontology more intensively. At the same time, he started his first investigations of the sea floor in the Bay of Naples, where Anton Dohrn's marine biology station was based. In 1886, Walther became a lecturer in Jena; in the following years he undertook extended geological expeditions abroad. During his time as lecturer Walther was almost without income, for his position had no salary, apart from the small lecture fees. He therefore had to make a living by writing (e.g., newspaper articles) and giving public lectures. However, in 1894 he was appointed associate professor (Haeckel Professor) at Jena, a post endowed by a wealthy Swiss admirer of Walther's teacher, the famous zoologist Ernst Haeckel.

In 1899, Walther married Janna Hentschel. They had two children, a son and a daughter. From 1906 until his retirement, he was Director of the Geological Institute of Halle University. His work comprised more than 120 publications, including a dozen books, most of which appeared in several editions. Walther received honorary degrees from Perth and Melbourne (the latter 2 weeks after the outbreak of World War I!). In 1928, he was Visiting Professor at John Hopkins University, Baltimore. From 1924 to 1931, he served as President of the German Academy of Scientists, *Leopoldina*, in Halle. He died at Hofgastein, Austria, on 4 May 1937.

Figure 1 Johannes Walther.

Early in his career, Walther set himself the aim of 'modernizing' Lyell's uniformitarianism. Thus he undertook journeys to study as many geological conditions and environments as possible, but he focused on two fields in particular: shallow marine environments (including reefs) and deserts. At the time when he wrote his principal works, palaeontology, stratigraphy, and tectonics dominated geology in the universities, but Walther did not focus on these fields. Sedimentary rocks were chiefly studied because of their fossil content, useful for stratigraphy, and not because they illustrated former environments. Strongly influenced by Haeckel, the German apostle of Darwinism, Walther applied his ideas to geology by looking at the mode of formation of sediments, and the processes that formed rocks, not just their characteristics. His credo was: "Aus dem Sein erklären wir das Werden" (From the present state [of a rock] we explain its origin). But he was well aware that some past processes do not occur today. Though physical laws remain the same, geological conditions vary and may even be unique.

With this dynamic approach, Walther was able to integrate climatological, sedimentological, palaeontological, and other aspects into Amanz Gressly's 'facies' concept (1838), which was the basis for Walther's comprehensive consideration of facies. His view of the succession of palaeoenvironments was expressed in the 'Law of Facies', which was subsequently named after him (even though it had already been found by Gressly): "Es ist ein Grundsatz von weittragender Bedeutung, dass primaer sich nur solche Facies und Faciesbezirke geologisch überlagern können, die in der Gegenwart nebeneinander zu beobachten sind" ("It is a principle of far-reaching importance that only the facies or facies areas that are at present adjacent to one another can be geologically superimposed upon one another") (Walther [1894], p. 979; see also Middleton [1972]). In other words, the relative horizontal distribution of sediments with their organic content will be transformed into a vertical distribution, having a chronological order. Gressly found the rule during his extended fieldwork for his admirable monograph on the Swiss Jurassic near the town of Solothurn, where he carefully studied the facies changes. Fifty years later, Walther discovered Gressly's study of 1838 anew, when he was working for his volume on lithogenesis. He explained and discussed this idea in detail in the first of three chapters on facies in his most interesting third volume (*Lithogenesis of the Present*) of his fundamental work, *Einleitung in die Geologie als historische Wissenschaft* (1894). This discussion of the 'law' was only a minor part of the total corpus of his work, but it is on this that his present reputation chiefly rests.

Walther stressed the importance of organisms in geological processes and *vice versa*: biogeology. The dependence of biocenoses (groups of organisms living together, forming natural ecological units) on their substrates means that lithology should have priority over palaeontology. He mentioned, for example, that the empty shells of index fossils can be transported over long distances and may, therefore, give false stratigraphic results. With his zoological background of comparative anatomy Walther thus advocated 'comparative lithology'. This concept improved stratigraphy substantially and Amadeus W. Grabau (1870–1946) paid tribute to it by dedicating his classic *Principles of Stratigraphy* (1913) to Walther.

The first of the other two volumes of Walther's *Introduction to Geology as Historical Science* (see above), *Bionomie des Meeres*, was a treatise on marine biogeology that dealt with the interactions between the sea floor, fauna, and flora; beyond that it also had chapters on oceanography. (He had previously published a popular booklet on general oceanography in 1893.) The second volume was mostly a collection of faunal lists.

All Walther's pioneering work was undertaken in his years in Jena. One of his later publications was the voluminous *Allgemeine Paläontologie* (*General Palaeontology*) (1927). He wanted to give a summary of all the topics he had thought and taught about, and this was a lot. The book contained a wealth of varied ideas, though regrettably a number of them were already outdated by the time the book appeared. With its 809 pages, it presents difficult, yet still inspiring reading. For Walther, theories were much more important than details.

Special Contributions

Shallow Marine

Walther's final study on the sediments of Dove Bank (*Taubenbank*, 1910), a shoal in the Gulf of Naples, was a classic. He compared his first maps of 1884 with the latest findings of 1910 to evaluate the changes caused by volcanic activity (Vesuvius produced a great ash-fall in 1906) and the unusual storms of the intervening years (an early look at event stratigraphy!). His former studies of the rock-forming calcareous algae (1885) enabled him to compare his old and new results. He showed that coralline algae spread rapidly, consolidating sediments within 25 years.

Perhaps even more interesting was Walther's application of biological experiments to explain bioturbation, which he stated could occur down to 15 cm. He also measured the digging velocity of mussels. Using a 'hyperactualist' method, he explained the formation of calcareous sands by keeping crayfish and mussels together in an aquarium and finding that four crayfish of 12–18 cm length worked a load of 580 g of mussels into detritus within 12 days. The final weight was 240 g.

A 'fossil example' was provided by his study of the fauna of the Solnhofen sediments (Upper Jurassic, Bavaria, South Germany) (1904), a contribution for a volume in honour of Haeckel's 70th birthday. Walther tried to show all features of a biotope by determining the frequency of marine and land fossils at various localities in the Solnhofen sediments. He noted the lack of freshwater and brackish species, and localized the presence of plant debris; by the decreasing frequency of occurrence of the planktonic crinoid *Saccocoma*, from a centre outwards in all directions, he concluded that the greatest depth of water had been in this central area. Besides his own observations, Walther used data from all available collections and fossil lists for his demonstration of the Solnhofen biotope. Unfortunately, this exemplary study was almost hidden among the zoological articles in the Haeckel volume and in consequence was overlooked by many geologists.

Reefs

During his early studies (1885) in the Gulf of Naples, Walther became interested in the growth of calcareous algae. He compared his marine findings with what could be found in Tertiary sediments in Sicily and with Triassic alpine rocks, and concluded that lithification takes place concomitantly with deposition, leading to the formation of structureless limestones.

A further step towards greater understanding of reefs was achieved as a result of his voyages to the south coast of Sinai (1886) and to the Palk Strait (India) (1888–1889). Walther also emphasized that reefs are traps for considerable amounts of sediments, which form up to 60% of the whole reef complex. Further, he noted the importance of tectonic movements, sea level changes, and topography for the reef growth (elevations are preferred places), laterally or vertically.

Deserts

The journey of 1886 along the Sinai coast, with a return along the Egyptian side of the Red Sea through the Galala Desert offered Walther splendid opportunities for the study of desert environments, which he was subsequently able to extend in the USA (1891), central Asia (1897), Egypt (1911), and Australia (1914). He soon recognized the importance of aeolian erosion, which previously had not been much taken into account. Consequently, his first publication on deserts had the provocative title *Die Denudation in der Wüste und ihre geologische Bedeutung* (1892) (*Desert Denudation and its Geological Importance*) and provoked many controversies. In fact, he tended to exaggerate the notion of desert erosion and neglected the influence of periods with strong fluviatile erosion.

Walther was the first to describe ventifacts and *Dreikanter.* The form of sand grains gave him indications of their aeolian origin, an approach that was systematically utilized by André Cailleux in the 1930s. Walther dealt with the phenomena of fossil deserts, which he classified as tropical, coastal, plantless volcanic, rain-shadow, and glacial deserts. The fact that the different types could be associated with younger or older sediments from other climatic zones was explained by changes in the Earth's axis of rotation. With this idea, he was well ahead of his time. In a special publication, Walther described the fauna of a lake in the *Buntsandstein* desert in Thuringia (1904). From his study of the Nubian Sandstone in Egypt he became convinced of the desert character of the *Buntsandstein.* His ideas about the formation of (minor) salt deposits are still accepted, along with Ochsenius's bar theory. Walther was the first to describe laterite profiles in the deserts of Western Australia, interpreting them as weathering products of a former climate (1915).

A revised and enlarged version of his first desert book was published in 1900 and went through four editions until 1924. The last edition has recently been translated into English with commentary (1997). This book was the main reason for his being invited to Australia by the British Association for the Advancement of Science in 1914. In Germany, he was called '*Wüsten* (desert) Walther'.

National Education

The nineteenth century was a golden age for national science education and Walther's teacher, Haeckel, was indefatigable (and extremely successful) in this field. Walther himself wrote a flowing, somewhat poetic, prose, which was very apt for such work. His books for a broader public enthused more than one generation and were popular with both school teachers and their students. The *Vorschule für Geologie* (1905) (*Elementary Course in Geology*), for example, sold 22 000 copies and was translated into Russian and Czech. Its last Russian edition appeared as late as 1940. Another successful book was the *Geologie von Deutschland* (1910). Walther saw an important task in the training of schoolteachers in geology and held many vacation courses for them.

However, his extensive work in this field affected his academic reputation in later years.

Impact

This may be one of the reasons why Walther's ideas were not fully appreciated by his contemporary fellow geologists. Another reason could be that he tended to go to extremes when proposing and defending a new idea. Walther was still young when he published his pioneering works and was then of low academic status. Many of his colleagues, especially in Germany, did not recognize or appreciate the views he introduced. They did not fit into the current academic trends. The recognition of his books in Austria and Switzerland was greater than at home. But some of the outstanding German geologists of the time (Wilhelm von Gümbel, Karl von Zittel, Hermann Credner, Edmund von Mojsisovics) appreciated his revolutionary steps in the direction of biogeology, as opposed to palaeontology or stratigraphy.

Walther had particular influence in Russia. His desert book was translated in 1911 and lithology became one of the main fields of Russian geology. In Britain, his reputation was mostly based on the desert book and was soon forgotten after his death. In the USA, only a few geologists, such as William Henry Twenhofel, mentioned his books. Walther's somewhat 'baroque' style of writing caused language difficulties. Also the World War I may have hindered the spread of his works and ideas. Interest in them was revived during the second half of the twentieth century by the facies research in the oil industry. After some delay, Walther eventually became better known in the USA than in Europe.

See Also

Fossil Plants: Calcareous Algae. **History of Geology From 1900 To 1962**. **Sedimentary Environments:** Depositional Systems and Facies; Carbonate Shorelines and Shelves; Deltas; Deserts; Reefs ('Build-Ups'). **Sedimentary Processes:** Aeolian Processes. **Stratigraphical Principles**.

Further Reading

Ginsburg RN, Gischler E, and Schlager W (eds.) (1994) *Johannes Walther on Reefs*. English translation with commentary. Miami: University of Miami, Rosenstiel School of Marine and Atmospheric Science (*Geological Milestones II*).

Gischler E and Glennie KW (eds.) (1997) *The Law of Desert Formation: Present and Past*. English translation, with preface and introduction, of Johannes Walther (1924). Miami: University of Miami, Rosenstiel School of Marine and Atmospheric Science (*Geological Milestones IV*).

Gressly A (1838) *Observations Géologique sur le Jura Solenrois. Nouvelles Mémoires de la Société Helvétiques des Sciences Naturelles.* Volume 2. Neuchâtel.

Middleton GV (1972) Johannes Walther's law of the correlation of facies. *Bulletin of the Geological Society of America* 84: 979 988.

Seibold I (1992) *Der Weg zur Biogeologie: Johannes Walther 1860 1937.* Berlin, Heidelberg and New York: Springer.

Vissotzky WR (1965) *Johannes Walther and his Role in the Progress of Geology.* Moscow: Nauka (in Russian).

Walther J (1885) Die gesteinsbildenden Kalkalgen des Golfes von Neapel und die Entstehung structurloser Kalke. *Zeitschrift deutsch Geologische Gesellschaft* 37: 329 357.

Walther J (1888) Die Korallenriffe der Sinaihalbinsel: Geologische und biologische Beobachtungen. *Abhandlungen der mathematisch physikalischen Classe der königlisch Sächsischen Gesellschaft der Wissenschaften zu Leipzig* 14: 435 506.

Walther J (1891) Die Adamsbrücke und die Korallenriffe der Palkstrasse: Sedimentstudien im tropischen Litoralgebiet. *Petermanns Geographische Mitteilungen* 22: 40.

Walther J (1891) Die Denudation in der Wüste und ihre Geologische Bedeutung. *Abhandlungen der mathematisch physikalischen Classe der königlisch Sächsischen Gesellschaft der Wissenschaften zu Leipzig* 16: 345 570.

Walther J (1893) *Allgemeine Meereskunde*. Leipzig: Weber.

Walther J (1893 1894) *Einleitung in die Geologie als historische Wissenschaft.* 3 vols. Jena: Fischer. I. *Bionomie des Meeres: Beobachtungen über die marinen Lebensbezirke und Existenzbedingungen*: 1 196; II. *Die Lebensweise der Meeresthiere: Beobachtungen über das Leben der geologisch wichtigen Thiere*: 200 531; III. *Lithogenesis der Gegenwart: Beobachtungen über die Bildung der Gesteine an der heutigen Erdoberfläche*: 535 1055.

Walther J (1900) *Das Gesetz der Wüstenbildung in Gegenwart und Vorzeit,* 1st edn. Berlin: Reimer.

Walther J (1904) *Die Fauna der Solnhofener Plattenkalke. Festschrift. 70. Geburtstag von Ernst Haeckel,* pp. 133 214. Jena: Fischer.

Walther J (1904) Über die Fauna eines Binnensees in der Buntsandsteinwüste. *Zentralblatt für Mineralogie, Geologie und Paläontologie* un numbered volume (for 1904): 5 12.

Walther J (1905) *Vorschule der Geologie*. Jena: Fischer.

Walther J (1910) Die Sedimente der Taubenbank im Golfe von Neapel. *Abhhandlungen der königlich Preussischen Akademie der Wissenschaften, Physikalisch Mathematische Classe* 3: 1 49.

Walther J (1910) *Lehrbuch der Geologie Deutschlands.* Leipzig: Quelle & Meyer.

Walther J (1915) Laterit in Westaustralien. *Zeitschrift der deutschen Geologischen Gesellschaft* 67: 113 132.

Walther J (1927) *Allgemeine Paläontologie. Geologische Fragen in biologischer Betrachtung.* Berlin: Borntraeger.

Wegener

B Fritscher, Munich University, Munich, Germany

Introduction

Alfred Wegener (Figure 1) was the 'father' of the theory of continental drift, one of the most influential theories in modern earth sciences. From a geophysical point of view, Wegener constructed a new picture of a dynamic Earth, postulating large-scale, ongoing horizontal movements of the continents, contrary to the theory of the permanence of continents and oceans that prevailed in the early twentieth century. First published in 1912, Wegener's theory had to wait for more than 50 years to become fully acknowledged, due in particular to the lack of a credible mechanism to explain, and direct empirical confirmation of, continental movement. The revival of Wegener's theory in the late 1950s came from studies of the ocean floors, an approach that Wegener had never considered. Though modern plate tectonics differs significantly from Wegener's original theory, there was nevertheless hardly any geological idea in the twentieth century that was subjected to greater scientific and public dispute than the idea of 'drifting' continents.

Meteorology and Polar Research

Born on 1 November 1880, Alfred Lothar Wegener studied astronomy, meteorology, and physics at Heidelberg, Innsbruck, and Berlin, earning a PhD in astronomy from the University of Berlin in 1905. Following his older brother Kurt Wegener (1878–1964), he was appointed an assistant at the Aeronautical Observatory at Lindenberg, near Berlin. There, he became acquainted with modern methods for the study of the higher atmosphere, including free balloon riding; together with his brother, he broke the world endurance record for balloon riding in 1906 by staying aloft for more than 52 h. That same year, Wegener joined a Danish expedition to Greenland under Ludvig Mylius-Erichsen (1872–1907); the goal was to map Greenland's north-east coast. Wegener became the first to use kites and tethered balloons to study the atmosphere in an Arctic climate. His Arctic research earned him a position at the University of Marburg, where he lectured on meteorology and practical astronomy from 1909. He published several papers on meteorological subjects, including a monograph on *The Thermodynamics of the Atmosphere* (1911), which became a standard textbook throughout Germany. By his studies on the chemical composition of the atmosphere, and its temperature distribution, Wegener pioneered the new science of aerology.

Wegener first presented his theory of continental drift to the public at the beginning of 1912. He found little time to reply to his numerous critics, for only a few month later he was in Greenland again. Together with the Dane, Captain Johan Peter Koch (1879–1928), Wegener became the first to winter on the icecap, and in the following spring, they undertook the longest crossing of the great ice sheet ever made up to that time. After his return from Greenland, Wegener married Else Köppen (1892–1992), daughter of Wladimir Köppen (1846–1940), a leading European meteorologist who became Wegener's lifelong mentor and collaborator.

In the summer of 1914, Wegener was drafted into the German army but was soon released from combat duty after being twice wounded. The fall of a meteorite in April 1916, near Marburg in Hesse,

Figure 1 Alfred Wegener in 1910. Reproduced from the Deutsches Museum, Munich.

turned his attention to these bodies and to the origin of the craters of the moon, which he thought were formed by impacts of bodies belonging to the solar system. In 1919, following his father-in-law, he became head of the department of theoretical meteorology at the German Marine Observatory at Grossborstel near Hamburg, and a 'professor extraordinary' (außerordentlicher professor) at the newly founded University of Hamburg. From 1924, Wegener held a professorship in meteorology and geophysics at the University of Graz. Soon he was preparing another expedition to Greenland, for a systematic study of the great icecap and its climate; the expedition departed in 1930. Wegener died at the beginning of November 1930, a day or two after his 50th birthday, while returning from a rescue expedition that brought food to a party of his colleagues camped in the middle of the Greenland icecap. His body was eventually recovered in May 1931.

A New Image of the Earth

Wegener's famous theory of continental drift was actually the work of just a few months. In the autumn of 1911, he became aware of a paper summarizing the evidence for the close relationship of the older fauna of South America and West Africa. He also remembered an earlier observation of the striking congruence of the coastlines on either side of the Atlantic Ocean; towards the end of the year, in 1911, he gave a preliminary account of his basic ideas in a letter to Wladimir Köppen. On 6 January 1912, Wegener presented his new theory at a meeting of the Geologische Vereinigung (Geological Society) in Frankfurt, and promptly published a preliminary paper on his ideas in *Petermann's Geographische Mitteilungen* (Figure 2).

For most earth scientists of Wegener's day, the hypotheses of wandering continents sounded rather fantastic, chiefly because of the lack of a satisfactory explanation of the moving forces. However, with the help of the German geologist Hans Cloos (1885–1951), Wegener extended his early paper and published his first book on his theory as *Die Entstehung der Kontinente und Ozeane* (*The Origin of Continents and Oceans*) in 1915, now also referring to some earlier (rather speculative) ideas on continental displacements by William Henry Pickering

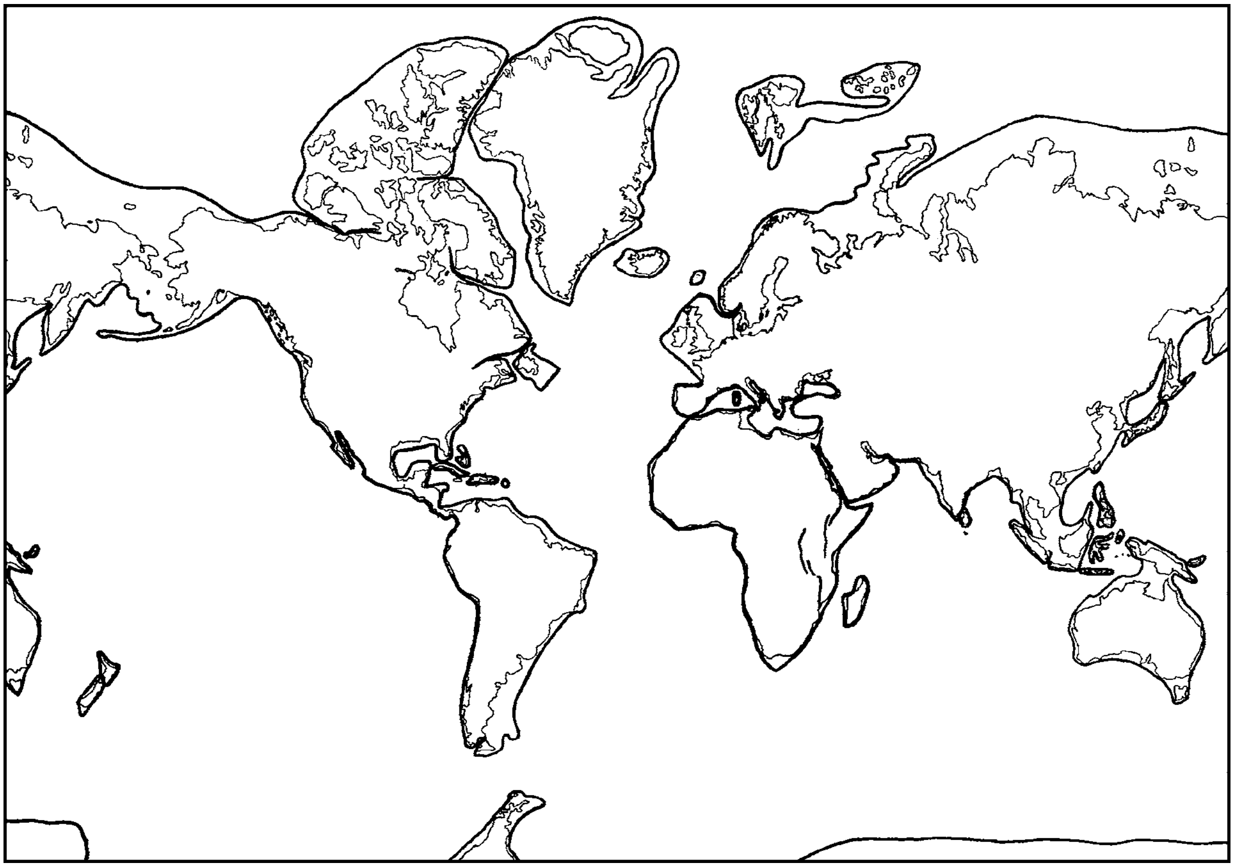


Figure 2 Map from Wegener's 1912 paper, *Die Entstehung der Kontinente*, showing the boundaries of the continental blocks (including the continental shelves).

(1858–1938) and Frank Bursley Taylor (1860–1938). After World War I, the second and third editions of his book, both revised and enlarged, were published (in 1920 and 1922, respectively). Palaeoclimatology was an essential empirical background of his theory, and, together with Köppen, Wegener published in 1924 a detailed discussion of the climates of the geological past (*Die Klimate der geologischen Vorzeit*). A fourth edition of *The Origin of Continents and Oceans* followed in 1929. From 1924 onwards, there were also translations (of the third edition) into English, French, and other languages, by which the theory became known internationally.

Wegener's original intention had been to give a genetic explanation of the large-scale features of Earth's surface (the continental blocks and the ocean basins) according to a single comprehensive principle. This principle of horizontal mobility, i.e., of the splitting off and drifting apart of continental blocks, did not emerge from new experimental research, but, rather, from a 'rearrangement' of known geophysical and geological information. This was set against two widely accepted and closely related assumptions in the earth sciences around 1900, namely, the hypotheses of former (now sunken) land bridges between continents (postulated due to striking geological and palaeontological conformities) and the theory of earth contraction due to its general cooling, advocated by leading geologists such as Eduard Suess (1831–1914) and Albert Heim (1849–1937).

Contraction would not have been sufficient to account for the large folds of Earth's crust. The great arching forces required to transmit the full shrinkage of a whole great-circle to one point of it have been proved to be physically impossible. The idea of a shrinking Earth, commonly illustrated by the simile of a drying apple, contradicted gravity measurements, i.e., the doctrine of isostasy (the rocks under the oceans are denser than are those under the continents, so altitudinal differences are compensated and equilibrium of pressure or 'isostasy' prevails; accordingly, the less dense continental blocks may be thought of as 'swimming' on the underlying mass, like an iceberg floating in the sea).

Wegener referred to Suess' distinction of the 'Sial' (silica/alumina-rich) and 'Sima' (silica/magnesia-rich) layers, emphasizing their different densities (2.5–2.7 for the Sial, or the continental blocks, and about 3.0 for the Sima of the ocean floors) and their different melting points (with the Sial's melting point being 200–300°C above that of the Sima). Concerning the temperature increase towards Earth's interior, the difference might not be sufficient to justify the assumption that solid Sialic blocks 'swim' in a fluid Sima (Figure 3). Rather, both layers had to be thought of as being viscous, i.e., plastic (with greater plasticity for the Sima). The characteristics of such viscous fluids are paradoxical in that the duration of the action of the forces determines whether the body behaves as a solid or a fluid. Consequently, within geological time, large horizontal displacements of the

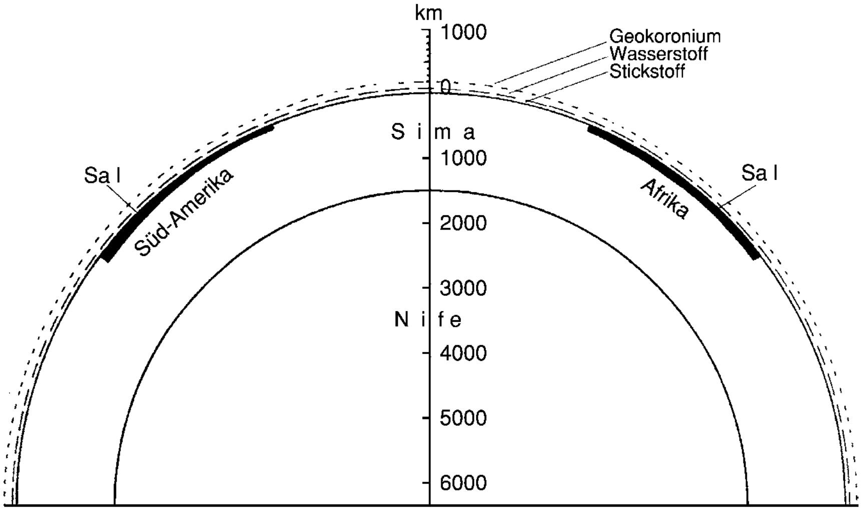


Figure 3 Cross section along a great circle through South America and Africa, from Wegener's 1912 paper, *Die Entstehung der Kontinente*, showing the 'swimming' continents within the Sima (silica/magnesia rich region). Also represented are the Nife (nickel/iron rich core of the Earth), and the atmospheric layers of nitrogen (Stickstoff), hydrogen (Wasserstoff), and 'geokoronium' (a hypothetical gas, which Wegener had introduced to account for the typical green spectrum line of the aurora borealis) (figure in correct scale).

continents would be possible, though the motion would be very slow.

Mountain Formation, Volcanism, and Rift Valleys

Wegener first discussed the origin of continents and oceans and the formation of folded mountain ranges. Relating to the nappe-fault structures of the Alps, where the original areas that are now thrust up as mountains were apparently up to four or five times wider than at present, Wegener thought that mountain formation was a unilateral, irreversible process: each pressure brings about an increase of the thickness and a shortening of the surface, while, on the other hand, tension leads to splitting of the continental blocks. The individual stages of perceived as mountain formation comprised continual processes of splitting and compression, whereby the original Sialic crust (for which Wegener assumed a thickness of about 30–35 km) gradually decreased in surface area, split into separate pieces, and increased in thickness. Along with the movement of the continental blocks, a hypothesized universal ocean ('Panthalassa') began to divide into a shallow sea and a deep sea.

Volcanism, for Wegener, was mainly related to the continental 'fronts'. Areas where tension prevailed, such as the Atlantic Ocean, and also opening faults, seemed to be relatively poor in volcanoes as compared with areas such as the Pacific Ocean, where pressure was increasing. The fronts of moving blocks made conditions more favorable to volcanism than did the 'backs'. Nevertheless, Wegener wondered whether the mid-Atlantic ridge might be considered as a zone where, with the continuing expansion of the Atlantic, the floor was continuously breaking up, making room for fresh, relatively fluid and high-temperature Sima from below! Moreover, increased volcanic activity in some periods of Earth history might be due to large displacements (as, for instance, during the Tertiary).

Trench faults (Grabenbrüche), i.e., rift valleys, acquired new meaning as representing the beginnings of new continental separations. Gravity measurements had shown that beneath such lines lay material of greater density, compared to that on either side. Therefore, these lines could be seen as incipient fissures within the continental blocks (into which the denser Sima was rising according to the principle of isostasy). The best examples of such separations were provided by the East African trenches and their continuation through the Red Sea. At the majority of the trenches, the measurable mass deficit was not compensated by greater density of the matter beneath it. Thus, the trenches must be youthful disruptions of a continental block.

Wegener's theory of mountain formation was further supported by the fact that the folding of the Andes seems to have been essentially simultaneous to the opening of the Atlantic Ocean. The American blocks, during their westward drifting, had encountered resistance at the presumably very old and relatively rigid floor of the Pacific Ocean. Thus, the extended shelf, with its mighty sediments, forming the western border of the continental block, was compressed to a range of fold mountains. For the Tertiary folds of the Himalayas, Wegener assumed that lower India had formed an extended peninsula prior to compression, the southern end of which lay next to that of South Africa. The folds had been produced by 'impact' of the Indian subcontinent and the main mass of Asia.

Geological and Palaeontological Evidence

The palaeontological evidence indicating a former connection between the organic components of different continents had already given rise to the doctrine of former land bridges. Among the most striking findings were the distributions of the *Glossopteris* flora on the southern continents and the occurrence of *Mesosaurus* at the turn of the Permian and the Carboniferous exclusively in south-eastern South America and the western parts of Africa; both of these discoveries suggested a former connection of the two continents. Using these relationships also allowed calculations of when the continents were separated (either by horizontal displacements or by sinking of the land bridges). South America and Africa had been connected during the Mesozoic, but were separated at the end of the Eocene or Early Oligocene. The connection between Europe and North America seemed to have been maintained during the older Tertiary period, but separation occurred in the Miocene, although it might have continued in the far north (over Scandinavia and Greenland) into the Pleistocene. The connection of Lower India with southern Africa, which Wegener had postulated based on his ideas on the formation of the Himalayan range, was also confirmed by palaeontological evidence. Zoogeographers had long assumed a former elongated Indian–Madagascan peninsula (called 'Lemuria'), separated from the African block by the Mozambique Channel.

The zoogeographic concept of Lemuria had given rise to Suess' notion of a great southern continent, Gondwana, comprising parts of South America, Africa, Lower India, Australia, and Antarctica. Assuming the unchanged positions of its present-day relics, however, required ascribing a huge extent to this continent. Wegener, by contrast, proposed a much reduced primeval continent, Pangaea. In the Permian, i.e., until some 300 Ma ago, all the continents were supposedly joined in one land mass extending from pole to pole. During the Triassic, about 200 Ma ago, Pangaea began to break up and the newly emerging continents started moving into their current positions. In the Jurassic, there were few remaining connections except at the northern and southern ends. Just as northern Europe and North America remained connected until the older Tertiary period, a connection of the southern continents seems to have persisted, running from the southern coast of Australia over Antarctica to South America. Later, the Antarctic block, like the South American block in the Tertiary, moved over from South Africa towards the side of the Pacific Ocean. Only in the Quaternary period, then, did the Australian block become detached (Figure 4).

For geological and tectonic evidence, Wegener referred particularly to Suess' *magnum opus*, published in three volumes during 1885–1909, *Das Antlitz der Erde* (*The Face of the Earth*). Considering the tectonic relations, Europe/Africa and both Americas seemed to represent the edges of an immense expanded fissure. In the north, for instance, the Greenland massif was matched by Scandinavia, both consisting of gneiss, and the less mountainous North America corresponded to the likewise less mountainous Europe. The most striking example, however, was the Carboniferous mountain range, called the Armorican mountains (Suess' 'transatlantic Altaides'), which made the coalfields of North America appear to be the direct continuation of the European ones.

Wegener's theory of mountain formation was also confirmed by remarkable differences between the Atlantic and the Pacific hemispheres, such as the distinction between Pacific and Atlantic types of coasts (marginal chains and ocean trenches in front

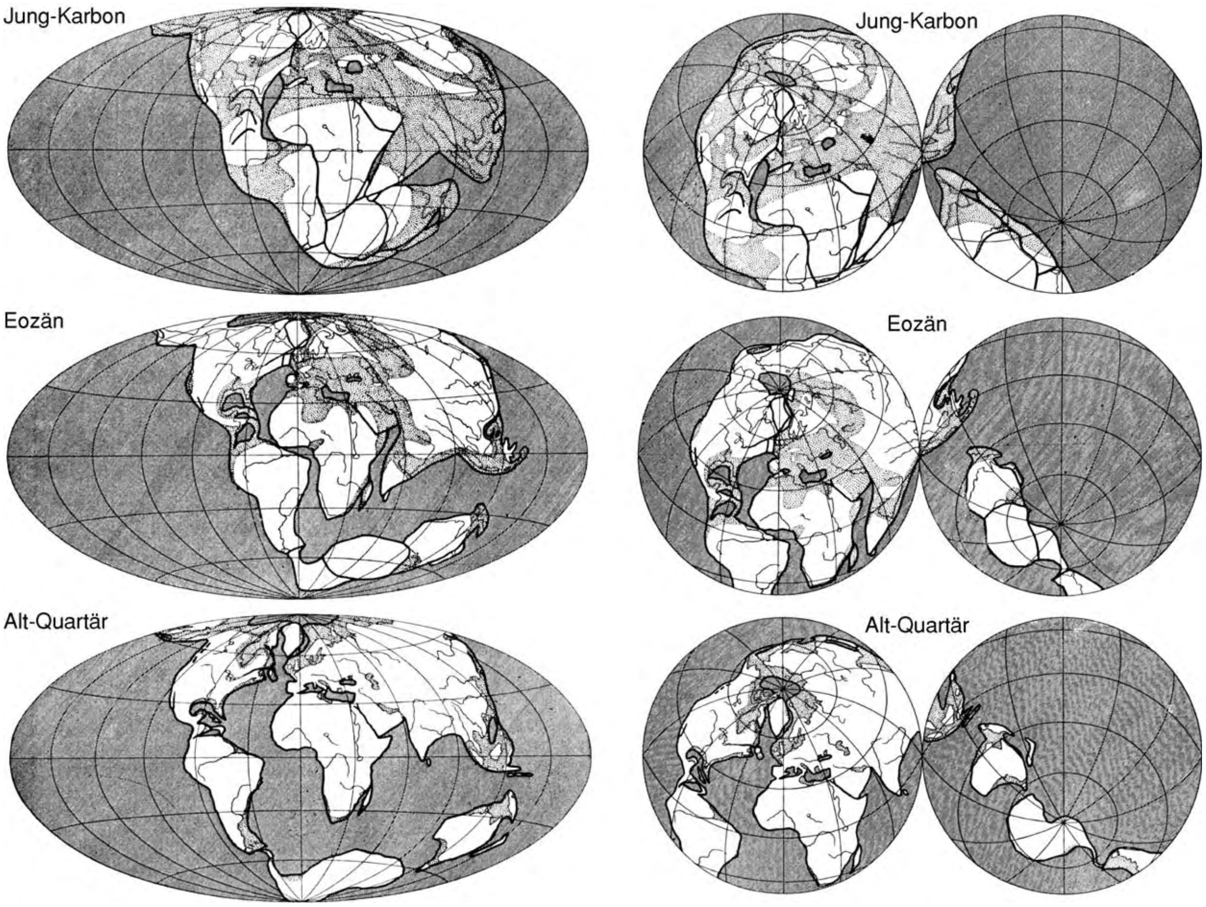


Figure 4 Wegener's reconstruction of the separation of the continents from the primeval Pangaea, from his 1926 paper *Paläogeo graphische Darstellung der Theorie der Kontinentalverschiebungen*, showing the relative positions of the continents during the Upper Carboniferous (Jung Karbon), Eocene (Eozan), and Lower Quaternary (Alt Quartar) (in two different projections). Cross hatching represents deep seas, dotted regions represent shallow seas; rivers, recent coastlines, and outlines are shown only for orientation.

of the Pacific coasts, as contrasted to the wild, irregular 'ria' Atlantic coastlines). There were also differences in the volcanic lavas of the two hemispheres, as emphasized by the Vienna petrographer Friedrich Becke (1855–1931) and others. The Atlantic lavas contained a greater proportion of sodium, whereas calcium and magnesium prevailed in the Pacific lavas. Such differences were intelligible according to the assumptions of continental movements. The opening of the Atlantic was matched by the general pressing of the continents against the region of the Pacific Ocean: pressure and compression prevailed at the coasts of the latter whereas tension and splitting occurred at the latter.

Palaeoclimatology

Traces of glaciation during the Permian (ground moraines lying on scratched bedrock) were to be found on the southern continents, e.g., in East India and Australia. If the present-day arrangement of the land masses had prevailed at that time, this 'Permian ice age' would have required an icecap of seemingly impossible size. And the north pole would have been in Mexico, where no trace of glaciation during that period was recorded. Following the idea of horizontal displacements, however, all regions subjected to glaciation came together concentric to the southern margin of Africa. And one had only to place the south pole in this much reduced glaciated area to give the Permian ice age a much more plausible form.

Wegener had discussed these palaeoclimatological features since 1912. In 1924, he gave a detailed description of the climatological changes from the Carboniferous through to recent times, following the traces of glaciations, swamps, and deserts, i.e., moraines, coal, salt, and gypsum, throughout Earth's history (Figure 5). In reconstructing the respective polar shifts, Wegener emphasized that they obviously took place along with the great displacements of the continental blocks. In particular, there was temporal coincidence of the best confirmed polar shift, in the Tertiary, and the opening of the Atlantic (Figure 6). Movement of the poles since the Pleistocene might also be related to the final separations of the continents in the north and the south.

Motive Forces

Wegener was very cautious about the forces that might have caused continental displacements. First, it was necessary to demonstrate the reality and the manner of the displacements before indulging in the hope of finding their cause. Nevertheless, he tentatively suggested two candidates: centrifugal forces caused by the rotation of Earth and tidal-type waves within Earth, generated by the gravitational pull of the sun and the moon. In the 1929 revision of Wegener's theory in

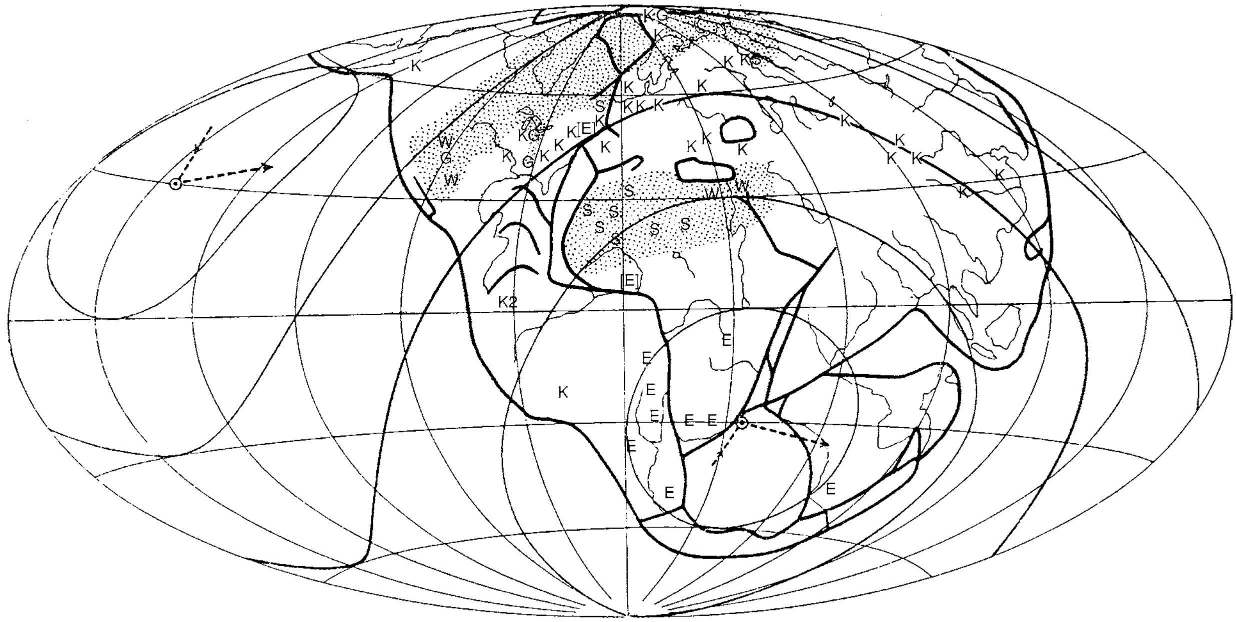


Figure 5 Wegener thought continental drift was the key to the climatic changes during Earth's history. This map, published in the 1924 book by Koppen and Wegener, *Die Klimate der geologischen Vorzeit*, shows traces of glaciation, swamps, and deserts for the Carboniferous. E, Traces of glaciation; K, coal; S, salt; G, gypsum; W, desert sandstone. Dotted regions indicate arid areas, dashed lines indicate the positions (i.e., the pathways) of the poles, and the bold curved line indicates the respective position of the equator.

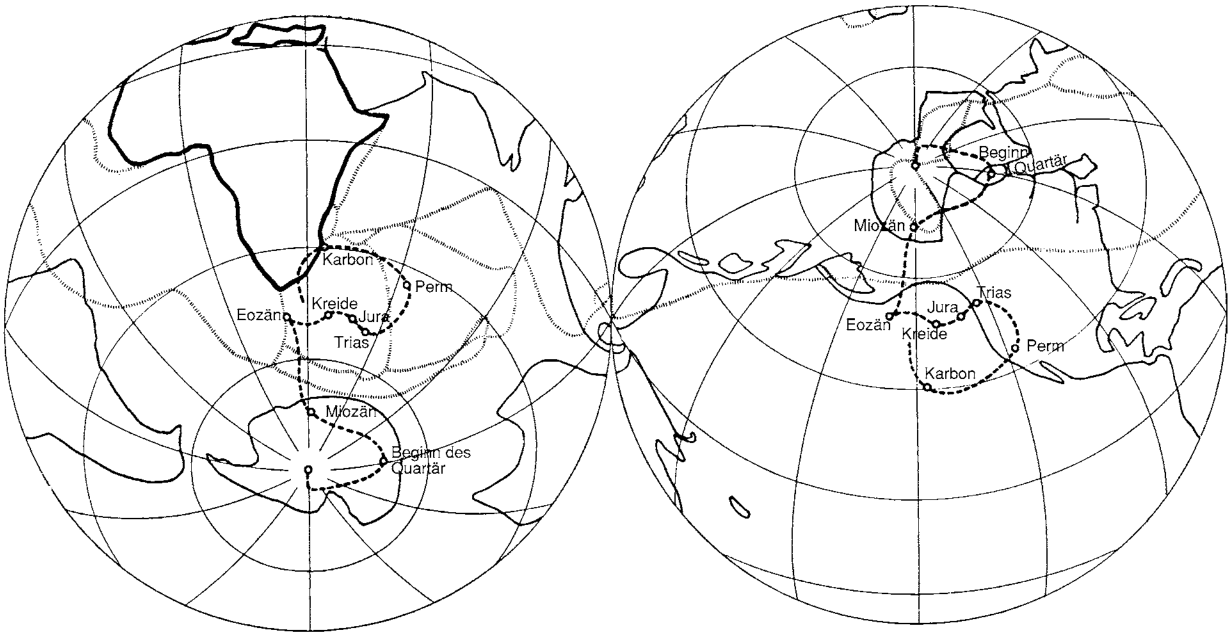


Figure 6 Map published in the 1924 book by Koppen and Wegener, *Die Klimate der geologischen Vorzeit*, showing polar shifts (dashed lines) from the Carboniferous to recent, related to the African table (left, south pole; right, north pole). Bold lines outline the continental blocks; hatched lines represent the Carboniferous (Karbon) period. Perm, Permian; Jura, Jurassic; Trias, Triassic; Kreide, Cretaceous; Eozan, Eocene; Miozan, Miocene; Beginn des Quartar, beginning of the Quaternary.

Die Klimate, he also mentioned convection currents within the Sima; these had been first discussed as a cause of mountain formation by the Vienna geologist Otto Ampferer (1875–1947) in 1906.

Wegener also endeavoured to calculate the recent velocity of the relative motion of the continents, though he was well aware that these values must be quite uncertain. In his 1912 paper, comparing various longitude determinations for Greenland, he had deduced an increase of the distance to Europe of 11 m year ^{1}. Referring to the lengths of transatlantic cables, he suggested that North America was drifting away from Europe at about 4 m year ^{1}.

From Continental Drift to Plate Tectonics

The theory of continental drift was long rejected by the majority of geologists. Among Wegener's few followers were the South African Alexander Du Toit (1878–1948), for whom continental drift provided the best explanation of the close similarities between the strata and fossils of Africa and South America, and the Swiss geologist Émile Argand (1879–1940), who saw continental collisions as the only means of producing the folded and buckled strata he had observed in the Alps (*see* **Famous Geologists:** Du Toit). Nevertheless, Wegener's explanation of the Permo-Carboniferous ice age impressed even his critics.

Wegener's reputation as a meteorologist and a polar explorer contributed to keeping his theory alive. His work was immediately remembered when, around 1960, surprising data were obtained from the ocean floor: palaeomagnetic patterns alongside the mid-ocean ridges clearly suggested the spreading of the seafloor. Within about two decades, Wegener's principle of horizontal displacements of parts of Earth's crust became almost universally accepted, although, ironically, the process still lacked a consensus as to its causes, though convection currents in the internal mantle are most commonly advocated.

It should be noted that Wegener's original concept differed from modern plate tectonics in essential points, particularly with regard to the Sial and the Sima. According to modern theory, the (Sialic) continents do not 'plough' through the (oceanic) Sima. Instead, both continents and ocean floor are regarded as forming solid plates, 'floating' on the asthenosphere, which, due to tremendous heat and pressure, behaves like an extremely viscous liquid (as Wegener had thought the Sima did). Therefore, the older term 'continental drift', still often used today, is not quite appropriate for the modern concept. Notwithstanding these differences, Wegener's basic ideas remain sound, and the lines of evidence that he used to support his theory are still valid. He first envisaged a dynamic Earth, connecting its major features and various geological processes – continental movements, folded mountain ranges, rift systems, earthquakes, volcanism, ocean transgressions, palaeoclimatological changes, etc. – on a global scale. In this sense, Wegener's theory was a true forerunner of plate tectonics.

See Also

Africa: Rift Valley. **Famous Geologists:** Du Toit; Suess. **Gondwanaland and Gondwana**. **History of Geology From 1900 To 1962**. **History of Geology Since 1962**. **Palaeoclimates**. **Pangaea**. **Plate Tectonics**. **Tectonics:** Mid-Ocean Ridges; Mountain Building and Orogeny.

Further Reading

Carozzi AV (1985) The reaction of continental Europe to Wegener's theory of continental drift. *Earth Sciences History* 4: 122 137.

Fritscher B (2002) Alfred Wegener's 'The origin of continents, 1912'. *Episodes* 25: 100 106.

Jacoby WR (2001) Translation of 'Die Entstehung der Kontinente, Dr Alfred Wegener, *Petermann's Geographische Mitteilungen*, 58 (1912)'. *Journal of Geodynamics* 32: 29 63.

Köppen V and Wegener A (1924) *Die Klimate der geologischen Vorzeit*. Berlin: Bornträger.

Lüdecke C (1994) Stratigraphische Methode der Rekonstruktion von Expeditionsergebnissen am Beispiel des Todes von Alfred Wegener während der Grönlandexpedition (1930 31). In: Fritscher B and Brey G (eds.) *Cosmographica et Geographica: Festschrift für Heribert M. Nobis zum 70. Geburtstag, Algorismus,* vol. 13, pp. 347 367. Munich: Institut für Geschichte der Naturwissenschaften.

Oreskes N (1999) *The Rejection of Continental Drift: Theory and Method in American Earth Science*. New York and Oxford: Oxford University Press.

Runcorn SK (ed.) (1966) *Continental Drift*. New York and London: Academic Press.

Schwarzbach M (1986) *Alfred Wegener: The Father of Continental Drift*. Madison. WI: Science Tech Publications.

Sengör AMC (1991) Timing of orogenic events: a persistent geological controversy. In: Müller DW, McKenzie JA, and Weissert H (eds.) *Controversies in Modern Geology: Evolution of Geological Theories in Sedimentology, Earth History and Tectonics,* pp. 403 473. London: Academic Press.

Wegener A (1912) Die Entstehung der Kontinente. *Petermann's Mitteilungen aus Justus Perthes' Geographischer Anstalt* 58: 185 195, 253 256, 305 309.

Wegener A (1926) Paläogeographische Darstellung der Theorie der Kontinentalverschiebungen. In: Dacqué E (ed.) *Paläogeographie,* pp. 171 189. Leipzig and Wien: F Deuticke.

Wegener A (1971) *The Origin of Continents and Oceans*. (Translation from the 4th revised German edition by J Biram, with an introduction by BC King.) London: Methuen.

Wegener A (1980) *Die Entstehung der Kontinente und Ozeane*. (Reprint of the 1st and 4th editions, edited by A Vogel.) Braunschweig: Vieweg.

Wegener E (1960) *Alfred Wegener: Tagebücher, Briefe, Erinnerungen*. Wiesbaden: Brockhaus.

Wutzke U (1998) *Kommentiertes Verzeichnis der schriftlichen Dokumente seines Lebens und Wirkens,* Berichte zur Polarforschung 288. Bremerhaven: Alfred Wegener Institut für Polar und Meeresforschung.

FLUID INCLUSIONS

A H Rankin, Kingston University,
Kingston-upon-Thames, UK

Introduction

Fluid inclusions are small droplets of fluid that have been trapped within crystals either during primary growth from solution or at some later stage, usually as a result of recrystallization along healed microfractures. They are ubiquitous in both naturally occurring minerals and in laboratory-grown crystals. To the chemist or materials scientist, these gross defects cause endless obstacles in their quest to grow near-perfect crystals. However, to the geologist, they provide a unique fossil record of the various fluids responsible for the formation and evolution of rocks and minerals throughout the history of the Earth.

Despite their small size (usually less than 20 μm), their chemical composition and physical properties can be readily determined, and the data may be used to estimate the temperatures, pressures, and physicochemical nature of the fluid at the time of trapping. This information has made an immense contribution to the development of modern theories of ore genesis, petrogenesis, diagenesis, and petroleum migration and accumulation, and to our understanding of the importance of the fluid phase in a wide range of geological processes.

Occurrence and General Characteristics

Formation and Genetic Classification of Fluid Inclusions

Small changes in the chemical or physical properties of fluids near to a growing crystal face can lead to perturbations in the stability of crystal growth and the development of gross defects, manifested as embayments, along crystal faces. These embayments will seal over during a period of greater stability, trapping a portion of fluids to form 'primary' (P) fluid inclusions. In many instances, the trapped fluid will be 'homogeneous' at the time of trapping. In others, where immiscible fluids are present or where mechanical entrapment of other coexisting crystalline phases has occurred, trapping will be 'heterogeneous'.

At some stage after primary growth, 'secondary' (S) fluid inclusions can form from later fluids, particularly as a result of recrystallization along microfractures. The chemical and physical properties of these inclusions may be very different from those of the earlier mineral-forming fluids. However, if fracturing and rehealing take place during primary growth, the fluids may be indistinguishable, and the terms 'pseudosecondary' or 'primary–secondary' (PS) appropriately describe such inclusions. A schematic representation of this genetic classification of inclusions is shown in Figure 1.

For most geological applications, it is necessary to establish whether the inclusions are primary, secondary, or pseudosecondary, and also whether heterogeneous trapping has occurred. Heterogeneous trapping may be recognized by the variable proportions of liquids and solids in a single group or generation of inclusions. Various criteria may be used to distinguish between P, PS, and S inclusions, but these may be difficult to apply and it may be difficult to identify primary inclusions in many samples.

Choice of Material for Study

The successful application of fluid inclusion studies depends partly on serendipity and partly on the type and quality of material available for study. Due to their small size, observations on fluid inclusions are carried out under a microscope using polished wafers around 1–2 mm thick. In most cases, clear, transparent minerals are needed, but it is also possible to study inclusions in some deeply coloured, semi-transparent minerals in very thin ($<50\,\mu$m) polished sections. Care must be taken with soft, easily cleaved minerals, such as calcite and fluorite, because of the possibility of leakage during sample preparation or analysis. Fluid inclusions are particularly difficult to observe in turbid or translucent minerals, such as feldspar. Quartz is usually the preferred host.

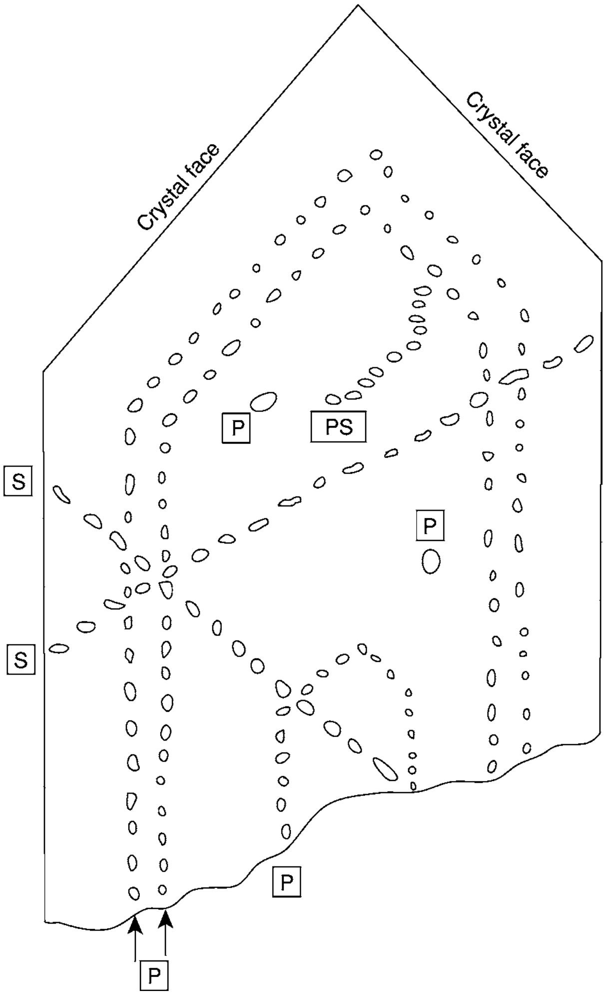


Figure 1 Schematic representation of the distribution of primary (P), secondary (S), and pseudosecondary (PS) fluid inclusions in a quartz crystal. Modified from Rankin AH (1989) Fluid inclusions. *Geology Today* 5: 21 24.

Size and Shape of Inclusions

Fluid inclusions seldom exceed 1 mm in size; most are less than 20 μm, and those greater than 1 cm are exceptionally rare and usually regarded as museum specimens. Fluid inclusions display a variety of shapes. They may be flattened and irregular, rounded, or regular with three-dimensional 'negative crystal' shapes mimicking the crystal symmetry of the host crystal (Figure 2).

Phases Present at Room Temperature

Fluid inclusions contain varying proportions of liquid (L), solid (S) and gas (G) depending on the composition (X), temperature (T), pressure (P), and volume (V) of the enclosed fluid at the time of entrapment.

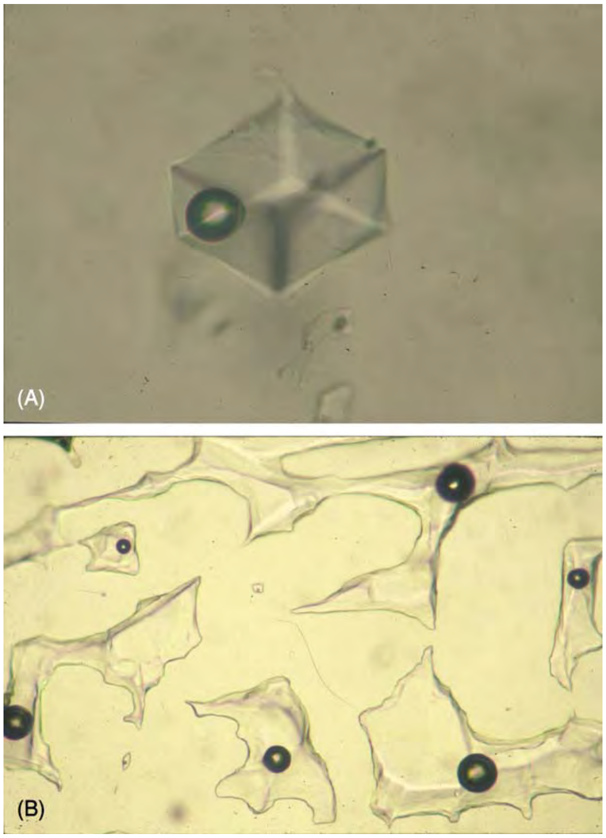


Figure 2 Contrasting shapes of two phase, liquid gas (L + G) aqueous inclusions in fluorite. (A) Negative crystal cavity mimicking the cubic symmetry of the host mineral. (B) Group of irregularly shaped inclusions showing evidence of necking down.

The nature of these phases will vary according to the geological conditions during formation.

At room temperature, inclusions in magmatic or hydrothermal minerals usually contain gas bubbles in addition to entrapped liquids (e.g., aqueous brines and silicate melts). In aqueous inclusions, the bubbles are mobile and often move in response to small temperature changes, e.g., through heat accompanying an increase in the intensity of microscope illumination. In silicate melt inclusions, the bubbles are immobile, either locked within a quenched silicate glass or squashed in bundles of crystals and crystallites. H. C. Sorby, often regarded as the 'father of modern petrography', explained the origin of these bubbles in his classic paper of 1858, which laid the foundations for the widespread application of fluid inclusions as geothermometers and geobarometers. He demonstrated that the gas bubbles in two-phase, gas–liquid inclusions were 'shrinkage bubbles' resulting from the differential thermal contraction of the contained liquids and host crystal on cooling. A simple analogue is the mercury-in-glass thermometer in which the headspace represents the shrinkage bubble.

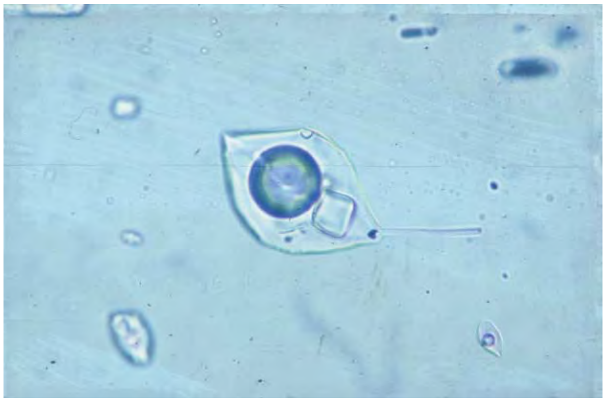


Figure 3 Three phase, liquid gas solid (L + G + S) aqueous inclusion in topaz containing a cubic halite daughter mineral and several smaller, unidentified daughter or captive minerals.

Figure 4 Three phase, liquid liquid gas (L + L + G) CO_2–H_2O inclusion with the characteristic 'double bubble' comprising an outer rim of liquid CO_2 enclosing a bubble of CO_2 gas.

Any dissolved salts that reach their saturation point on cooling from the trapping temperature will crystallize *in situ* as daughter minerals in aqueous inclusions (Figure 3). These can sometimes be identified on the basis of their optical and crystallographic properties and behaviour on heating and cooling. The most common daughter minerals are halite (NaCl), sylvite (KCl), nahcolite ($NaHCO_3$), gypsum/anhydrite ($CaSO_4$), and calcite ($CaCO_3$). Dissolved gases, such as CO_2, CH_4, and N_2, will also partition preferentially on cooling into the shrinkage bubble. In gas-rich aqueous inclusions, a 'double bubble' containing CO_2-rich liquid and CO_2-rich gas is sometimes apparent at temperatures below the critical point of CO_2 at 31.1°C (Figure 4).

The phases present in silicate melts include bubbles and daughter crystals, but this will vary depending on the cooling history and whether the host rock is plutonic or volcanic/subvolcanic. They may be glassy or crystalline and vapour bubbles may or may not be present.

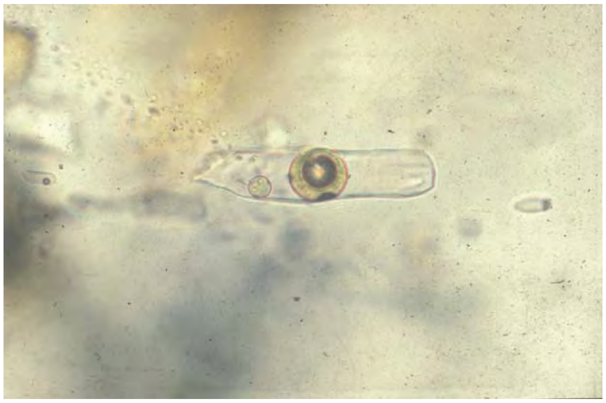


Figure 5 Aqueous inclusion containing two immiscible droplets of brown coloured oil in fluorite. Note the presence of a gas bubble in the larger oil droplet.

Heterogeneous trapping of coexisting liquids, melts, and volatiles is also evident in some geological environments. For example, immiscible droplets of oil- or methane-rich vapour may be trapped in aqueous inclusions from low-temperature sedimentary/diagenetic environments (Figure 5). Recognition of these immiscible assemblages provides important information on magmatic differentiation processes, such as the exsolution of brines from silicate melts in granitic rocks, and silicate–carbonate immiscibility in alkaline and carbonatitic environments. In hydrothermal environments, the coexistence of low-density gas-rich and high-density aqueous inclusions in hydrothermal minerals can be used to infer boiling conditions in a hydrothermal system.

Heterogeneous trapping of coexisting solid and liquid can also lead to the development of inclusions containing varying proportions of captive rather than true daughter minerals.

Geological Applications – Information Available

Prerequisites and Assumptions

There are a number of potential pitfalls in the application of fluid inclusion studies to geological problems.

Is the material trapped within an inclusion a representative sample? For most applications, the assumption that the fluid was trapped homogeneously is necessary. Although this is not always the case, it is often possible to infer homogenous trapping from the consistency of the liquid, solid, and gas phase proportions in a given generation of inclusions. Varying proportions will indicate either heterogeneous trapping, multiple fluid generations, or necking down.

A more fundamental issue is whether the compositional differences known to occur within the submicroscopic boundary layer between the crystal and growth solution are significant. Under most geological growth conditions, this layer is generally assumed to be small (less than 1 μm) and the effect is only likely to be significant in very small inclusions, less than a few micrometres, which are too small to study anyway. A rule of thumb is that larger inclusions are much less likely to suffer from these boundary layer effects. This is backed up by evidence from studies of synthetic inclusions, where compositions usually match those in the parent growth solutions.

Have components been lost or added since entrapment? There has been much debate in the past as to whether fluid inclusions have retained their compositional integrity over geological time. The consensus, based on laboratory studies of natural and synthetic inclusions, is that in many situations bulk leakage is unlikely. In others, it can often be recognized or avoided. Bulk leakage and decrepitation (the explosive release of inclusion contents usually as a result of rapid heating) are relatively easy to recognize in natural samples, either because the inclusions appear ‘empty’ or are surrounded by characteristic decrepitation haloes of smaller inclusions. In soft, easily cleaved minerals, such as halite and calcite, the inclusions are more susceptible to leakage. It is partly for this reason that fluid inclusion researchers usually prefer to study inclusions in more resilient minerals, such as quartz. Bulk leakage and decrepitation are also likely when pre-existing, low-temperature inclusions are subjected to higher temperatures (and pressures), e.g., in low-temperature mineral veins or evaporite minerals in the contact zone of igneous intrusions. Experimental and theoretical studies have shown that, under high thermal pressure and chemical gradients, components such as H_2O and H^+ may move in or out of inclusions via diffusion rather than bulk leakage. Experts have argued that the high gradients required to maintain an effective rate of diffusion are insufficient over the time-scales involved in many geological processes, except under extreme P–T conditions, e.g., during high-grade metamorphism.

Has the volume of the inclusion remained constant since trapping; what are the effects of recrystallization? This question is particularly important if fluid inclusions are to be used as effective geothermometers and geobarometers. Irregularly shaped fluid and melt inclusions have a tendency to change towards energetically more stable shapes, such as negative crystals or spheres, via a process of internal recrystallization of the inclusion walls. This usually occurs without any change in volume.

Crystallization of material on the walls of inclusions in hydrothermal and metamorphic minerals after trapping is usually considered to be insignificant because of the low solubilities of host minerals, such as quartz, in aqueous or CO_2-rich fluids. In highly water-soluble minerals, such as halite, or in melt inclusions formed in slow cooling plutonic environments, significant crystallization may occur on the walls of the cavity. Although there will be some apparent change in volume after trapping, the situation can be reversed by heating the sample at an appropriate rate in the laboratory.

The internal recrystallization processes responsible for changing the shape of large irregular inclusions can also cause the inclusion to split up or 'neck down' into a number of smaller inclusions. This is not usually a problem if necking down has occurred before the development of vapour bubbles or daughter phases in the inclusions, because each necked-down inclusion will still retain the original *P*–*V*–*T*–*X* properties at the time of trapping. Problems arise when necking down has taken place after heterogenization, in which case inclusions with variable phase proportions can develop (Figures 2B and 6).

Can a particular fluid inclusion assemblage be linked to a specific stage of crystal development or geological process? If primary (or pseudosecondary) inclusions are present, given the caveats above, they should represent the primary crystallization conditions and processes responsible for the formation of the host mineral. If only secondary inclusions are present, all that can be inferred are the *P*–*V*–*T*–*X* conditions during some recrystallization processes. Several generations of secondary inclusions, each representing a different stage in the evolutionary history of a rock or mineral sample, may be present, and careful petrographical examination is required to distinguish between them. Cathodoluminescence studies of quartz or carbonate hosts can sometimes be used to establish relative fluid inclusion chronologies. With care and patience, it is sometimes possible to link them to particular stages of crystallization or fluid activity (Figure 7).

Fluid inclusion studies on hydrothermal ore deposits are usually carried out on transparent gangue minerals because the associated ore minerals are usually opaque. Even when primary inclusions are present, it is critically important to establish a clear mineral paragenesis to ensure that a particular generation of transparent gangue is coeval with a particular stage of ore mineralization.

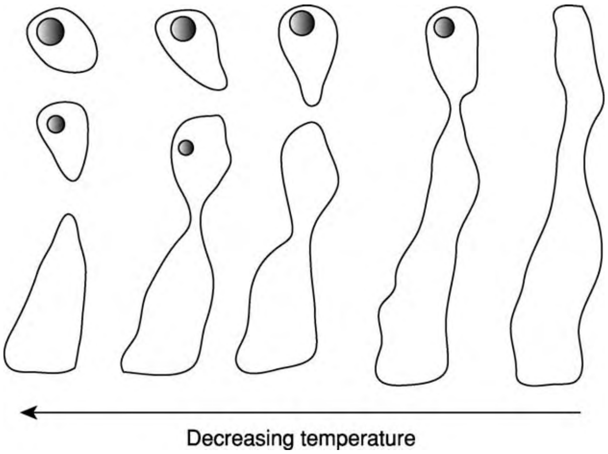


Figure 6 Schematic representation of necking down processes in aqueous fluid inclusions after trapping, showing the development of contraction bubbles at different stages in the necking down process. Modified from Roedder E (1984) Fluid inclusions. *Mineralogical Society of America, Reviews in Mineralogy* 12, and Shepherd TJ, Rankin AH, and Alderton DHM (1985) *A Practical Guide to Fluid Inclusion Studies*. London: Blackie and Sons.

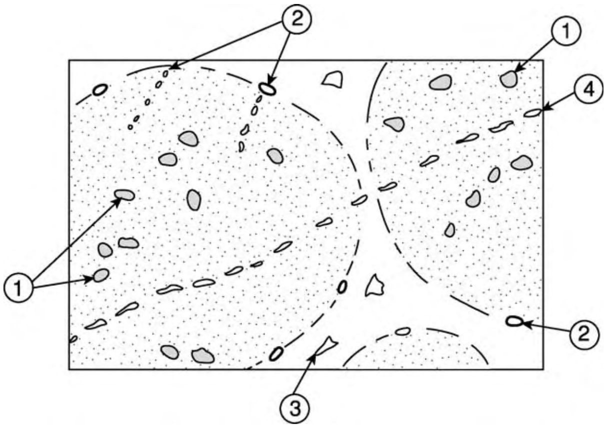


Figure 7 Schematic representation of four different generations of fluid inclusions in a quartz cemented sandstone. The earliest inclusions (1) are contained within the quartz grains and represent fluid processes prior to sedimentation and diagenesis. The latest inclusions (4) cross cut both the sand grains and the quartz cement and are also unrelated to the cementation process. Inclusions related to early cementation (2) may occur at the quartz grain boundary and along healed microfractures in the quartz grains. Inclusions related to later cementation processes (3) occur within the centre of the cement.

Principles of Fluid Inclusion Geothermometry and Geobarometry

One of the most important applications of fluid inclusion studies is their use as palaeogeothermometers and palaeogeobarometers. This involves the careful measurement of the temperatures at which the heterogeneous contents of the inclusion became homogeneous on heating, referred to as the homogenization temperature (T_h). This involves direct observations, under the microscope, using specially constructed heating (and cooling) stages covering the temperature range from −196°C (the boiling point of liquid nitrogen used as a coolant) up to about 1200°C. In the case of high-temperature melt inclusions, an alternative is to carry out observations on the inclusions after a series of successive heat/quench runs using conventional muffle furnaces.

The homogenization temperature may be used to place constraints on the P–T trapping conditions and, hence, the formation conditions if the inclusions in the host rocks are primary, provided that:

1. The inclusions have maintained their integrity over geological time.
2. Homogeneous versus heterogeneous trapping conditions can be recognized.
3. The bulk composition of the inclusions (aqueous brine, silicate melt, H_2O fluid, liquid petroleum, etc.) is known and the P–V–T data for the relevant fluid systems are available from the literature or can be calculated from thermodynamic data.

Aqueous inclusions and gaseous inclusions The principles are illustrated in Figure 8 with reference to two-phase aqueous inclusions modelled on the simple H_2O system. The homogenization temperature defines a unique point on the vapour–liquid or boiling point curve and represents the minimum P–T trapping conditions. As the composition (X) is known, this point also uniquely defines the density ($1/V$) of the fluid. The line of equal density emanating from this point is known as an isochore. The true P–T trapping conditions must have taken place at some point along this isochore. If depths or pressures are known, the trapping temperature (T_t) can be estimated. Reciprocally, if pressure-independent methods, such as oxygen isotope geothermometry, are available, the pressure or depth can be established.

The same principles apply to gas-rich inclusions in the system CO_2 N_2 CH_4, which characterize many metamorphic environments and associated mineral deposits. Homogenization will take place either to liquid or vapour at temperatures below 31°C, the highest critical point for an end-member gaseous component in this system (CO_2). By studying different generations of gas-rich inclusions, it may be possible to establish a reasonable idea of the P–T evolution and peak conditions during the metamorphic history of a particular sample.

Coexisting aqueous and gas-rich or petroleum-bearing inclusions In cases in which it can be demonstrated that aqueous and gaseous fluid inclusions represent the homogeneously trapped end members of a 'boiling fluid', trapping must have occurred on the G + L curve. In these cases, T_h will be the same as the trapping temperature (T_t). An extension of this principle is the 'intersecting isochore' method of geothermobarometry which can be applied to homogenization temperature data for coexisting aqueous and hydrocarbon inclusions, using the relevant bubble point or methane saturation curves, as illustrated in Figure 9.

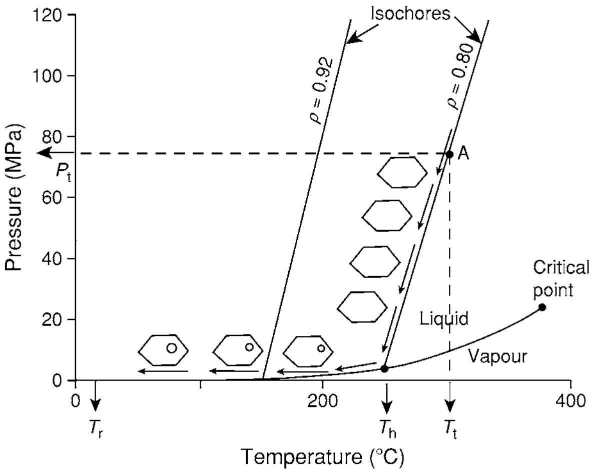


Figure 8 P T phase diagram for pure water illustrating the principles of fluid inclusion geothermometry and geobarometry. An inclusion trapped at P_t and T_t will follow the P T cooling path defined by a line of equal fluid density (an isochore). The inclusion remains homogeneous until the isochore meets the vapour liquid equilibrium curve, at which point a vapour bubble nucleates. Continued cooling takes place along this curve, causing further shrinkage of liquid and growth of the bubble down to room temperature. On heating, the liquid expands, completely filling the inclusion at a point on the vapour liquid curve referred to as the 'homogenization temperature' (T_h), which represents the minimum P T trapping conditions. However, trapping could have occurred anywhere within the single phase region above this curve along the P T path defined by the isochore. An estimate of P_t (and salinity) is required to determine T_t. The difference between T_t and T_h is often referred to as the 'pressure correction'.

Melt inclusions It is generally accepted that the vapour–liquid homogenization temperatures of silicate melt or glass inclusions in minerals from intrusive or subvolcanic igneous rocks approximate to their trapping temperature without the need for pressure corrections. This is because silicate melts are far less compressible than other fluids, such as water or CO_2. In deeper seated intrusions, care is needed to ensure that equilibrium conditions have been maintained during the homogenization of multicomponent crystalline melt inclusions by using slow heating rates.

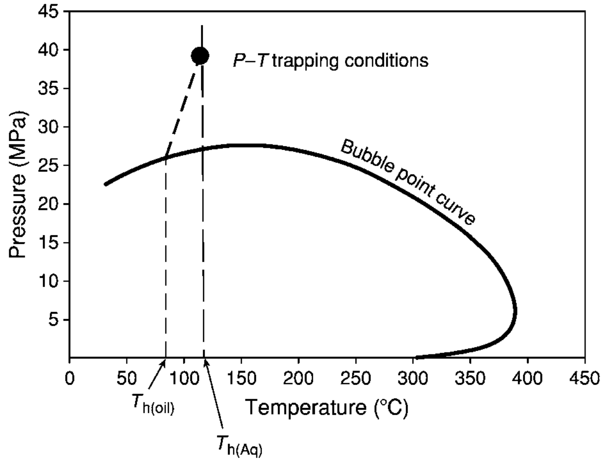


Figure 9 Principle of the isochore intersection method applied to coexisting oil and aqueous inclusions. The bubble point curve can be calculated from estimated oil compositions (e.g., from gas chromatography analysis of the inclusions). The homogenization temperature of the oil, $T_{h(oil)}$, can be used to calculate the P T trapping path of the oil inclusion in the single phase region above this curve. The point at which the isochore intersects the homogenization temperature line for coexisting aqueous inclusions, $T_{h(Aq)}$, defines the P T trapping conditions. See Munz In: Andersen T, Frezotti M L, and Burke EAJ (eds.) (2001) Petroleum inclusions in sedimentary basins: systematics, analytical methods and applications. *Lithos* 55: 185 212.

The T_h values of melt inclusions can be used effectively as independent geobarometers in cases in which silicate melts and associated fluid inclusions have been trapped coevally. The isochores for these associated aqueous or gaseous inclusions can be used to fix the pressure or depth of formation.

Chemical Compositions of Geological Fluids from the Analysis of Inclusions

A wide variety of methods are available to determine the chemical compositions of fluid and melt inclusions.

Simple optical methods Simple optical methods are often very effective. For example, the Becke line test can be used to distinguish between aqueous fluid and silicate glass phases because of their contrasting refractive indices. Ultraviolet (UV) fluorescence microscopy can be used to identify petroleum inclusions, and the identification of daughter minerals can sometimes be based on their optical and crystallographic properties.

Microthermometry Microscope heating and freezing stage studies (microthermometry) are frequently used to identify CO_2 and associated volatiles and to estimate salinities of aqueous inclusions. Pure CO_2 should melt at $-56.6°C$ (the triple point). Other volatiles, such as CH_4 and N_2, will lower this melting temperature. Similarly, the melting point of pure water is lowered by the addition of salts, such as NaCl, KCl, and $CaCl_2$, in solution. By measuring this depression and referring to published freezing point depression curves, it is possible to estimate the salt content or salinity of aqueous inclusions. Mostly, this is performed with reference to the simple NaCl H_2O system, and the results are usually expressed as equivalent weight per cent NaCl to take into account the fact that components other than NaCl are also usually present. In practice, because of supercooling effects, the melting temperatures of frozen inclusions, rather than their freezing temperatures, are used to estimate salinities.

Bulk methods A range of instrumental methods with the required sensitivity (parts per million to parts per billion range) are available for the analysis of inclusions in bulk samples (100 mg to 1000 g). Gas chromatography and mass spectrometry are widely used to determine volatile and hydrocarbon contents and the stable isotopic compositions of inclusion fluids released by crushing or heating. The analysis of cations and anions in aqueous leachates from crushed samples has often been used to determine the chemical compositions of aqueous inclusions. However, in common with other bulk methods, there may be significant problems of contamination from the host or admixed minerals, and it is not always possible to obtain a large enough sample with a single or dominant generation of inclusions.

Point methods Various spectroscopic and microbeam methods have been applied to the *in situ* analysis of fluid and melt inclusions. These include electron microprobe analysis, which is particularly suited to the analysis of melt inclusions, and proton-induced X-ray emission (PIXE) and synchrotron X-ray fluorescence (SXRF) spectroscopy, which are particularly suited to the analysis of trace elements and ore metals in melt and brine inclusions. However, the methods which have had the largest impact in recent years are laser Raman microprobe analysis (LRM) and laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS). LRM is non-destructive and is widely used to identify the solid, liquid, and volatile components in unopened inclusions down to a few micrometres in size. It is particularly useful for determining the molar volumes of CO_2, N_2, CH_4, and other Raman-active volatiles in aqueous inclusions and H_2O in silicate melt inclusions. It has also been used to identify a wide variety

Figure 10 Crystallized melt inclusion in flux grown syn thetic ruby. Note the deformed vapour bubble (transparent) and crystalline flux material.

of daughter minerals in inclusions. The LA-ICP-MS method is destructive, but is capable of multielemental and isotopic analysis of individual inclusions down to about 20 μm in size. It is particularly suited to the analysis of trace elements (e.g., ore metals and rare earth elements) in inclusions.

Fluid Inclusion Studies as Tools in Exploration

The distribution and abundance of particular types of fluid inclusion have been used, with variable degrees of success, in the exploration for mineral, petroleum, and geothermal energy resources. Examples include:

1. The logging of the distribution of oil-bearing inclusions in exploration wells, which can be used to infer the migration of oil in particular lithologies and the former presence of oil in an otherwise dry well.
2. The mapping of the overall abundance and distribution of fluid inclusions in crystalline rocks, which often show regional and localized increases (on metre to kilometre scales) near to zones of intense hydrothermal alteration and mineralization.
3. The recognition of 'boiling' assemblages, which may be used to help locate high-grade bonanza gold areas in epithermal gold veins or the depths of boiling zones in modern geothermal systems.

Fluid inclusion studies are useful in assessing the suitability of proposed sites for nuclear waste disposal, because they serve as indicators of the fluid processes that have affected the proposed repository site throughout its geological history.

Finally, in gemmology, the nature and distribution of various solid and fluid inclusions are often used to help distinguish natural from artificial gems or simulates (Figure 10). They can be used to characterize the source or geographical location of a particular gem mineral, and also to assess whether colour has been enhanced through artificial heat treatment.

See Also

Diagenesis, Overview. **Gemstones**. **Gold**. **Igneous Processes**. **Mining Geology:** Hydrothermal Ores; Magmatic Ores. **Petroleum Geology:** The Petroleum System. **Tectonics:** Hydrothermal Activity.

Futher Reading

Andersen T, Frezotti M L, and Burke EAJ (eds.) (2001) A series of review papers in this special issue of *Lithos* dealing with all aspects of fluid inclusion studies. *Lithos* 55: 1 321.

Goldstein RH and Reynolds TJ (1994) Systematics of fluid inclusions in diagenetic minerals. Short Course No. 31, Society of Economic Palaeontologists and Mineralogists, Tulsa, USA.

Gubelin EJ and Koivula JI (1986) *Photoatlas of Inclusions in Gemstones*. Zurich: ABC Edition.

Lowenstern JB (1995) Applications of silicate melt inclu sions to the study of magmatic volatiles. In: Thompson JFH (ed.) *Magmas, Fluids and Ore Deposits. Short Course,* vol. 23. pp. 71 99. Mineralogical Association of Canada (Vancouver, B.C.).

Rankin AH (1989) Fluid inclusions. *Geology Today* 5: 21 24.

Richards JP and Larson PB (eds.) (1998) Techniques in hydrothermal ore deposits: geology. *Reviews in Economic Geology* 10 (contains reviews on Fluid Inclusion Tech niques of Analysis (T. J. Shepherd & A. H. Rankin) and Fluid Inclusion Modeling for Hydrothermal Systems (P. E. Brown)).

Roedder E (1972) The composition of fluid inclusions. *US Geological Survey, Professional Paper* 440 JJ.

Roedder E (1984) Fluid inclusions. *Mineralogical Society of America, Reviews in Mineralogy* 12.

Roedder E and Bodnar RJ (1980) Geologic pressure deter minations from fluid inclusion studies. *Annual Review of Earth and Planetary Sciences* 8: 263 301.

Roedder E and Bodnar RJ (1997) Fluid inclusion studies of hydrothermal ore deposits. In: Barnes HL (ed.) *Geochemistry of Hydrothermal Ore Deposits,* 3rd edn. ch. 13, pp. 657 697. New York: Wiley and Sons.

Shepherd TJ, Rankin AH, and Alderton DHM (1985) *A Practical Guide to Fluid Inclusion Studies*. London: Blackie and Sons.

Sorby HC (1858) On the microscopical structure of crystals, indicating the origin of minerals and rocks. *Quarterly Journal of the Geological Society of London* 14: 453 500.

FORENSIC GEOLOGY

K Pye, Royal Holloway, University of London,
Egham, UK

Introduction

Forensic geology is concerned with the application of geological data and techniques to issues that may come before a court of law. It is closely related to environmental forensics, forensic engineering, and forensic archaeology. Environmental forensics is somewhat broader in scope than forensic geology and involves a wider range of environmental data, knowledge, and expertise. It frequently involves investigations of environmental problems such as water and air pollution. Forensic engineering also overlaps with environmental forensics and is typically concerned with such issues as ground stability, the failure of buildings and other engineering structures, flooding, wind damage, fires, and explosions.

All subdisciplines of the geosciences have potential forensic applications, but sedimentology, mineralogy, petrology, geochemistry, palaeontology, and geophysics have so far made the greatest contributions. Shallow geophysical prospecting methods have been widely used by forensic archaeologists and others to locate and characterize clandestine graves and buried objects such as drugs and weapons (*see* **Engineering Geology:** Geophysics). However, probably the most widely recognized application of forensic geology is the use of geological materials as trace evidence, which can be of value in linking a suspect to a crime scene. In the wider forensic and legal literature, sediment, soil, dust, and rock fragments have often been grouped together under the loose term ‘soil’ evidence.

Some of the earliest users of geological and soil evidence were not, in fact, geologists. As early as 1893 Hans Gross, an Austrian professor of criminology, had pointed out the value of examining ‘dirt’ on a suspect’s shoes as a possible indicator of their movements, and the German chemist Georg Popp is widely credited with being the first to undertake systematic ‘soil’-comparison studies in the early 1900s. Around this time, the English writer Sir Arthur Conan Doyle, literary creator of the investigator Sherlock Holmes, also used soil comparison in one of three cases in which he became a real-life investigator. The work of these men had an important influence on Edmund Locard, initially a student of forensic medicine, who later went on to be Director of the Technical Police Laboratory in Lyons, France. Locard developed the first detailed scientific procedures for the analysis of dust traces and established the famous Locard exchange principle:

> Whenever two objects come into contact, there is always a transfer of material. The methods of detection may not be sensitive enough to demonstrate this, or the decay rate may be so rapid that all evidence of transfer has vanished after a given time. Nonetheless, the transfer has taken place. Locard (1930a Part1), cited by Murray and Tedrow 1992, p7)

In the USA, the Federal Bureau of Investigation Laboratory at Quantico, Virginia, was using soil and mineral analyses extensively in criminal cases as early as 1935 and has maintained a strong interest in this area. Other important forensic work relating to particulates was undertaken in the 1960s and 1970s at McCrone Associates in Chicago, where the first Particle Atlas was developed. Outside the USA, scientists at the UK Home Office Forensic Laboratory at Aldermaston, now closed, made extensive use of soil evidence in the 1960s and 1970s. In Japan, extensive use has been made of geological and soil evidence since the early 1980s by scientists at the National Research Institute of Police Science.

Since the early 1990s the true potential of geological and soil evidence, and that of other related subdisciplines such as forensic botany, forensic entomology, and forensic anthropology, has become much more widely recognized amongst police forces and forensic scientists worldwide. A much broader range of geological techniques and approaches is now being applied in the context of both criminal and civil-law investigations. Many of these techniques also have extensive application in other areas, including the war against terrorism, international drug smuggling, and broader environmental-quality and public-health campaigns. There is considerable current interest in applying forensic-geological techniques and principles to such issues as illegal trading in ivory and rhino horn, archaeological artefacts and works of art, and in traceability studies related to a wide range of foodstuffs and other commodities.

Despite this long-standing interest, there is still only a relatively small specialist published literature in forensic geology, and until recently Murray and Tedrow’s book *Forensic Geology*, originally published in 1975 and reprinted in 1992, was the only one available in the field.

Types of Information Provided by Geological and Soil Evidence

There are many questions and issues to which forensic-geological techniques and information can be applied. These include:

whether an individual, motor vehicle, or other specified item was present at a particular location (e.g. crime scene),
the sequence and possible timing of a visit to that location and possibly others,
the location of buried objects (e.g. bodies, arms caches, drugs),
the source of imported or smuggled items,
the cause of death (especially in cases of possible drowning or suffocation),
the geographical origin of unidentified human remains, and
the length of time a body has been present at a location and the length of the post-mortem interval.

Traces of rock, sediment, soil, and dust can be present on a whole variety of items of interest, but amongst those most frequently submitted to the crime laboratory for examination are footwear, clothing, vehicles, flooring materials, digging implements, washing-machine filters, polythene bags in which items have been stored, firearms, and knives (Figures 1 and 2). Samples associated with the human body are also sometimes subject to examination. These include tapings of the skin, finger-nail scrapings, washings from the hair, nasal passages, trachea, and lungs, the contents of the gastrointestinal tract, and faeces.

With modern analytical techniques, very small traces of mud on an exhibit can provide ample sample material for a whole battery of tests to be undertaken (Figure 3). However, the greater the amount of material that is available for analysis, the wider the options in terms of the analytical techniques that can be employed. If only small amounts of mud or particulates are present, it is usually necessary to preserve as much of the evidence as possible for possible re-examination. Consequently, non-destructive tests, or those that are minimally destructive, are preferable to destructive tests that require a relatively large amount of sample material.

Figure 1 Boots and spade seized from an individual suspected of digging illegal treasure trove from a national heritage site.

Even exhibits that on first examination might seem unpromising from the viewpoint of preservation of forensic evidence, such as a burnt-out car (Figure 4), can in fact be a source of useful geological evidence. Many geological materials are not affected by the temperatures that typically occur in standard vehicle fires, and in many instances particles of gravel and sand can be recovered from suitable parts of such vehicles, such as the footwells and suspension arms.

Comparison of Suspect Samples with Crime-Scene Samples and Other Reference Samples from Known Locations

One of the simplest situations faced by the forensic geologist is to compare mud or soil on a suspect's footwear or clothing with reference samples taken from a known crime scene. In most cases, comparison is also made with reference samples from one or more other locations (e.g. the suspect's home address or place of work) for elimination purposes. Comparisons of the samples should be based on several criteria, chosen from the list shown in Table 1, the choice depending on a number of factors such as the amount and type of material present in the forensic samples, any available background information that might indicate which criteria are likely to provide the best discrimination, and time, cost, and equipment-availability limitations. Clearly, the greater the number of lines of comparison that can be used, the greater is the potential for discrimination between samples. In some circumstances a single method of comparison may be sufficient to screen samples and to identify those that can be eliminated from further consideration. However, great care needs to be taken in selecting the screening method, since some soil and sediment properties, such as colour and particle size, may vary considerably over short distances and with time. As a general rule, at least three independent methods should be used for sample comparison, and considerably more may be required where an apparent similarity between 'suspect' and crime-scene samples is identified.

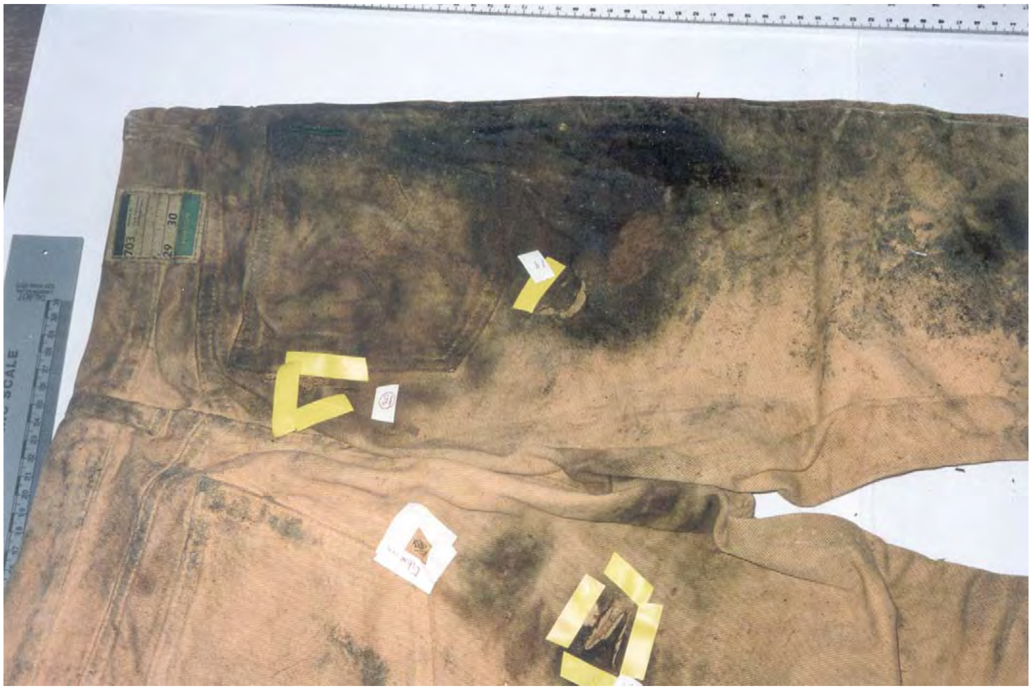


Figure 2 Extensive soil staining on a pair of jeans taken from the victim of a multiple stabbing dumped in woodland.

Figure 3 Mud spots on the jersey of an individual suspected of having buried the body of a murder victim.

Figure 5 shows a soil-covered house brick that was recovered from a hold-all containing dismembered body parts found in a London canal. The identity of the victim was unknown, and forensic examination of the hold-all showed it to be of a type widely available in the UK. Initial questions raised by the police therefore included whether anything could be said about the origin of the brick based on an analysis of the brick itself and also of the adhering soil. The approach in this case was to remove the soil and to analyse both the bulk material and different size fractions using a combination of techniques. These included quantitative colour analysis, particle-size analysis, mineralogical analysis, major and trace element analysis, and pollen analysis. The results indicated that the probable source was a garden adjacent to a domestic property in a part of north London, and subsequent comparison with reference samples taken from the garden of a suspect who came into the enquiry showed a very high degree of similarity.

Figure 6 shows a boot with extensive red soil staining taken from the body of an illegal immigrant who was found in the wheel-well of a Boeing 747 aircraft that landed at London's Heathrow airport. The man carried no formal identification or indication of his origin. Since the plane had made a number of flights to several different countries since the wheel-well had last been subject to detailed examination, there were three possible places where the unidentified individual could have managed to stow away. In order

Figure 4 A burnt out car used by a man who was later convicted of having stabbed his wife to death.

Table 1 Aspects of sediments and soils that have been used for the purposes of forensic comparison

(A) Bulk sample properties

Bulk sample properties	*Main techniques and equipment used*
Rock/sediment/soil texture	X radiography, microtomography, optical and scanning electron microscopy, image analysis
Particle size distribution	Dry and wet sieving, laser granulometry
Particle shape properties	Image analysis
Surface area	Nitrogen gas adsorption
Colour	Colour charts, spectrophotometry
pH	pH electrode, colorimetry
Water soluble cations and anions	Atomic absorption, ion chromatography
Enzymes	Enzymatic extraction
Bacteria	Culture experiments, microscopy
Lipid biomarkers	Gas chromatography mass spectrometry
Carbon, nitrogen, and sulphur contents	Wet chemistry, CHNOS elemental analyzer
Bulk organic matter content	Walkley Black colorimetric method, Fourier transform infrared spectroscopy, pyrolysis gas chromatography mass spectrometry
Polyaromatic hydrocarbons	Gas chromatography mass spectrometry, high pressure liquid chromatography
Calcium carbonate content	Collins calcimeter, Chittick apparatus
Thermoluminescence characteristics	Heat induced photon emission
Fluorescence characteristics	Fluorescence microscopy
Major and trace element composition	X ray fluorescence, inductively coupled plasma spectrometry, neutron activation
Bulk mineralogy	Optical microscopy, point counting, automated scanning electron microscopy, X ray chemical microanalysis, X ray diffraction
Clay mineralogy	X ray diffraction, infrared spectroscopy
Mineral magnetics	Magnetic susceptibility, frequency dependent susceptibility, isothermal remanent magnetization
Stable carbon, nitrogen, and sulphur isotopes	Continuous flow mass spectrometry, laser fluorination mass spectrometry

Table 1 Continued

(B) Individual particle type properties and assemblages

Individual particle type	*Main techniques and equipment used*
Radiogenic isotopes	Thermal ionization mass spectrometry, quadrupole mass spectrometry, laser ablation mass spectrometry
Radioactive isotopes	Alpha counting, beta counting, gamma counting
Opal phytoliths	Optical and scanning electron microscopy, supplemented by energy dispersive X ray chemical analysis
Foraminifera	As above
Coccoliths	
Coralline particles	
Molluscs	
Gastropods	
Ostracods	
Diatoms	
Insect remains	
Pollen and spores	
Plant seeds	
Leaf and stem fragments	
Coal fragments	
Charcoal fragments	
Wood fragments	
Quartz sand grain surface textures	
Gravel surface textures	
Coatings on mineral grains	
Light fraction mineral grains	As above, plus cathodoluminescence microscopy, laser Raman spectroscopy, electron probe analysis, ion probe analysis, laser ablation inductively coupled plasma spectrometry, microspectrophotometry, dating by Ar Ar and U Pb series methods
Heavy minerals	As above
Slag and ash	As above, with exception of dating
Spherules	As above
Brick	
Concrete	
Pottery	
Glass	
Alloys and pure metals	
Fibres	Optical microscopy, fluorescence microscopy, microspectrophotometry, scanning electron microscopy, X ray chemical microanalysis, Fourier transform infrared spectroscopy and microscopy, ultraviolet spectroscopy
Paint	As above
Paper	As above

to identify the location, red soil from the boots was examined using a combination of techniques including quantitative colour analysis, bulk sample and clay-fraction mineralogy by X-ray diffraction, chemical analysis by inductively coupled plasma spectrometry, and pollen analysis. The results clearly indicated a source in a wet tropical country. Comparison was made with control samples taken from adjacent to the airport in the country (Ghana) from which the plane had last departed prior to its arrival at Heathrow, and a very high degree of similarity was obtained in terms of all comparison criteria (Figure 7).

The first stage in any forensic comparison of soil or other geological samples is to determine whether or not a possible 'match' can be excluded. If it can, then no further attention need be given to that sample. If it cannot, then further investigation may be warranted. A conclusion of an exact match can sometimes be drawn with virtual certainty when the samples in question make a physical fit and have the same texture and chemical composition. This may occur, for example, with two halves or several broken pieces of rock or ornamental stone. In other circumstances a physical fit may be observed, for example, between a shoe impression in mud and a shoe seized from a suspect that is of the same size and has the same tread pattern as the shoe that made the impression. However, there may be several thousand such shoes in circulation, and a specific 'match' with an individual shoe often cannot be made. In this instance, analysis of mud adhering to the shoe, if shown to be indistinguishable from that in which

Figure 5 House brick with soil staining recovered from a hold all containing the dismembered remains of a prostitute dumped in a canal.

Figure 6 Soil stained boot from a deceased male found in the undercarriage stowage space of a Boeing 747.

the shoe impression was found, may provide strong supportive evidence that the particular shoe under consideration made the mark.

However, there are many cases where mud-stained footwear is recovered during an investigation but an exact spot at the crime scene where it may have been acquired has not been identified. In such cases, comparisons of the soil on the shoe with several different reference samples from the crime scene, and usually elsewhere, have to be made on the basis of statistical and graphical comparisons, and the results can be interpreted only in probabilistic terms. The degree of similarity between samples can be expressed in several semi-quantitative and quantitative ways, but meaningful statistics about the likelihood of such a degree of apparent 'match' being due to chance are often difficult to provide. This is because the full range of variation that exists in natural soils is impossible to determine and can be estimated only on the basis of sampling. The availability of database information relating to suitable comparison samples varies greatly from one region to another, and there may be a total absence of pre-existing information in some parts of the world. In such cases, it is necessary to undertake a suitable background investigation, involving collection and analysis of a sufficiently large number of reference samples, in order to provide adequate contextual information for interpretation.

Where no physical fit has been identified, the nearest thing to a definitive connection between two questioned samples is usually provided by the identification of one, or more commonly several, highly unusual (or 'exotic') particle types in both samples. These may be naturally occurring particles or they may be of human or animal origin. They need not be considered 'unique' in themselves, but should be sufficiently rare, either alone or in combination with other unusual particles in the same sample, that the chance of them occurring in any two samples under investigation is extremely low. Examples of two particles that fall into this category are shown in Figures 8 and 9. Waste-dumps, industrial premises, and roadside verges are examples of locations that often contain mixtures of particles that have a more restricted distribution than natural soils. The assemblages of particles present in such locations often show considerable local variation, and it may be possible to limit a potential source area to just a few square metres.

Persistence of Geological Evidence

Geological evidence may persist for a considerable period of time after it has been picked up from the source location. For example, gravel, sand, or mud that enters

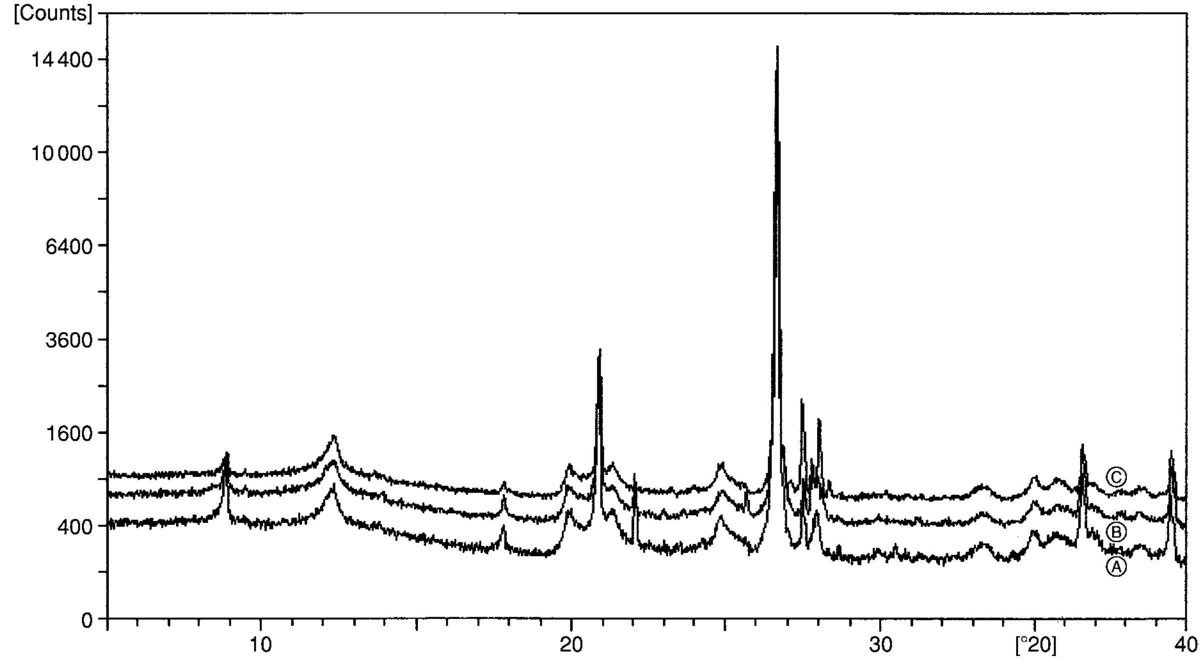


Figure 7 Comparison of X ray powder diffractograms for the <150 μm fraction of soil samples from the right and left boots (A and B) of a deceased stowaway found in the undercarriage stowage space of a Boeing 747 jet after landing in London, compared with a control sample from Accra airport, Ghana (C).

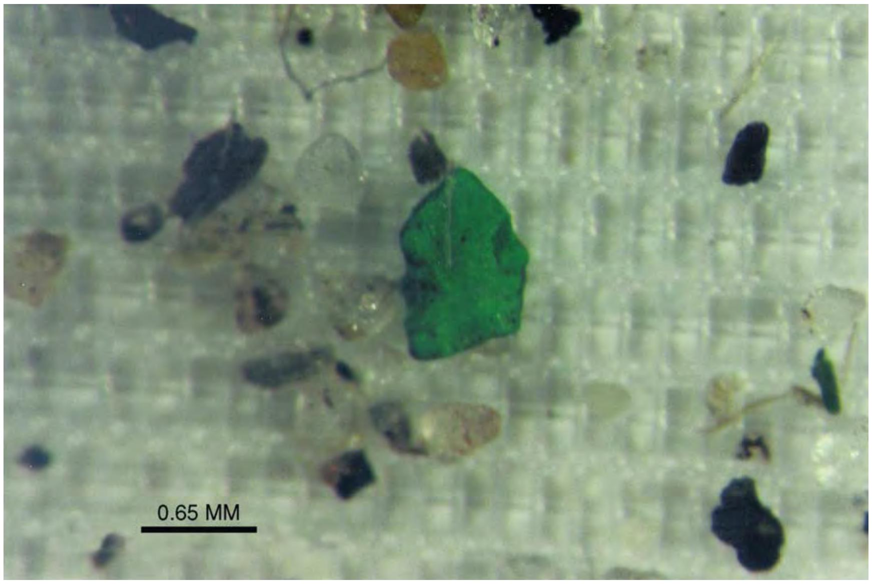


Figure 8 Example of an 'exotic' synthetic chromium rich particle, identified by optical microscopy, which was interpreted to be of 'industrial' origin.

the interior of a car on footwear, clothing, or other items (such as a spade) will stay there, subject to some loss due to outward transfer on the footwear or clothing of later occupants, until such time as the vehicle is thoroughly cleaned. Even after cleaning, traces may remain in certain hard-to-access locations. Whereas some forms of forensic botanical evidence degenerate as a result of oxidation or fungal decomposition, most inorganic sediment particles are very resistant and may be immune to changes over time. Consequently, they provide useful clues years or even decades after a crime was committed, providing that exhibits have been retained and suitably stored to prevent environmental and cross-exhibit contamination.

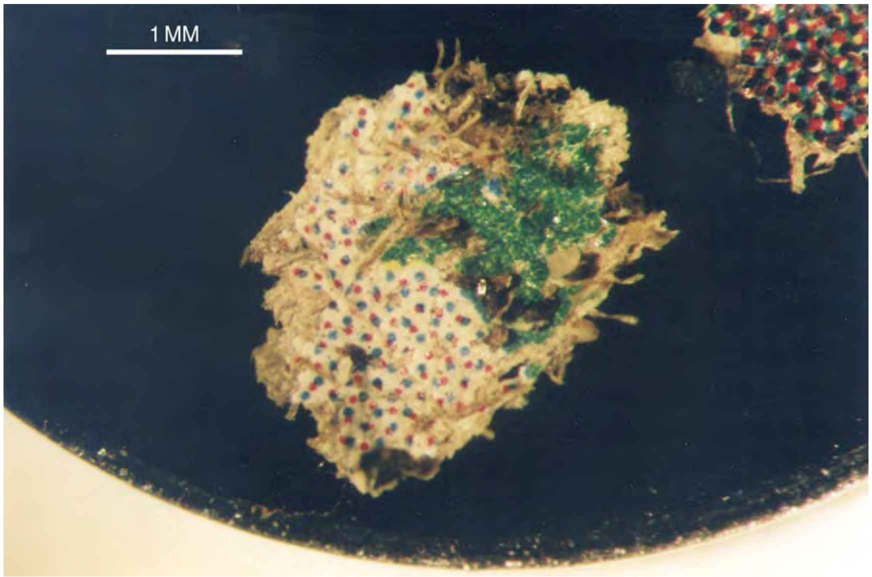


Figure 9 A further example of 'unusual' particles found in soil on the boots of a suspect who was later convicted of murder. The particles are from the cover of a particular issue of a glossy magazine, copies of which had been dumped at the murder scene.

Modification of Primary-Transfer Soil Evidence

It should always be borne in mind that material initially picked up from a location, a process referred to as primary transfer, may subsequently be modified as additional particles are picked up from other locations or some of the primary particles are lost during subsequent movement. During primary transfer, the material transferred may not exactly reflect the nature of the material at the source point, depending on the nature of the material and the nature of the contact involved. For example, if a person sits or lies on wet ground there is frequently a tendency for the finer particles to be selectively transferred and retained on the clothing. In other circumstances, only a certain size range of coarse particles may be retained, for example coarser particles trapped within the detail of footwear sole treads or gravel particles trapped in tyre treads. For this reason, it is important that sample comparisons are undertaken on narrowly defined size fractions as well as on bulk samples.

Secondary Transfer

The possibility of secondary transfer of soil and other geological evidence should always be borne in mind when exhibits are examined. For example, consider a case where Person A walks across a muddy car park and gets into the passenger seat of a vehicle, which is driven by another person (B) to another location, where A gets out. During this process, mud from the car park is transferred via the footwear of Person A to the front passenger footwell of the vehicle. The driver of the car (B) then drives to a third location and picks up another person, C, who also sits in the front passenger seat. The shoes of Person C come into contact with mud in the front passenger footwell deposited by Person A, and this is then transferred out of the car, via Person C's footwear, onto the hallway carpet at Person C's home address. If it subsequently emerges that a crime has been committed in the car park, and Person C becomes a suspect, simple comparison of mud on Person C's footwear and hall carpet with control samples from the car park might lead to the spurious suggestion that Person C had been present at the crime scene. For this reason, great care needs to be taken by the forensic geologist to document the amount, distribution, layering, and nature of any mud or similar evidence present on items, including footwear, submitted for analysis. In this connection police and scenes-of-crime examiners have an important responsibility to provide the forensic geologist with necessary background information and to ensure that comparison samples are taken from all locations and exhibits of possible relevance for comparative examination.

Location of Crime Scenes, Buried Bodies, and Weapons and Drugs Caches

A frequent problem that the forensic geologist is asked to address concerns the identification of the location where a crime has taken place, perhaps involving the disposal of a body, or where weapons, money, and drugs have been stored or buried. In many instances one or more suspects are identified and their vehicles seized for possible identification of evidence that might indicate the location of the deposition or burial site. Both the inside and outside of the vehicle are then examined in detail, in parallel with searches for blood, other forms of DNA, fibres, hairs, and fingerprints. Numerous samples are normally taken from the footwells, wheel arches, mud flaps, and other parts of the bodywork and chassis to build up as detailed a picture as possible of the vehicle's recent movements. Similar examination and sampling is often undertaken on associated items such as petrol cans, car jacks, mats, spades, footwear, and clothing belonging to the suspect. The full spectrum of the samples' physical, chemical, and biological make-up is examined, the objective being to create an environmental profile of the samples, which may assist the direction of further police enquiries.

In this type of work, individual particle types, which may be either inorganic or biological, can be highly diagnostic. Particular pollen types may indicate specific ecological habitats, such as moorland, coniferous forest, deciduous broad-leaf woodland, or salt-marsh. Diatom assemblages may indicate saline, brackish, or freshwater environments. Highly diverse assemblages of particles of industrial or human origin may indicate waste dumps or industrial estates. Natural rock particles, such as chalk, coal, slate, and basalt, or mineral assemblages may indicate particular areas within specific geological outcrops. Even the relative abundance of different morphological types and surface textures of common minerals such as quartz may suggest specific localized areas with a particular surface geology and soil type (Figure 10). Viewed in polished section under an optical or scanning electron microscope, rock types that have very similar chemical compositions can be seen to have quite different depositional and diagenetic textures, which may be specific to particular lithostratigraphic units only a few tens of millimetres in thickness (Figure 11). Such precise identification is often aided by the analysis of microfossil assemblages, including foraminifera and shell debris.

Figure 12 shows a number of gravel- and coarse-sand-sized particles that were recovered at post mortem from the trachea and bronchi of a murder victim who had been shot in the head and whose burning body was found dumped on a farm track outside Edinburgh. The lithological assemblage of the gravel clasts, which were characterized by numerous *in situ* freshwater diatoms, indicated that they had originated from a river or river-marginal setting in the Airdrie area near Glasgow. The large size of the particles and the depth of penetration into the lungs meant that

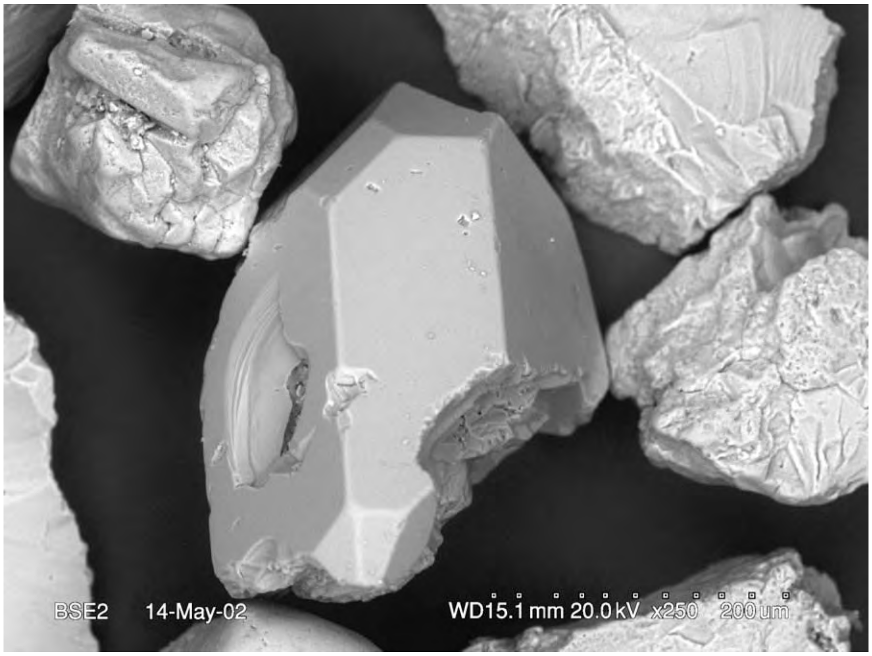


Figure 10 Scanning electron microscope image of a very fresh, unabraded, and uncorroded euhedral quartz grain. Such particles are normally found only very close to the rock source.

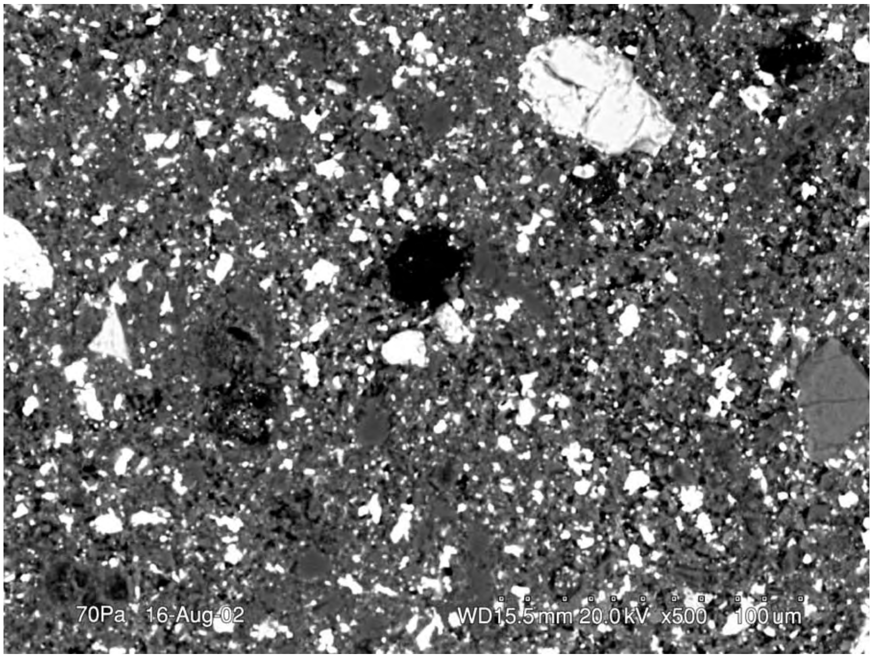


Figure 11 Scanning electron microscope image of a polished section of a piece of Chalk recovered from the suspension of a car owned by an individual who was later convicted of murdering two young girls and using the car to deposit their bodies.

they could have been sucked in only while the victim was alive, possibly being held face down with his head in water just prior to being shot.

Studies of Human Remains

Unidentified human remains are not infrequently found dumped by the roadside or washed up on the coast or in rivers and lakes. The remains may consist of whole bodies, with or without clothing, or of torsos or even isolated limbs. If identification proves impossible using dental records, fingerprints, or DNA, alternative means must be used to determine the geographical origin and identity of the victim. There are several ways in which the forensic geologist may contribute. First, studies of sand, mud, and dust particles present on clothing or on the outside of the body may indicate the area from which the body came (Figure 13). Particles that are exotic to the site where the body was discovered, such as coralline algae on a body found in the UK, would clearly indicate a recent tropical or subtropical association. Studies of pollen may provide further information about botanical exposure.

Studies of particles within the body may provide indications of environmental exposure in the hours or days immediately prior to death. Washings of the nasal passages, lungs, and hair may be useful in this regard, as may fingernail scrapings and particulates contained within the gastrointestinal tract and faeces. The absence of exotic pollen and inorganic particles within the gastrointestinal tract of a known recent immigrant may provide significant evidence that he or she had been present in the country where they were found for at least several days.

In cases where the cause of death cannot be determined with certainty by conventional post-mortem examination, examination of particulates in the lungs and other body tissues may assist in determining whether death was due to true drowning or some other cause prior to, or during, immersion. Analysis of diatoms present in the lungs, liver, spleen, blood, and bone marrow has for many years been undertaken as a confirmatory test in possible drowning cases (Figure 14). However, the diatom test has been controversial because numerous cases of false negative and false positive results have been documented. However, if the test is combined with studies of other particles that are known to be associated with a particular water body or type, its reliability is potentially much improved. As previously noted, quite large quantities of particles up to fine gravel in size can be aspirated into the trachea and lungs through the open mouth in the presence of water. Such particles may also be swallowed and carried into the stomach of a drowning person. Water is not always required, however: death may quite frequently be due

Figure 12 Group of 23 gravel sized particles recovered from the trachea and upper bronchii of a murder victim who had been shot in the head and his body set on fire, apparently after having his head immersed in a freshwater stream.

to suffocation when dry silt, sand, soil, or even fine gravel is forcibly or otherwise involuntarily inhaled. Careful analysis of diatoms and other environmentally sensitive biotic indicators associated with the inorganic particles is a useful means of determining whether death occurred while the face was submerged in water. Similar environmental discrimination can be achieved using sediment and soil sometimes found in other body orifices of deceased persons, including the anus and vagina.

Analysis of the trace-element and isotopic compositions of nail, hair, bones, and teeth can provide information about environmental exposure and diet over time periods ranging from a few days to an entire lifespan. Stable isotopes of carbon, oxygen, hydrogen, and nitrogen provide information about diet (including drinking water) and hence about climate; radiogenic isotopes of strontium, lead, and neodymium reflect aspects of diet, geological source terrain, and atmospheric exposure. In general, hair and nails provide information about time periods of a few days to a few months, bones provide information relating to the last few years of life, and tooth enamel provides information relating to the first few years of life, from *in vivo* until approximately age fifteen. By comparing data for teeth and parts of several different bones, information can be gained about human and animal migration during life.

Analysis of radioactive-isotope ratios in human remains may help to determine the post-mortem interval (i.e. the time since death). In the case of bodies that are between about a year and several decades old, lead and polonium isotopes provide the most useful information, while for longer time periods other isotopes, such as caesium and radiocarbon, can help to distinguish modern from ‘archaeological’ bone.

Increasingly, the work of forensic geologists is being used not only for intelligence purposes in criminal investigations but also as expert-witness evidence presented in court. Geological evidence has made a

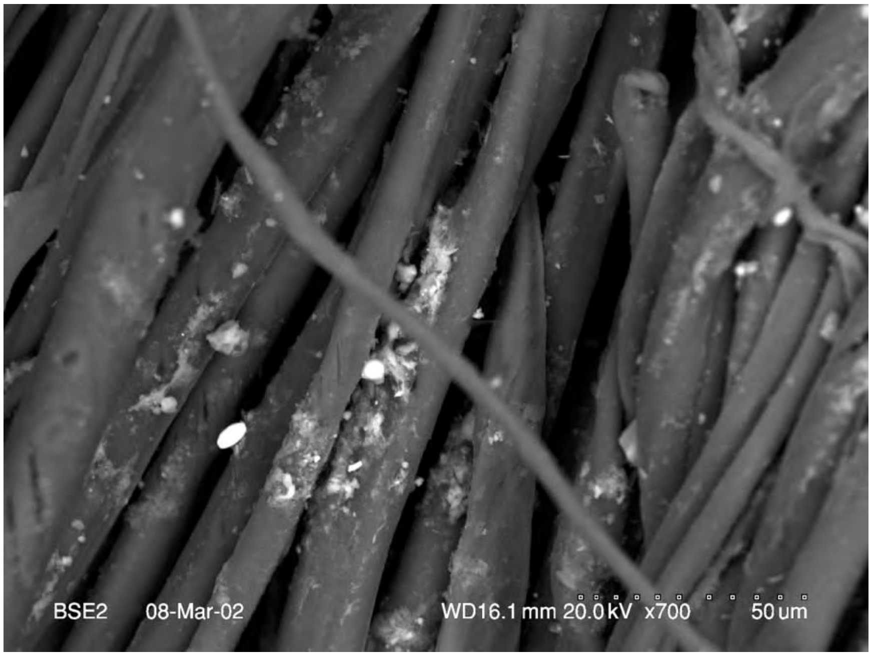


Figure 13 Scanning electron microscope image showing sediment particles adhering to fibres in a pair of orange shorts that had been placed on the torso of a murdered child found in the River Thames.

Figure 14 Scanning electron microscope image of a pennate diatom found in the lung tissue of a suspected drowning victim.

significant contribution in recent years to a number of high-profile trials involving murder, terrorism, international drug smuggling, and people trafficking. Other areas of frequent geological expert-witness testimony include various aspects of environmental forensics (e.g. contamination studies), engineering-geology failures, and traceability studies relating to food-stuffs and raw materials.

See Also

Analytical Methods: Geochemical Analysis (Including X-Ray). **Engineering Geology:** Geophysics. **Geoarchaeology**. **Microfossils:** Palynology. **Soils:** Modern.

Further Reading

Brown A (2000) Going to ground. *Police Review* 4 February: 18 20.

Croft DJ and Pye K (2004) Multi technique comparison of source and primary transfer soil samples: an experi mental investigation. *Science and Justice* 44: 21 28.

Demmelmeyer H and Adam J (1995) Forensic investigation of soil and vegetable materials. *Forensic Science Review* 7: 119 142.

Foster IDL (ed.) (2000) *Tracers in Geomorphology.* Chichester: Wiley.

Hall DW (1997) Forensic botany. In: Haglund WD and Sorg MH (eds.) *Forensic Taphonomy,* pp. 353 363. Boca Raton: CRC Press.

Hunter J, Roberts C, and Martin A (1997) *Studies in Crime: An Introduction to Forensic Archaeology.* London: Routledge.

Kubic T and Petraco N (2002) Microanalysis and examin ation of trace evidence. In: James SH and Nordby JJ (eds.) *Forensic Science. An Introduction to Scientific and Inves tigative Techniques,* pp. 251 296. CRC Press: Boca Raton.

Locard E (1930) The analysis of dust traces. Part I. *American Journal of Police Science* 1930: 276 278.

Locard E (1930) The analysis of dust traces. Part II. *American Journal of Police Science* 1930: 401 418.

Locard E (1930) The analysis of dust traces. Part III. *American Journal of Police Science* 1930: 496 514.

Marumo Y, Sugita R, and Seta S (1999) Soil as evidence in crime investigation. *International Criminal Police Review* 474 475: 75 84.

McCrone WC, Delly JG, and Palenik S (1973) *The Particle Atlas. Volumes I to VI. An Enclopedia of Techniques for Small Particle Identification.* Chicago: Ann Arbor Science.

Morrison RD (2000) *Environmental Forensics. Principles and Applications.* Boca Raton: CRC Press.

Munroe R (1995) Forensic geology. *Royal Canadian Mounted Police Gazette* 57: 10 17.

Murphy BL and Morrison RD (eds.) (2002) *Introduction to Environmental Forensics.* San Diego: Academic Press.

Murray RC (2000) Devil in the details: the science of forensic geology. *Geotimes* 45: 14 17.

Murray R and Tedrow JFC (1992) *Forensic Geology,* 2nd edn. Englewood Cliffs, NJ: Prentice Hall.

Pollanen MS (1998) *Forensic Diatomology and Drowning.* Amsterdam: Elsevier.

Pye K and Blott SJ (2004) Particle size analysis of sediments, soils and related particulate materials for forensic purposes using laser granulometry. *Forensic Science International* 144: 19 27.

Pye K and Croft DJ (eds.) (2004) *Forensic Geoscience Principle, Techniques and Applications.* Special Publica tion 232. London: Geological Society.

Sabine PA (1991) Geologists at war: a forensic investigation in the field of war time diplomacy. *Proceedings of the Geologists Association* 101: 139 143.

Shuirman G and Slosson JE (1992) *Forensic Engineering. Environmental Case Histories for Civil Engineers and Geologists.* San Diego: Academic Press.

FOSSIL INVERTEBRATES

Contents

Arthropods
Trilobites
Insects
Brachiopods
Bryozoans
Corals and Other Cnidaria
Echinoderms (Other Than Echinoids)
Crinoids
Echinoids
Graptolites
Molluscs Overview
Bivalves
Gastropods
Cephalopods (Other Than Ammonites)
Ammonites
Porifera

Arthropods

L I Anderson, National Museums of Scotland, Edinburgh, UK

Introduction

Arthropods are the most diverse group of fossil and living invertebrate animals. They have colonized land, sea, and air due to a relatively plastic and adaptable body plan organization. Many different phyla of metazoans, including the arthropods, show repetition of structure along their long axis. This is generally termed metamerism.

Arthropod bodies are characteristically segmented, and the different major groups are defined by various patterns of tagmosis (segment fusion) to form heads, bodies, and tails. The phylum derives its name from the characteristic jointed appendages found in all members (Greek: *αρθροσ* '*arthros*', jointed; *ποδοσ* '*podos*', foot). These jointed appendages, which consist of individual elements called podomeres, take the form of walking legs, gill branches, or antennae. Broad divisions can be made within the arthropods depending on the exact nature of the leg appendages. Uniramians, or arthropods with single-branched appendages, retain only a walking or swimming leg. Biramians have an appendage which consists of two separate elements: a leg branch for movement and a gill branch used in respiration.

Arthropods all possess a rigid exoskeleton which provides support and protection, and acts as a buffer between the external environment and internal body processes. The exoskeleton of arthropods is composed of a carbohydrate polymer called chitin. In some arthropod groups, the chitin superstructure is reinforced with mineral salts, such as calcium carbonate or calcium phosphate. Arthropods with this biomineralization show a higher preservation potential than non-mineralized forms, and consequently have a more complete fossil record through time. For non-mineralized forms, we have to look to sites of exceptional preservation (Fossil Konservat-Lagerstätten) to trace their evolutionary history and discover aspects of their palaeobiology. Localities, such as the Middle Cambrian Burgess Shale, the Ordovician Soom Shale, the Devonian Rhynie chert, the Carboniferous Mazon Creek nodule hosted biota, the Jurassic Solnhofen Lithographic limestone, and Baltic, Dominican, and Mexican ambers, all provide windows onto the arthropod faunas at stratigraphical points in geological time in which preservational failure would otherwise obscure our view of evolution of the various groups (*see* **Lagerstätten**). Furthermore, there is a gradation of fidelity of preservation even in these special examples. To this end, it is entirely possible to study the exquisite three-dimensional outer surfaces of Tertiary flies in amber (Figure 1) or the internalized respiratory systems of Devonian arachnids.

A consequence of the possession of a rigid exoskeleton composed of cuticle is that, in order for the arthropod to increase in body size over time, the rigid cuticle has to be shed or moulted and a new cuticle grown below the surface to replace it. Arthropods differ from many other animal groups in that they go through the process of ecdysis (literally shedding the exoskeleton as a moult) in order to attain greater body size. Ecdysis is facilitated by the presence of sutures in the exoskeleton, lines of weakness which can be broken through so that the arthropod can struggle free of the old skin. The ontogeny of ecdysis allows segment addition and tagmosis to be identified from the first instars right up to the adult forms. Unfortunately, for arthropods, moulting of the old skin and the time taken to harden the new exoskeleton leaves them vulnerable to predator attack or adverse environmental conditions. For this reason, many arthropods (both in aquatic and terrestrial environments) seek out a 'moulting refugia' in which the process of ecdysis and subsequent hardening of the new exoskeleton can take place in relative safety. Where mortality occurs during this post-moult stage, the cuticle of the arthropod fossil formed can display wrinkles, creases, or even tears in the structure.

Figure 1 A fungus gnat preserved in Baltic amber from the Upper Oligocene Lower Eocene of Palanga, Lithuania (×55). © The Trustees of the National Museums of Scotland, Edinburgh, UK (NMS G.2002.6.17).

The cuticle itself takes the form of hardened sclerites, one per body segment. Dorsal sclerites are termed tergites, ventral sclerites are termed sternites. Where the tergites and sternites are fused together, they form a simple ring structure enclosing the arthropod soft tissues. Lying between these discrete segments are areas of connective and unmineralized cuticle. The presence of this less robust construction material explains why, upon death, well-mineralized arthropods, such as trilobites, can disarticulate to a greater or lesser degree into their constituent parts (cephalon, thorax, thoracic segments, pygidium, etc.). The leg appendages in arthropods are similarly constructed with cylindrical reinforced cuticular podomeres with intervening, less well-sclerotized arthrodial membranes connecting them. Muscle attachments required to move the individual leg podomeres are attached to small boss-like projections from the inside of the exoskeleton, called apodemes.

The cuticle of arthropods explains in part their success in the colonization of such a wide variety of different environments with widely varying challenges. The cuticle provides a physical barrier between the internal organs and the external environment (which helps to reduce desiccation through water loss in terrestrial forms) and a chemical barrier in aquatic conditions (where hypo- or hypersalinity could upset the osmotic balance, leading to dehydration of the organism and ultimately death). As well as a physical barrier, waxes produced by the cuticle layer help to waterproof this material. Some other more unusual properties of arthropod cuticle have been identified recently, including the fluorescence under ultraviolet (UV) light of 'hyaline' cuticle within the exoskeleton of scorpions. It happens that this 'hyaline' layer is actually very resistant to bacterial degradation and is therefore often the main remaining constituent material of fossil scorpions. For example, the Early Carboniferous East Kirkton biota of West Lothian, Scotland, contained a number of different scorpion taxa which were recovered by hydrofluoric acid maceration of the enclosing limestone matrix.

The level of morphological detail made available through this technique was as good as studying extant species.

Classification Schemes

There still remains scope for significant disagreement in the higher level taxonomy of this group of animals. Historically, the main dichotomy relates to whether the phylum Arthropoda is a monophyletic or polyphyletic group of organisms. Some variation occurs in the higher level classification of this group, with some authors recognizing Arthropoda at phylum level and the constituent members variously as subphyla or orders. A monophyletic grouping for the arthropods would imply that all of the lower level groupings originated from a single common ancestor. This was the largely historic viewpoint based on the striking common appearance of the various groups. In order to counter what seems to be the logical conclusion based on many shared characters in common, the argument for polyphyly has to promote plausible explanations for the observed similarities between groups as well as providing proof of separate points of origin. The fossil record may provide just such evidence, allowing the evolution of the various arthropods to be considered with the benefit of deep time hindsight. However, since the work of Sidnie Manton, who suggested that three major groups, the Crustacea, the Chelicerata, and the Uniramia, would more comfortably be accommodated as phyla of their own, the consensus view has swung back to the likelihood of the monophyletic status of the group being the correct conclusion. This has been due in part to the advent of molecular sequencing techniques. Novel (non-morphological) hypotheses linking crustaceans and hexapods were primarily based on the analysis of single or a handful of genes. However, as the science advances and the techniques of extraction improve, the case for monophyly appears to be more solid. Indeed, one study not only supports the crustacean–hexapod clade, but also recognizes pycnogonids as the sister group to other euarthropods. The historical convention has been to subdivide the Arthropoda into the following equal-biased groupings:

- Trilobita (containing all trilobites, but excluding naroiid arthropods);
- Chelicerata (Arachnida, Scorpionida, Xiphosura, Eurypterida, Chasmataspida);
- Crustacea;
- Uniramia (including both insects and myriapods).

However, as more and more examples of basal arthropods and ‘oddball’ taxa are discovered, primarily from Fossil Konservat-Lagerstätten, other less conventional higher taxon groupings can be identified. Some examples include:

- Atelocerata (Diplopoda, Chilopoda, Pauropoda, Symphyla, Hexapoda);
- Arachnomorpha/Arachnata (Trilobites, Chelicerates, Trilobitomorpha).

Arachnomorpha is a clade which basically consists of all taxa more closely related to Chelicerates than crustaceans. Panarthropoda is again a clade, but this time containing all true arthropods and their soft-bodied relatives, such as onychophorans and tardigrades.

To this end, the term ‘arthropod’ should probably be used to represent a grade of organization in which a soft-bodied ancestor (presumably a segmented worm) developed a toughened cuticle. Gene sequencing and molecular markers may provide fascinating new insights into the interrelationships of the various high-level arthropod groups.

What the Fossil Record Says

Tracing arthropods back through geological time indicates that the divisions between the major groups extend almost as far back as the fossil record of Metazoa. Convergence of form and channelling by functional morphology of multiple arthropodization events through forming a rigid cuticle could have led to polyphyly; however, the more complete our record of the group becomes, the less likely this seems.

A Brief History

The first arthropods appear in the fossil record of life on Earth in the Lower Cambrian some 540 Ma. The arthropod faunas of the Chengjiang biota of China demonstrate that, even at this early stage, a diverse range of arthropod body plans had evolved. By Middle Cambrian times, rapid and explosive radiation and evolution of arthropods led to the development of many different body plans, including the majority of those still extant. This ‘Cambrian Explosion’ in the number and diversity of arthropodan groundplans was one of the first major radiations of the group and took place in the marine environment. By the Upper Cambrian, trilobites had diversified markedly and this carried through into Ordovician times.

Mid-Silurian and Lower Devonian rocks yield evidence of the next major step in the evolution of the arthropods: the colonization of the terrestrial environment, a process termed terrestrialization. Different terrestrial arthropod groups probably took different terrestrialization pathways. Although this is difficult to determine from the fossil forms, extant members of the ancient lineages can lend clues as to what these pathways might have been. For example, osmotic concentrations of body fluids in myriapods suggest that this group terrestrialized straight from the sea without an intermediate step in freshwater.

One particular locality which has yielded a disproportionately large amount of data on early terrestrial ecosystems is the Devonian Rhynie Chert Lagerstätte in Aberdeenshire, Scotland. Here, the silica-mineralizing action of hot spring and geyser fluids captured, in three-dimensional detail, early terrestrial arthropods, such as arachnids, insects, and centipedes, as well as aquatic forms, such as freshwater crustaceans and euthycarcinoids. The remarkable morphological stasis in early terrestrial arthropods, in relation to extant forms, suggests that arthropod-dominated terrestrial ecosystems stabilized early on in the Earth's history and that soil and litter habitats and inhabitants have changed very little ever since.

One of the novel innovations of arthropods was the evolution of muscle-powered flight in the insects (*see* **Fossil Invertebrates:** Insects). The earliest insects or apterygotes (insects without wings) first appear in sediments of Devonian age in the form of collembolans and thysanurans. In earliest Carboniferous times, the major expansion and evolution of land-based flora provided new ecological niches available to this highly adaptable group. The advent of insects and the development of flight led to the colonization of an entirely new ecospace. Interestingly, the earliest non-scorpion arachnids appear to have relied solely on poisoned fangs to capture and subdue prey items. Orb web-weaving using spider silk could only have developed in response to the presence of aerial insects.

By Lower Carboniferous times, the first winged insects had developed and an explosion in the diversity of insects occurred due to coal swamp forest ecosystems. Preservation and discovery are aided by the presence of siderite concretions in the roof shales to many coal seams which, in turn, are commercially exploitable. During the Middle Carboniferous, insects became diverse and widespread for the first time in their fossil record. This trend continued through the Carboniferous and into the Permian period, whereupon an explosion in the total diversity of fossil insects took place. The depositional environments in which these fossils are found do not differ significantly, suggesting a real increase in diversity. During the Jurassic and Cretaceous, favourable shallow (epieric) sea conditions saw the emergence and dominance of decapod crustaceans as a major player in nearshore marine ecosystems. This development within the Crustacea continues to the present day, with arthropods still developing new and successive waves of colonization, particularly in terms of the terrestrialization of some decapod crustaceans, such as the land crabs.

Respiration

From the same basic body plans, different solutions to the challenges of respiration in aquatic and terrestrial environments have evolved. In the smallest of microarthropods, cutaneous respiration, by which the skin acts as a gas exchange surface, takes place. The ratio of surface area to internal volume limits the maximum size of arthropod which can utilise this method. In larger aquatic arthropods, respiratory gas exchange is through structures termed 'book gills'. These thin lamellate cuticular structures are bathed in oxygen-containing water, either wholly external to the body, as in the gill branches of trilobites, or in a specialized cavity, as in the horseshoe crabs (Figure 2). Terrestrial arthropods show adaptation to the particular challenges of living in an environment in which the constant danger of desiccation means that

Figure 2 The xiphosuran, *Rolfeia* sp., from the Lower Carboniferous of Mumbie Quarry, Glencartholm, Dumfries and Galloway, Scotland (×1.1). © The Trustees of the National Museums of Scotland, Edinburgh, UK (NMS G.1998.35.3).

book gills would cease to function. In terrestrial arthropods, respiration is via book lungs (analogous and probably derived in part from book gills) or by a tracheal system. The lamellate structure of book lungs requires additional supporting struts so that they do not collapse in air. Furthermore, the book lungs tend to be positioned internally within the body of the animal as a means of further reducing water loss through desiccation. The tracheal system is a passive method of diffusing oxygen into the body of the arthropod via a branching system of tubes opening on the surface of the exoskeleton at a tracheole. The insects show the greatest development of the tracheal system, related no doubt to the oxygen demands of muscle-powered flight. Other uniramians, such as millipedes, also respire in a similar manner through the presence of a series of paired spiracles along the length of the body. The earliest example of this form of respiration is recorded in the mid-Silurian millipede, *Pneumodesmus newmani*, from Stonehaven, Scotland.

The diffusive process involved in tracheal respiration places an upper limit on the maximum size of terrestrial arthropods which utilize this method, unless a means of secondary 'pumping' derived from muscle contraction can improve efficiency. Book gills, on the other hand, operate in water and are not subject to the same constraints on size which are faced by a land-based book lung system. In some respects, this may help to explain why the largest arthropod fossils found are those of aquatic organisms, such as the extinct eurypterids. Indeed, some very active arachnids, such as solfugids (camel spiders), possess both book lung and tracheal systems, allowing them to meet the oxygen demands of their highly active and voracious life style. The antiquity of the development of tracheal respiratory systems in terrestrial arthropods stretches back as far as the earliest opilionid (harvestman spider) found in the Early Devonian Rhynie Chert, Aberdeenshire, Scotland.

Reproduction

Aquatic and terrestrial environments present different challenges to reproductive strategies in arthropods. Terrestrial arachnids tackled the problem of the desiccation of reproductive material by producing a spermatophore, which is either passed from male to female using specially adapted leg appendages or pedipalps (spiders), or left on the ground attached to a stalk (scorpions).

Internal fertilization is practised by a few groups, including the opilionids (harvestman spiders), and the preservation of internalized organs in arachnids, such as the Rhynie Chert opilionids, in which penis and ovipositor structures are known in separate specimens, demonstrates morphological conservatism.

Direct morphological evidence of sexual dimorphism, or reproductive structures, has long been identified in the chelicerate arthropod group, the eurypterids. In these arthropods, Type A and Type B appendages have been identified on the ventral surfaces. A recent study went further and demonstrated that the longer Type A appendage was probably female, whilst the Type B appendage belonged to the male. An elaborate system of flexure along predefined sutures in the cuticle allowed the female to retrieve an unstalked spermatophore from the substrate prior to storage in the spermathecae.

Feeding

An effective demonstration of how the understanding of the form and feeding function of extant arthropods could be 'reverse engineered' to understand those of the fossil forms was provided by one particular example: the Lipostracan crustacean, *Lepidocaris*, known from three-dimensional fossils in the Early Devonian Rhynie Chert.

The most recent synthesis of feeding in trilobites (*see* **Fossil Invertebrates:** Trilobites) suggested that, as a group, the various trilobites employed different strategies. Indeed, it was argued that much of the variety of form seen in the exoskeletal structure of trilobites was probably driven in response to specific feeding modes (Figure 3). In the main, the structures most responsive to this driving factor were the cephalon (or head region) and the hypostome, a small plate lying anterior to the mouth cavity and variously free or fixed depending on the feeding strategy. The primitive mode of feeding in trilobites was identified as predatory or scavenging, with cephalon and accompanying hypostome morphology (coupled with overall body shape) used as a guide to indicate possible feeding strategies within different trilobite groups. However, this may even be a relict from the arthropods ancestral to the trilobites.

On the other hand, detritivorous trilobites possessed a hypostome detached from the doublure of the cephalon. The glabella (the axially aligned raised area atop the cephalon), where the stomach was thought to reside, would be relatively small in these forms, reflecting the small quantities of food processed at any one time. Filter-feeding trilobites exhibit a different overall morphology again, and often possess a highly vaulted cephalon which would have allowed the leg appendages to stir into suspension food-bearing sediment particles. The demonstration of an elegant scheme such as this in a reasonably

Figure 3 The trilobite, *Phacops* sp., from the Devonian of Morocco, North Africa (×1.7). © The Trustees of the National Museums of Scotland, Edinburgh, UK (NMS G.1987.20.4).

well-known fossil group should be applied to the less well-known Palaeozoic arthropods.

Relevance

Arthropod fossils are used by palaeontologists and geologists in a number of different ways for different reasons. In rocks of Palaeozoic age, the trilobites have long been used as biostratigraphical marker fossils. They find application in this particular field of study as they are readily fossilized (carbonate-impregnated exoskeleton), numerically abundant (each individual moults many times over a lifetime, 'amplifying' their number of potential fossils), and show sufficiently rapid speciation events to make them useful as index fossils. In Early Palaeozoic rocks, nektonic species are useful in palaeogeographical reconstructions of former oceans, seas, and, by default, landmasses; the nektonic life style has led to a widespread occurrence of the resultant fossil remains, ideal for broad stratigraphical correlations. Another use for the numerically abundant trilobites in a stratigraphical framework is the evaluation of evolutionary rates within well-defined lineages. Such frameworks allow trilobites to be utilized in studies of evolutionary changes within a species, up until the point at which a new species evolves, a field of study known as micro-evolution.

Finally, many micro-palaeontological samples, such as fossils of ostracodes (*see* **Microfossils:** Ostracoda), are similarly numerically abundant and readily preserved, and have found application in palaeontological and archaeological studies of past environments due to their environmental sensitivity. The calcium carbonate-impregnated shells of these small bivalve-carapaced crustaceans are a common constituent in rocks of all ages from the Cambrian onwards.

Visual Systems

Two main types of eyes are recognized in the arthropods: compound (multifaceted) and simple ocellar eyes (Figure 4). Compound eyes probably developed separately in a number of different arthropod groups, giving rise to analogous if not homologous structure. Many arthropods possess both compound and simple eyes, as demonstrated by the extant chelicerate horseshoe crab, *Limulus*. The ocellar eyes are positioned on the top of the dorsal surface of the carapace and are primarily used to detect light levels, indicating whether or not the top of the carapace is covered by sand; this information is of particular use to burrowing arthropods. The compound eyes of trilobites are further subdivided into holochroal and schizochroal forms, with the schizochroal structure developing as a result of paedomorphosis. Holochroal trilobite eyes consist of many round or polygonal lenses, the edges of which are all in contact and covered by a single corneal membrane. Schizochroal eyes are a unique system restricted to certain trilobite groups (Phacopina). The lenses are large and separated from one another. Each separate lens has its own separate corneal covering. Mounting the eye lens on upstanding and curved turrets may have allowed certain trilobite species a full 360° field of vision.

Trace Fossils

The combination of rigid appendages and the ability to be motile in a wide range of ecological habitats has provided a rich and diverse trace fossil record which can be assigned to the activities of arthropods. Trace fossils can indicate the presence of arthropods in depositional environments in which preservation would normally fail, and body fossils are non-existent. A good example of this is the Permian Coconino Sandstone Formation of Arizona. Various studies of the desert sandstones have revealed the presence of scorpion and tarantula trackways, the body fossils of which are absent from these environments.

Figure 4 The enigmatic arthropod, *Ainiktozoon loganense*, from the Silurian Fossil Konservat Lagerstatten of Lesmahagow, Lanarkshire, Scotland (×1.5). © The Trustees of the National Museums of Scotland, Edinburgh, UK (NMS G.1996.40.1).

Distinctive trace fossils, such as *Kouphichnium*, can reveal the presence of xiphosuran arthropods (horseshoe crabs) and, by analogy with their extant relatives, such as *Limulus*, can be used to imply nearshore or freshwater palaeoenvironments in otherwise sediments devoid of body fossils. Trace fossils, when found in conjunction with body fossil assemblages, may allow behaviour to be interpreted, such as the speed of movement or different actions (nesting, foraging, etc.). Of particular importance in the fossil record of insects is the formation of leaf mines and bite marks in accompanying floral assemblages, giving a clear indication of an escalating arms race between plants and arthropods.

Final Remarks

In conclusion, the Arthropoda can be marked out as one of the most important groups of organisms on the Earth. Their fossil record spans from the earliest recognizable invertebrate faunas right up until the present day. Their numerical abundance, both in the present and in the past, positions them as the major players in ecological structure: predators and prey in all three major Earth environments (land, sea, and air). The group provides excellent opportunities to investigate the effects of extinction events, adaptive radiation, macro- and micro-evolution, and the consequences of palaeobiology in extinct groups. Understanding the relationships between arthropods and other plant and animal groups remains important. Arthropods were an important part of the food chain in the past, just as they are today. Further advances in DNA sequencing and other types of molecular data will no doubt increase our knowledge of the inter-relationships of this diverse and fascinating group.

See Also

Evolution. **Fossil Invertebrates:** Trilobites; Insects. **Lagerstätten**. **Microfossils:** Ostracoda. **Palaeoecology**. **Trace Fossils**.

Further Reading

Allen KC and Briggs DEG (1989) *Evolution and the Fossil Record*. London: Belhaven Press, Pinter Publishers.

Benton MJ (ed.) (1993) *The Fossil Record 2*. London: Chapman and Hall.

Braddy SJ and Dunlop JA (1997) The functional morphology of mating in the Silurian eurypterid, *Baltoeurypterus tetragonophthalmus* (Fischer, 1839). *The Zoological Journal of the Linnean Society* 121: 435 461.

Briggs DEG and Crowther PR (eds.) (1990) *Palaeobiology: A Synthesis*. Cambridge: Cambridge University Press.

Edgecombe DG, Wilson GDF, Colgan DJ, Gray MR, and Cassis G (2000) Arthropod Cladistics: combined analysis of histone H3 and u2 snRNA sequences and morphology. *Cladistics* 16: 155 203.

Fortey RA and Owens RM (1999) Feeding habits in trilobites. *Palaeontology* 43: 429 465.
Fryer G (1985) Structure and habits of living branchiopod crustaceans and their bearing on the interpretation of fossil forms. *Transactions of the Royal Society of Edinburgh: Earth Sciences* 76: 103 113.
Rasnitsyn AP and Quicke DLJ (2002) *History of Insects*. Dordrecht, Boston, London: Kluwer Academic Publishers.
Willmer P (1993) *Invertebrate Relationships: Patterns in Animal Evolution*. Cambridge: Cambridge University Press.

Trilobites

A W A Rushton, The Natural History Museum, London, UK

Introduction

The trilobites are a large group of extinct marine arthropods, the Class Trilobita, characterized by the longitudinal division of the exoskeleton into three lobes – a convex central (axial) lobe flanked on each side by a flatter pleural area (Figure 1) – hence their name. Trilobites were widely distributed throughout the Palaeozoic era, first appearing in the later part of the Early Cambrian period (at about 520 Ma) and disappearing just prior to the end-Permian mass extinction (*ca.* 250 Ma). During their 270 million year existence, trilobites spread worldwide, occupied a wide range of marine habitats, and evolved a great variety of morphologies.

Trilobites are a relatively well-studied group and a good deal is known, or inferred, about their life, growth, and activities. Their extensive fossil record has provided illuminating case histories in evolution and ecological adaptation, and, in addition, they have proved to be of great value in stratigraphy, especially in the Cambrian and Ordovician periods, and in interpreting Palaeozoic palaeobiogeography.

Thanks to their striking appearance, strange-looking yet evidently organic in origin, trilobites have always been attractive fossils. Specimens have been recorded in prehistoric burial sites and they have found a place in folk-lore; in more recent times, avid collectors have built up magnificent collections of fine specimens, some of which have commanded high prices. The scientific study of trilobites commenced at the end of the seventeenth century, and, during the eighteenth and nineteenth centuries, European workers, including Linnaeus, published descriptions and illustrations. The benchmark was Barrande's illustrated account of the Bohemian trilobites, published in 1852, that showed their astonishing variety and stimulated interest in the group amongst all scientific communities; this interest continues to this day, such that more than 5000 genera have now been described and a huge number of species named, perhaps approaching 20 000.

Form of the Exoskeleton

The general features of the exoskeleton are labelled in Figure 2. The exoskeleton typically has an oval outline in plan view, and consists of a cephalon (headshield), furnished with a pair of compound eyes, a jointed thorax of few to many segments, and a pygidium (tail-piece) made up of fused segments. Typically the animal was 2–10 cm in length, but a species of *Acanthopleurella* is just over 1 mm in length when fully grown, and the largest known *Isotelus* is 72 cm in length.

The trilobite exoskeleton, unlike that of many arthropods, was mineralized by calcite and phosphate, thus greatly increasing its potential for preservation

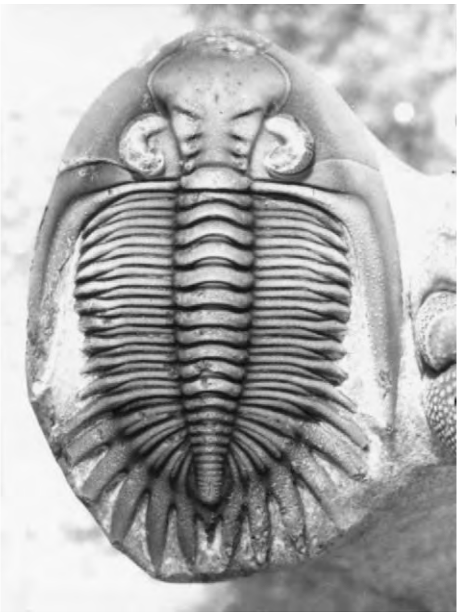


Figure 1 The dorsal exoskeleton of *Neometacryphaeus* (sub order Phacopina) from the Devonian of Morocco. About natural size. Courtesy of Professor R. A. Fortey.

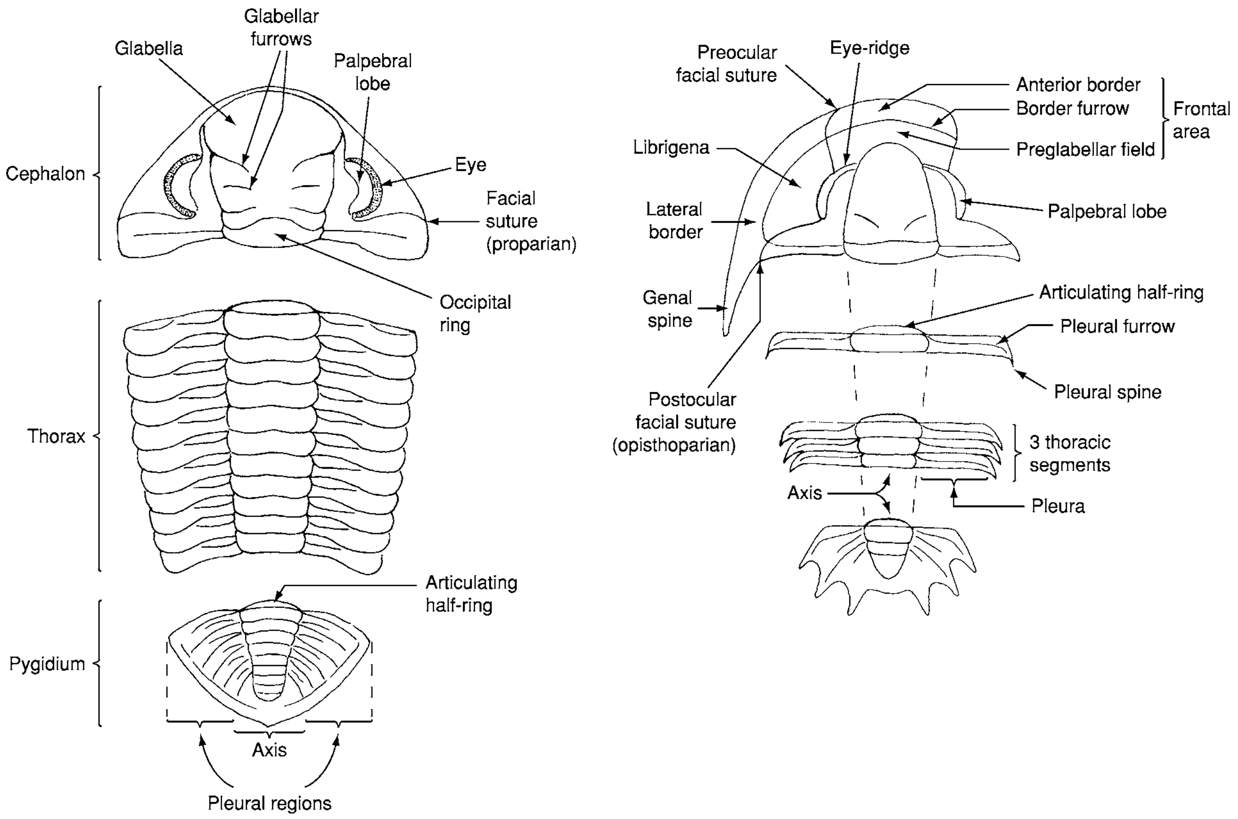


Figure 2 Terminology based on the phacopine *Acaste* and the ptychopariid *Marjumia*.

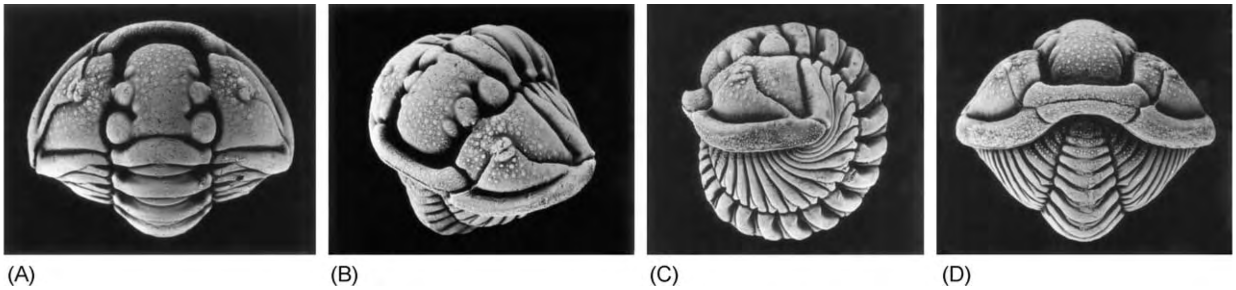


Figure 3 An enrolled *Calymene* from the Silurian. Width of cephalon about 3 cm. Courtesy of Dr Derek Siveter.

and leading to the good fossil record of the class. In contrast, the appendages (antennae and limbs) were not mineralized and are found only rarely, in conditions of special preservation, and the ventral membrane is scarcely known. The depressed form of many early trilobites suggests that they originated as benthic forms, with the unprotected ventral side on the substrate, and the dorsal shield protecting the animal from above. They subsequently evolved a great variety of forms that allowed a range of feeding strategies to be employed and enabled the occupation of many marine environments. So great is their variety that almost every generalization given here could be modified by some exception.

The dorsal exoskeleton is more or less vaulted, especially the axial lobe, which housed most of the musculature, the alimentary tract, and the main circulatory and nervous conduits. The pleural lobes are less convex and served to cover the limbs, namely the walking legs and the gill-filaments. The thoracic segments articulate with each other, so that many trilobites could roll up, thus protecting the ventral side (Figure 3). The outer edge of the exoskeleton is recurved under the body to form a rim, the doublure (Figure 4), on which the animal could rest on the seafloor.

In most trilobites, the cephalon is divided by the facial sutures into three main sclerites: the cranidium,

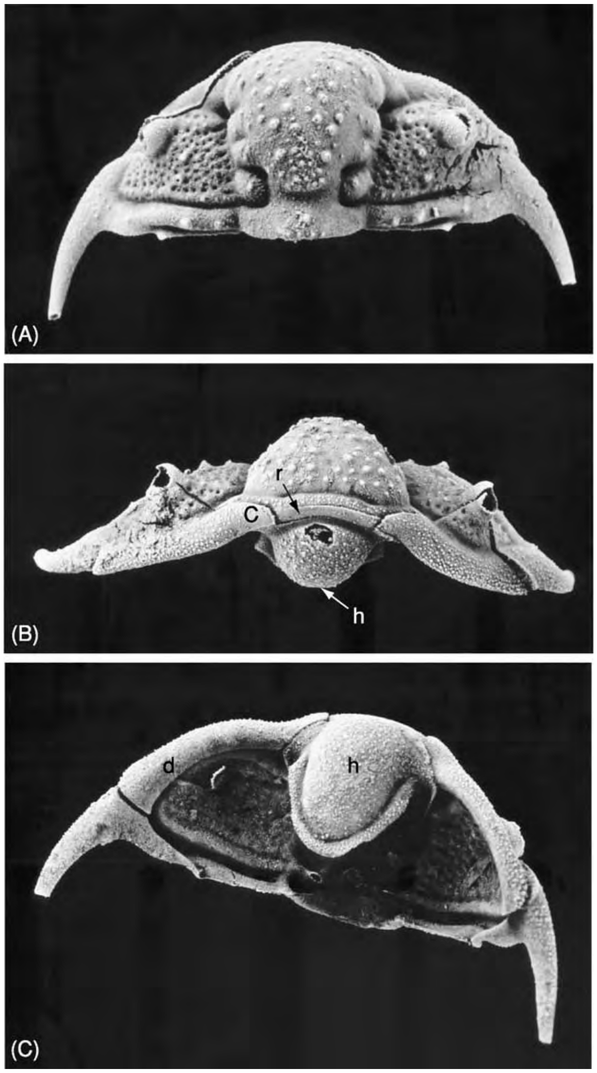


Figure 4 A silicified cephalon, about 8 mm in length, of *Ceraurus* (Cheirurina) from the mid Ordovician: views from top (A), front (B), and below (C). The hypostome (h) underlies the fore part of the glabella and its anterior edge is set against a narrow rostrum (r). Reproduced with permission from Whittington HB (1992) *Fossils Illustrated 2. Trilobites.* Woodbridge: Boydell Press.

flanked by a pair of librigenae. The glabella, whose shape and lateral furrows are important features for classification, housed the fore-gut and stomach. The posterior part of the glabella, the occipital ring, articulates with the thorax and commonly mimics features of the thoracic axis. The palpebral lobes overlay the eyes, but the ocular surface was commonly attached to the librigena, and the facial suture almost invariably passed around the upper edge of the eye to facilitate moulting of the exoskeleton, whilst avoiding damage to the ocular surface. The course of the preocular and postocular sutures is very varied. The preocular sutures curve inwards to meet in front of the cranidium and commonly bound a separate mineralized plate, the rostrum (not always present), that underlies the frontal margin (Figures 4 and 5). Underneath the glabella is the hypostome, a mineralized plate lying anterior of the mouth and under the fore-gut. The anterior end of the hypostome was either set rigidly against the rostrum or doublure (the conterminant condition, Figures 4 and 5B; sometimes modified to impendent, Figure 5C), or was separated from the rostrum and supported only by musculature and the ventral membrane (the natant condition; Figure 5A). In some instances, the hypostome was docked against the rostrum, but was not rigidly attached to it; it may have been joined by a sort of hinge of unmineralized integument.

The thorax is made up of several more or less similar segments, commonly 8–12 in number, but as few as two in some forms and more than 40 in others. Sometimes one or more segments show special adaptations, such as axial spines or macropleurae. Each axial ring has an articulating half-ring that fits under the segment in front and, during enrollment, they protect parts of the axial lobe from being exposed; at the same time, the pleural tips of each segment can slide under the pleurae in front (Figure 3).

The pygidium is composed of few to many fused segments and its size varies from tiny to larger than the cephalon (Figures 6A and 6B). In some pygidia, the segmentation remains obvious, but others are well fused so that the segmentation is not visible. During enrollment, the underside of the pygidium could, if it was small, be tucked into the cephalic chamber; alternatively, it could be pressed against the underside of the cephalon, the margin of which commonly shows pits or grooves – coaptive structures – that allow a perfect fit.

The trilobite exoskeleton may be smooth, granulose or tuberculate, pitted (Figure 4A), or covered with terrace lines (the latter being especially characteristic of the ventral doublure; Figure 7). Spines may be developed almost anywhere and some species are extravagantly spiny (Figure 17B). Many exoskeletal features, such as tubercles, were perforated with pores that might have housed setae and presumably had sensory functions. There is a trend in many major trilobite groups for some species to become 'effaced', whereby the axial, glabellar, and border furrows are smoothed out and the 'trilobed' appearance is lost (Figure 6).

Eyes

Most trilobites had well-developed eyes, those from the Early Cambrian being amongst the earliest eyes known in the fossil record. The eyes are generally placed high on the fixigenae and, in certain forms,

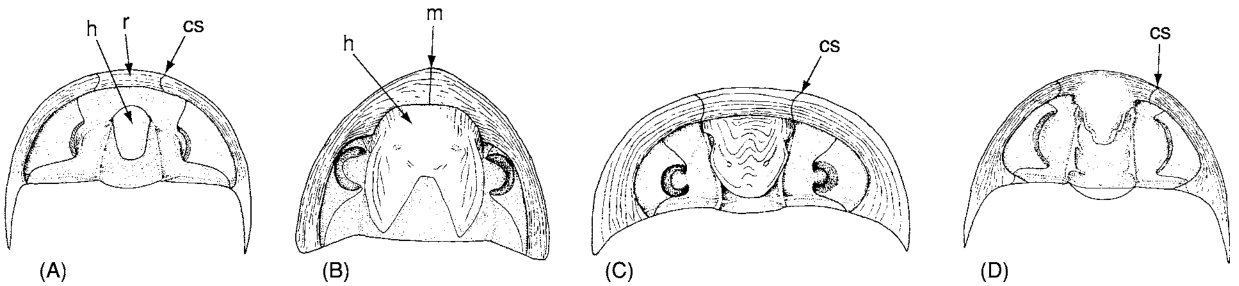


Figure 5 Ventral views of cephala. (A) *Elrathia* (Ptychopariina), with simple natant hypostome (h) and rostrum (r) separated from the librigenae by connective sutures (cs). (B) *Isotelus* (Asaphina), with forked conterminant hypostome (h) and a median suture (m) in place of the rostrum and connective sutures. (C) *Raymondaspis* (Styginina), with impendent hypostome, similar to the conterminant condition, but the anterior edge of the glabella does not correspond to that of the hypostome. (D) *Fieldaspis* (Corynexochida), in which the hypostome and rostrum are fused into a single plate; cs, connective sutures. Reproduced with permission from Fortey RA (1990) Ontogeny, hypostome attachment and trilobite classification. *Palaeontology* 33: 529 576.

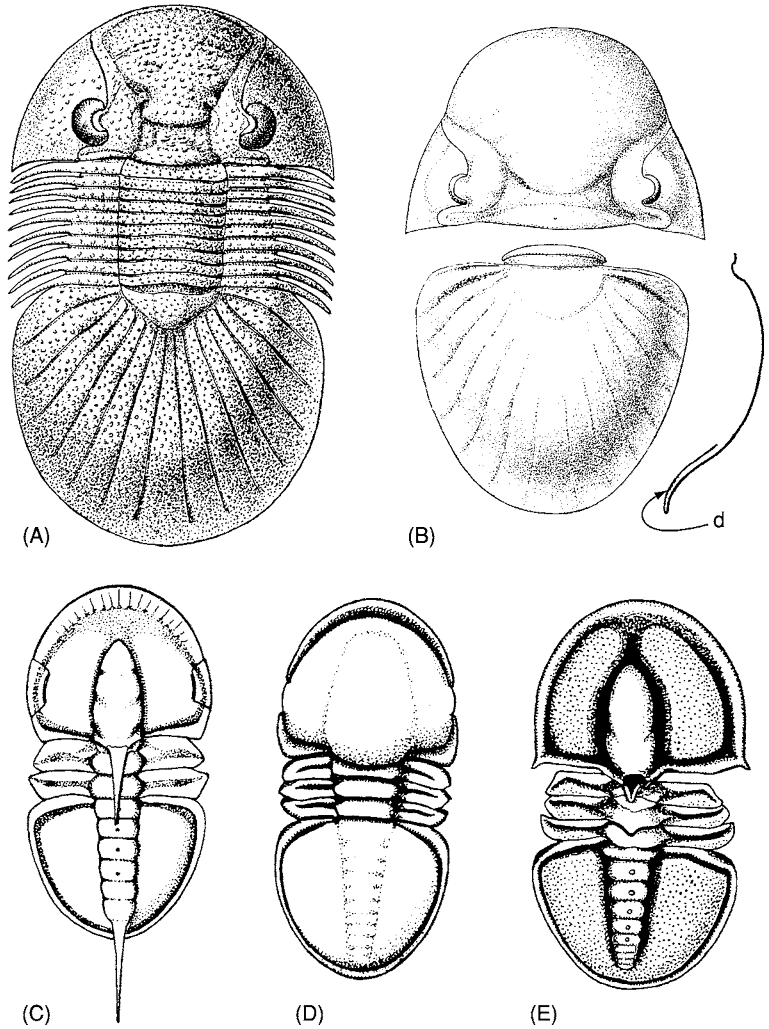


Figure 6 Effacement of external features in Styginina (A, B) and Eodiscina (C E). (A) *Scutellum*, Devonian, 10 cm in length. (B) *Paralejurus*, Devonian, cephalon (5 cm in length) and pygidium with longitudinal profile, showing doublure (d). (C) *Pagetia*, mid Cambrian. (D) *Pagetiellus*, Early Cambrian, cephalic and pygidial axis effaced, eyes marginal. (E) *Eodiscus*, mid Cambrian, not effaced, but eyes absent. (C E) each about 8 mm in length.

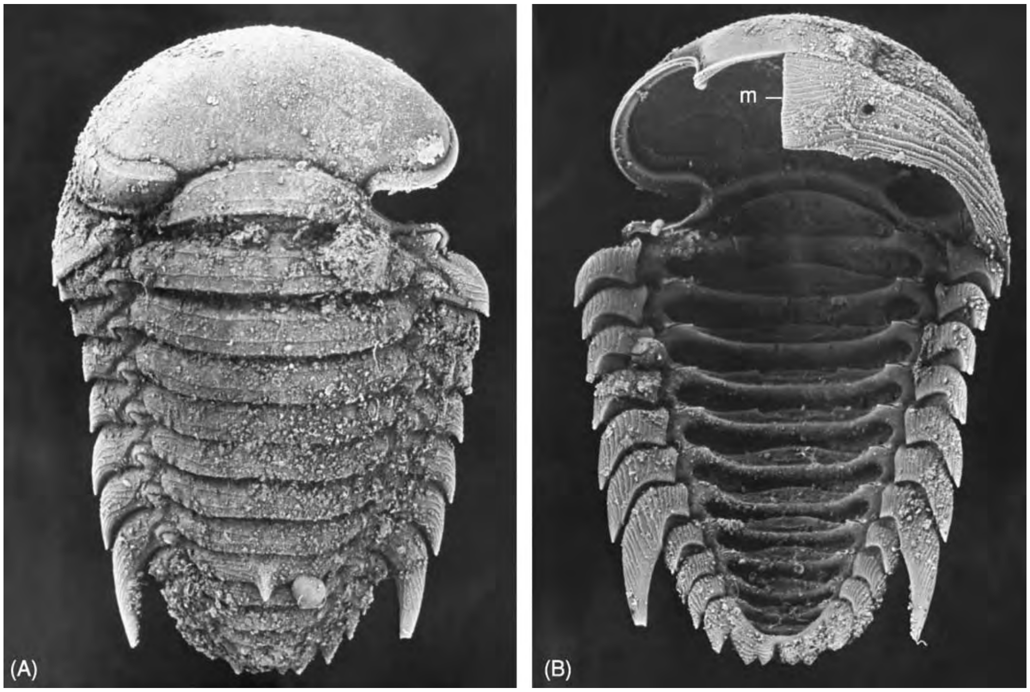


Figure 7 Dorsal and ventral views of a silicified dorsal exoskeleton of *Remopleurides* (Asaphina), with one librigena in place. Length about 20 mm. The ventral view shows the median suture and terrace lines striating the doublure. Reproduced with permission from Whittington HB (1992) *Fossils Illustrated 2. Trilobites.* Woodbridge: Boydell Press.

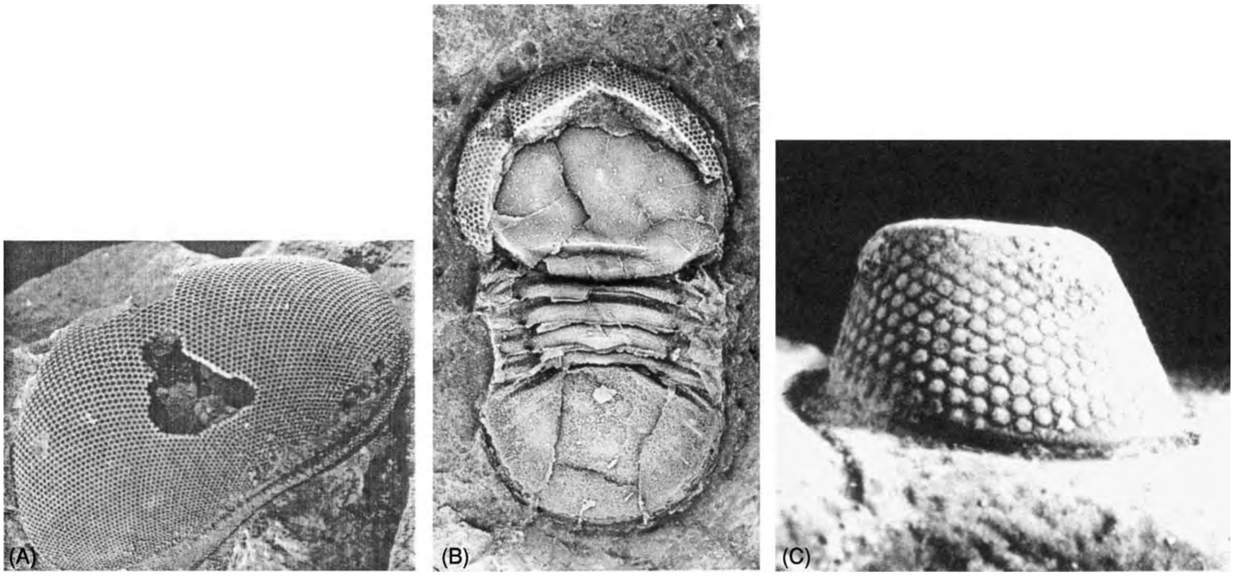


Figure 8 (A, B) Holochroal eye in Cyclopygidae (Asaphina). (A) *Pricyclopyge* left eye, about 10 mm in length. (B) *Microparia lusca*, a species in which the eyes have grown forward and joined up in front. (C) Schizochroal eye of *Dalmanites* (Phacopina), side view of right eye, about 5 mm in length; each lens is about 0.25 mm in diameter. Reproduced with permission from Treatise O, revised vol. 1, Owens RM (2002) Cyclopygid trilobites from the Ordovician Builth Llandrindod inlier, Central Wales. *Palaeontology* 45: 469–485, and Clarkson ENK (1966) Schizochroal eyes and vision in some phacopid trilobites. *Palaeontology* 9: 464–487.

lie at the summit of eye-stalks. Being arcuate in plan, they afforded a wide field of view. Most trilobites had eyes consisting of many small contiguous lenses covered by a cornea of calcite (holochroal eyes; Figure 8A), whereas the suborder Phacopina had highly evolved eyes with fewer but larger biconvex lenses, each with their own cornea, that were separated by interlensar sclera (schizochroal eyes;

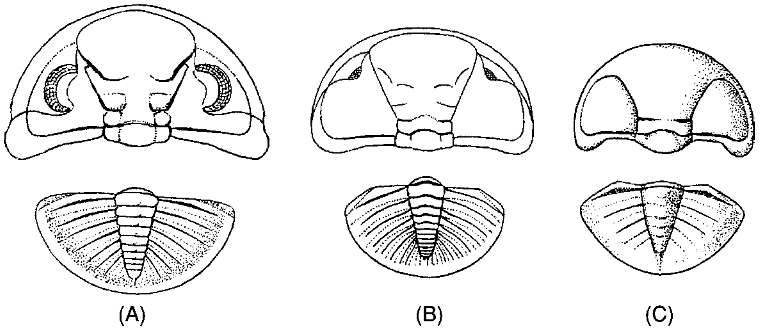


Figure 9 Reduction and loss of eyes in some Phacopidellinae (Phacopina) from the Silurian and Devonian. (A) *Phacopidella*, ×1. (B) *Denkmannites*, ×1, with reduced eyes. (C) *Ductina*, ×3, blind.

Figure 8C). These lenses were composed of calcite and a study of their optics has shown that they had features to overcome birefringence and chromatic aberration.

In pelagic or mesopelagic trilobites, the holochroal eyes became enlarged and bulbous, even merging in front of the glabella, giving all-round vision (Figure 8B). In contrast, secondary eye reduction or loss is present in various groups of trilobites, and is commonly associated with species from deep-water settings. Some lineages have been found that show progressive reduction of eyes, leading to their loss, and blindness is a feature of species in many major clades (Figures 6E and 9).

Limbs

The ventral limbs of trilobites are known only in about a dozen taxa, but these exemplify several major groups of various ages from the Cambrian to Devonian. The anterior appendages were a pair of uniramous antennae, all the succeeding appendages being deeply biramous and uniform in structure, apart from a pair of posterior antenniform cerci known only in *Olenoides*. There were generally three pairs of biramous limbs under the posterior part of the cephalon (but four pairs have been claimed in one or two instances), and one under each of the thoracic and pygidial segments (Figure 10). The coxa at the base of the limb is spinose and presumably acted as a gnathobase. Attached to the coxa was a lower branch, the endite, of seven podomeres (joints), including a small terminal claw; this branch was used for walking and digging. It is inferred that the hinge between the coxa and the ventral side of the exoskeleton allowed forward and backward movement of the endite, whilst the joints between the podomeres allowed the limb to be curled in under the body. The upper branch, the exite, was filamentous and is assumed to have been a gill. It occupied the small space above the endite and under the pleural region (Figure 10).

Other Features

Sites of the insertion of muscles are seen on parts of the exoskeleton, notably as paired furrows on the glabella, but also as apodemes on the ventral side of the axial furrows and, more sporadically, as pits on the pygidial axis; they may also appear as colour patterns, and may thus be visible in effaced trilobites in which the glabellar furrows are obscure. Many trilobites, especially in the Cambrian, show caeca, i.e., traces of fine anastomosing ridges that cross the preglabellar and genal fields (Figure 11). They have been regarded as an exterior expression of part of the respiratory system or, alternatively, diverticula of the digestive tract.

Two trinucleid trilobites have been found that, thanks to freakish weathering, show the alimentary tract extending back from the glabella, along the axial region to the pygidium. Wonderfully preserved juvenile examples from Sweden of the highly derived form *Agnostus pisiformis* have revealed not only the peculiar limbs of that species, but also features whose presence had been inferred in trilobites but not otherwise seen, e.g., the mouth, gut, anus, ventral membrane, and minute sensory pits and setae. Several trilobites have been found that show exoskeletal modifications caused by teratology, parasites, healed injury, or partial failure in moulting. Exoskeletal features attributable to sexual dimorphism have not been positively identified. However, bulbous structures present in front of the glabella in certain species have been interpreted as brood-pouches.

Growth

Trilobites, in common with other arthropods, grew by ecdysis (moulting), so that any animal that reached

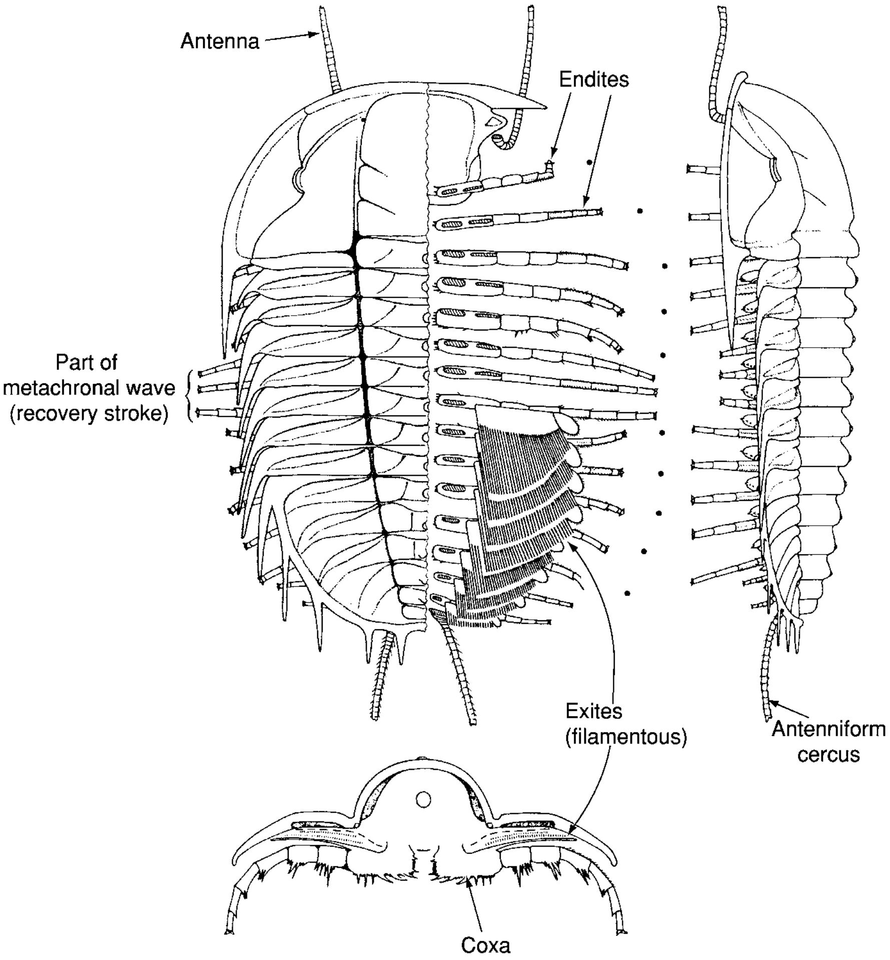


Figure 10 Reconstruction of *Olenoides* (Corynexochida), showing limbs and supposed gait when walking. Sagittal length about 7 cm.

Figure 11 Cranidium and librigena of *Parabolina* (Olenina), showing genal caeca. Length of cranidium, 10 mm. Reproduced from Rushton AWA (1982) The biostratigraphy and correlation of the Merioneth Tremadoc Series boundary in North Wales. In the Cambrian Ordovician Boundary: © National Museum of Wales, Geological Series No. 3, Cardiff.

maturity could have left remains of several exoskeletons; there are records of bedding planes with large accumulations of moulted remains. Most often the exuvia became disarranged, being preserved merely as scattered sclerites, but many examples have been found that show arrangements suggestive of the moulting strategies employed by various species. In some instances, the facial sutures opened and the librigenae were sloughed off separately from the cranidium; in others, such as the phacopids, the thoracopygon (thorax + pygidium) may be found only a little disarranged, whilst the cephalon or cranidium is found nearby, separated and overturned; in such a case, the animal crept out from the cast shell by way of the split between the cephalon and thorax (Figure 12).

The earliest growth stage, the protaspis, is a single shield commonly a fraction of a millimetre in length. Later moults show the developing cranidium and the addition of a protopygidium with fused thoracic segments (Figure 13C); successive segments were generated at the posterior end of the pygidium and, as they grew forward, were released into the thorax (meraspid stages) until the full complement of segments was achieved (holaspid stage), as shown in Figure 13. Full or partial series of growth stages

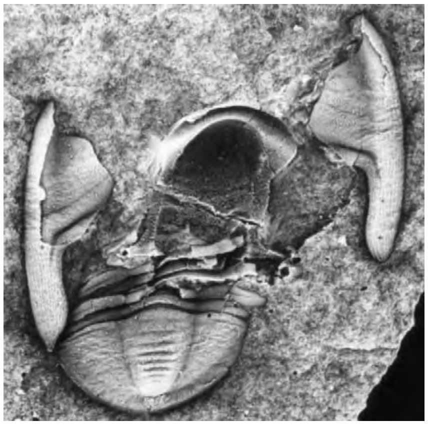


Figure 12 Moulted exoskeleton of *Spatulina* (Proetida) from the Early Carboniferous. The librigenae are cast off, the thorax is slightly disarranged, and the cranidium is inverted in relation to the other parts; the hypostome is missing. Sagittal length, as preserved, about 10 mm. Reproduced, with permission from Cambridge University Press, from Owens RM and Tilsley JW (1995) An atheloptic trilobite assemblage from the Carboniferous of North Devon. *Geological Magazine* 132: 713 728.

have been reconstructed for a considerable range of species, and some of the growth patterns are characteristic of major groups, e.g., the Asaphida, in which advanced members of the group have relatively large globular protaspides (Figures 13I and 13J). With all trilobites, there were great changes in the shape as well as the size of the exoskeleton throughout growth, and such changes provided opportunities for the derivation of new morphologies by the operation of heterochronic processes, whereby forms developed sexual maturity whilst still showing juvenile features (paedomorphosis).

Activities

Walking

Tracks made by arthropods, including trilobites, in unconsolidated sediments may be preserved as trace fossils, although it is generally impossible to be sure which animal was responsible for which traces (*see* **Trace Fossils**). It is thought that trilobites, like other arthropods with many walking limbs, crawled forwards using their limbs in a metachronal wave (whereby groups of limbs push forwards whilst intervening groups of limbs make a return stroke, ready to give their next push, as represented in Figure 10). Tracks that are of a size compatible with contemporaneous trilobites have been analysed and are indicative of such motion. There are also trace fossils which show that trilobites could progress by making a series of short hops or jumps.

Digging

Trilobites used their endites to excavate resting 'nests' or to probe the substrate for prey. The bilobed burrowing traces, *Rusophycus*, have been positively associated with trilobites because they sometimes show the outline of the cephalic shield or genal spines, and one example has been found with a complete *Flexicalymene* above it. *Rusophycus* shows transverse scratch marks that were made by the tips of the endites. Complete specimens of calymenids and homalonotids found with the thorax extended in a concave curve may represent animals that died whilst occupying such a resting burrow. The longer bilobed trackways, known as *Cruziana*, are thought by some to represent digging or ploughing by an advancing trilobite, although the means by which the locomotion and digging activities could have been combined has not been worked out.

Swimming

A number of trilobites have been identified on the basis of several lines of evidence as having been adapted to pelagic or mesopelagic habits, all the best documented being Ordovician in age (Figure 14). All have huge eyes with a wide field of vision. Some, like *Carolinites*, have narrow pleural areas and a vaulted axis that gives space for strong musculature; they are found in association with benthic faunas of all facies, being independent of any of them. These are pelagic forms. Other less vaulted and muscular forms, typified by the cyclopygids (Figure 14B), are likewise associated with various faunas, but in their case only in outer-shelf settings. These are regarded as mesopelagic. Certain highly streamlined genera are interpreted as rapidly swimming predators (Figure 14C), but as the limbs are not known in any pelagic form, the actual action of swimming is conjectural.

Feeding

Various morphotypes associated with particular feeding strategies were developed independently at different times in different major trilobite groups. Predators and scavengers tended to be relatively large forms with well-developed eyes and a big conterminant hypostome that is braced against the rostrum, and sometimes fused with it, as in Corynexochids (Figure 5D) and *Paradoxides*. The posterior end of the hypostome, near the mouth, commonly has special features, such as a notch or forked and serrated margin, presumably adaptations to manipulate food (Figure 5B). A large and forwardly expanded glabella and an impendent hypostome characterize advanced forms like *Phacops* and its relatives (Figure 17D).

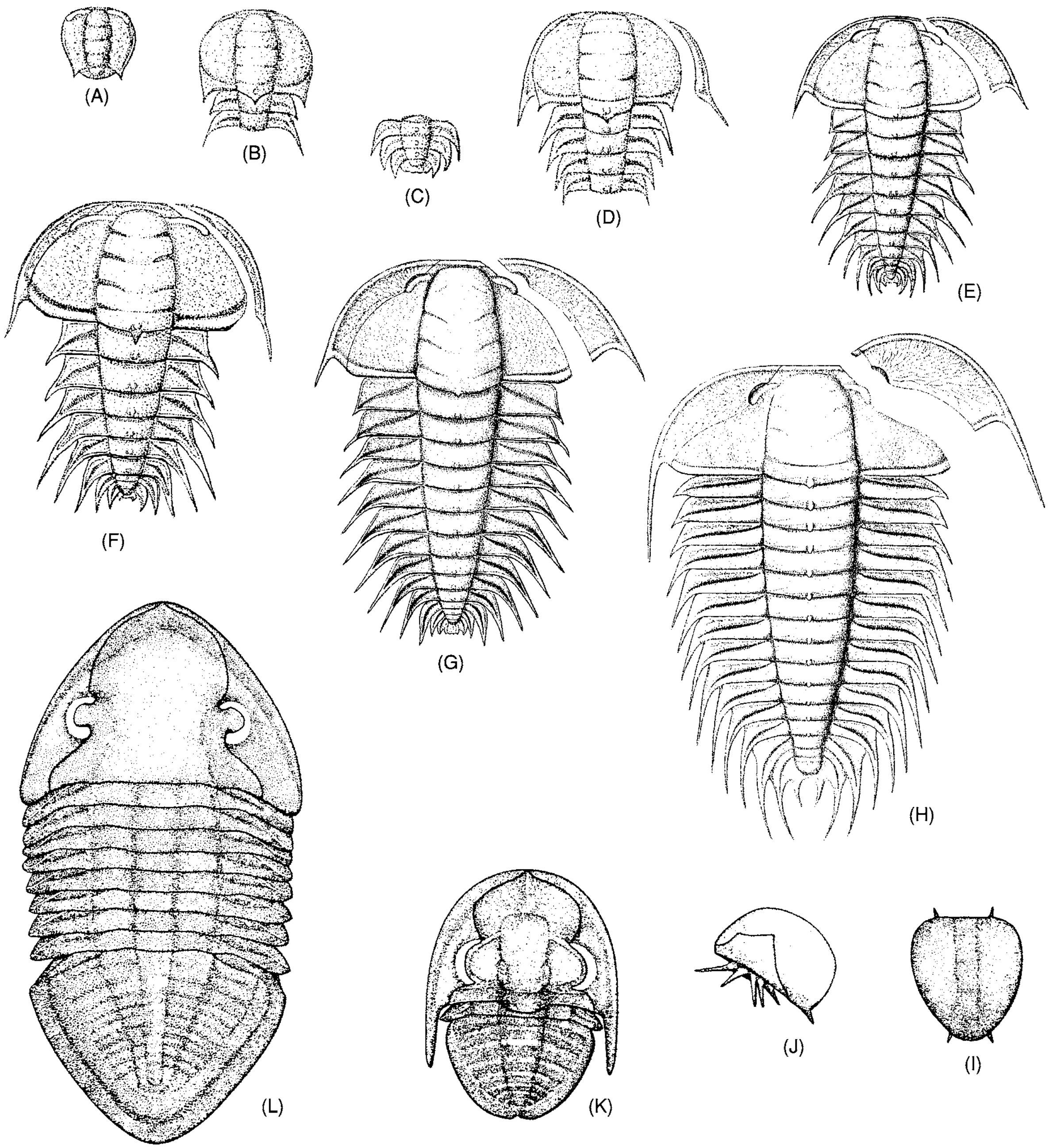


Figure 13 Growth of trilobites, exemplified by *Parabolina* (Olenina) (A H) and *Isotelus* (Asaphina) (I L). *Parabolina*: (A) protaspis, 0.5 mm in length; (B) meraspid cranidium, 0.6 mm in length with two free segments; (C) transitory pygidium (to same scale) with unreleased thoracic segments; (D, E) larger meraspides (to same scale) with four and six free segments; (F, G) meraspides with eight and 10 segments, respectively, 3.3 and 4.5 mm in length; (H) holaspis form with 12 segments, 16 mm in length. *Isotelus*: (I, J) globose protaspis, 0.6 mm in length, top and side views; (J) shows the spinous hypostome; (K) meraspides degree 1, 3 mm in length; (L) holaspis, 60 mm in length. A H Reproduced by permission of the Royal Society of Edinburgh and E. N. K. Clarkson, C. M. Taylor, and P. Ahlberg from 'Ontogeny of the trilobite *Parabolina spinulosa* (Wahlenberg, 1818) from the upper Cambrian Alum Shales of Sweden' by Euan N. K. Clarkson, Cecilia M. Taylor, and Per Ahlberg in *Transactions of the Royal Society of Edinburgh: Earth Sciences*, volume 88 (1997) pp. 69 89. I L Treatise O, revised, vol. 1.

In contrast, detritus feeders tended to be smaller, less active animals with a smaller glabella and relatively wide preglabellar and pleural areas. The hypostome was commonly natant and conservatively ovoid in shape (Figure 5A). An enormous number of Cambrian species of this type have been described and presumably they signify a subtle partitioning of resources; fewer are known from the Ordovician and later periods when such trilobites had to compete with newly evolved Mollusca. An offshoot of

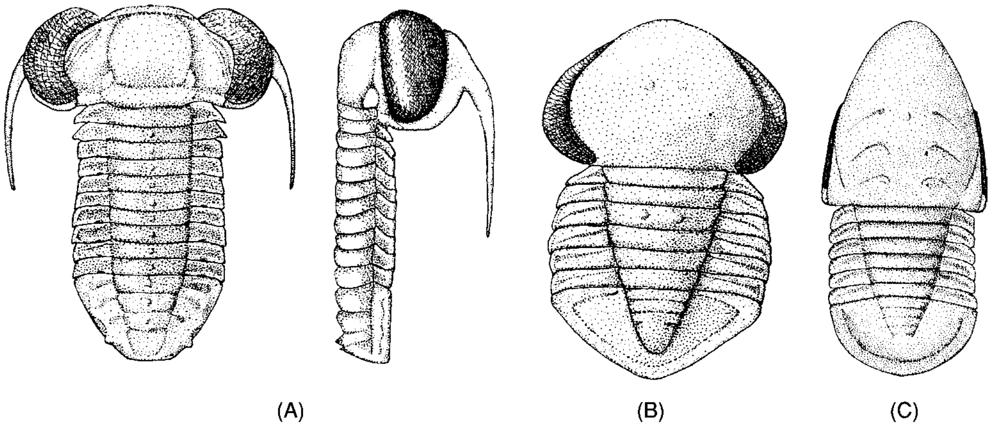


Figure 14 Large eyed trilobites. (A) *Carolinites* (Proetida), top and side views of a pelagic form (about 3 cm in length). (B) *Pricyclopyge*, a broad mesopelagic cyclopygid (Asaphina) about 3 cm in length. (C) *Degamella*, a streamlined cyclopygid, 5 cm in length. Reproduced by permission of the Royal Society of Edinburgh and R. A. Fortey from 'Pelagic trilobites as an example of deducing the life habits of extinct arthropods' by R. A. Fortey in *Transactions of the Royal Society of Edinburgh: Earth Sciences*, volume 76 (1985) pp. 219 230.

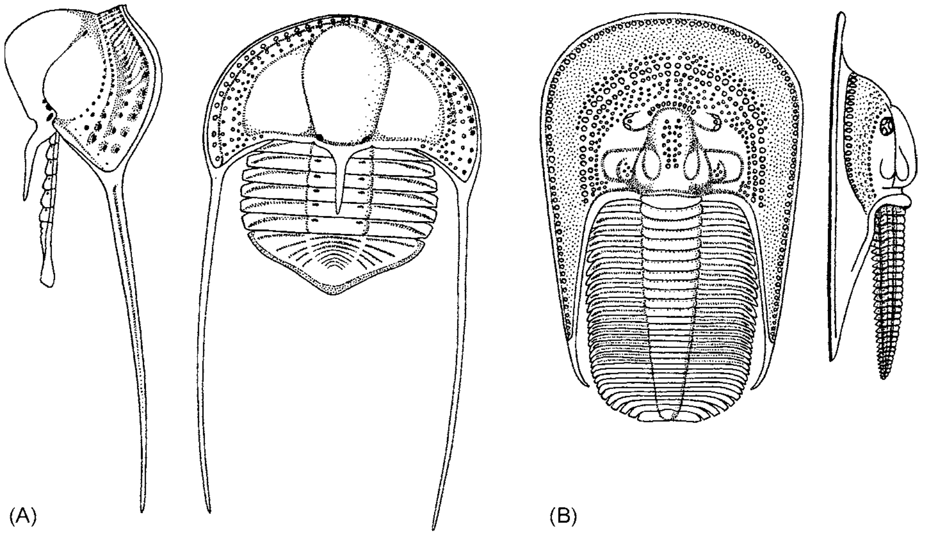


Figure 15 Top and side views: (A) *Cryptolithus* (axial length, 3 cm), an Ordovician trinucleoid in the Order Asaphida; (B) *Harpes* (axial length, 6 cm), a Devonian member of the Harpina, Order Ptychopariida; both show the large cephalon and relatively small thorax that are taken to characterize a chamber filter feeding mode of life. Although not closely related, these groups show a remarkable convergence, both having developed a broad bilamellar fringe on the cephalon. The upper lamella is part of the cranidium; the lower lamella includes the genal spines and is separated from the upper lamella by the facial suture, which extends around the margin of the cephalon.

the detritus feeder type is the possible development in the Olenidae (e.g., Figure 13H) of a chemoautotrophic metabolic strategy. The olenids had a multiplicity of thoracic segments and conservative hypostomes that were degenerate in some instances, and they survived in conditions of unusually low oxidation and high sulphur content. It has been suggested that they fed on symbiotic sulphur bacteria that they cultivated on their exites.

A group of trilobites with large head-shields is regarded as chamber filter feeders (Figure 15). In these, the cephalon is considerably deeper than the thoracopygon and the hypostome is elevated well above the substrate. It is thought that the animal stirred up the sediment in the chamber beneath the cephalon and sorted edible particles from suspension. The best known of these trilobites are the harpetids, whose massive genal prolongations extended the filter chamber backwards alongside the thorax, and the trinucleids. These groups independently developed broad cephalic fringes composed of an upper and a lower lamella that were separated by a facial suture

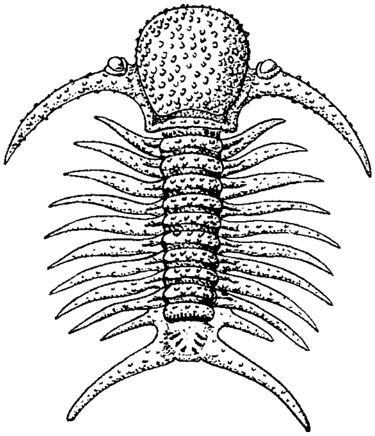


Figure 16 *Deiphon* (Cheirurina), an enigmatic Silurian trilobite, about 3 cm in length.

extending around the margin of the cephalon. Of the trinucleid group, *Cryptolithus* is significant because the limbs are known and *Rusophycus* excavations made by the animal have also been recognized.

Of the pelagic trilobites mentioned above, most were probably plankton feeders; the large streamlined forms (*Degamella*, *Parabarrandia*) were likely to have been predatory upon smaller plankton feeders. Besides the foregoing, there is a host of other forms whose feeding strategies are undetermined, including many species that occupied specialist niches within carbonate mounds, as well as bizarre forms such as *Deiphon* (Figure 16).

Classification

Cladistic analysis shows that Class Trilobita is not primitive in relation to other early arachnomorph arthropods, and the trilobites are taken to be a natural group characterized by the calcified exoskeleton and ocular surface, the facial sutures, and features of the hypostome and pygidium. However, a comprehensive natural classification of the trilobites has not yet been devised. Thousands of genera have been grouped into hundreds of subfamilies and families, but the content and limits of many of these are subjective, and it is debated how they should be grouped into higher categories. Several major natural groups are widely accepted, especially amongst the post-Cambrian trilobites, but the origins of some of them are cryptic and how the various groups are interrelated remains uncertain. Obstacles to classification have included the relatively poor knowledge of the ontogenies of many trilobites, their limbs (and ventral features generally), the tendency for iterative evolutionary trends to yield misleading homeomorphic forms, and the unresolved questions that surround the great mass of plesiomorphic taxa in the Cambrian. None the less, great advances in the description and interpretation of trilobites in the latter half of the twentieth century has led to an improved understanding of the problems, and has yielded such provisional classifications as that presented in outline in Table 1. Cladistic and morphological analyses have provided some unexpected, although fundamentally satisfying, results, e.g., the transfer of the trinucleid group to the Asaphida.

Table 1 Major orders and suborders of trilobites. Some examples are shown in Figure 17

Order	*Suborders and some other major groups*
Agnostida	Agnostina, Eodiscina
Redlichiida	Redlichiina, Olenellina
Corynexochida	Corynexochina, Styginina, Leiostegiina
Lichida	(including superfamily Odontopleuroidea)
Phacopida	Phacopina, Cheirurina, Calymenina
Asaphida	(including superfamilies Asaphoidea, Remopleuridioidea, Dikelokephaloidea, Anomocaroidea, Cyclopygoidea, Trinucleoidea, etc.)
Proetida	
Ptychopariida	Ptychopariina, Olenina, Harpina

Distribution in Time and Space

The distribution of trilobites was determined by the availability of environmental settings with appropriate water depth, temperature, turbulence, substrates, oxygenation, etc. Barriers to migration included oceans at unfavourable temperatures, or too deep and wide to cross. Hence, faunas from shallow continental shelf areas are more endemic than those from continental slopes. This kind of evidence, especially from the most endemic types of trilobites (and other fossils), is exploited to provide lines of evidence, independent of plate tectonics, for reconstructing continental palaeogeographies. Thus, the occurrence of the Gondwanan trilobite *Plaesiocomia* in subsurface Florida, USA, indicated that, during the Ordovician, that state was a fragment of Gondwana, and was only later accreted on to Laurentia.

Cambrian

The earliest history of the trilobites is unknown. The oldest known species appeared in shallow marine settings around several palaeocontinents and are referable to a number of separate clades, indicating that the trilobites had an earlier history (probably as unmineralized or weakly mineralized forms) that is not yet known. One of the earliest to appear is the olenelloid *Fallotaspis* (Figure 17M), a genus

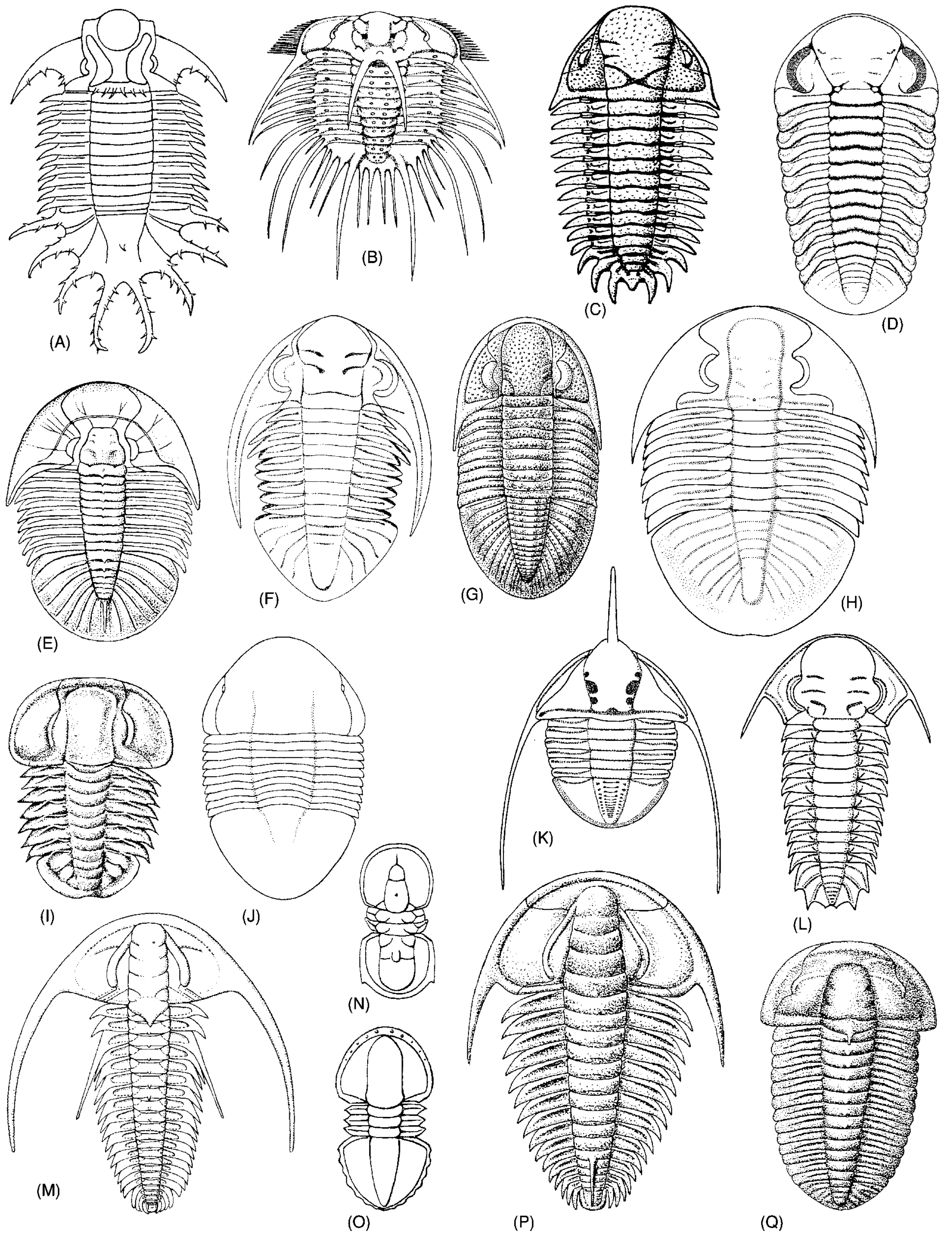


Figure 17 Representatives of some of the principal groups of trilobites. Typical sagittal lengths of each exoskeleton are given. (A, B) Order Lichida this includes the lichids and odontopleurids, which have distinctive cephalic and pygidial structures; although probably related to the Dameselloids, the derivation of the whole group is unknown. (A) *Terataspis* (Lichoidea), Devonian, 500 mm. (B) *Odontopleura* (Odontopleuroidea), Silurian, 40 mm. (C, D) Order Phacopida this order contains the suborders Phacopina,

that makes an early appearance in Siberia, Laurentia, and on the edge of Gondwana (Morocco). The Olenellina went on to radiate particularly in the Early Cambrian of Laurentia. They are not found in south China, where Redlichiida (Figure 17P) dominated, accompanied by oculate eodiscoids, and later by Ptychopariida (e.g., Figure 17Q). The redlichiids extended to Morocco, where they are associated with endemic olenelloids and opisthoparian forms, such as bigotinids and ellipsocephalids. The provinciality of the Early Cambrian shallow-water faunas continued into the Middle Cambrian, but as trilobites colonized deeper water environments, there arose taxa with the capacity to cross from one continental margin to another, e.g., *Arthricocephalus*, certain Agnostida, and *Centropleura*. The Late Cambrian saw an expansion of ptychopariid deposit feeders, with endemic faunas in Laurentia, north and south China, Australia, Kazakhstan, and north-west Siberia, whilst the dysaerobic environment of the ‘olenid sea’ around Baltica provided conditions that allowed the Olenidae to flourish. Some very widely distributed taxa, such as *Glyptagnostus* and *Irvingella* (Figure 18), which reached all these sites, may have been pelagic.

Ordovician

During the Ordovician, the trilobites reached their acme; the exoskeleton developed an even greater range of modifications than was seen in the Cambrian: well-documented pelagic forms evolved, schizochroal eyes appeared, and some benthic lineages evolved cephalic fringes as an aid to chamber filter feeding. Many Cambrian groups disappeared, but at least as many new groups appeared to replace them, and spread to occupy new niches. The Ordovician is characterized especially by trilobites of the Asaphida, including such groups as the remopleuridids, trinucleids, and cyclopygids (Figures 17H, 17K, and 17L). As in the Cambrian, Early Ordovician trilobite faunas on the shallow platformal areas were highly endemic; the bathyurids (Figure 17F) of Laurentia and Siberia, dikelokephalinids of eastern Gondwana (south China and Australia), and calymenoideans of west Gondwana show little or no intermixing, whereas the deeper shelf and planktonic forms are less exclusive. During the mid- to later Ordovician, there was some blurring of this provinciality as Laurentia, Baltica, and Avalonia approached one another, and, before the onset of the major Late Ordovician glaciation, latitudinal disparities in temperature seem to have been less marked. The extinction associated with that glaciation affected many groups of trilobites (e.g., agnostoids, bathyurids, the last ptychopariids, very nearly all the asaphoids), especially those that exploited the open ocean at some stage in their development, and greatly reduced the diversity of the class.

Silurian

The trilobites that survived the end-Ordovician extinction (Phacopida, Proetida, styginids, lichids, and odontopleurids; see Figure 17) mostly survived into the Late Devonian, giving these faunas a distinctive stamp. Many of the genera were cosmopolitan, spreading throughout the tropical and temperate parts of the world, thanks to the equable conditions and the closeness of the major continental masses during the period. However, in the cold waters of the southern hemisphere, a new province (the Malvinokaffric Province) was becoming established.

Cheirurina, and Calymenina; they are united partly by the form of their protaspides, which suggests that they may be related to an ancestral ptychopariid. (C) *Cheirurus*, Silurian, 50 mm. (D) *Ananaspis* (Phacopina), Silurian, 30 mm (for *Calymene*, see Figure 3). (E G) Order Proetida compact opisthoparian trilobites with a natant hypostome (conterminant in some derived taxa) and a distinctive form of protaspis. (E) *Tropidocoryphe* (Proetoidea), Devonian, 20 mm. (F) *Bathyurus* (Bathyuroidea), Ordovician, 50 mm. (G) *Phillipsia* (Phillipsiidae), Carboniferous, 20 mm. (H, K, L) Order Asaphida a very large group that includes several suborders, probably derived from a ptychopariid ancestor. A globular protaspis is characteristic, and a median suture is generally present and the rostrum absent in the asaphoid, remopleuridoid, and cyclopygoid groups; in the trinucleoid groups, the eyes are lost and the suture has become marginal. (H) *Ogygiocaris* (Asaphoidea), Ordovician, 70 mm; (K) *Ampyx* (Trinucleoidea (see also Figure 15A)), Ordovician, 40 mm; (L) *Robergia* (Remopleuroidea), Ordovician, 20 mm (for Cyclopygoidea, see Figure 14). (I, J) Corynexochida conterminant Cambrian forms typically with a forward expanding glabella and a large pygidium, which is thought to have given rise to post Cambrian Styginina, including the Illaenoidea. (I) *Amphoton* (Corynexochida), Cambrian, 25 mm; (J) *Ectillaenus* (Illaenoidea), Ordovician, 70 mm (for *Scutellum* (Styginina), see Figure 6). (M, P) Redlichiida large eyed, commonly conterminant forms with many thoracic segments. The Redlichiina are primitive opisthoparian forms, but the Olenellina primitively lacked facial sutures. (M) *Fallotaspis* (Olenellina), Early Cambrian, 60 mm. (P) *Redlichia* (Redlichiina), Cambrian, 40 mm. (N, O) Agnostida small highly derived forms, treated here as trilobites, although excluded from the class by some workers. The Eodiscina, indubitably trilobites, are here included in the Agnostida. (N) *Agnostus* (*Homagnostus*), Cambrian, 5 mm. (O) *Serrodiscus* (Eodiscina), Cambrian, 15 mm. (Q) Ptychopariida a huge and intractable paraphyletic group, primitively opisthoparian and with a natant hypostome, mainly Cambrian in age. It includes the suborders Olenina and Harpina, and may have given rise to other orders, e.g., Asaphida, Phacopida, Proetida. (Q) *Parasolenopleura* (Ptychopariina), Cambrian, 25 mm (for Olenina, see Figure 13; for Harpina, see Figure 15B).

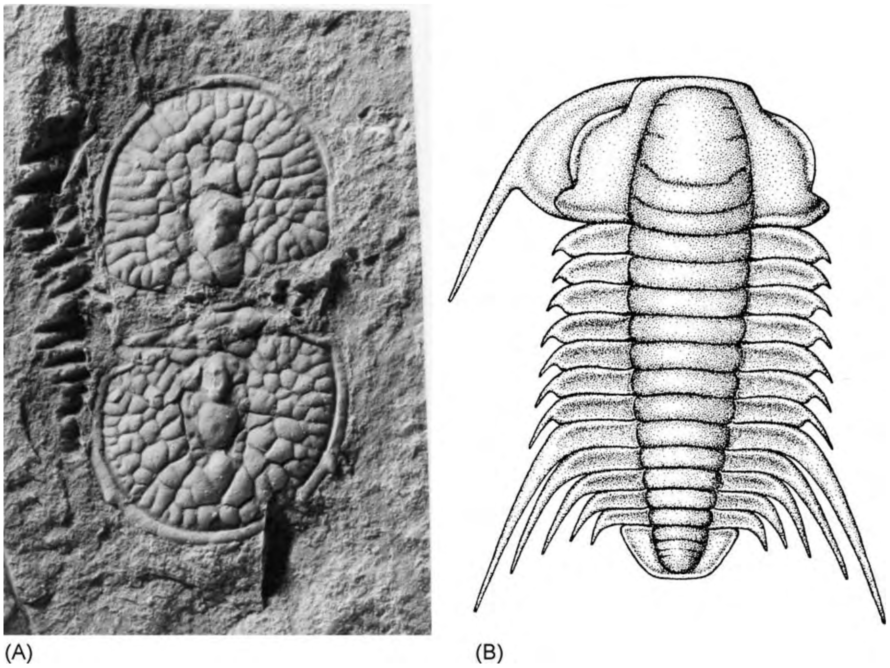


Figure 18 Two globally distributed trilobites that are useful for intercontinental correlation. (A) *Glyptagnostus reticulatus*, characteristic of an early Late Cambrian age; length, 6 mm. (B) *Irvingella*, Late Cambrian; length, 20 mm. Reproduced with permission from Rushton AWA (1983) Trilobites from the Upper Cambrian *Olenus* zone in central England. *Spec. Pal. Paleont.* 30: 107–139, and Rushton AWA (1967) The Upper Cambrian trilobite *Irvingella nuneatonensis*. *Palaeontology* 10: 339–348.

Devonian

Trilobite faunas of the Malvinokaffric Province were widespread in cool-water clastic rocks in the southern hemisphere and were dominated by radiation of the calmoniids (a clade of phacopids), accompanied by homalonotids, aulacopleurids, and odontopleurids. In temperate and tropical latitudes, other families that formerly characterized the Silurian now diversified in the warmer, more varied conditions obtaining during the earlier Devonian, only to vanish later during extinction events precipitated by eustatic changes of sea-level; the cheirurids, calymenids, and lichids became extinct in the Givetian Stage; odontopleurids, harpetids, and styginids in the Frasnian; and the last phacopids at the end of the Famennian. Only Proetida survived to the Carboniferous.

Carboniferous

Despite the extinction of so many clades, the remaining proetid trilobites, represented by four families, diversified to occupy a range of niches in inner-shelf, carbonate mound, and outer-shelf settings during the Early Carboniferous (Mississippian). The morphology of certain outer-shelf taxa is convergent on some of the unrelated, small-eyed (atheloptic) species that characterized comparable environments in the Ordovician and Devonian, and several phillipsiids (Figure 17G) developed features supposedly typical of predators or scavengers – forwardly expanding glabellas and conterminant hypostomes. After the mid-Carboniferous crisis, many forms disappeared and the Late Carboniferous (Pennsylvanian) and Permian trilobites mainly lived in shallow shelf settings.

Permian

Three families of Proetida survived in the Permian, and these, especially the phillipsiids, diversified somewhat in the mid-Permian. However, the major marine regression towards the end of the Permian so limited the habitats available to trilobites that the group finally became extinct just before the end of the Palaeozoic era.

Biostratigraphy

Despite the endemic character of many taxa, trilobites have proved to be of great importance in biostratigraphy, especially for the Cambrian period. The superposition of faunas, empirically determined, has provided one of the most widely applicable means for correlating sedimentary successions, both locally and worldwide. As trilobites dominated the faunas of the Cambrian and are rich in morphological features, they are the obvious choice for characterizing local and regional successions. The Early Cambrian in Morocco, Siberia, and China is divided into about 10 trilobite biozones in each area; likewise, the Middle Cambrian of Australia and China is divided into some 8–10 biozones. In the Late Cambrian of Australia, 17 trilobite zones are recognized, and about 15 in Kazakhstan. In the olenid facies of Scandinavia, there are eight zones divided into about 30 subzones. Correlation between the successions around different continents is difficult, especially in the earlier Cambrian, but becomes more secure in the later half of the Cambrian. The occasional taxa that crossed the barriers between continents (e.g., *Irvingella* and some Agnostida, such as *Glyptagnostus*) are important because they provide key correlative ties (**Figure 18**). In the Ordovician and Silurian, graptolites are used widely for intercontinental correlation, but the pelagic trilobite, *Carolinites* (**Figure 14A**), is of equal value in the earlier Ordovician. Other trilobites retain their value in local successions, especially those with few graptolites, e.g., the Early Ordovician of the Laurentian platform and the Middle Ordovician of Baltica. Elucidation of the complex patterns of pits in the fringes of trinucleid trilobites has enabled detailed correlations locally, e.g., in the Middle and Late Ordovician of Britain. Trilobites also have biostratigraphical value in certain facies of the Devonian and Carboniferous.

See Also

Biozones. **Evolution**. **Fossil Invertebrates:** Arthropods. **Palaeoecology**. **Palaeozoic:** Cambrian; Ordovician; Silurian; Devonian; Carboniferous; Permian. **Trace Fossils**.

Further Reading

Clarkson ENK (1979) The visual systems of trilobites. *Palaeontology* 22: 1 22.

Fortey RA (2000) *Trilobite! Eyewitness to Evolution*. London: HarperCollins.

Fortey RA and Owens RM (1999) Feeding habits in trilobites. *Palaeontology* 42: 429 465.

Harrington HJ, Henningsmoen G, Howell BF, *et al.* (1959) Trilobita. In: Moore RC (ed.) *Treatise on Invertebrate Paleontology*, Part O, *Arthropoda* 1. Lawrence, KS: Geological Society of America and University of Kansas Press.

Jell PA and Adrain JM (2003) Available generic names for frilobites. *Memoirs of the Queensland Museum* 48: 331 553.

Levi Setti R (1993) *Trilobites*, 2nd edn. Chicago and London: University of Chicago Press.

Šnajdr M (1990) *Bohemian Trilobites*. Prague: Czech Geological Survey.

Whittington HB (1992) *Fossils Illustrated 2. Trilobites*. Woodbridge: Boydell Press.

Whittington HB, Chatterton BDE, Speyer SE, *et al.* (1997) *Treatise on Invertebrate Paleontology*, Part O, *Arthropoda* 1, *Trilobita*, revised, vol. 1. Boulder, CO and Lawrence, KS: Geological Society of America, Inc., and University of Kansas Press.

Insects

E A Jarzembowski, University of Reading, Reading, UK and Maidstone Museum and Bentlif Art Gallery, Maidstone, UK

Introduction

Insects, which belong to the taxonomic group Hexapoda (‘six legs’; Table 1), are the most successful organisms on Earth, if biodiversity is measured as a count of the number of species. More than 1.4 million species of insects have been described in the past 250 years, comprising 65% of all known species of life on Earth (Figure 1A). Insects are also the most successful group in the fossil record, if palaeodiversity is measured as a count of the number of families documented, as many palaeontologists do (Figure 2).

How many insects are there? Nobody knows for sure, although all estimates suggest that millions of insect species remain to be described. The great majority of these live in exotic places. It seems unlikely that description will keep pace with global habitat loss and extinction brought about by human

Table 1 Insects systematic position

Taxonomic division	*Group/number*
Kingdom	Animalia
Phylum	Arthropoda
Superclass	Hexapoda (insects in the broad sense)
Order	About 43
Family	Over 1500
Genus	?
Species	3 20 million

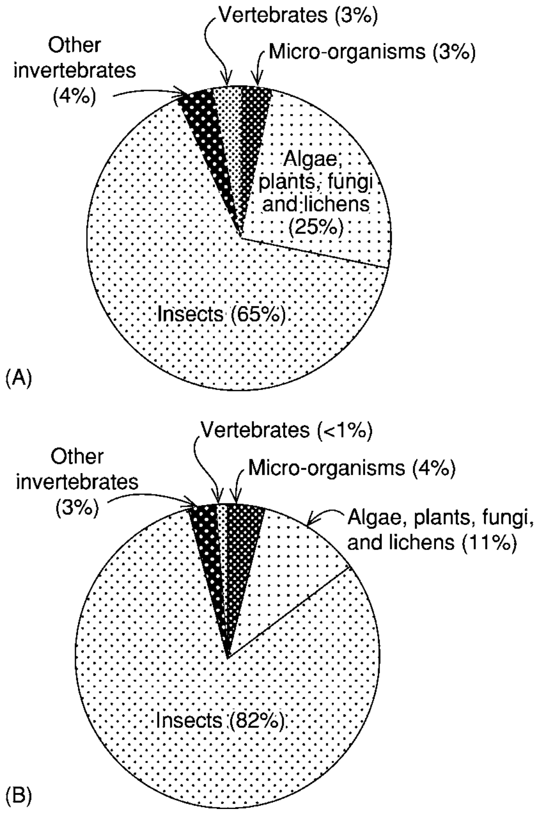


Figure 1 Global biodiversity, showing percentage breakdown for insects and other organisms. (A) Percentages of the app roximate number of known species worldwide (1 454 000). (B) Percentages of the estimated total number of species worldwide, which includes those thought to be undiscovered (65 654 000).

development. Simple extrapolation from the fossil record suggests that the total number of insect species is probably less than 20 million, although some biologists prefer a higher figure. Everyone is agreed, nevertheless, that insects represent over 50% of all known species and that they belong to an exclusive group of hyperdiverse organisms.

Origins

Where do insects come from? Insects are undoubtedly a class of arthropods, or 'joint-legged' animals (Table 1). Arthropods (*see* **Fossil Invertebrates:** Arthropods) also include trilobites, crustaceans (prawns, etc.), chelicerates (spiders, etc.), and myriapods (millipedes and centipedes). Insects show closest relationships to crustaceans (e.g., in the structure of their compound eyes) and myriapods (e.g., in their tubular or tracheate respiratory system) (Figure 3). The exact relationships are currently a subject of debate.

Classification

Insects are divided into two main groups – winged and wingless hexapods (Figure 4). The wingless (apterygotan) insects are a mixed group and only some (silverfish) are thought to share a common ancestor with winged (pterygotan) insects. The pterygotes are divided into two main groups (Figure 5) – those that can fold their wings over the body (Neoptera) and those that cannot (Paleoptera). The neopterans, in turn, can be divided into two groups – those that undergo complete metamorphosis (Holometabola) and those that undergo incomplete metamorphosis (cockroach and grasshopper orders and bug orders, or Polyneoptera and Paraneoptera, respectively); the Polyneoptera and Paraneoptera are also known as exopterygotes, because the wings develop on the outside in the young stages (Figure 6). In contrast, the wings develop inside holometabolous insects, or endopterygotes. For holometabolans, the chrysalis, or pupa, is the 'resting' stage between the caterpillar, maggot, or grub stage and the flying adult stage. The holometabolans are the most diverse insects and apterygotes are the least diverse (Figure 7). Some pterygotes have, however, lost their wings (e.g., fleas). Very high diversities (100 000 or more species) are reached in only four (holometabolous) orders: Coleoptera (beetles), Lepidoptera (moths and butterflies), Hymenoptera (wasps, ants, and bees), and Diptera (true flies).

Geological History

The origin of insects is a mystery, the Cambrian Burgess Shale arthropods being too early to cast any light on the subject. The oldest definite hexapod is *Rhyniella praecursor* from the Early Devonian Rhynie Chert. *Rhyniella praecursor* is a springtail belonging to the living apterygote order Collembola. The earliest true insect is currently considered to be *Rhyniognatha hirsti*, also from the Rhynie Chert. In the latest Lower Carboniferous and Upper Carboniferous there is evidence of the radiation of the pterygotes, including paleopterans and polyneopterans. These insects were the world's first flying animals, long before vertebrates took to the air. In the succeeding Permian, the paraneopterans and holometabolans became established. Insects are essentially terrestrial organisms, but the first definite freshwater forms appeared in the Permian. After a setback in the Early Triassic extinction (Figure 8), insects regained their ordinal strength by the Tertiary, establishing some new innovations on the way, e.g., evolving parasitic and parasitoid forms as well as insect societies (Figure 9). The extinction at the start of the Mesozoic seems to have

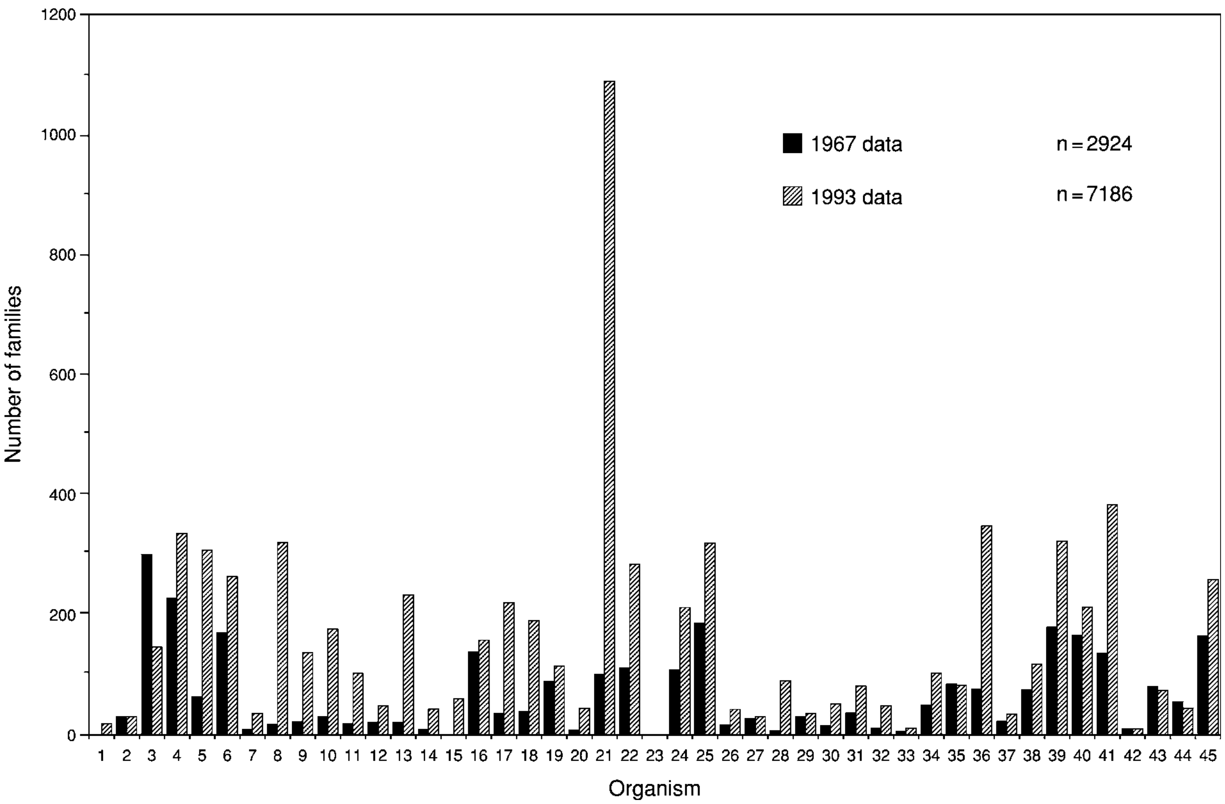


Figure 2 Palaeodiversity of organisms. Key to numbers: 1, bacteria and blue green algae; 2, fungi; 3, other algae; 4, single cell organisms; 5, sponges; 6, corals, etc.; 7, chitons, etc.; 8, snails; 9, nautiluses; 10, ceratites; 11, ammonites; 12, belemnites; 13, bivalves and tusk shells; 14, uncertain molluscs; 15, segmented worms; 16, trilobites; 17, spiders, etc.; 18, crustaceans (excluding seed shrimps); 19, seed shrimps; 20, millipedes, etc.; 21, insects; 22, lamp shells; 23, phoronids; 24, moss animals; 25, sea urchins, etc.; 26, primitive chordates; 27, graptolites; 28, problematica; 29, miscellaneous; 30, conodonts; 31, lampreys, etc.; 32, cyclostomes; 33, primitive fish; 34, sharks, etc.; 35, primitive bony fish; 36, advanced bony fish; 37, more bony fish; 38, amphibians; 39, reptiles; 40, birds; 41, mammals; 42, mosses, etc.; 43, ferns, etc.; 44, seed plants (excluding 45, flowering plants).

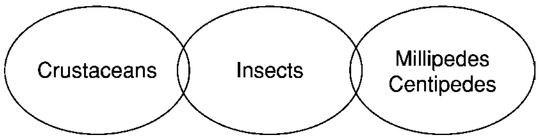


Figure 3 Relationship of insects with other arthropods.

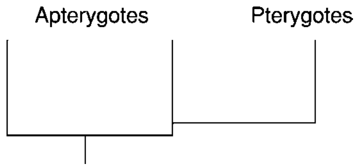


Figure 4 Major insect groups; apterygotes include springtails and silverfish.

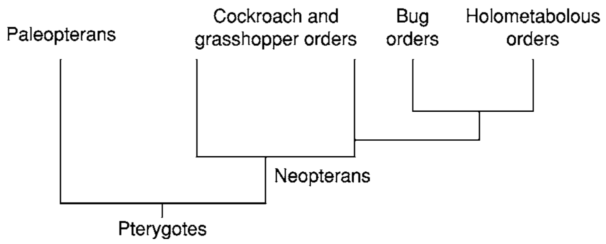


Figure 5 The two main groups of pterygotes are the paleopterans and the neopterans, which are further divided into subgroups. Paleopterans include dragonflies and mayflies. Cockroaches and grasshoppers, along with stoneflies, stick insects, crickets, locusts, earwigs, termites, and praying mantises comprise the polyneopteran orders. Paraneopteran orders include bugs and lice. Holometabolous orders include beetles, lacewings, wasps, ants, bees, caddisflies, moths, butterflies, flies, and fleas.

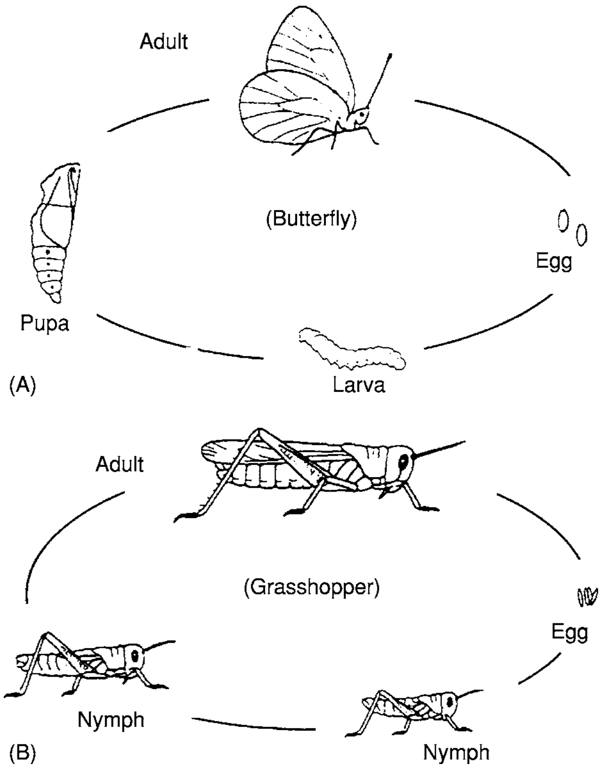


Figure 6 Life cycles of insects. (A) Endopterygotes undergo a complete metamorphosis. (B) Exopterygotes undergo an incomplete metamorphosis.

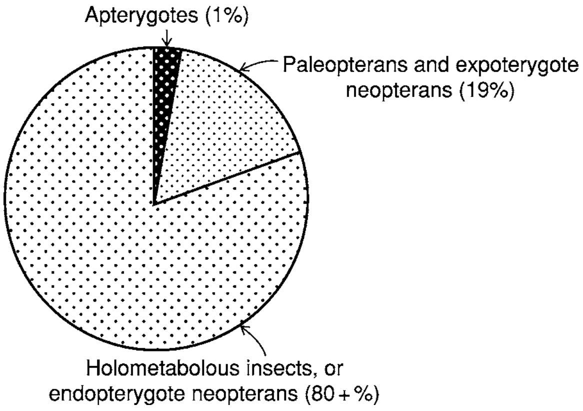


Figure 7 Approximate biodiversity of major insect groupings, showing percentages of primitively wingless insects (apterygotes), primitively winged insects with incomplete metamorphosis (paleopterans and exopterygote neopterans), and primitively winged insects with complete metamorphosis (holometabolous insects, or endopterygote neopterans).

been the biggest in insect history, although losses were not really catastrophic. Indeed, the successful order Diptera (true flies), which were opportunists ('specials'), arose in the Triassic (Figure 8). The extinction showed, however, that even hyperdiverse organisms are affected by global environmental change. The principal groups (orders) of insects in the fossil record are outlined in Table 2.

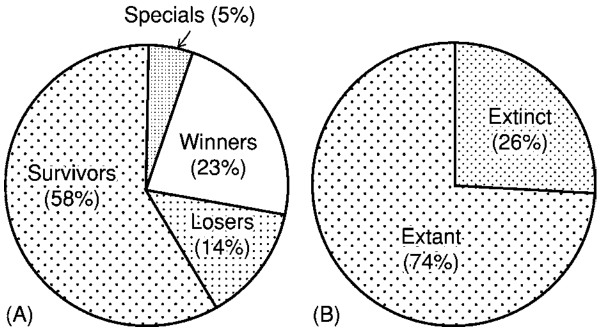


Figure 8 Insect orders during the (A) Triassic extinction; (B) Phanerozoic composition. Opportunists are categorized as 'specials'.

Collecting and Documentation

Fossil insects are more common than is generally supposed, especially their disarticulated remains. They occur in a variety of sedimentary environments, including marine and non-marine deposits, in both organic and fine-grained clastic rocks. Insects often occur in early diagenetic concretions, including ferruginous and phosphatic ones, and in calcareous mudstones. Some of the best preserved insects occur in amber from the Lower Cretaceous onwards. More unusual modes of preservation include pyritized or silicified insects and inclusions in gypsum crystals. Insects are found as trace fossils as well as body fossils. Fossil insects have been used in palaeoenvironmental reconstruction in addition to phylogenetic analysis, especially in Quaternary deposits.

Just as there are millions of insects to be described in today's hot countries, so there are thousands of fossil species to be described in places that were once warmer. The process of collection and documentation of fossil insects is of scientific as well as cultural value. It is possible to find more new fossil insect species on one field trip than in an entire lifetime of collecting of the more popular fossil groups (e.g., vertebrates). Fossil insects are thus ideal for satisfying the goal of finding something new (and they usually require less storage space on account of their small size). Knowledge of the pre-Quaternary insects has largely been forgotten since the era of the pioneer Victorian geologists and naturalists. There is now, however, a revival of interest, the wider search for early (Carboniferous pre-Namurian)

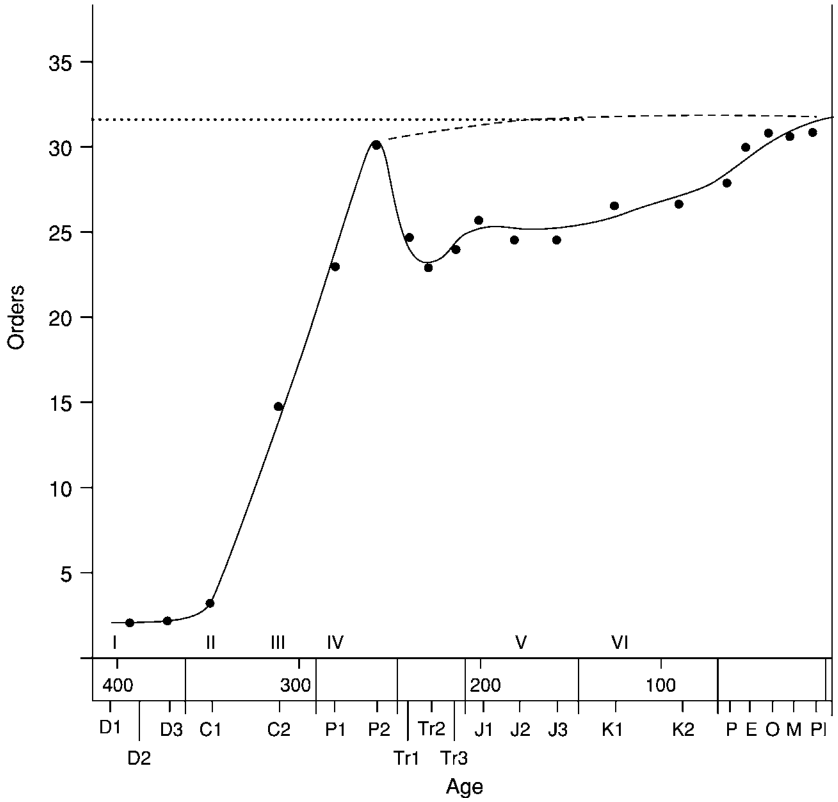


Figure 9 Orders through time and key events. Dashed line represents extrapolation, dots represent equilibrium (saturation) value. Key: Pl, Pliocene; M, Miocene; O, Oligocene; E, Eocene; P, Paleocene; K, Cretaceous; J, Jurassic; Tr, Triassic; P, Permian; C, Carboniferous; D, Devonian. Numerals 1, 2, and 3 represent Lower, Middle, and Upper subperiods and epochs; I, earliest hexapods; II, earliest pterygotes; III, near modern range of plant feeding strategies; IV, evolution of holometabolous insects; V, earliest insectan parasites and parasitoid radiation; VI, earliest amberized and definite social insects.

Table 2 Principal groups of insects in the fossil record

Principal group	*Order*	*Age/description*
Apterygota	Collembola (springtails)	Lower Devonian Recent
	Diplura (two pronged bristletails)	Upper Carboniferous Recent
	'Thysanura' (three tailed bristletails)	This order is split into two groups: Archaeognatha (Middle Devonian? Recent) and Zygentoma (Upper Carboniferous Recent; e.g., silverfish). The extinct order Monura (Upper Carboniferous Permian) with single 'tails' is probably related to the Archaeognatha, whereas Zygentoma is related to the Pterygota
Pterygota		
Paleoptera	Palaeodictyopteroid group	An Upper Carboniferous Upper Permian group of three or four extinct orders (Palaeodictyoptera, Permothemistida, Megasecoptera, Diaphanopterodea) with beak like mouthparts and an additional pair of 'winglets'
	Ephemeroptera (mayflies)	Upper Carboniferous Recent
	Protodonata	Upper Carboniferous Triassic, including the giant dragonflies with wingspans up to 70 cm, the largest insects of all time
	Odonata (dragonflies, damselflies)	Upper Carboniferous Recent
Neoptera	'Protorthoptera'	Lower Carboniferous Triassic. A taxonomic wastebasket of early neopterans

Continued

Table 2 Continued

Principal group	*Order*	*Age/description*
Polyneoptera	Plecoptera (stoneflies)	Lower Permian Recent
	Embioptera (web spinners)	Lower Permian Recent
	Phasmatodea (stick insects)	Upper Permian Recent
	Orthoptera (crickets, grasshoppers, katydids, locusts)	Upper Carboniferous Recent
	Titanoptera	Extinct Triassic order allied to Orthoptera
	Grylloblattodea (ice bugs)	Lower Permian Recent
	Mantophasmatodea	Eocene Recent
	Protelytroptera	Extinct Permian earwiglike insects
	Dermaptera (earwigs)	Lower Jurassic Recent
	Miomoptera	Extinct Upper Carboniferous Lower Jurassic insects
	Blattodea (cockroaches)	Upper Carboniferous Recent
	Isoptera (termites)	Lower Cretaceous Recent
	Mantodea (praying mantises)	Lower Cretaceous Recent
	Caloneurodea	Extinct Upper Carboniferous Permian insects
Paraneoptera	Zoraptera (angel insects)	Oligocene Recent
	Psocoptera (bark and book lice)	Lower Permian Recent
	Phthiraptera (lice)	Eocene Recent
	Thysanoptera (thrips)	Lower Permian Recent
	Hemiptera (true bugs)	Upper Carboniferous Recent
Holometabola (Oligoneoptera)	Glosselytrodea	Extinct Lower Permian Upper Jurassic insects
	Strepsiptera (stylopids)	Eocene Recent
	Coleoptera (beetles)	Lower Permian Recent
	Raphidioptera (snake flies)	Upper Permian Recent
	Megaloptera (alder flies)	Lower Permian Recent
	Neuroptera (lacewings)	Lower Permian Recent
	Hymenoptera (wasps, ants, bees)	Upper Triassic Recent
	Trichoptera (caddis flies)	Lower Permian Recent
	Lepidoptera (moths, butterflies)	Lower Jurassic Recent
	Diptera (true flies)	Lower Triassic Recent
	Siphonaptera (fleas)	Lower Cretaceous Recent
	Mecoptera (scorpionflies)	Lower Permian Recent

insects being of paramount importance. For logistic reasons, the study of fossil insects (palaeoentomology) relies on international co-operation; to facilitate this objective and to promote knowledge, the International Palaeoentomological Society was founded in 2001.

See Also

Fossil Invertebrates: Arthropods.

Further Reading

In addition to the print literature, several groups have web sites that are sources of information about insects: the Arthropod Laboratory of the Russian Academy of Sciences (http://www.palaeoentomolog.ru), the International Palaeoentomological Society (http://www.cwru.edu/affil/fossilinsects), and the University of Barcelona's Meganeura Palaeoentomological Newsletter (http://www.ub.es/dpep/meganeura/meganeura.htm).

Benton MJ (ed.) (1993) *The Fossil Record 2*. London: Chapman & Hall.

Carpenter FM (1992) *Superclass Hexapoda. Treatise on Invertebrate Paleontology, Part R, Arthropoda 4, 3 & 4.*

Earl of Cranbrook (1996) The scientific value of collections. *Sarawak Museum Journal* 50(71): 73 86.

Jarzembowski EA (2001) Insect "bioerosion". *Acta Geologica Leopoldensia* 26(52/53): 161 164.

Jarzembowski EA (2003) Palaeoentomology: towards the big picture. *Acta Zoologica Cracoviensia, Krakow* 46(suppl.): 25 36.

Jarzembowski EA and Ross A (1993) The geological record of insects. *Geology Today* 9(6): 218 223.

Rasnitsyn AP and Quicke DLJ (eds.) (2002) *History of Insects*. Dordrecht: Kluwer Academic Publishers.

Wilson EO (1992) *The Diversity of Life*. Cambridge, MA: Harvard University Press.

Brachiopods

D A T Harper, Geologisk Museum, Copenhagen, Denmark

Introduction

The brachiopods or lamp-shells form a distinctive and diverse group of marine, mainly sessile, benthic invertebrates with a long and varied geological history dating back to the Early Cambrian. Over 12 000 fossil species and approximately 350 living species have been reported, belonging to nearly 6000 genera. Brachiopods consist of two shells or valves, the dorsal and ventral, which grow by accretion, recording the ontogenetic history of the animal. The group is distinguished by the possession of a ciliated feeding organ (the lophophore), together with, in most taxa, a fleshy attachment stalk or pedicle. The phylum has a characteristic set of muscles that act in opposition to open and close the valves, together with a variety of skeletal structures that support both the lophophore and musculature. Despite their minimalist metabolism, requiring little food and oxygen, the phylum has adopted a huge range of morphologies and a wide range of ecological strategies during a history of nearly 600 million years.

Brachiopod Animal

The brachiopod animal is enclosed by two, morphologically different shells or valves, opened and closed by a variety of muscles (Figure 1). In contrast with the bivalves (*see* **Fossil Invertebrates:** Bivalves), where the right valve is a mirror image of the left, the plane of symmetry bisects both valves perpendicular to the commissure. The larger of the two valves is generally named the ventral or pedicle valve; in many brachiopods, the fleshy stalk or pedicle emerges from the apex of this valve and attaches the animal to the seabed. The pedicle can vary from a thick, fleshy stalk to a bunch of delicate, thread-like strands, which can anchor the brachiopod in fine mud. Some brachiopods lost their pedicles during ontogeny and adopted a free-living mode of life, lying recumbent on or partially in the sediments on the seafloor. The dorsal or brachial valve contains the extendable food-gathering organ or lophophore, together with its supports. A number of styles of lophophore have evolved. The earliest growth stage, the trocholophe, is an incomplete ring of filaments, still retained by the paedomorphic microbrachiopod *Gwynia*; by the schizolophe stage, a bilobed outline has developed, which probably characterized many of the smaller Palaeozoic taxa. The more complex plectolophe, ptycholophe, and spirolophe lophophore types are characteristic of the more advanced articulated brachiopods.

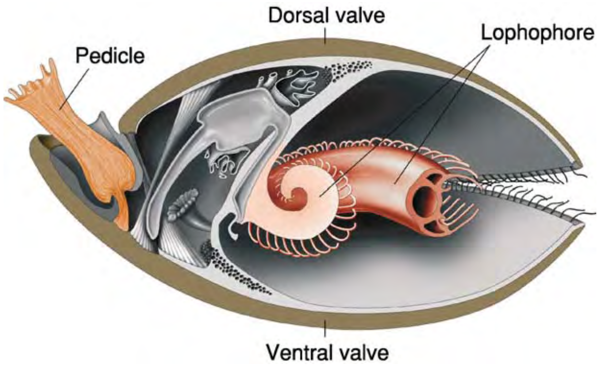


Figure 1 The brachiopod animal showing the key apomorphies of the phylum: the lophophore, pedicle, and dorsal and ventral valves. Redrawn and modified from Kaesler RL (1997).

The linguliformeans (Figure 2A) have organophosphatic shells with pedicles that either emerge between both valves or through a foramen. The shells develop from a planktotrophic larval stage without mantle reversal; the group is characterized by an alimentary tract ending in an anus. In the lingulates, the opening and closing of the valves is achieved by a complex system of muscles and the pedicle emerges between both valves. Some authors have suggested that the withdrawal of the soft parts posteriorly causes a space problem that can force the valves apart; relaxation allows the animal to expand again forwards, allowing the valves to close. The paterinates are the oldest group of brachiopods, appearing in the lowest Cambrian Tommotian Stage. Although linked to the other linguliformeans on the basis of an organophosphatic shell substance, the shell structure of the group is quite different, and the shells have true interareas, delthyria, notothyria, and apparently had functional diductor muscle systems.

The craniiformeans (Figure 2B) contain a diverse, yet probably monophyletic, group of morphologies centered on *Crania*, but including *Craniops* and the bizarre trimerellids. The shells consist of organocarbonate, and the animal developed separate dorsal and ventral mantle lobes after the settlement of a nektobenthonic larval stage.

The rhynchonelliformeans (Figure 3) have a pair of calcitic valves that contain a fibrous secondary layer,

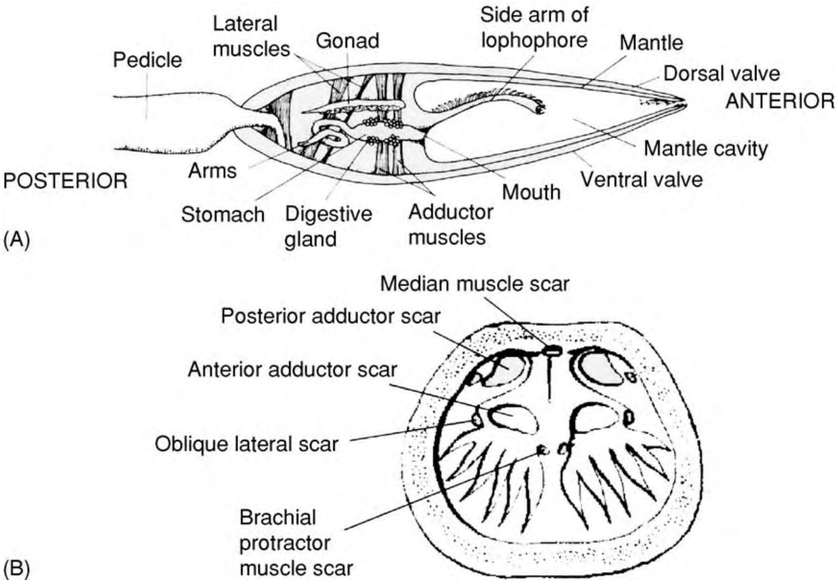


Figure 2 (A) Linguliformean morphology based on *Lingula*. (B) Craniiformean morphology based on *Neocrania*.

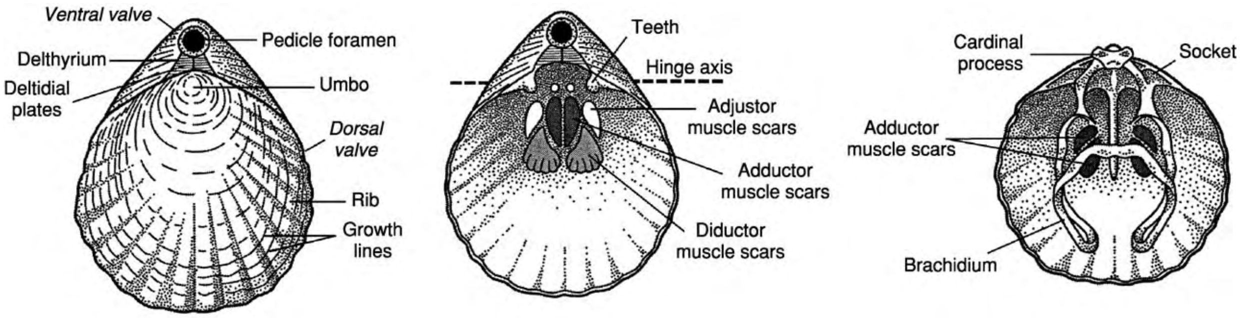


Figure 3 Rhynchonelliformean morphology showing, from left to right, features of the exterior, ventral, and dorsal valves; based on the living genus *Magellania*. Adapted from Clarkson ENK (1998) *Invertebrate Palaeontology and Evolution*, 4th edn. Cambridge: Blackwell Science.

with variable convexity, hinged posteriorly and opening anteriorly along the commissure. The mantle lobes are fused posteriorly, where the interareas are secreted; their margins form the hinge between the ventral and dorsal valves. Articulation was achieved by a pair of ventral teeth and dorsal sockets and the valves were opened and closed by opposing diductor and adductor muscle scars. In the majority of rhynchonelliformeans, the valves were attached to the substrate by a pedicle, developed from a larval rudiment, and emerging through a foramen in the delthyrial region. The subphylum contains five classes: the Chileata, the Obolellata, the Kutorginata, the Strophomenata, and the Rhynchonellata. Already by the Early Cambrian, representatives of four of the five classes were present. However, the last two classes, containing over 1500 and 2700 genera, respectively, dominated Phanerozoic brachiopod faunas.

Brachiopod Shell

The brachiopod shell is a multilayered complex of both organic and inorganic material that has proven to be of fundamental importance in the classification of the phylum. The shells of most rhynchonelliformean brachiopods consist of three layers (Figure 4). The outer layer (periostracum) is organic, whereas underneath are the mineralized primary and secondary layers. These layers are sequentially secreted by cells within the generative zone of the mantle, forming first a gelatinous sheath, followed by the organic periostracum, and then the granular calcite of the primary layer; the subsequent secondary layer is thicker, composed of calcite fibres and, in some brachiopods, a third prismatic layer is secreted. There are a number of variations of this basic template. The linguliformeans, for example,

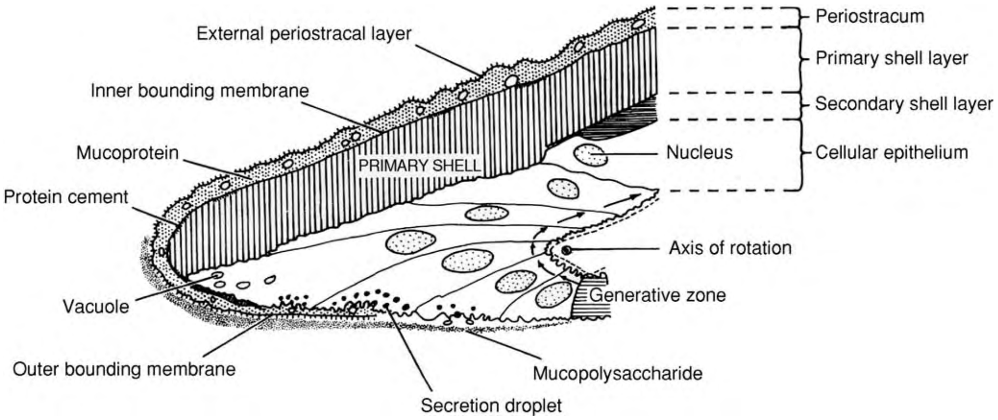


Figure 4 Brachiopod shell structure based on sections through living *Notosaria*. Adapted from Clarkson ENK (1998) *Invertebrate Palaeontology and Evolution*, 4th edn. Cambridge: Blackwell Science.

have phosphatic material combined into the shell fabric. The rhynchonelliformean brachiopods are composed of low-magnesian calcite; these organocarbonate shells may have fibrous, laminar, or cross-bladed laminar shell fabrics in their secondary layers.

Many shells are perforated by small holes or punctae, in life holding finger-like extensions of the mantle or caeca. Some strophomenides have pseudopunctae, with fine inclined rods or taleolae embedded in the shell fabric.

Classification and Phylogeny

The traditional split of the phylum into the Inarticulata and Articulata has been discarded in favour of three subphyla: the Linguliformea, Craniiformea, and Rhynchonelliformea (Table 1). All three have quite different body plans and shell fabrics (Figure 5). The linguliformeans contain five orders united by organophosphatic shells; the inclusion of the paterinides is the most problematic, as the group shares some morphological characters with the rhynchonelliforms. The craniiformeans include three rather disparate groups with quite different morphologies, but which together possess an organocarbonate shell. Most scientists accept eight articulated orders, mainly based on the nature of the cardinalia and the morphology of the other internal structures associated with the attachment of muscles and the support of the lophophore; recently, the more deviant chileides, obolellides, and kutorginides have been added to the subphylum. In addition, the articulated taxa have been split into those with deltidiodont (simple) and cyrtomatodont (complex) dentitions; the former group includes the orthides and strophomenides, and the latter includes the spire bearers.

Cladistic-based investigations have developed a phylogenetic framework for the phylum, supporting the three subphyla (Figure 6); their defining characters are based on shell structure and substance. The mutual relationships between these groups are still unclear, as are the relationships between the many primitive articulated and nonarticulated groups that appeared during the Cambrian explosion. During and following the Ordovician radiation, the relationships between taxa are more clearly defined.

Origin and Affinities

To date, no definite brachiopods have been described from Precambrian rocks. Molecular data and the inclusion of the brachiopods within the other lophophorates (the bryozoans (*see* **Fossil Invertebrates:** Bryozoans) and phoronids) suggest that the phylum belongs to the protostomes. This would support the evolution of the Brachiopoda from a slug-like ancestor, such as *Halkieria* (with anterior and dorsal shells), by the folding of the mantle along a line perpendicular to the length of the animal. On the other hand, morphological evidence suggests that brachiopods are in fact deuterostomes. The ancestral brachiopod may have possessed a planktotrophic larva with a straight gut with the adult form attached by a pedicle. The group diverged to form the organophosphatic linguliformeans with a displaced anus, and the craniiformeans and rhynchonelliformeans with calcitic shells; the rhynchonelliformeans lost the anus and acquired complex hinging mechanisms.

Ecology

Living and fossil brachiopods have developed a wide range of life styles (Figure 7). The majority were attached by a pedicle cemented to a hard substrate or rooted into soft sediment. A number of quite different inarticulated and articulated taxa were cemented to the substrate, whereas some groups

Table 1 The Linguliformea, Craniiformea, and Rhynchonelliformea

Subphylum	*Order*	*Key characteristics*	*Stratigraphical range*
Linguliformea	Lingulida	Spatulate valves with pedicle usually emerging between both shells	Cambrian Recent
	Acrotretida	Micromorphic forms with conical ventral valve; dorsal valve with platforms	Cambrian Devonian
	Discinida	Subcircular shells with conical ventral valve and distinctive pedicle foramen	Ordovician Recent
	Siphonotretida	Subcircular, biconvex valves with spines and elongate pedicle foramen	Cambrian Ordovician
	Paterinida	Strophic shells with variably developed interareas	Cambrian Ordovician
Craniiformea	Craniida	Usually attached by ventral valve; dorsal valve with quadripartite muscle scars	Ordovician Recent
	Craniopsida	Small oval valves with internal platforms and marked concentric growth lines	Ordovician Carboniferous
	Trimerellida	Commonly gigantic, aragonitic shells, with platforms and umbonal cavities	Ordovician Silurian
Rhynchonelliformea	Chileida	Strophic shells lacking articulatory structures but with umbonal perforation	Cambrian
	Dictyonellida	Biconvex valves with large umbonal opening commonly covered by a colleplax	Ordovician Permian
	Naukatida	Biconvex shells with articulatory structures and apical foramen	Cambrian
	Obolellida	Oval valves with primitive articulatory structures	Cambrian
	Kutorginida	Strophic valves with interareas but lacking articulatory structures	Cambrian
	Orthotetida	Biconvex shells, commonly cemented, with bilobed cardinal process	Ordovician Permian
	Billingsellida	Usually biconvex with transverse teeth and simple cardinal process	Cambrian Ordovician
	Strophomenida	Concavoconvex, usually bilobed cardinal process; recumbent life mode; cross laminar shell structure with pseudopunctae	Ordovician Permian
	Productida	Concavoconvex valves with complex cardinalia; recumbent or cemented life mode; often with external spines	Ordovician Triassic
	Protorthida	Well developed interareas, primitive articulation and ventral free spondylium	Cambrian Devonian
	Orthida	Biconvex, usually simple cardinal process; pedunculate; delthyria and notothyria open	Cambrian Permian
	Pentamerida	Biconvex, rostrate valves with cruralia and spondylia variably developed	Cambrian Devonian
	Rhynchonellida	Usually biconvex, rostrate valves with variably developed crurae	Ordovician Recent
	Atrypida	Biconvex valves with dorsally directed spiralia and variably developed jugum	Ordovician Devonian
	Athyridida	Usually biconvex valves with short hingeline and posterolaterally directed spiralia	Ordovician Jurassic
	Spiriferida	Wide strophic valves with laterally directed spiralia; both punctate and impunctate taxa	Ordovician Jurassic
	Thecideida	Small, strophic shells with complex spiralia including brachial ridges and median septum	Triassic Recent
	Terebratulida	Biconvex valves with variably developed long or short loops	Devonian Recent

evolved clasping spines and possibly extended mantle fibres to help stabilize their shells. In a number of groups, the pedicle atrophied during ontogeny. Many taxa thus developed strategies involving inverted, pseudoinfaunal, and recumbent life modes; a number lived in cosupportive clusters and others mimicked corals. Not all brachiopods were sessile; a few, such as *Lingula*, adopted an infaunal life style, whereas the articulated forms, *Camerisma* and *Magadina*, were semi-infaunal.

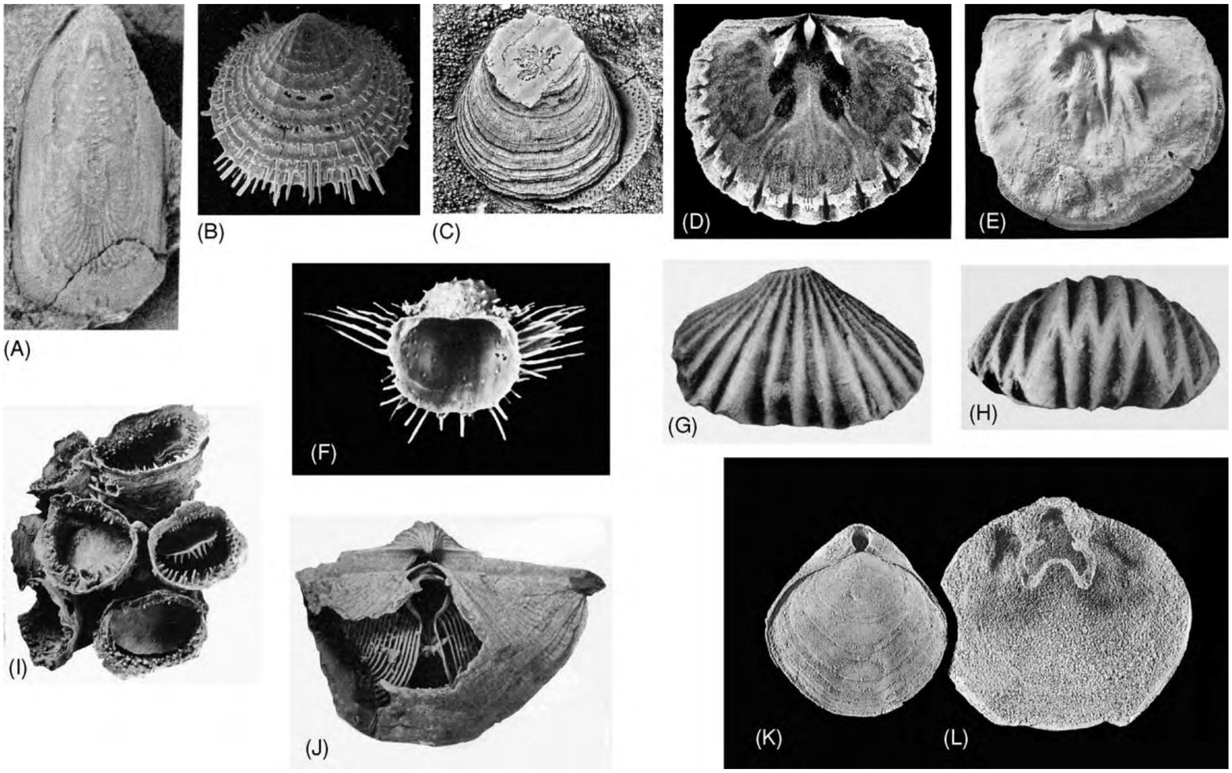


Figure 5 Range of brachiopod morphologies: (A) *Pseudolingula* (Ordovician lingulide: ×2); (B) *Nushbiella* (Ordovician siphonotretide: ×10); (C) *Crania* (Palaeogene craniide: ×2.5); (D) *Sulevorthis* (Ordovician orthide: ×3); (E) *Rafinesquina* (Ordovician strophomenide: ×1/3); (F) *Grandaurispina* (Permian productide: ×1); (G, H) *Rostricellula* (Ordovician rhynchonellide: ×3); (I) *Cyclacantharia* (Permian richthofeniid: ×0.5); (J) *Neospirifer* (Permian spiriferide: ×2/3); (K,L) *Seymourella* (Palaeogene terebratulide: ×1).

Throughout the Phanerozoic, the brachiopods have participated in a spectrum of level-bottom, benthic palaeocommunities. Pioneer studies on Silurian brachiopods suggested that their palaeocommunities were depth related (Figure 8A). The onshore–offshore assemblages of the *Lingula*, *Eocoelia*, *Pentamerus*, *Stricklandia*, and *Clorinda* palaeocommunities have been amplified and modified to form the basis of Benthic Assemblage (BA) zones 1–5, ranging from intertidal environments to the edge of the continental slope; more basinal environments are included in an extra BA 6. Parallel studies on Mesozoic brachiopods have, on the other hand, suggested that brachiopod-dominated palaeocommunities were controlled by substrate rather than depth (Figure 8B). Clearly, a combination of these and other factors in reality controlled the distributions of the Brachiopoda in a complex system of suspension-feeding guilds.

Brachiopods have also acted as substrates for a variety of small epifaunal animals. The progressive and sequential colonization of Devonian spiriferids by *Spirorbis*, *Hederella*, *Paleschara*, and *Aulopora* marks the development of sere and climax palaeocommunities on a brachiopod shell. It has been suggested that such animals congregated adjacent to the inhalant currents on the median parts of the anterior commissure. An alternative hypothesis reverses the direction of flow through the brachiopod mantle cavity, and thus these commensal organisms took advantage of waste being ejected from the brachiopod.

Geographical Distribution

The biogeographical patterns of the linguliformean brachiopods were quite different from those of the craniiformeans and rhynchonelliformeans. The former had planktotrophic larval phases with a facility for wide dispersal; in contrast, the lecithotrophic larvae of the latter were short-lived. Brachiopods dominated the benthos of the Palaeozoic evolutionary fauna. Nevertheless, Cambrian brachiopods were organized into tropical and natal realms, where linguliformeans developed widespread distributions in shelf and slope settings, and rhynchonelliformeans were more diverse in the tropics, preferring shallow-water carbonate and mixed carbonate–siliciclastic environments. Ordovician brachiopods generally showed a decreased provincialism during

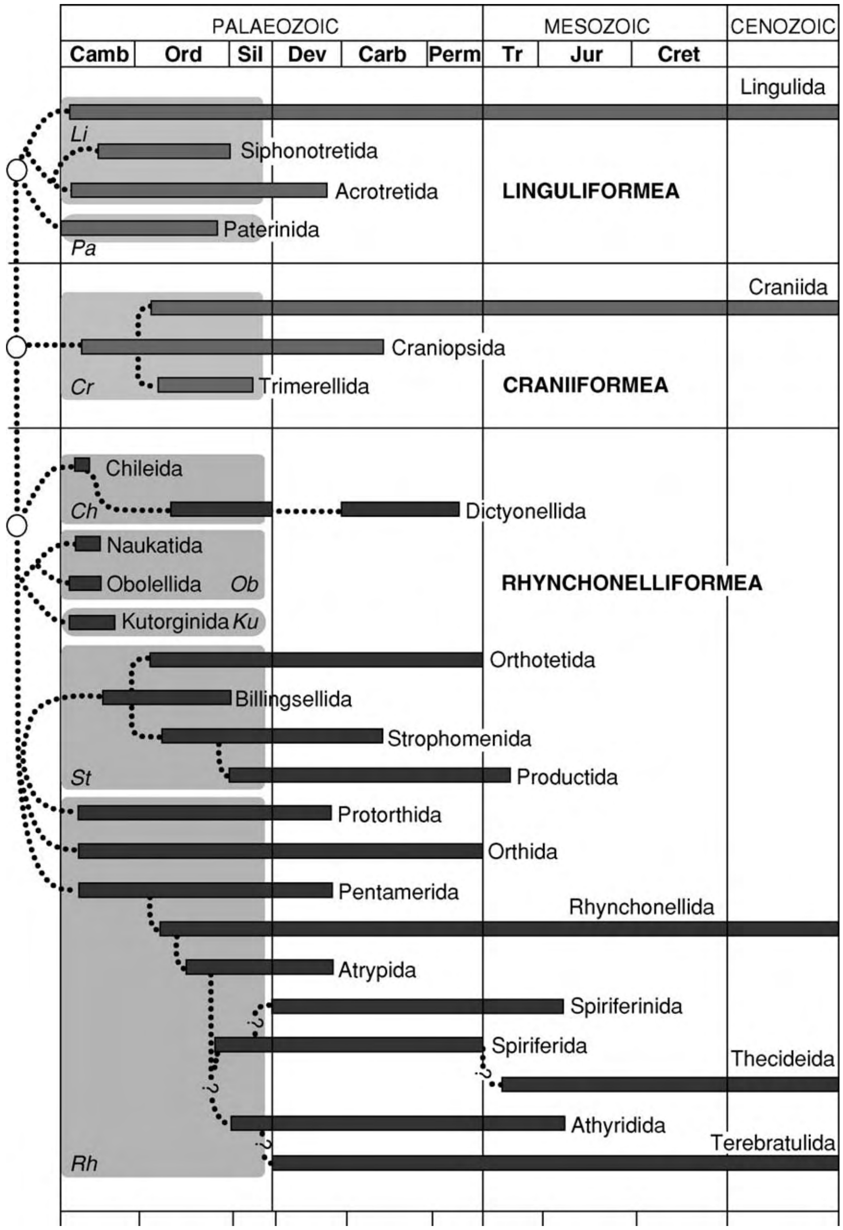


Figure 6 Phylogeny and stratigraphical distribution of the brachiopod orders.

the period; provinciality was most marked during the Early Ordovician, when a range of platform provinces associated with the continents of Baltica, Gondwana, Laurentia, and Siberia were supplemented by loci of endemism associated with a range of microcontinents and volcanic arcs and island complexes. Provincialism was reduced during the Silurian with the close proximity of many major continents; by the Wenlock, however, two broad provinces, the cool-water *Clarkeia* and the mid-latitudinal *Tuvaella* faunas, emphasize an increasing endemism, climaxing during the Ludlow and Prídolí epochs. Provinciality was particularly marked during the mid-Devonian, coincident with peak diversities in the phylum. Although clear biogeographical patterns continued into the Carboniferous, the Permian was characterized by high degrees of provinciality probably associated with steep climatic gradients.

During the Triassic, brachiopod faunas, following an interval of cosmopolitan disaster taxa, were organized into Boreal (high-latitude) and Tethyan (low-latitude) realms. This pattern continued throughout the Mesozoic with loci of endemics and occasional modifications due to ecological factors, such as the circulation of ocean currents and the

LIFE STYLE	BRACHIOPOD TAXA
ATTACHED BY PEDICLE	
Epifaunal - hard substrate[1] (plenipedunculate)	Orthides, rhynchonellides, spiriferides and terebratulides
Epifaunal - soft substrate[2] (rhizopedunculate)	*Chlidonophora* and *Cryptopora*
Cryptic	*Argyrotheca and Terebratulina*
Interstitial	Acrotretides and *Gwynia*
CEMENTED	*Craniops* and *Schuchertella*
ENCRUSTING[3]	Craniids and disciniids
CLASPING SPINES[4]	*Linoproductus* and *Tenaspinus*
MANTLE FIBRES	Orthotetoids
UNATTACHED	
Cosupportive[5]	Pentamerids and trimerellids
Coral-like[6]	Gemmellaroids and richthofeniids
Recumbent	Strophomenides
Pseudoinfaunal[7] and Inverted[8]	*Waagenoconcha* and *Marginifera*
Free-living[9,10]	*Cyrtia, Chonetes, Neothyris* and *Terebratalla*
MOBILE	
Infaunal[11]	Linguloids
Semi-infaunal[12]	*Camerisma* and *Magadina*

ADAPTATIONS

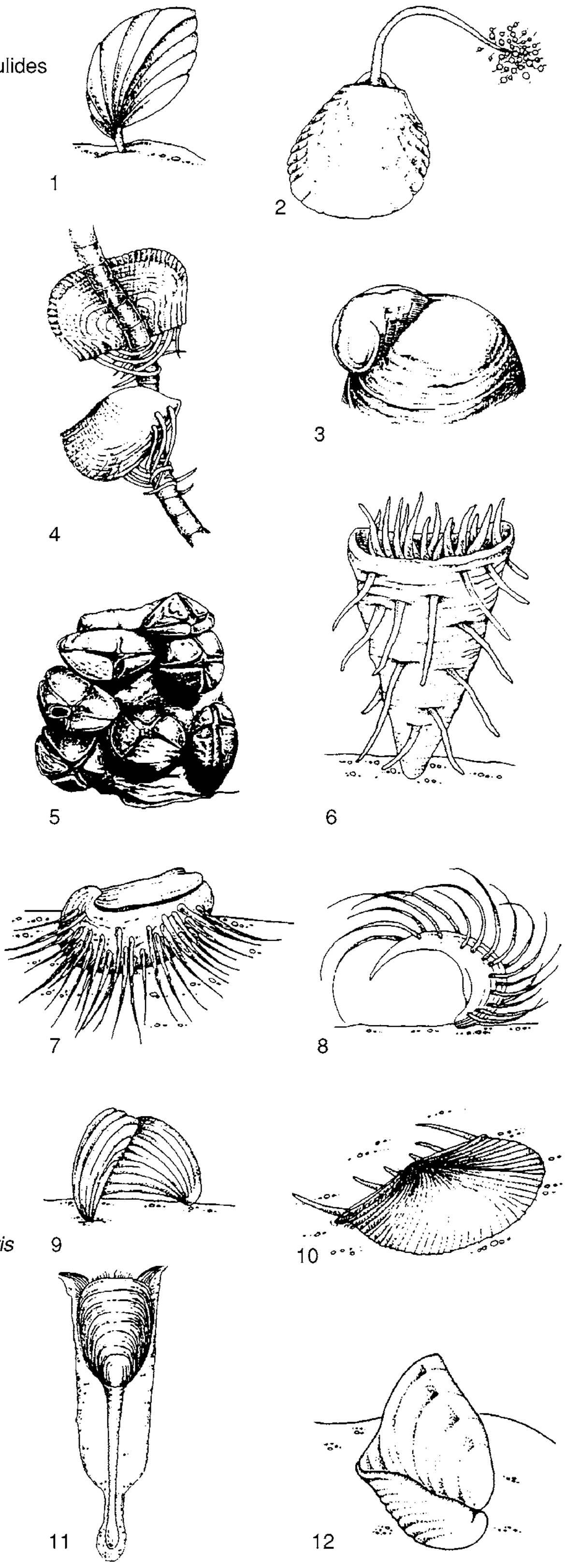


Figure 7 Brachiopod life styles. Reprinted from Harper DAT and Moran R (1997) Brachiopod life styles. *Geology Today* 13: 235 238.

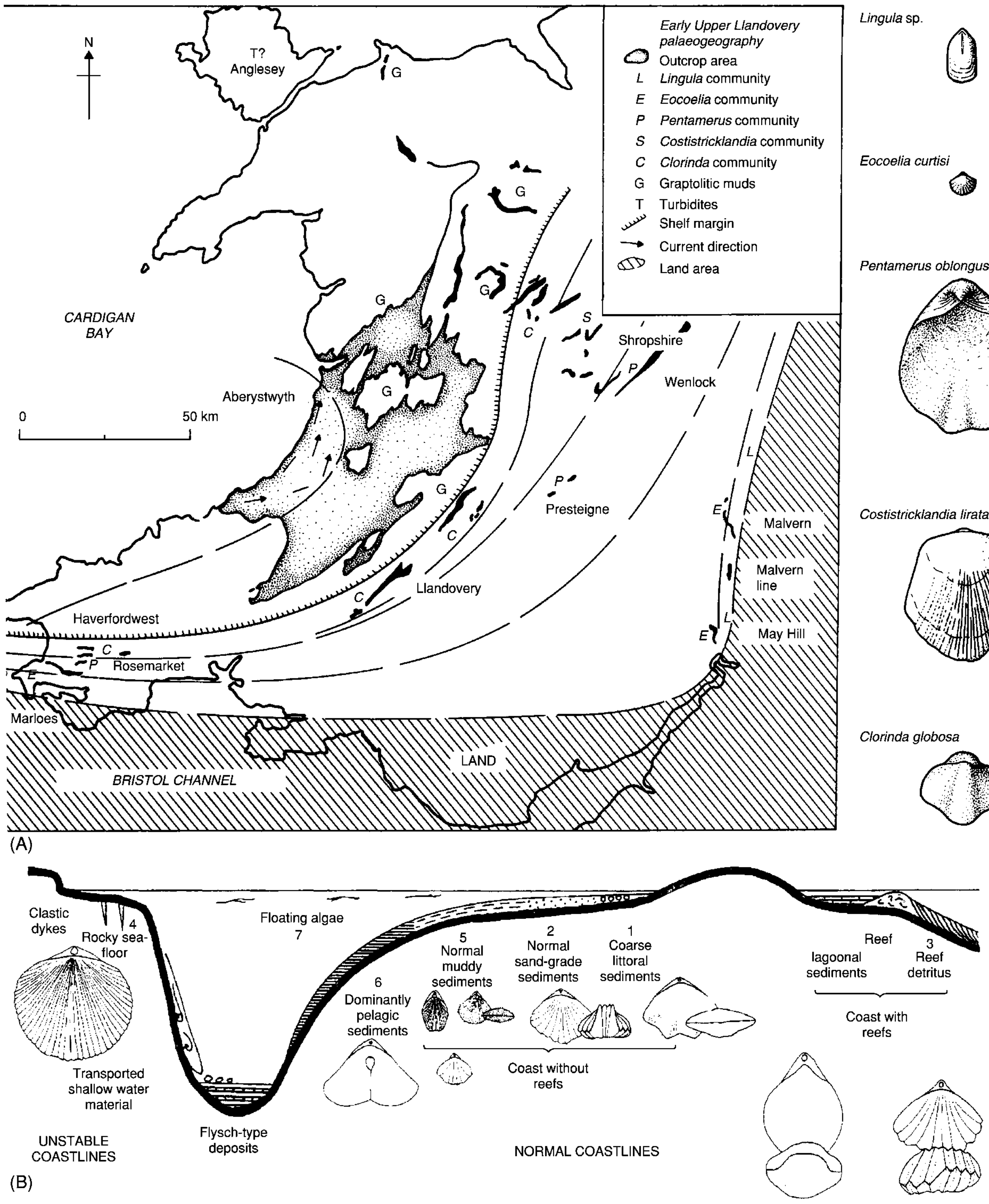


Figure 8 (A) Depth related brachiopod dominated palaeocommunities, first established in the Silurian palaeoenvironments of the Welsh Basin. Modified from Ziegler AM, Cocks LRM and McKerrow WS (1968) *Palaeontology*. Cambridge: Blackwell Science. (B) Substrate related brachiopod dominated palaeocommunities from the Mesozoic, established across the Alpine region. Adapted from Benton MJ and Harper DAT (1997) *Basic Palaeontology*. Harlow: Addison Wesley Longman.

local development of chemosynthetic environments. Biogeographical patterns amongst living forms reflect their Cenozoic roots; a Southern Area, Northern Pacific, and a Northern Area (Atlantic, Mediterranean, North Sea, and the circumpolar northern oceans) are based on a variety of articulated brachiopod associations; the linguliformeans have more widespread, near-cosmopolitan distributions.

Stratigraphical Distribution

In terms of the Phanerozoic evolutionary faunas, the Cambrian, Palaeozoic, and Modern brachiopod faunas are fundamentally different. Cambrian faunas were dominated by a range of nonarticulated groups, together with groups of disparate articulated taxa, such as the chileides, naukatides, obolellides, kutorginides, billingsellides, protorthides, orthides, and pentamerides. These groups participated in a variety of loosely structured, nearshore palaeocommunities. During the Ordovician radiation, the deltidiodont orthides and strophomenides dominated faunas (Figure 9), many first generated around Early Ordovician island complexes and later dominating the platforms, where they participated in an offshore movement of palaeocommunities. Following the end-Ordovician extinction event, spire-bearing brachiopods reached their dominance, particularly in the carbonate environments of the mid-Palaeozoic. The Carboniferous and, particularly, the Permian were intervals of spectacular experimentation: some brachiopods mimicked corals or developed extravagant clusters of spines, whereas a number of groups reduced their shells, thus presenting soft tissues to the outside environment. The end-Permian extinction removed this diversity. Mesozoic and Cenozoic brachiopods were and are overwhelmingly dominated by cyrtomatodont rhynchonelliformeans with either crurae (rhynchonellides) or loops (terebratulides).

Extinctions and Radiations

The brachiopods experienced five main extinction events followed by recoveries and radiations of varying magnitudes (Figure 9). The end-Ordovician event occurred in two phases against a background of glaciation and accounted for almost 80% of the existing brachiopod families. The recovery and subsequent radiation are marked by the decline of deltidiodont groups, such as the orthides and strophomenides, whereas the spire-bearing atrypides, athyridides, and spiriferides, together with the pentamerides, achieved a greater dominance, particularly in carbonate environments. Late Devonian events, at the Frasnian–Famennian stage boundary, also associated with climate change, removed the atrypides and pentamerides, and severely affected the orthides and strophomenides, whereas the spiriferides and rhynchonellides survived in deeper water environments and staged an impressive recovery. A particular feature of the post-Frasnian fauna is the diversity of recumbent brachiopod megaguilds, dominated by the productides.

The end-Permian event (*see* **Palaeozoic:** End Permian Extinctions) has been ascribed to a range of different causes and was associated with the disappearance of over 90% of all living species, including some of the most ecologically and taxonomically diverse brachiopods. The post-extinction fauna was first dominated by a variety of disaster taxa, including lingulids; nevertheless, the brachiopod fauna later diversified within a relatively few clades dominated by the rhynchonellides and terebratulides. The end-Triassic event removed the majority of the remaining spiriferides and the last strophomenides. The agenda set by the end-Permian event, involving the subsequent dominance of rhynchonellide and terebratulide groups, was continued after the end-Triassic event. The end-Cretaceous event (*see* **Mesozoic:** End Cretaceous Extinctions) may have been responsible for the loss of about 70% of Chalk brachiopod faunas in North-West Europe; nevertheless, the subsequent radiation in the Danian limestone facies involved many of the pre-extinction taxa at the generic level.

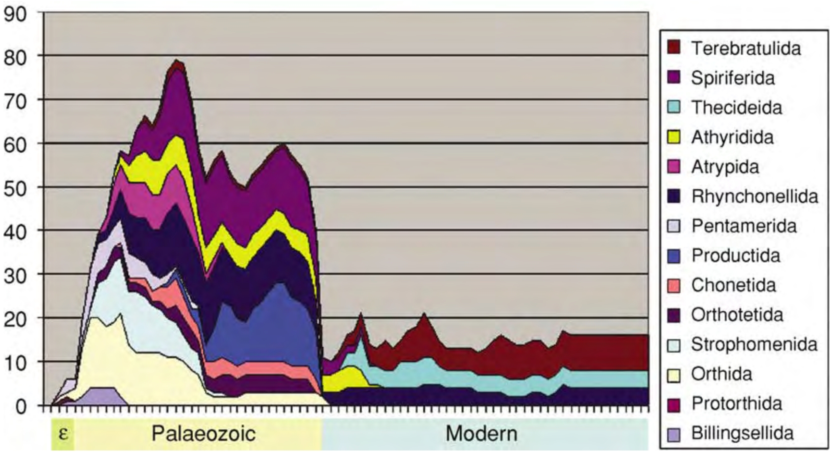


Figure 9 Stratigraphical abundance and distribution of the main orders of rhynchonelliformean brachiopod.

Modern Brachiopod Faunas

Despite their relative rarity today as compared with the Palaeozoic, living brachiopods are actually widespread, represented mainly by pedunculate forms attached to a variety of substrates across a spectrum of water depths. At high latitudes, brachiopods range from intertidal depths to basinal environments at depths of over 6000 m. They are most common in fjord settings in Canada, Norway, and Scotland and in the seas around Antarctica and New Zealand. The association between the horse mussel, *Modiolus modiolus*, and *Terebratulina retusa* is particularly widespread in the Northern Hemisphere. In the tropics, however, many species are micromorphic, exploiting cryptic habitats in reef crevices or in the shade of corals and sponges. Larger forms live in deeper water environments, evading groups of predators that might graze on meadows of newly attached larvae.

See Also

Biological Radiations and Speciation. **Fossil Invertebrates:** Bryozoans; Bivalves. **Mesozoic:** End Cretaceous Extinctions. **Palaeozoic:** Cambrian; Ordovician; End Permian Extinctions. **Sedimentary Environments:** Carbonate Shorelines and Shelves.

Further Reading

Benton MJ and Harper DAT (1997) *Basic Palaeontology.* Harlow: Addison Wesley Longman.
Brunton CHC, Cocks LRM, and Long SL (eds.) (2001) *Brachiopods Past and Present. Systematics Association Special Volume Series 63*. London and New York: Taylor and Francis.
Carlson SJ and Sandy MR (eds.) (2001) *Brachiopods Ancient and Modern. A Tribute to G. Arthur Cooper. Paleontological Society Papers 7*. New Haven: Yale Uni versity Reprographics.
Clarkson ENK (1998) *Invertebrate Palaeontology and Evolution,* 4th edn. Cambridge: Blackwell Science.
Harper DAT and Moran R (1997) Brachiopod life styles. *Geology Today* 13: 235 238.
Harper DAT and Rong J (2001) Palaeozoic brachiopod extinctions, survival and recovery: patterns within the rhynchonelliformeans. *Geological Journal* 36: 317 328.
Kaesler RL (ed.) (1997 2002) *Treatise on Invertebrate Paleontology,* Part H, *Brachiopoda* (revised), vol. 1 4. Boulder, CO and Lawrence, KS: The Geological Society of America and the University of Kansas (continuing).
Nielsen C (2002) *Animal Evolution: Interrelationships of the Living Phyla,* 2nd edn. Oxford: Oxford University Press.
Williams A, Carlson SJ, Brunton CHC, Holmer LE, and Popov LE (1996) A supra ordinal classification of the Brachiopoda. *Philosophical Transactions of the Royal Society, Biological Sciences* 351: 443 481.

Bryozoans

P D Taylor, The Natural History Museum, London, UK

Introduction

The Bryozoa are a phylum of colonial invertebrates occasionally referred to as Ectoprocta or by the obsolete name Polyzoa. Because the great majority of bryozoan species possess calcareous skeletons, they are among the commonest groups of macrofossils found in the post-Cambrian marine fossil record. Bryozoans are active suspension feeders. In common with most other suspension feeders, they do not need to move in search of food and, with a few exceptions, are sessile, living permanently anchored to a hard surface, such as a rock, shell, or seaweed. Although present-day bryozoans can be found living in freshwater lakes and slow-flowing rivers, the majority are marine, inhabiting depths from the intertidal to the abyssal. Most modern species are stenohaline, and fossil bryozoans are almost exclusively marine owing to the absence of mineralized skeletons in freshwater species.

All bryozoans are colonial. Each colony comprises a group of genetically identical modular units called zooids. Zooid size enables bryozoans to be distinguished from colonial corals (*see* **Fossil Invertebrates:** Corals and Other Cnidaria), with which they are occasionally confused: bryozoan zooids seldom exceed 2 mm in maximum surface dimension, whereas the equivalent coral structures (corallites) are typically centimetric in scale. Colony growth occurs mostly by the addition of new zooids, a process termed budding, supplemented in some groups by lengthening of existing zooids. Fully grown colonies of some species contain fewer than 10 zooids, whereas other species may develop colonies containing many thousands of zooids, which may on occasion reach 50 cm in diameter.

Anatomy and Feeding

The basic bryozoan zooid consists of a body wall enclosing a fluid-filled cavity (coelom or pseudocoel) in which is suspended the digestive system and various other organs (Figure 1). The body walls of most species incorporate a calcareous layer, which forms the hard skeleton of the zooid and is sometimes termed the zooecium (the hard skeleton of the colony as a whole is termed the zoarium). Bryozoans have a U-shaped gut with separate mouth and anus. They feed actively on particles in suspension, mainly phytoplankton, using an inverted-V-shaped or bell-shaped lophophore comprising a ring of between 8 and 30 or more tentacles. Cilia on the tentacles beat in unison to create a current of water that enters the open end of the lophophore and exits at the sides between the tentacles. Some food particles are thrown directly towards the mouth at the bottom of the lophophore, while others are sieved by the cilia and /or batted back into the fast flow at the centre of the lophophore by tentacular flicking. Zooids in some species feed more or less independently, but in others the feeding currents of groups of zooids combine. Spaced a few millimetres apart over the colony surface, excurrent chimneys may develop to which zooids channel their filtered water. These chimneys are often marked by raised areas on colony surfaces termed monticules, which can be clearly distinguished in fossil bryozoans. In other species with mesh-like erect colonies, zooidal lophophores create a one-way flow through the holes or fenestrules.

When not feeding, the lophophore can be withdrawn into the safety of the body walls, where it is further

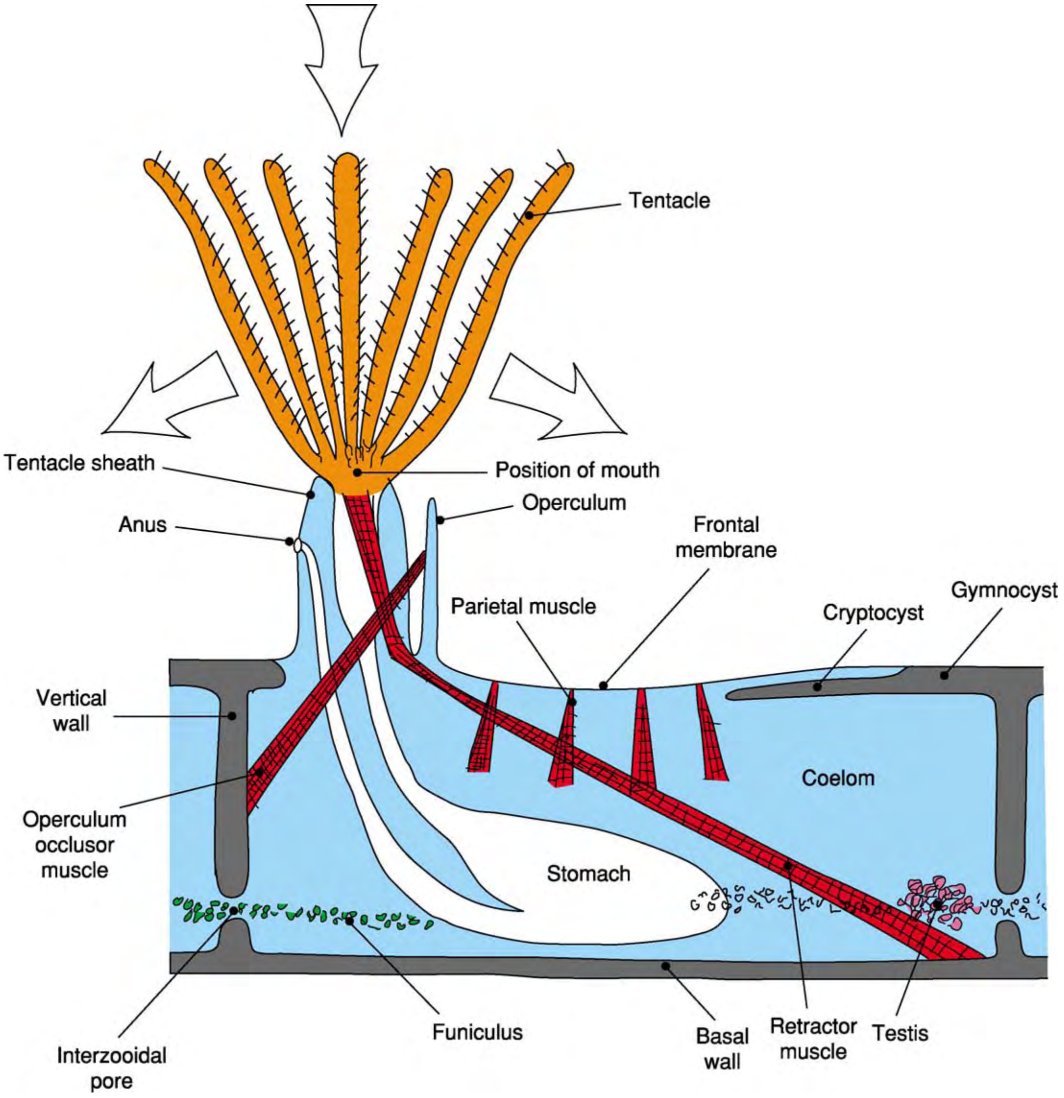


Figure 1 Anatomy of a basic bryozoan zooid illustrated by a diagrammatic vertical section through a cheilostome zooid with the lophophore extended. The gut and tentacle crown (lophophore) are shown in orange, the coelom in blue, the muscles in red, the funiculus in green, the testis in purple, and the calcified skeleton in dark grey. Arrows above and to the sides of the lophophore indicate feeding current flow. The parietal muscles are attached to the basal wall outside the plane of section.

protected by a hinged flap (operculum) in some species. Withdrawal is accomplished by the fast contraction of a retractor muscle running between the base of the lophophore and a point on the body wall. Different taxonomic groups employ different mechanisms for protruding the lophophore. However, all use a hydrostatic system entailing muscles acting on body walls to drive coelomic fluid into the tentacle sheath, forcing it to evert and push the lophophore out through the aperture of the zooid.

Bryozoans have a primitive nervous system. Nervous activity can be demonstrated by touching a lophophore, thereby inducing its immediate retraction into the zooid. The simultaneous retraction of the lophophores of adjacent zooids shows that the nervous systems of zooids are interconnected. Zooids can also be linked by a system called the funiculus. This is thought to be a homologue of the circulatory systems found in many other animals, but it does not carry blood or oxygen or have a heart. Instead, the funiculus, which is linked to the stomachs of the zooids, transports lipids and other metabolites both within and between zooids. Movement of food resources across the colony is undoubtedly important in the provisioning of nonfeeding zooids, developing larvae, and newly budded zooids at the growing edges of the colony. Pores in the skeletal walls allow both nervous and funicular connections between zooids. In some species soft-tissue linkages also occur over the outer ends of the skeletal walls between the zooids.

Reproduction and Growth

The life cycles of bryozoans are complex and varied. Colony growth occurs mostly by zooidal budding, an asexual process involving mitotic cell division. However, the formation of new colonies is usually a sexual process, with meiotic cell division occurring in the testes and ovaries, which are located alongside the funiculus. Bryozoan colonies are hermaphroditic, with each zooid usually producing sperm and eggs sequentially, although some species have separate male and female zooids. Released through tiny pores in the tips of the tentacles, sperm are carried away in the exhalent feeding currents and ambient flow, and a small proportion survive to fertilize an egg in another colony. Fertilized eggs develop into larvae. In a minority of bryozoan species the larvae are planktotrophic – they feed while in the plankton for weeks or months. However, most bryozoan species have non-planktotrophic larvae, which are brooded by the parent colony before being released into the plankton for a short period (hours or days). In both cases, the larvae eventually settle on a firm or hard surface (e.g. a shell or stone) and undergo metamorphosis to form the first zooid (ancestrula) of a new colony. Budding from the ancestrula produces the first generation of asexual zooids, which in turn bud further zooids, and so on.

New zooids are usually budded at specific locations, such as branch tips in tree-like colonies and the peripheral growing edge in sheet-like encrusting colonies (Figure 2B). The location and growth orientation of new buds largely determines the shape of the colony. Modular zooids of similar shape can be 'assembled' into disparate colony forms, many of which have evolved in parallel in unrelated groups. Various schemes have been devised to classify these growth forms, some using geometric terms (e.g. reticulate) and others based on genera having the growth form (e.g. reteporiform). A simple division is into encrusting, erect, and free-living colonies. Encrusting colonies tend to be either sheet-like (Figure 2B), with zooids arranged multiserially, or

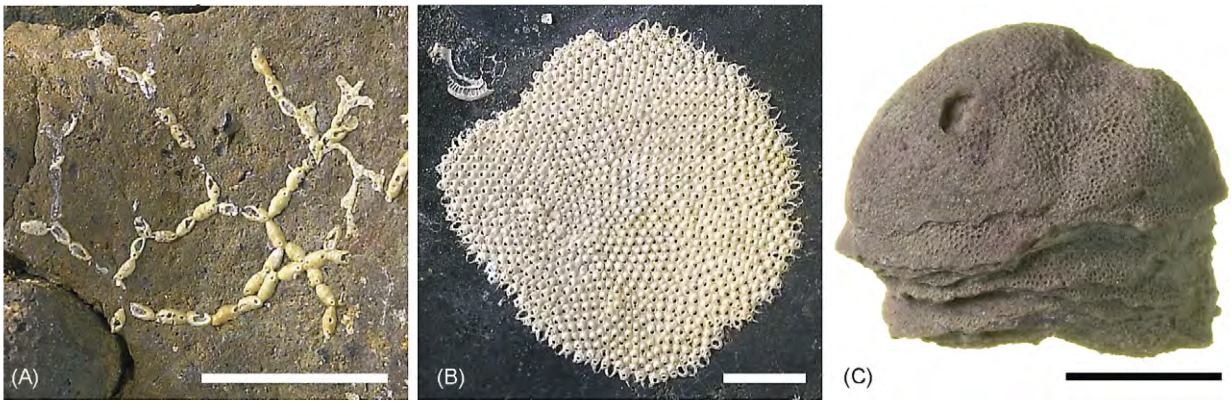


Figure 2 Bryozoan colony forms. (A) Encrusting runner (cyclostome *Voigtopora*, Cretaceous, England, scale bar 5 mm), (B) encrusting sheet (cheilostome *Porella*, Plio Pleistocene, New Zealand, scale bar 2 mm), and (C) dome shaped colony (trepostome '*Dianulites*', Ordovician, Russia, scale bar 10 mm).

runner-like, with uniserially or pauciserially arranged zooids forming a branching colony (Figure 2A). These two kinds of encrusting colonies represent alternative survival strategies. Sheet-like colonies are compact and place a premium on the defence of the living space that they occupy (phalanx strategy), whereas runner-like colonies disperse their zooids widely across the substrate, maximizing the chances that some will survive (fugitive or guerrilla strategy). New zooids are sometimes budded on top of older zooids to give a multilayered colony. Among erect bryozoans, common colony forms are tree-like (Figure 3A), frondose (Figure 3B), and net-like or reticulate (Figure 3C). Some erect colonies are rigid, others flexible, by virtue of having either elastic joints linking stiff internodes or a weakly mineralized skeleton. Flexible colonies are less likely to snap under the bending stresses imposed by ambient currents. Free-living colonies outgrow their substrates (e.g. sand grains) and rest directly on the sediment surface. They include massive dome-shaped colonies (Figure 2C), commonest in the Lower Palaeozoic, and smaller concavo-convex colonies (lunulites) found in the Cretaceous–Holocene. The latter support themselves above the seabed using structures called setae and in one genus (*Selenaria*) are even able to 'walk'.

Fragmentation provides an alternative mode of colony formation, especially in erect species with fragile branches that are easily broken off, transported away and able to resume colony growth elsewhere. Colonies formed in this way are clones. The ability to produce new colonies by fragmentation depends on colonies being able to survive after the death of some of their constituent zooids. The death of zooids within still-living colonies (partial mortality) is routine in bryozoans. The oldest zooids within large long-lived colonies are frequently dead (or at least dormant and no longer feeding), and predators can kill individual zooids in living colonies. If the loss of zooids to predators or other causes exceeds the budding of new zooids at the growing edges of the colony, then 'negative growth' occurs, and the colony actually decreases in size with time and age. Another complexity in bryozoan life cycles is introduced by the potential for different colonies of the same species to fuse with one another if they come into contact during growth. The fused colony will be a genetic chimera unless the original colonies were clonal, which may be the case in cyclostome bryozoans where embryos in brood chambers divide into many genetically identical larvae, a process called polyembryony.

Cyclical polypide degeneration and regeneration is characteristic of bryozoans. This process occurs when the polypide – essentially the gut, lophophore, and certain associated tissues – breaks down, giving a mass called a 'brown body' that can be expelled. A new polypide is subsequently formed within the body walls and feeding recommences. Individual zooids may pass through several polypide cycles during their lifespan. The 'brown deposits' found in some fossil bryozoans with zooidal chambers filled by diagenetic cement are probably remnants of non-expelled brown bodies.

Polymorphism and Within-colony Zooidal Variations

One of the most intriguing features of bryozoans is zooidal polymorphism. This is where the colony contains more than one type of zooid, each having a distinctive morphology and fulfilling a particular

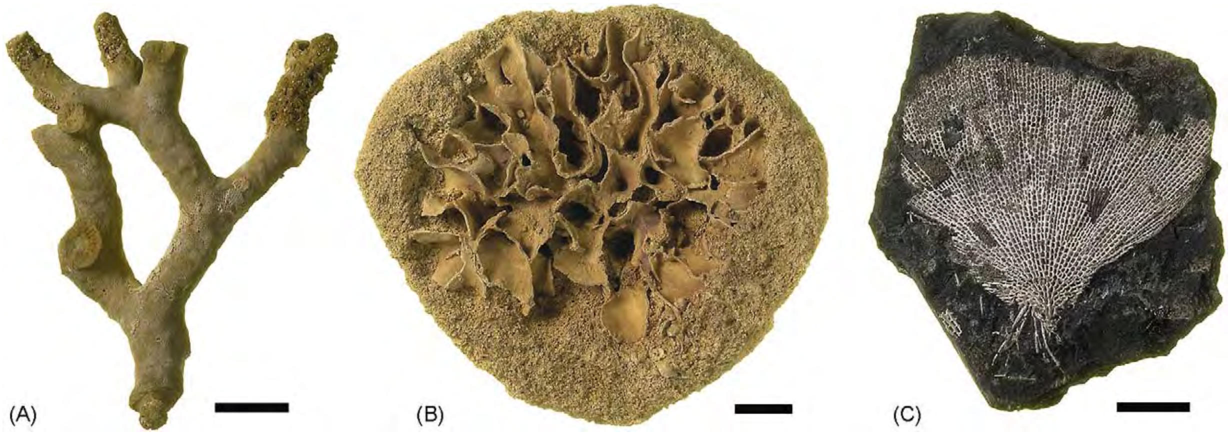


Figure 3 Erect bryozoan colony forms. (A) Tree like (cyclostome *Heteropora*, Pliocene, England, scale bar 10 mm), (B) frondose (cheilostome *Pentapora*, Pliocene, England, scale bar 20 mm), and (C) reticulate (fenestrate *Fenestella* s.l., Carboniferous, Wales, scale bar 10 mm).

functional role. Feeding zooids are known as autozooids, and non-feeding polymorphs are known as heterozooids. Heterozooids include the avicularia found in the majority of cheilostome bryozoans. Relative to their overall size, avicularia have large opercula called mandibles, which act like jaws and have been observed capturing and holding onto predators until they die. Some avicularia replace autozooids in the normal budding sequence of the colony and are of similar size, whereas others are much smaller and are budded onto the surfaces of the autozooids, often positioned so as to guard the vulnerable aperture. A subcategory of avicularia, vibracula, have elongated mandibles (setae) apparently used for cleaning detritus from the colony surface and also forming the appendages of lunulites. Polymorphs called gonozooids function as larval brood chambers in cyclostome bryozoans. Other heterozooids found in extant bryozoans include rhizozooids (roots) and kenozooids (space fillers). Palaeozoic stenolaemate bryozoans with long tubular zooids may contain polymorphs of smaller diameter between the autozooids. If these polymorphs possess diaphragms (cross partitions), they are called mesozooids, otherwise they are termed exilazooids. It is likely that both mesozooids and exilazooids functioned passively as space fillers to prevent overlap between the lophophores of the feeding zooids.

Aside from polymorphism, there are three other sources of variation in zooids within colonies: ontogeny, astogeny, and microenvironment. Ontogenetic variation occurs because zooids within a colony are of different ages and hence differ in their developmental stage: newly budded zooids may have thinner skeletons than older zooids and can look very different. Astogeny refers to the development of the colony: zooids budded early in astogeny are often smaller and simpler than those budded later during colony growth. Additional variations within colonies can be related to differences in the microenvironment experienced by the zooids; for example, obstacles on the substratum surface may cause localized distortion of zooids in encrusting colonies.

Skeleton

The mineralized skeleton of bryozoans consists of an extracellular secretion of tiny crystallites of calcium carbonate set in an organic matrix. There are clear similarities with the shells of brachiopods and molluscs in the calcareous composition of the skeleton, the morphology of the crystallites, and the mode of secretion. However, bryozoan skeletons are typically more intricate and comprise two fundamentally different types of wall. Exterior walls resemble brachiopod and mollusc shells in growing on an organic template (cuticle) and thickening through accretion on one side only. Interior walls develop initially as invaginations of the epithelia forming the exterior walls, lack a cuticular layer, and thicken through accretion on both sides. Exterior walls usually occur at the interface between the bryozoan and the external environment, whereas interior walls typically form the boundaries between zooids or intrazooidal structures. Some stenolaemate bryozoans have skeletal rods (styles) embedded in their interior walls, often with a cone-in-cone internal structure, and projecting from the colony surface as spines.

Whether exterior or interior, each skeletal wall typically contains several layers of contrasting fabric. Three basic categories of wall fabric are: finely granular; lamellar, with crystallite long axes subparallel to the wall surface; and perpendicular, with crystallite axes at a high angle to the wall surface. Granular fabrics have very small crystallites and in some fossils are apparently neomorphic in origin. Lamellar fabrics often comprise lath-like crystallites imbricated at low angles or stacks of platey crystallites. Perpendicular fabrics are usually fibrous.

Mineralogically, most bryozoan skeletons are made of calcite, but an increasing proportion of cheilostomes from the Upper Cretaceous onwards use aragonite, either alone or overlying a calcite basal skeleton. The amount of magnesium in bryozoan calcite varies from low (less than 4% wt) to high (more than 10% wt). High-magnesium calcite skeletons are found in some post-Palaeozoic cheilostomes and may also have been present in a few Palaeozoic trepostomes and cystoporates with granular neomorphized walls containing microdolomites. Slight deviations from equilibrium in carbon and oxygen isotopes have been detected in bryozoan skeletons, possibly due to kinetic effects associated with rapid growth rates.

Classification and Evolutionary History

Three classes of bryozoans are generally recognized: Phylactolaemata, Gymnolaemata and Stenolaemata (Table 1). Extant phylactolaemates live entirely in freshwater lakes and slow-moving rivers. They are of relatively low diversity (about 80 species) and lack mineralized skeletons. Body fossils of phylactolaemates are unknown, but the chitinous disc-shaped statoblasts manufactured by phylactolaemates for overwintering and dispersal, like plant seeds, are very occasionally recorded in the fossil record as far back as the Permian. Anatomical evidence indicates

Table 1 Major bryozoan groups with their main skeletal features and geological ranges

Class	*Order*	*Skeletal morphology*	*Geological range*
Phylactolaemata	[none]	(Soft bodied)	Permian Recent
Gymnolaemata[a]	Ctenostomata[a]	(Soft bodied)	Ordovician Recent
	Cheilostomata	Box shaped zooids often with complex frontal morphology and intricately shaped apertures; avicularia; ovicells; very diverse colony forms; occasionally with monticules	Jurassic Recent
Stenolaemata	Cyclostomata[a]	Cylindrical zooids often with exterior frontal walls; porous interior walls; gonozooids; diverse colony forms; occasionally with monticules	Ordovician Recent
	Trepostomata[a]	Cylindrical zooids without exterior frontal walls; nonporous interior walls; diaphragms and styles common; some with mesozooids or exilazooids; tree like, sheet like or mound like colony forms, some frondose colonies; monticules common	Ordovician Triassic
	Cystoporata	Cylindrical zooids without exterior frontal walls; interior walls nonporous in most; extrazooidal vesicles and/or lunaria present; some with exilazooids; sheet like or mound like colony forms, some tree like and frondose colonies; some with monticules	Ordovician Triassic
	Cryptostomata	Box shaped to cylindrical zooids without exterior frontal walls, regularly arranged; interior walls nonporous; tree like or frondose colony forms, some articulated	Ordovician Triassic
	Fenestrata	Box shaped zooids without exterior frontal walls, regularly arranged; interior walls nonporous; narrow branched erect colonies, often reticulate or pinnate with zooids opening on only one side of the branch	Ordovician Permian

[a] Probable paraphyletic taxa.

that phylactolaemates are the most primitive of bryozoans. Gymnolaemates are a paraphyletic grouping within which the third bryozoan class – stenolaemates – is nested. Two orders of gymnolaemates are recognized, the soft-bodied ctenostomes and the calcareous cheilostomes.

Ctenostomes

Ctenostomes are predominantly marine, but a small number of the approximately 300 extant species are found in freshwater. Their colonies tend to be inconspicuous and 'weedy', often forming delicate encrustations with prostrate zooids sometimes aligned along a colonial stolon, although more robust encrusting and erect colonies occur in some gelatinous species. Despite their lack of hard parts, ctenostomes can be preserved as fossils in two ways. Some species bore into calcareous substrates, especially mollusc shells. These leave a trace fossil comprising branching arrays of thread-like tunnels bearing zooids at regular intervals, which communicate with the surface of the substrate via a small circular aperture. Although most examples of ctenostome borings are of Mesozoic or Cenozoic age, the earliest examples come from the Arenig Stage of the Ordovician. The second mode of preservation of ctenostomes is bioimmuration. This is where organic overgrowth, by an oyster for example, leaves a natural mould of the bryozoan on the underside of the overgrowing organism. Bioimmured ctenostomes occur quite commonly in the post-Palaeozoic. Most have oval zooids arranged in branching uniserial chains. Many bioimmured ctenostomes from the Jurassic have very similar zooids and colony growth patterns to cheilostomes, and some also possess an operculum, normally regarded as diagnostic of cheilostomes, supporting the hypothesis that cheilostomes originated from a ctenostome ancestor that acquired a mineralized skeleton in the Late Jurassic.

Cheilostomes

Over 1000 genera of cheilostomes have been recognized, and, with an estimated 4300 extant species, cheilostomes are the dominant group of present-day bryozoans. As noted above, however, this gymnolaemate order made a relatively late appearance about 155 Ma ago and even then did not become abundant in the fossil record until the mid-Cretaceous. Cheilostomes underwent an explosive radiation during the Late Cretaceous. The onset of this spectacular diversification, which was accompanied by equally impressive increases in morphological disparity and abundance, coincided with the appearance of ovicells, hood-like chambers for brooding larvae (Figure 4A). This correlates with the evolution of short-lived brooded larvae from the long-lived nonbrooded larvae characteristic of more primitive cheilostomes. Such a switch in larval type is predicted to have had

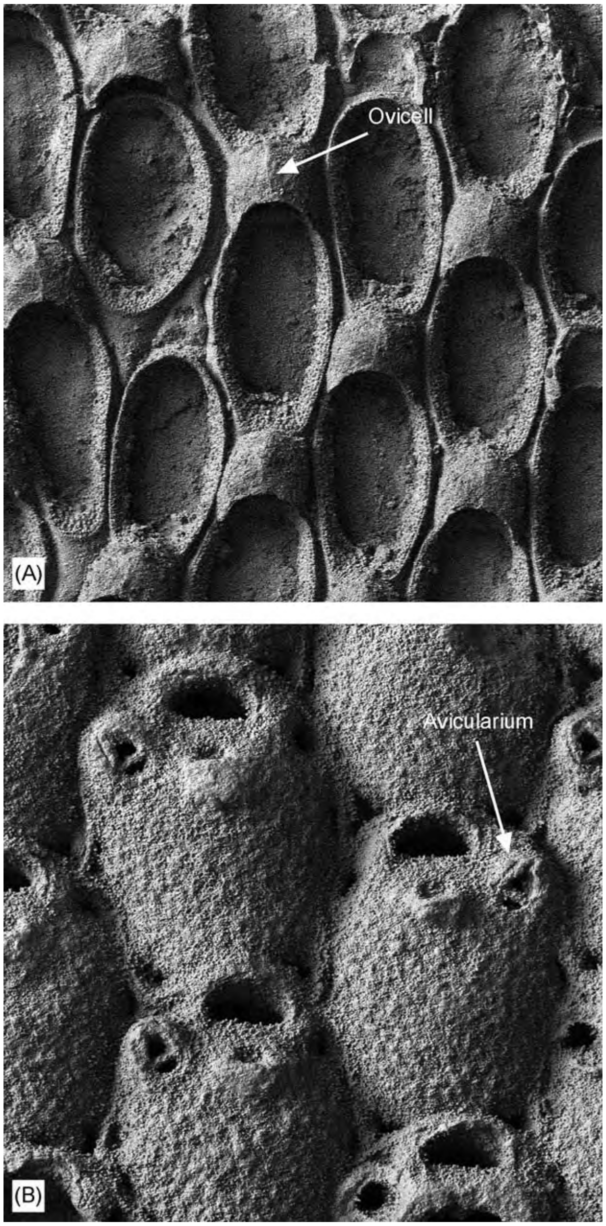


Figure 4 Fossil cheilostome bryozoan zooids (scanning electron micrographs). (A) Group of zooids from a colony of *Wilbertopora*, an anascan cheilostome with a mostly uncalcified frontal area; note the larval brood chambers (ovicells) (Cretaceous, Texas, USA; magnification × 55). (B) Zooids of *Microporella*, an ascophoran cheilostome with a cryptocystal frontal shield and a small semielliptical orifice; note the tiny adventitious avicularia (Pliocene, New Zealand; magnification ×65).

profound consequences for dispersal, population genetic structure, and speciation rate. The good larval dispersal and gene flow of nonbrooders is predicted to result in genetically homogeneous species that are unlikely to become fragmented into incipient new species, whereas poor larval dispersal and gene flow in brooders should give rise to genetically heterogeneous species, which are more prone to speciate.

At the colony level, cheilostomes exhibit a vast range of forms, some homeomorphic with colony forms seen in stenolaemate bryozoans but others unique to the cheilostomes. The most primitive cheilostomes have simple encrusting colonies with zooids arranged one after the other along ramifying branches or irregularly in lobate sheets. Somewhat more advanced cheilostomes show a more regular patterning of the zooids in coherent sheets (Figure 2B). While the majority of cheilostome species have encrusting colonies, erect species can be visually striking and are more likely to be noticed as fossils. Some have tree-like colonies, others are frondose (Figure 3B), and yet others are reticulate with lace-like colonies. Articulated colonies and flexible erect colonies with light calcification (and consequently poor fossilization potential) also occur among cheilostomes. Small spindle-shaped colonies anchored into particulate sediments by rootlets are found in some species, while the cap-shaped concavoconvex colonies of free-living lunulites are without close parallel among the other bryozoan orders.

The wide range of zooid-level skeletal features in cheilostomes provides useful characters for taxonomy. Indeed, species identification is often possible on the basis of just a few zooids, while colony level features are generally of less importance than in the stenolaemates. Cheilostome zooidal skeletons are typically box shaped and are rounded-rhombic or rectangular in frontal view (Figure 4). The vertical walls of the zooid are completely calcified, except where penetrated by pores that allow communication between adjacent zooids (Figure 1). Basal walls may be completely calcified or have a central uncalcified window. However, it is the frontal walls of the zooid that provide most of the variation useful in taxonomy. These may be largely uncalcified, as in most anascan cheilostomes (Figure 4A) where the frontal surface of the zooid is occupied by a nonmineralized frontal membrane with, at its distal end, the orifice. Through time an increasing proportion of cheilostomes, including the ascophorans, which predominate at the present day, have evolved protective frontal shields covering the frontal membrane (Figure 4B). Frontal shields are constructed of three main types of skeleton: gymnocyst is a planar exterior wall; spinocyst is another type of exterior wall comprising hollow spines that grow centripetally to overarch the frontal membrane; and cryptocyst is an interior wall, often with a pustulose surface texture. Many cheilostome zooids have erect spines, which are often hollow, articulated basally, and located near the orifice or around the inner edge (mural rim) of the gymnocyst.

Avicularian polymorphs are widespread among cheilostomes.

Stenolaemates

Stenolaemates were the dominant bryozoans throughout the Palaeozoic, where they are represented by five orders – Trepostomata, Cystoporata, Cryptostomata, Fenestrata, and Cyclostomata – of which only the cyclostomes are extant. The oldest stenolaemates are trepostomes reported from the Tremadoc of China. By the Arenig, representatives of the other four orders had appeared, initiating a phase of exponential diversification that continued into the Caradoc. The origin of the mineralized skeleton from a ctenostome-grade ancestor may have been the primary trigger for stenolaemate radiation, opening the way for the evolution of a much greater variety of colony forms. A plateau of stenolaemate family diversity characterizes the remainder of the Palaeozoic, before the catastrophic decline in diversity marking the end-Permian mass extinction (Figure 6). Although only one order (Fenestrata) is thought to have become extinct at this time, stenolaemates were rare in the Triassic, and the Palaeozoic 'holdover' orders (trepostomes, cystoporates and cryptostomes) did not survive beyond this period. Bryozoan faunas of Jurassic and Early Cretaceous age, before the rise of the cheilostomes, consist predominantly of cyclostomes.

Trepostomes Trepostomes (approximately 200 genera) are the 'stony bryozoans' that abound in Palaeozoic deposits such as the Upper Ordovician around Cincinnati in the USA. A few trepostome species have sheet-like encrusting or frondose erect colonies, but most have dome-shaped colonies (Figure 2C) or, more commonly, robust tree-like colonies, which are typically broken into individual branches during fossilization. Trepostome zooids are characteristically long and tubular, terminating in a polygonal aperture (Figure 5A). Many species have

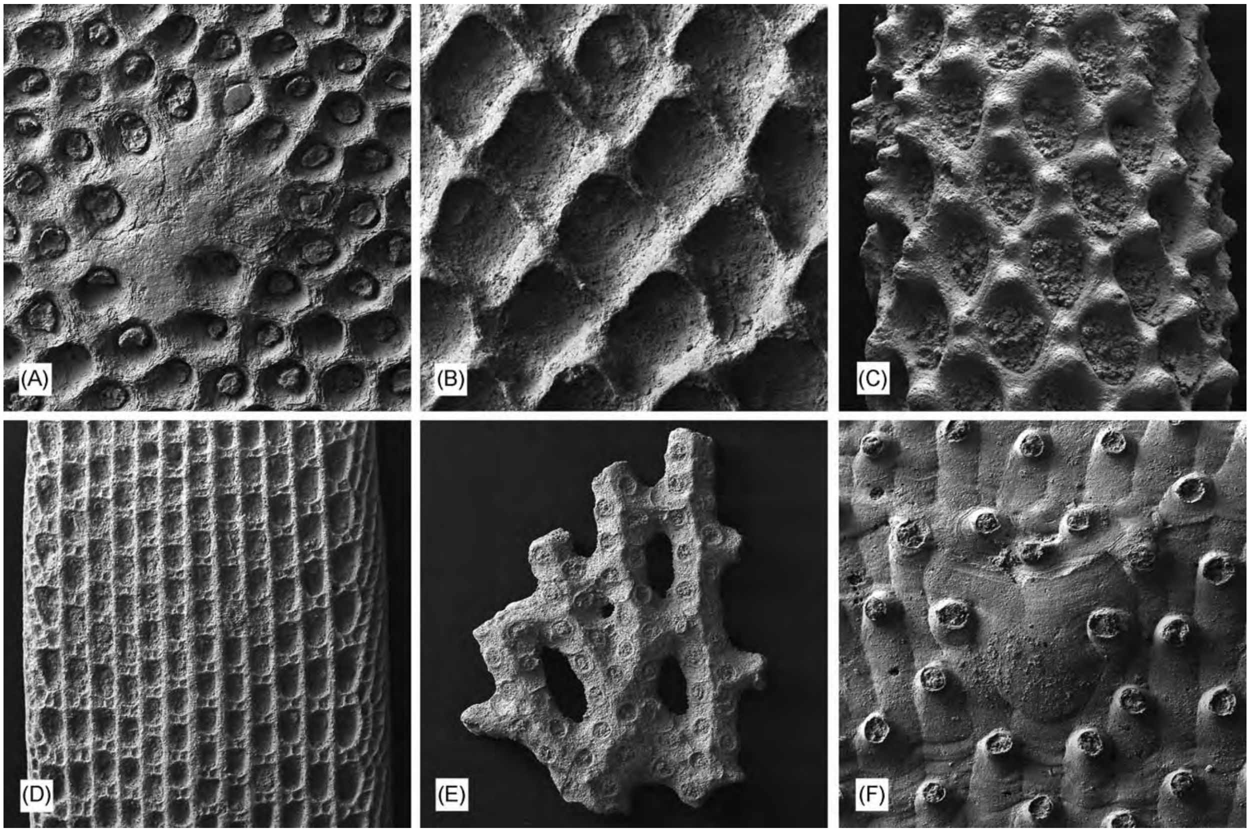


Figure 5 Fossil stenolaemate bryozoans (scanning electron micrographs). (A) Unidentified trepostome with polygonal zooidal apertures surrounding a monticule (Ordovician, Utah, USA, magnification ×30). (B) Diamond shaped zooidal apertures with lunaria in the cystoporate *Favositella* (Silurian, England, magnification ×30). (C) Branch of the rhabdomesine cryptostome *Rhabdomeson* with styles protruding as blunt spines (Carboniferous, Scotland, magnification ×50). (D) Flattened frond of the ptilodictyine cryptostome *Phaenopora* showing autozooids arranged in well defined rows and some smaller polymorphs (Silurian, Sweden, magnification ×15). (E) Fragment of a colony of the fenestrate bryozoan *Fenestella* s.1. with branches bearing two rows of apertures, short dissepiments and fenestrules (Carboniferous, Scotland, magnification ×15). (F) Autozooids and a gonozooid (centre) in the cyclostome *Hyporosopora* (Jurassic, England, magnification ×25).

mesozooids or exilazooids, and monticules are widespread. Styles are often present, as are diaphragms, which may be planar and complete, partial (hemiphragms), or curved (cystphragms). Like those of the other dominant Palaeozoic stenolaemate orders, trepostome skeletons lack frontal exterior walls. Branches of tree-like species usually have an inner thin-walled region (endozone) where zooidal budding is concentrated and zooidal axes are subparallel to the branch axis. This is surrounded by an outer thick-walled region (exozone) where zooid axes are almost perpendicular to the branch surface.

Cystoporates Cystoporates (approximately 90 genera) resemble trepostomes but can be distinguished from them by the presence of one or both of the following features: lunaria – hood-like projections over zooidal apertures (Figure 5B); and extrazooidal vesicles – 'bubbly' skeletal tissue between the zooids. They also often have monticules, sometimes with centres filled by vesicular tissue. Most cystoporates are sheet-like encrusters or dome shaped, but frondose and tree-like forms can also occur. A few have irregular pores in the walls between the zooids.

Cryptostomes Cryptostomes (approximately 80 genera) differ from trepostomes and cystoporates in having zooidal budding focused in restricted zones and zooids that tend to be shorter, giving colonies of a generally more delicate appearance with a regular pattern of apertures on the surface. All cryptostomes have erect colonies, and these are divided into two main sorts, the tree-like colonies of rhabdomesines (Figure 5C) and the frondose or flat-branched colonies of ptilodictyines (Figure 5D), which have a median lamina from which the zooids are budded. Articulations occur in some cryptostomes, either at the base of the colony in certain ptilodictyines or at regular intervals throughout the bushy colonies of one family of rhabdomesines. Styles are frequently present, and there may be more than one type of style per colony, some protruding from the colony surface as spines (Figure 5C). Short diaphragms (hemisepta) occur close to the aperture in many species.

Fenestrates Fenestrates (approximately 150 genera) are typified by short and often almost box-shaped zooids budded in a precise pattern. Colonies always have narrow erect branches bearing between two and about 10 series of zooids with subcircular apertures opening on only one side of the branch; the side without apertures may be reinforced by thick calcification. Often branches bifurcate and coalesce to form a planar mesh, but a similar mesh can be produced as a result of solid skeletal dissepiments linking branches at intervals (Figures 3C and 5E). In both cases, the holes in the mesh would have functioned as conduits for water flow during feeding. Pinnate colonies also occur. Probably the best-known and most diverse group of fenestrates is the Family Fenestellidae (fenestellids), which dominates many Carboniferous and Permian bryozoan assemblages. Fenestellid colonies have two rows of zooids along the branches, which are linked by dissepiments, and often develop nodes along the crests of the branches between apertures. In some genera, these nodes are greatly enlarged to form a superstructure affording protection to the zooids beneath. The peculiar screw-like colonies of *Archimedes* make it the most distinctive of all fenestellids. The screw is a massively thickened central axis that supports a helically wound mesh. Many populations of *Archimedes* relied on asexual propagation, with broken-off screws coming to rest prostrate on the seabed before producing daughter screws on their upper sides that restored vertical growth.

Cyclostomes Cyclostomes (approximately 370 genera) are first recorded in the Arenig but were uncommon until the Jurassic (Figure 6). Palaeozoic cyclostomes differ from contemperaneous orders in having exterior frontal walls. All Palaeozoic species have delicate 'weedy' colonies, which can be erect or encrusting and runner-like or sheet-like. The evolutionary zenith for cyclostomes was between the mid-Jurassic and the Late Cretaceous when a much greater spectrum of colony forms evolved. Species with exterior frontal walls (Figure 5F) were joined by others without frontal walls that are convergent with the extinct orders of stenolaemates of the Palaeozoic (Figure 3A). Cyclostomes possess regularly porous skeletal walls, permitting interzooidal communication, and nearly all species have bulbous gonozooids (Figure 5F) for brooding the polyembryonic larvae. Apertures of autozooids can be polygonal or subcircular and are often surrounded by a tubular prolongation (peristome).

Occurrence

Sedimentary environments most favourable to bryozoans are marine, shallow water, fully oxygenated with good circulation furnishing planktonic food resources, well provisioned with stable hard or firm substrates, and lacking a high influx of particulate sediment. Although bryozoans do occur intertidally, the majority of species are subtidal and inhabit the mid- to outer continental shelf. While

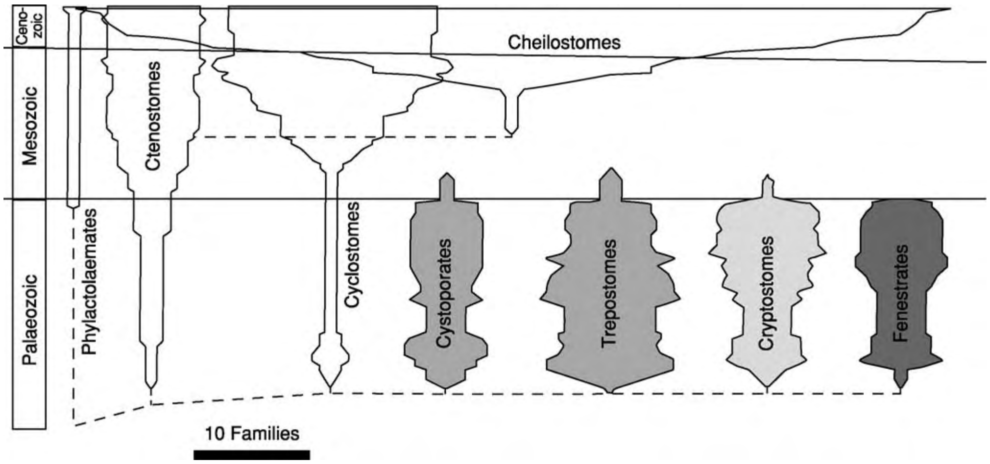


Figure 6 Diagram of changing family diversities in the bryozoan orders. Tentative phylogenetic relationships are indicated by dashed lines.

a few modern species have restricted bathymetric distributions, many have depth ranges exceeding 100 m. High diversities of bryozoans can be found in all major climatic zones, but tropical species tend to be inconspicuous, typically occurring as small encrusting colonies in cryptic habitats such as the undersides of stones and shells. Bryozoans are a major contributor to modern cool-water carbonates, such as those forming on the southern Australian continental shelf.

In the geological record bryozoans are often found in association with carbonate facies. The best collecting opportunities are in unconsolidated fine-grained clastics, where surface picking, bulk sampling, and gathering of shells that may be encrusted with bryozoans can yield diverse faunas. In some instances, bryozoans are present in rock-forming abundance, usually comprising branch fragments of parautochthonous erect colonies (Figure 7) and less often small bioherms of robust in-situ colonies. Bryozoan limestones in the Cenozoic are typically clean-washed cross-bedded calcarenites or calcirudites. Mud mounds, including the Waulsortian reefs of the Carboniferous, often contain bryozoans. It has been suggested that the bryozoans assisted in mud sedimentation by acting as baffles and through the production of faecal pellets.

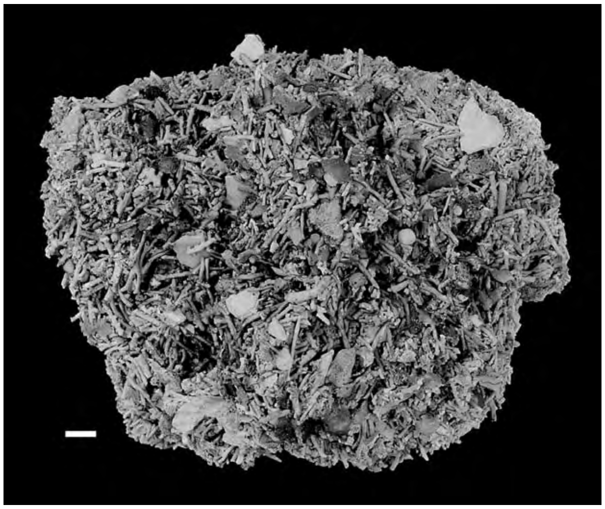


Figure 7 Bryozoan limestone from the Eocene of North Carolina, USA. A poorly cemented facies of the Castle Hayne Limestone consisting mainly of broken branches of tree like cyclostome and cheilostome colonies. Scale bar 10 mm.

Bryozoan-rich deposits show a striking change in latitudinal distribution through time, from a pan-latitudinal pattern in the Palaeozoic to an extra-tropical distribution in the post-Palaeozoic. The apparent displacement from the tropics of bryozoans contributing significant amounts of carbonate sediment may be related to the end-Permian mass extinction event, which had a severe effect on stenolaemate orders, and/or to increasing levels of colony destruction by grazing predators and exclusion by fast-growing corals and calcareous algae in the post-Palaeozoic.

The continuous presence of bryozoans through thick sequences deposited over long intervals is unusual. Instead, they tend to be patchily distributed, which is one of the factors that has hindered their use in biostratigraphy. Another limitation arises from typically long species durations; e.g. for cheilostomes currently living in the Mediterranean Sea and represented in the fossil record, average species longevity is about 15 Ma.

Palaeoecology

Much interest has focused on the use of bryozoan growth forms in palaeoenvironmental analysis. Pioneering work suggested that particular growth forms were adapted to particular hydrodynamic and sedimentational regimes. However, these findings are being re-evaluated in the light of the frequent occurrence of many different growth forms together. Nonetheless, the ratio of encrusting to erect species does decline with water depth in modern environments, while a dominance of either erect articulated colonies or free-living colonies usually points to relatively high sedimentation rates and/or a lack of sizeable substrates.

Shallow-water colonies of some erect species have been shown to possess thicker branches than deeper water colonies, giving the potential for detecting depth gradients in the geological record. The plasticity in colony form evident within some species is known to be ecophenotypic. For example, in the Mediterranean species *Schizoporella errata*, mound-like colonies develop in fast flow, and hollow branching colonies develop in slow flow. Variation in zooid size within species is inversely proportional to temperature. Two palaeoclimatic applications result from this relationship. First, comparisons of mean zooid size between colonies can yield information on the relative temperatures in which they grew, data that may be calibrated to an absolute value if the species concerned is extant. Second, zooid size within perennial colonies tracks seasonality – the higher the coefficient of variation in zooid size, the greater the annual range of temperature experienced by the colony.

With the exception of small circular borings possibly made by tiny gastropods, fossil bryozoans seldom show clear evidence of having suffered predation. Other biotic interactions are, however, frequently fossilized. Symbioses between bryozoans and hermit crabs are recorded back to the Jurassic. The crabs themselves are not fossilized, but the shelters provided by their bryozoan symbionts are diagnostic, forming thick encrustations on the surfaces of gastropod shells and extending the helicospiral coiling beyond the original shell aperture. Competitive interactions for substrate space between bryozoans and other encrusters with hard skeletons are evident on many fossil shells and clasts. Such competition between cyclostome and co-occurring cheilostome bryozoans has been documented by tallying overgrowths when sheet-like colonies of the two orders meet. The results have shown that cheilostomes have maintained a remarkably constant competitive advantage over cyclostomes for the past 100 Ma, on average winning about two-thirds of encounters.

See Also

Biological Radiations and Speciation. **Evolution**. **Fossil Invertebrates:** Brachiopods; Corals and Other Cnidaria. **Palaeoecology**. **Sedimentary Environments:** Carbonate Shorelines and Shelves; Reefs ('Build-Ups').

Further Reading

Boardman RS and Cheetham AH (1987) Phylum Bryozoa. In: Boardman RS, Cheetham AH, and Rowell AJ (eds.) *Fossil Invertebrates,* pp. 497 549. Oxford: Blackwell.

Boardman RS, Cheetham AH, Blake DB, *et al.* (1983) Bryozoa (revised). Volume 1. In: Moore RC and Robison RA (eds.) *Treatise on Invertebrate Paleontology. Part G,* p. 1 625. Boulder, Colorado and Lawrence, Kansas: Geological Society of America and University of Kansas Press.

Hageman SJ, Bock PE, Bone Y, and McGowran B (1998) Bryozoan growth habits: classification and analysis. *Journal of Paleontology* 72: 418 436.

Jackson JBC and Cheetham AH (1994) Phylogeny reconstruction and the tempo of speciation in cheilostome Bryozoa. *Paleobiology* 20: 407 423.

McKinney FK (1986) Historical record of erect bryozoan growth forms. *Proceedings of the Royal Society of London* B 228: 133 148.

McKinney FK (1995) One hundred million years of competitive interactions between clades: asymmetrical but not escalating. *Biological Journal of the Linnean Society* 56: 465 481.

McKinney FK and Jackson JBC (1989) *Bryozoan Evolution.* London: Unwin Hyman.

O'Dea A and Jackson JBC (2002) Bryozoan growth mirrors contrasting seasonal regimes across the Isthmus of Panama. *Palaeogeography, Palaeoclimatology, Palaeoecology* 185: 77 94.

Ryland JS (1970) *Bryozoans.* London: Hutchinson.

Smith AM (1995) Palaeoenvironmental interpretation using bryozoans: a review. *Geological Society of London Special Publication* 83: 231 243.

Taylor PD (1999) Bryozoa. In: Savazzi E (ed.) *Functional Morphology of the Invertebrate Skeleton,* pp. 623 646. Chichester: Wiley.

Taylor PD and Allison PA (1998) Bryozoan carbonates in space and time. *Geology* 26: 459 462.

Woollacott R and Zimmer RL (eds.) (1977) *Biology of Bryozoans.* New York: Academic Press.

Corals and Other Cnidaria

C T Scrutton, Formerly University of Durham, Durham, UK

Introduction

The phylum Cnidaria encompasses a varied group of primitive, dominantly marine, diploblastic metazoans of which the corals are by far the most important in the fossil record. Other cnidarians generally have poor preservation potential but nevertheless exist as scattered fossils extending back probably to the Late Precambrian.

Cnidarian Organization and Classification

All cnidarians consist fundamentally of a sac-like body with a mouth surrounded by tentacles (Figure 1). There are two variants: when the sac rests on its aboral surface with the mouth uppermost, it is a polyp; when it floats in the water with the mouth facing downwards, it is a medusa. When both forms exist in the same species, medusae are formed on the polyp, from which they separate (bud off) and produce male and female gametes. These fuse to form a free-floating larval stage, a planula, which settles and metamorphoses into a new polyp. The phylum is also characterized by the presence of stinging cells for the capture of prey (nematocysts), which are particularly abundant in the tentacles and which give the phylum its name.

There are three main subdivisions (classes) of the Cnidaria, which differ principally in the interior structure of the sac, the enteron or gut, and in the prominence of the two body forms (Figure 2). The Hydrozoa have no internal partitions in the gut, and in most species both polyps and medusae are well developed. The Scyphozoa have four radial flanges extending towards the centre of the gut, and the medusoid, or jellyfish, stage is dominant. The polyp is reduced to a specialized factory for producing medusae (scyphistoma). In the Anthozoa, the gut is subdivided in a more complex and varied way by radial partitions called mesenteries, whilst the medusoid stage is entirely absent, and gametes are produced directly by the polyp. The polyp stage may become modular by budding in both hydrozoans and anthozoans. Anthozoan colonies are simple, usually with no differentiation of the modules. However, in the Hydrozoa, modules may become specialized for different functions such as feeding and reproduction. The ultimate expression of this is the formation of complex colonies of specialized modules, not only of polypoid but also of medusoid origin, which function as superindividuals. These are the chondrophores and siphonophores, represented respectively in modern oceans by *Vellela* (by-the-wind-sailor) and *Physalia* (Portuguese Man-o'-War).

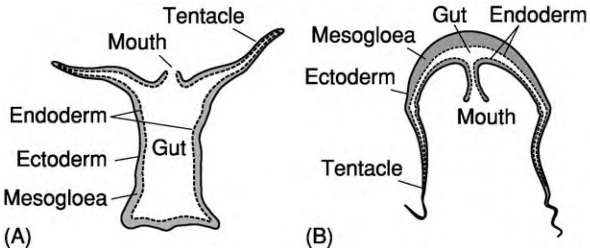


Figure 1 Basic cnidarian anatomy. (A) Polyp. The polyp is shown without partitions in the gut, as in the Hydrozoa. (B) Medusa. The jelly like appearance of jellyfish is due to the thickening of the mesogloea, gelatinous connective tissue separating the ectoderm and endoderm in the roof of the medusoid bell.

The corals are anthozoans that have secreted a calcium carbonate skeleton beneath the polyp or polyp colony. A few hydrozoan polyp colonies also secrete a carbonate skeleton. Some other anthozoans secrete spicules or rarely coral-like skeletons, but most are soft-bodied – for example, the sea anemones. Some hydrozoan polyps and scyphozoan scyphistomas may have an external organic periderm, but the medusae are essentially soft-bodied. The bias in the cnidarian fossil record is directly related to the relative skeletonization of the different groups.

Rare solitary coral-like fossils from the Lower Cambrian of Siberia have been assigned to a fourth, extinct, class of cnidarians, the Hydroconozoa.

Geological History

Precambrian Origins

All three of the principal classes have been claimed to be present in the Late Precambrian Ediacaran fauna (**Precambrian:** Vendian and Ediacaran). Here, internal and external moulds of nonmineralized organisms are preserved in siltstones and fine sandstones. They are controversial, and alternative interpretations have been advanced for these fossils. However, some are convincingly cnidarian. As well as unassigned jellyfish, there are others of likely scyphozoan affinity (*Kimberella*; Figure 3B) and structures reasonably interpreted as the floats of hydrozoan chondrophores (*Ovatoscutum* and *Chondroplon*). The most striking anthozoan is the

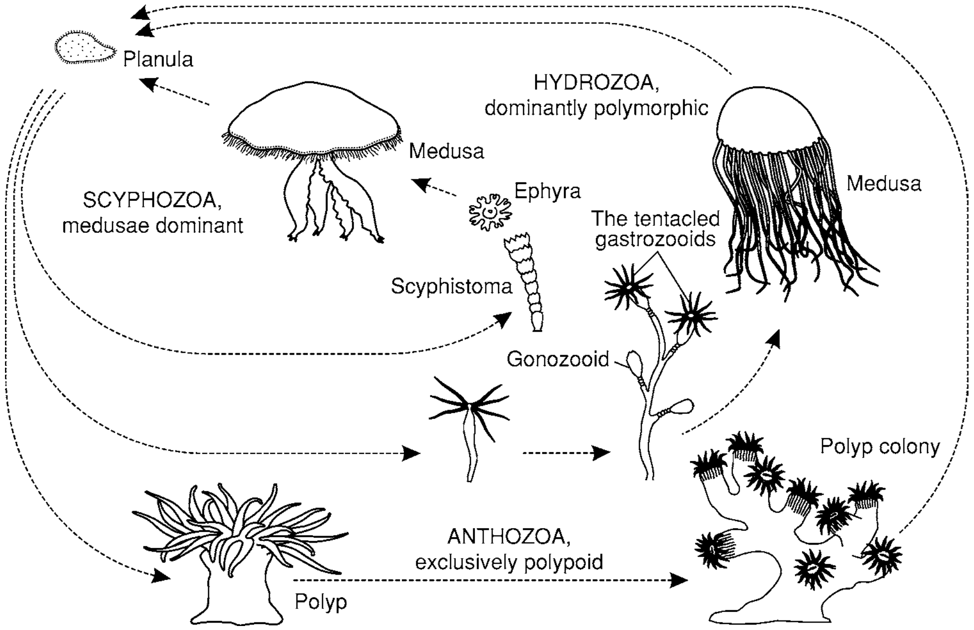


Figure 2 Life cycles of the three cnidarian classes. In the polymorphic Hydrozoa, the polypoid stage is usually colonial, often with specialization of individuals. The tentacled gastrozooids are specialized for feeding and the gonozooids for the production of medusae. The polyp stage in Scyphozoa is reduced to a benthonic structure that buds off medusae from its distal surface. Anthozoans are exclusively polypoid, and those that secrete a calcium carbonate skeleton constitute the corals. Adapted from Hill D and Wells JW (1956) Cnidaria general features, p. 7, Figure 4. In: *Treatise on Invertebrate Paleontology, part F*. Courtesy of and © 1956, The Geological Society of America and The University of Kansas Press.

sea pen *Charniodiscus* (Figure 3A). Other structures may represent the remains of cnidarians that are now extinct.

Unfortunately, even assuming their correct interpretation, these remains throw no light on the relationships within the Cnidaria, but suggest an earlier period of diversification not represented in the fossil record, at least to date. The origin of the classes has to be inferred by indirect means and has been the subject of much discussion. The internal structure of the gut and the character of their nematocysts suggest that the Hydrozoa are the most primitive cnidarians and the Anthozoa are the most advanced. The most favoured scenario envisages a hydrozoan medusa, derived from a planuloid ancestor, as the ancestral cnidarian, with the polyp evolving later as a prolonged juvenile stage.

Phanerozoic Diversity

The type of preservation represented by the Ediacaran fauna is not known after the Precambrian. This is attributed to the rise of active macrophagous predators and scavengers, disruption by burrowing organisms, and possibly the increasing influence of the meiofauna all contributing to the early breakdown of soft tissue in comparable environments. The fossil record of soft-bodied cnidarians becomes sparse and patchy.

Hydrozoa The Hydrozoa are represented by a very few, sometimes doubtfully assigned, medusae such as the questionable *Crucimedusina* (Late Carboniferous) and the more convincing *Hydrocraspedota* (Middle–Upper Jurassic) and *Kirklandia* (Lower Cretaceous). Hydroid polyp colonies are better represented. A rich fauna is known from the Ordovician and Silurian of Scandinavia, some so well preserved that the scleroprotein periderm can be recovered from the calcareous matrix by acid digestion. In younger rocks, hydroid colonies commensal with serpulid worms have been preserved as moulds and casts by overgrowth of the basal stoloniferous network by the calcareous worm tube (*Protulophila gestroi*, Mesozoic – Early Tertiary, Europe and the Middle East; Figure 3D). *Drevotella*, a freshwater hydroid (rare even among living hydroids), is recorded from the Upper Carboniferous Mazon Creek Lagerstätte of Illinois.

Structures interpreted as the floats of chondrophores are relatively common in the Early and Middle Palaeozoic. Examples are *Scenella* (Lower Palaeozoic; previously referred to the Mollusca) and *Plectodiscus* (Lower Devonian). Preservation is helped by the stiff scleroprotein cover surrounding the float. After the Carboniferous, however, for reasons unknown, only a single further example is recorded. The siphonophores appear to have left no fossil record at all.

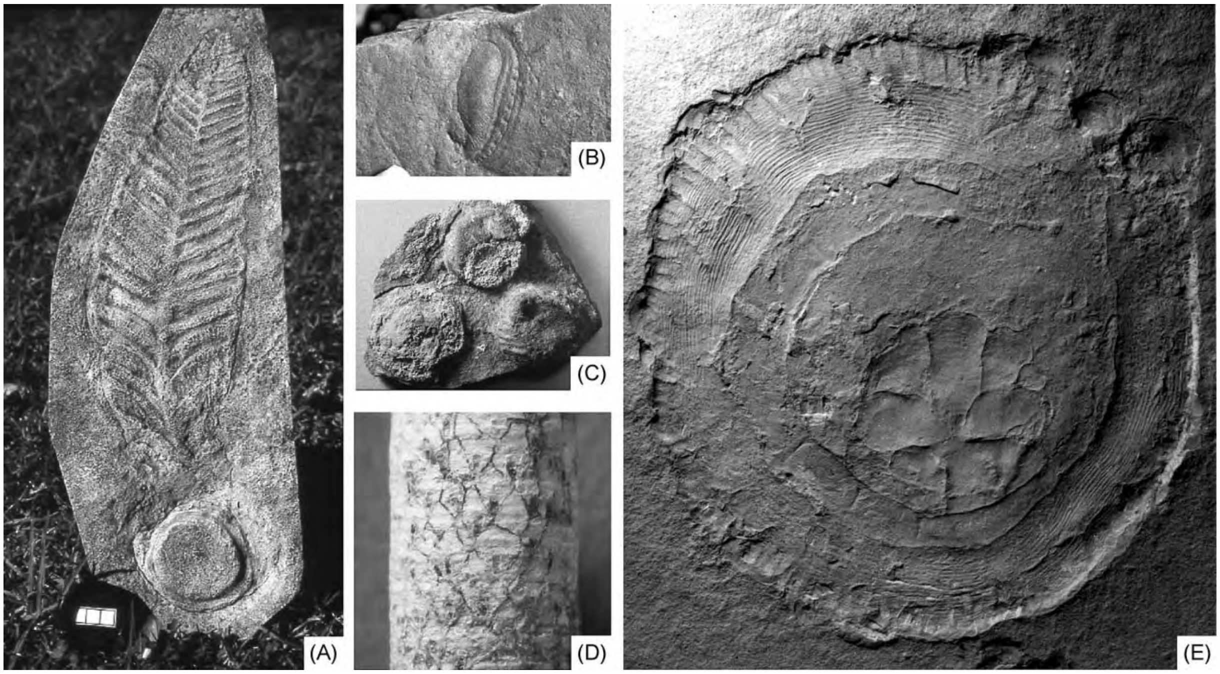


Figure 3 Representative non coralline cnidarians. (A) *Charniodiscus arboreus*; Precambrian, Vendian, Flinders Ranges, South Australia, ×0.2. (B) *Kimberella quadrata*, flank of bell with serial gonads; Precambrian, Vendian, Ediacara, South Australia, ×0.4. (C) *Bergaueria* sp., undersurface of fine sandstone bed with pedal impressions of anemones; Cambrian, Herrería Formation, Cantabria, Spain, ×0.5. (D) *Protulophila gestroi*, pyritized stolonal network and polyp bases exposed by the decortication of the worm tube *Protula* sp.; Cretaceous, Gault Clay, Ford Place, Kent, UK ×3. (E) *Rhizostomites admirandus*, a scyphomedusa; Upper Jurassic, Solenhofen Limestone, Solenhofen, South Germany, ×0.3.

Finally, there are a few calcified hydrozoan polyp colonies in the fossil record (excluding the Stromatoporata and Chaetetida, both formerly assigned to the Cnidaria but now referred to the Porifera). The Milleporina (*Millepora* is found on living reefs, particularly in the Caribbean) and Stylasterina can be traced back to the Late Cretaceous. The hydroid *Hydractinia*, which is now most commonly found encrusting gastropod shells, extends back to the Eocene and possibly the Mesozoic.

Scyphozoa Convincing post-Ediacaran fossil scyphomedusae are extremely rare and virtually restricted to the Solenhofen Lagerstätte (Upper Jurassic, southern Germany). Here beautiful impressions of *Rhizostomites admirandus* reach a diameter of 0.5 m (Figure 3E). Scyphopolyps (scyphistoma) are also rare, the most convincing being the conical scleroprotein tube of *Byronia* (Upper Cambrian–Middle Ordovician, Canada, Poland), which is very similar to the living scyphopolyp *Stephanocyphus*.

A group of relatively common chitinophosphatic tubular fossils from the Lower Ordovician to the Triassic, the Conulata, have been assigned more or less closely to the Scyphozoa by some, although others favour their separation as a discrete phylum. There is evidence of attachment possibly benthonic as juveniles and a pelagic adult stage, although interpretation of their lifestyle is controversial. Their cnidarian affinity is based on a commonly occurring quadrate cross-section and rare examples of internal ribs or septa in multiples of four, comparable to the tetrameral symmetry of scyphozoans.

Anthozoa The exclusively polypoid anthozoans have a much more complex partitioning of the gut than do other cnidarians. Beneath the mouth, a tube (stomodaeum) extends down into the gut, which is subdivided by radial partitions called mesenteries (Figure 4). Those that extend to meet the stomodaeum are termed complete. The mesenteries are the site of digestion, absorption, excretion, and gonad development, and at least some anthozoans can extrude long mesenterial filaments from the mouth to aid feeding.

The three subclasses are distinguished by the pattern of mesenteries in the gut. Relationships between them are not clear from the fossil record so are instead based on the morphology and development of living forms. The ancestral anthozoan is thought to have been a polyp with six unpaired complete mesenteries, which gave rise to the Ceriantipatharia on the one hand and, with the addition of two more complete mesenteries, to the Octocorallia on the other. The latter stock, with the addition of four incomplete

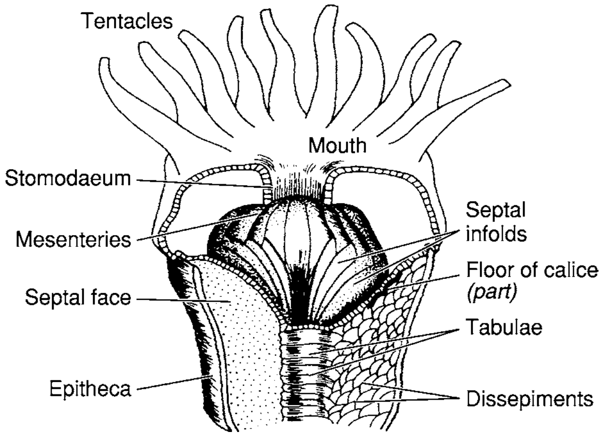


Figure 4 Hypothetical relationship between polyp and skeleton in a solitary rugose coral. Modified from Scrutton (1978, Figure 2A) with the permission of Springer Verlag. Scrutton, CT (1978) Periodic growth features in fossil organisms and the length of the day and month. In: Brosche P and Sundermann J. *Tidal Friction and the Earth's Rotation.* Berlin: Springer Verlag.

mesenteries and their arrangement into six pairs, gave rise to the Zoantharia, the subclass that includes the corals and true anemones.

The Ceriantipatharia has few living representatives and one doubtful fossil from the Miocene. However, it includes solitary, burrowing, anemone-like polyps, which may have left as yet unidentified trace fossils. The Octocorallia is much more important as a living group and is slightly better represented by fossils. The Pennatulacea (sea pens), convincingly present in the Precambrian Ediacaran fauna, have a few doubtful records in the Palaeozoic but more credible remains in the Mesozoic. Gorgonians (sea fans) are also sparsely and doubtfully represented in the Palaeozoic, apart from Ordovician material such as *Pragnellia*, but from the Jurassic onwards scattered fossils of the solid calcified horny axes of the polyp colonies are recorded. Alcyonarian spicules have been recorded from the Silurian of Sweden. One group of octocorals, the Coenothecalia, distinguished by a solid calcareous skeleton, includes the 'blue coral' *Heliopora* and has a record stretching back to the Upper Cretaceous. *Heliopora* and the living stoloniferous octocoral *Tubipora* are homoeomorphic with the Palaeozoic tabulate corals *Heliolites* and *Syringopora*, respectively, leading some formerly to classify the Tabulata as octocorals. The similarities are superficial, however, and, although they are anthozoans, the Tabulata are not considered to be closely related. Of the Zoantharians, the fossil record of anemones is extremely poor. The oldest appears to be a recent record from the Cambrian of China. *Mackenzia costalis* (Middle Cambrian Burgess Shale, British Columbia) is now considered to be an actinian anemone but is in need of revision, and Palaeactinia is another questionable actinian (Middle Ordovician, New York). Trace fossils of actinian burrows have also been claimed, including *Intrites* (Vendian) and the relatively common *Bergaueria* (Vendian–Middle Ordovician; Figure 3C). *Dolopichnus* (Early Triassic), if correctly interpreted, is the only record of a zoantharian from that interval and has a special significance in the evolution of the corals.

The corals The corals are essentially anemones that have secreted a calcium carbonate skeleton beneath the polyp (Figure 4), which then occupies a cup or platform at the top of the skeleton (the calice). The skeleton (corallum) is basically discoidal, conical, or cylindrical, with or without an external wall (epitheca) and radial disposed vertical plates or spines (septa), and usually with horizontal partitions that range from simple flat plates to different combinations of variously shaped plates and vesicles (tabulae and dissepiments). When the coral is a modular colony, each module secretes its own skeleton, the form of which is dependent on whether the modules separate completely, remain in contact, or are only partially separated.

Eight orders of corals are now recognized, of which three are by far the most important: the Rugosa, Tabulata, and Scleractinia (Figure 5). They are divided into two broad stocks on the basis of their known or inferred mesenterial and septal development and skeletal mineralogy. The aragonitic Scleractinia and the closely related actinian/corallimorpharian anemone group have a primary arrangement of six paired mesenteries, with subsequent sets of mesenteries inserted cyclically between the pre-existing cycles (Figure 6C and 6D). This leads to a sequence of 6, 6, 12, 24, 48, etc., paired mesenteries in successive cycles, which is generally reflected in the insertion of septa in the coralla of scleractinian corals. This pattern of septal insertion can be recognized in fossil scleractinians and in two minor extinct orders of aragonitic corals: the Kilbuchophyllida (Figure 7E) and less certainly the Numidiaphyllida. The two remaining major groups of corals, the Rugosa and Tabulata, are both extinct. The calcitic Rugosa (Figure 8) mostly have well-developed septa, which show a pattern of insertion quite distinct from that in the Scleractinia (Figure 6A). They also have six protosepta, reflecting an initial set of six paired mesenteries, but subsequent insertion of major septa is serial in only four of the loci thus defined. In addition, minor septa are usually present between the major septa. There is some dispute concerning the significance and interpretation of this, but by far the

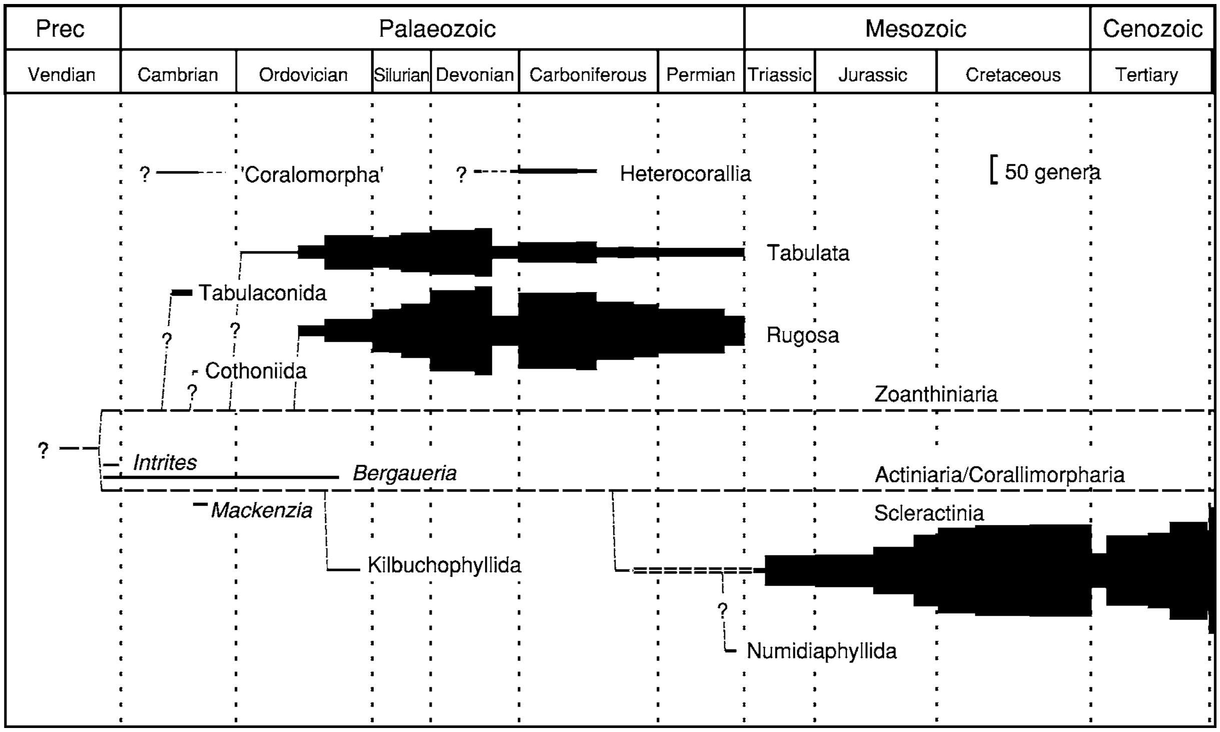


Figure 5 Geological ranges of the zoantharian corals. Their relationships to the two stocks of living anemones are indicated. The double dashed line leading to the Scleractinia notionally indicates its polyphyletic origins in the Middle Triassic. The Hetercorallia are sufficiently dissimilar from the other corals to suggest that they might have evolved from a quite separate anemone stock, which is now extinct. The Coralomorpha is an informal grouping of coral like fossils, the affinities of which are not yet determined. *Intrites* and *Bergaueria* are thought to be actinian trace fossils, and *Mackenzia* is a doubtful fossil actinian anemone. The generic diversities of the Tabulata, Rugosa and Scleractinia are indicated. Modified from Scrutton (1997, Figure 2) The Palaeozoic corals, I: origins and relationships. *Proceedings of the Yorkshire Geological Society*, 51: 177 208 with the permission of the Yorkshire Geological Society.

majority opinion is that these corals have an inferred pattern of mesenterial development similar but not identical to that in a small group of living anemones, the Zoanthiniaria (Figure 6B). Septa in the Tabulata (Figure 9) are also calcitic and are often poorly developed, but one small group has a pattern of insertion similar to that in the Rugosa. Assuming the integrity of the Tabulata as a monophyletic clade, which has been controversial, this suggests that they may, like the Rugosa, be descended from a zoanthiniarian ancestor. Not only have the tabulates been regarded as a mixed bag of unrelated corals, but some if not all of them have been claimed, among other things, to be calcareous sponges. Indeed, some fossils formerly assigned to the Tabulata, such as the Chaetetida, have been properly reclassified as sponges, but the group as currently understood is generally accepted as a valid coral taxon, and favositid tabulates have been found with convincing polyps preserved in the calices (Silurian, Anticosti Island; Figure 9C). Of the remaining orders, which are all apparently calcitic, the most important is the Heterocorallia, in which septal insertion is disputed but is quite unlike either pattern described above (Figure 7F). Its relationships to other corals are unclear, and it may have evolved from a separate polyp stock that is now extinct. The other minor orders, the Cothoniida and Tabulaconida, have not yielded identifiable patterns of septal insertion but are generally considered to be closest to the zoanthiniarian stock. It seems likely that the two polypoid stocks, zoanthiniarian and actinian/corallimorpharian, had already diverged by the beginning of the Cambrian (Figure 5).

Corals first appear in the Lower Cambrian. The Tabulaconida (*Moorowipora*, *Arrowipora*, *Tabulaconus*) have features similar to those of tabulate corals, to which some have referred them. However, the origination of new corallites (increase), at least in *Moorowipora* (Figure 7A–C), is more like that found in rugosans. Altogether, they possess features that do not suggest that they are ancestral to any of the Ordovician corals, in addition to which they disappear with the collapse of the archaeocyathan reef system, with which many are associated, at the end of the Lower Cambrian. These reefs and other carbonate environments also yield a range of other coralline organisms, which are doubtfully true corals and in some cases doubtfully cnidarians. *Flindersipora*, for example, has more in common with the enigmatic

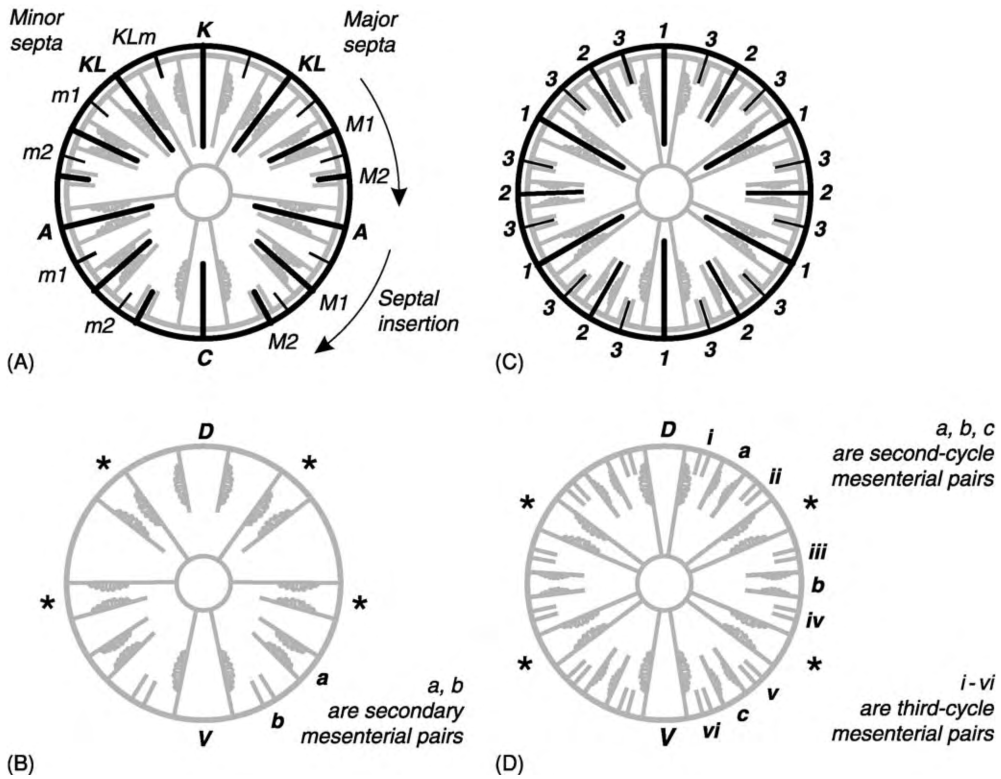


Figure 6 Relationship between septal insertion in (A) the Rugosa and (C) the Scleractinia and mesenterial insertion in (B, D) their presumed ancestral anemone stocks. (C, D) Scleractinian polyps and corallimorpharian anemones appear to be virtually identical except for the skeleton secreted by the former. (A, B) The relationship between the Rugosa and a zoanthiniarian like ancestor is more speculative. Modified from Scrutton (1997, Figure 7) The Palaeozoic corals, I: origins and relationships. *Proceedings of the Yorkshire Geological Society*, 51: 177 208 with the permission of the Yorkshire Geological Society.

Ordovician tetradiids, formerly grouped with the Tabulata but now considered of uncertain affinity. The Cothoniida from the Middle Cambrian (*Cothonion*; Figure 7D) are clustered cup-shaped corals with opercula, on the inside of which septa are well developed. However, no insertion pattern can be deduced. They are most similar to Rugosa, but with no features suggesting direct descent to the first true rugosans, which did not appear until the Mid-Ordovician. All the Cambrian corals are considered to be short-lived experiments in skeletonization, probably from the same anemone stock that later gave rise to the Rugosa and Tabulata.

The earliest accepted tabulate coral, *Lichenaria*, appeared in the Lower Ordovician. Diversity remained low until the Middle Ordovician, when the first Rugosa appeared and both orders, but particularly the Tabulata, underwent a modest diversification. By the time of the Late Ordovician extinction event, some 60 genera of Tabulata and 40 genera of Rugosa were established. The Kilbuchophyllida (*Kilbuchophyllia*; Figure 7E) also appeared in the Middle Ordovician but remained restricted to a small area of the UK and quickly became extinct.

The Tabulata were more seriously affected than the Rugosa by the Late Ordovician extinction event, after which the Rugosa were always generically the more diverse. Both peaked in diversity in the Middle Devonian, and both suffered severely in a series of extinction pulses beginning at the end of the Middle Devonian and continuing until the end of the period. Thereafter, the Tabulata remained a much-reduced presence, whilst the Rugosa recovered almost to peak diversity in the Lower Carboniferous. Both were affected by early Upper Carboniferous extinction before beginning a long, slow decline until final extinction in the Late Permian. The enigmatic Heterocorallia appeared in the Devonian, reached their acme in the Lower Carboniferous, and became extinct in the early Upper Carboniferous (Figures 7F and 7G). In addition, the Numidiaphyllida (*Numidiaphyllum*) made a brief and local appearance in the Late Permian.

The Scleractinian corals (Figures 10 and 11) first appear in the Middle Triassic. Formerly it was proposed that the scleractinians evolved from the rugosans of the Late Palaeozoic, but this is now generally discounted. Several lines of evidence support this

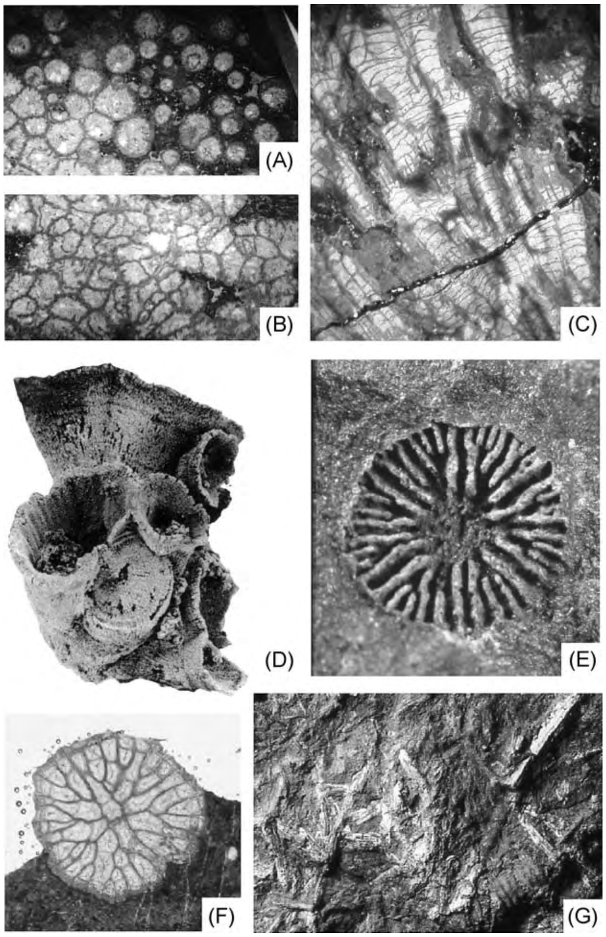


Figure 7 Representative Cambrian and minor group corals. (A–C) *Moorowipora chamberensis*; Lower Cambrian, Moorowie Formation, Flinders Ranges, South Australia, ×5: (A, B) cross sections and (C) longitudinal section; photographs courtesy of Margaret Fuller. (D) *Cothonion sympomatum*; early Middle Cambrian, Coonigan Formation, New South Wales, Australia, ×4.5; photograph courtesy of John Jell and reproduced by permission of the Australasian Association of Palaeontologists; figure 2 in Jell PA and Jell JS (1976) Early Middle Cambrian corals from western New South Wales. Alcheringa 1: 181–195. (E) *Kilbuchophyllia discoidea*, mould of a juvenile specimen showing clear cyclic septal insertion; Ordovician, Caradoc Series, Kirkcolm Formation, Southern Uplands, Scotland, UK, ×3.5. (F) *Heterophyllia grandis*, cross section showing distinctive heterocoral septal pattern; Carboniferous, Dinantian, Northumberland, England, UK; ×5. (G) *Heterophyllia ornata*, scattered specimens on a bedding plane; note the spines on the edges of the corallite top right; Carboniferous, Dinantian, Northumberland, England, UK; ×1.

conclusion: the lack of Early Triassic corals; a change in skeletal mineralogy from calcite (Rugosa) to aragonite (Scleractinia); contrasting patterns of septal insertion between the two orders; and proof of the antiquity of the scleractinian septal-insertion pattern with the discovery of *Kilbuchophyllia* in the Ordovician. This implies that the ancestral anemone stock survived from the Palaeozoic, and the Early Triassic trace fossil *Dolopichnus* may be proof of that. Furthermore, the earliest scleractinians suggest several already divergent stocks, which became mineralized at about the same time, the origins of which recent DNA work on living descendants suggests can be traced back to the Carboniferous (Figure 5). Following severe extinctions at the end of the Triassic, scleractinians steadily diversified through the Mesozoic, with a peak of reef-building activity in the Late Jurassic and less significant losses at the end of that period. The end-Cretaceous extinction event severely affected coral stocks, but from the survivors evolved the highly diverse and ecologically successful scleractinian faunas of the present day. No other group of zoantharian corals is recognized after the end of the Palaeozoic.

Coral Structure and Taxonomy

Coral taxonomy and classification is based principally on the skeletal characteristics of the corallum, the skeletal microstructure, and growth structure and form. Coral skeletons are often highly variable in internal structure and growth form. In addition, similar structures often reoccur, and it can be difficult to separate genetic relationships from convergent homoeomorphy. Given that coral structure is relatively simple to start with, the result is that considerable problems exist with both taxonomy and classification within the major orders.

Many species were erected in the past with little or no regard to intraspecific variation. Only in recent years have more serious attempts been made to take variation into account, with the result that many species and genera are now being placed in synonymy. Rather more stability is apparent in the grouping of genera at the family level, at least in Palaeozoic corals. Here the *Treatise on Invertebrate Paleontology* and its revision have been very influential, although now much modified by more recent work. However, evolutionary relationships within the Rugosa particularly are rather poorly understood. The *Treatise* has also been a major influence in Scleractinian classification, but here major revisions are being made as a result of more recent studies of better-preserved material, particularly from faunas in which the original aragonitic mineralogy has been preserved, allowing septal microstructure to contribute to unravelling relationships. Microstructure has yet to play a very significant role in classifying Palaeozoic corals. Their calcitic skeletons contain up to 8 mol% $MgCO_3$, and there is great controversy over which textures are original and which are secondary as a result of recrystallization.

The Kilbuchophyllida are solitary corals, whilst the Heterocorallia, the Rugosa and the Scleractinia contain both solitary and modular growth forms. All tabulate corals are modular colonies, as are all of the accepted Cambrian corals and the Numidiaphyllida. Overall, solitary corals show a wide variety in form, but this may be consistently developed within many species. Discoidal (Figures 8B and 10A–C) and similar broad, flat growth forms (tympanoid, cupolate) and those with a noncircular cross-section (calceoloid, flabellate, cuneiform, etc.) tend to be the least variable at the species level. Conicocylindrical solitary corals (Figure 8C) may show more irregular forms as the result of disturbance during growth, at least among the Rugosa where adult solitary corals are generally free-living. Colonial growth form in the skeleton varies from branching to massive, ramose, and foliose. In many branching (fasciculate) forms, the polyps separate completely, and the clonal group is held together only by the corallites they secrete beneath themselves (Figure 8H and 8I). The cluster of corallites may be loose and divergent (dendroid; Figure 10E) or close and parallel (phaceloid), and the arrangement appears to be under genetic constraint in many cases. Corallites may remain attached on opposite sides, like the palisades in a fence, the ranks joined together to enclose lacunae in the skeleton (cateniform; Figure 9G and 9H). An external wall may enclose the ranks but along them adjacent corallites may be in contact or separated by colonial tissue, suggesting neural connectivity between the polyps. A range of structural relationships between the modules exist in massive, ramose, and foliose growth forms. Corallites may be packed more or less tightly together, becoming polygonal in cross-section, each retaining their own wall (rugosan cerioid, suggesting complete separation of the polyps; Figure 8J–M). However, various levels of integration of the polyps may be reflected in the loss of individual corallite walls (for example scleractinian cerioid, Figure 10D)

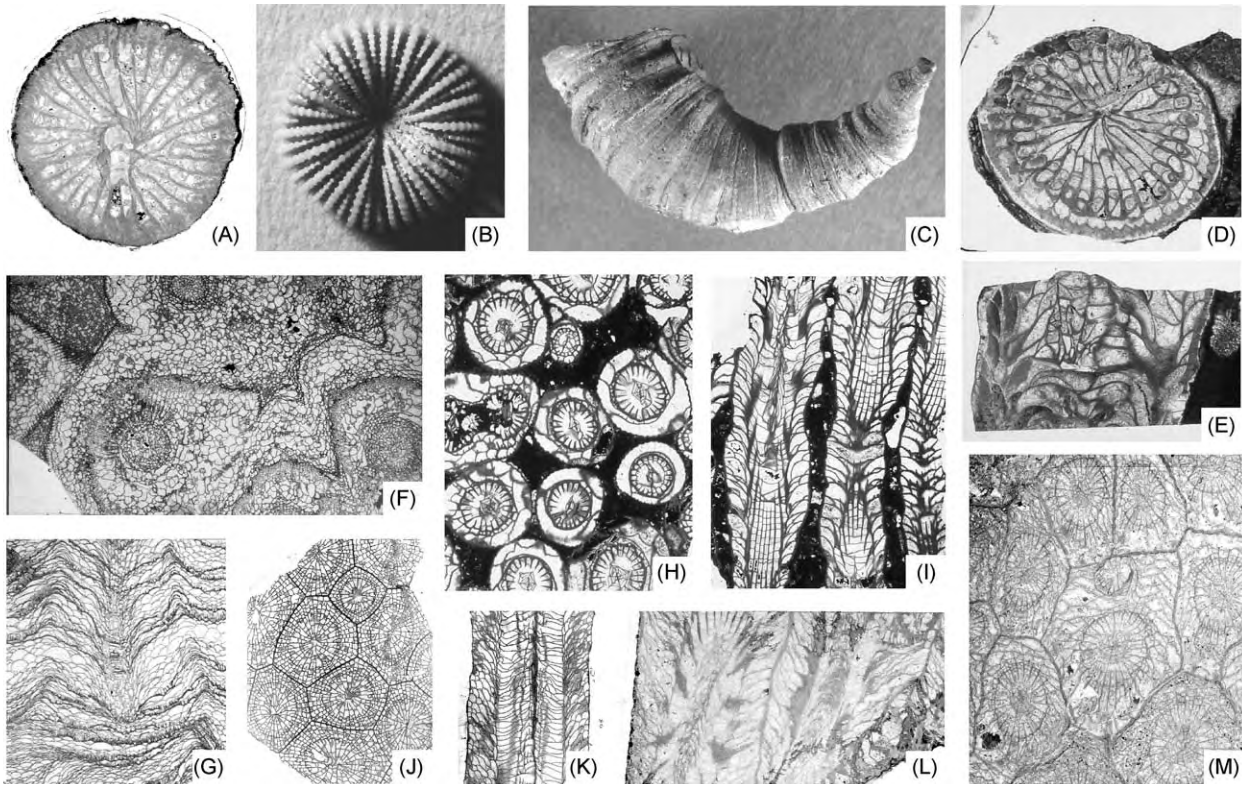


Figure 8 Representative Rugose corals. (A) *Amplexizaphrentis* sp. showing a pinnate arrangement of septa reflecting the pattern of insertion; Lower Carboniferous, Dinantian, Northern England, UK; ×1.5. (B) *Palaeocyclus porpita*, discoidal corallum showing a pinnate arrangement of septa particularly about the very short cardinal septum, bottom centre; lower Silurian, Llandovery Series, Gotland, Sweden ×2.5. (C) Typical solitary horn coral, shown in growth orientation; Carboniferous, locality unknown; ×2. (D, E) *Tabulophyllum* sp., showing well developed septa, dissepiments, and tabulae; Upper Devonian, Frasnian, East Ogwell Limestone, south Devon, UK; ×1.5: (D) cross section and (E) longitudinal section. (F, G) *Arachnophyllum murchisoni*, an amural colony with (G) pseudo walls formed of stacks of dissepiments; Silurian, Much Wenlock Limestone, Shropshire, UK; ×1: (F) cross section and (G) longitudinal section. (H, I) *Lonsdaleia duplicata*, phaceloid fasciculate colony; Carboniferous, Dinantian, Cumbria, UK; ×1: (H) cross section and (I) longitudinal section. (J, K) *Lithostrotion araneum*, massive cerioid colony; Carboniferous, Dinantian, northern England, UK; ×1.3: (J) cross section and (K) longitudinal section. (L, M) *Actinocyathus floriformis*, massive cerioid colony showing corallites with well developed axial structures; Carboniferous, Namurian, Great Limestone, Co. Durham, UK; ×1.3: (L) longitudinal section and (M) cross section.

and, ultimately, by calical centres set in continuous colonial tissue (amural (Figure 8F and 8G), coenenchymal (Figure 9I and 9J) and plocoid (Figures 10G, 11A and 11B in ramose growth forms)). In the Rugosa and Tabulata, individuals are still sufficiently well defined to suggest that polyp integration was restricted to neural connectivity in most cases and only rarely was there confluence of body cavities. However, some massive and branching tabulates have tubes or pores connecting adjacent corallites (Figure 9A and 9B), suggesting interconnected gastric cavities. In some Scleractinia, separation of polyps may be more or less incomplete, resulting in small clusters (polycentric phaceloid; Figure 11B) or linear series (flabellate, meandroid; Figure 10F) of mouths opening into a single gut and surrounded by a single series of tentacles.

Corallite structure in colonial corals is either relatively invariant or restricted to a well-defined and narrow range at species level and often also at generic level. Growth form is more variable. In the past, it has been considered to reflect environmental influences, but increasingly the importance of underlying genetic controls has become apparent, at least for some corals.

Coral Ecology and Palaeoecology

Scleractinia

Living scleractinian corals range from the tropics to high latitudes (<75°) and from the surface to abyssal depths (Figure 12). However, two ecological groups can be distinguished, one with symbiotic algae (zooxanthellae), the other without (Table 1). Zooxanthellate scleractinians dominate living coral reefs and are restricted in their distribution by their algal symbionts to water temperatures of 16–40°C and shallow water depths, generally less than 190 m. Azooxanthellate scleractinians can exist through a much broader range of temperature and depth. They are minor components of shallow-water tropical reefs but build structures of their own (coral thickets and banks) in deeper and cooler waters.

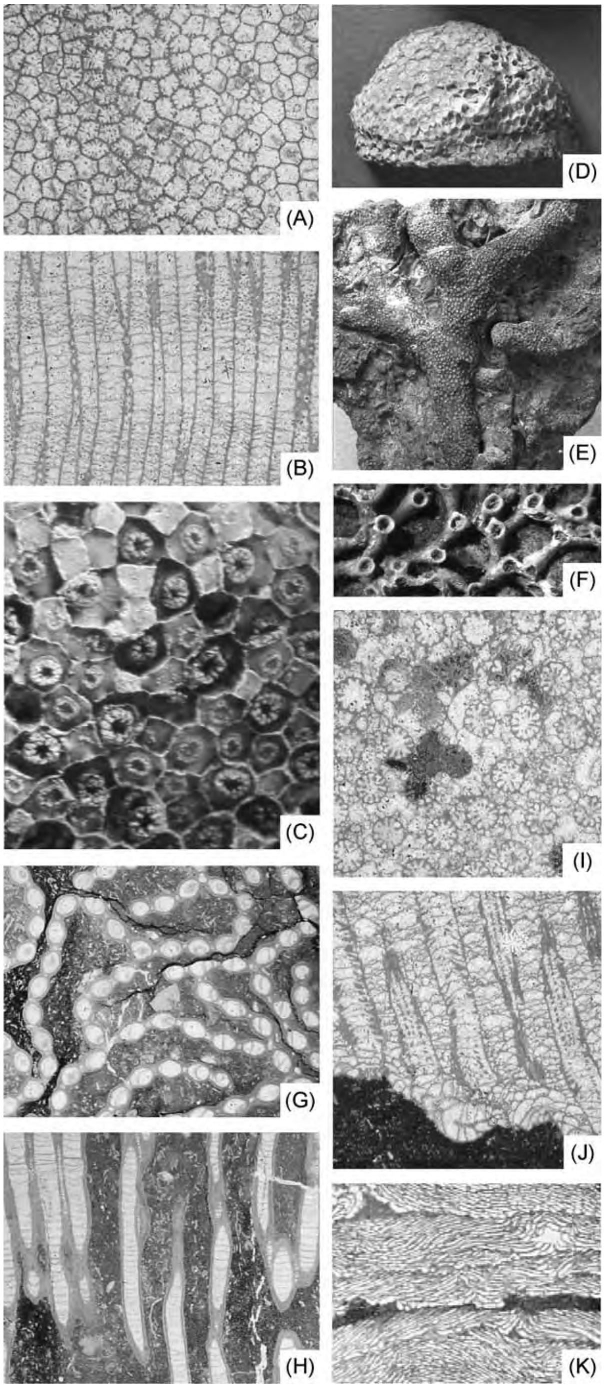


Figure 9 Representative Tabulate corals. (A, B) *Paleofavosites asper*, massive cerioid colony with well developed pores connecting adjacent corallites; septa are well developed in denser bands, with more crowded tabulae (reflecting an annual growth cycle), and weak to absent elsewhere; Silurian, Much Wenlock Limestone, Shropshire, UK; ×3: (A) Cross section and (B) longitudinal section. (C) *Favosites* sp., with polyps preserved in many calices; Silurian, Llandovery Series, Jupiter Formation, Anticosti Island, Canada; ×4. Photograph courtesy of Paul Copper. (D) *Paleofavosites rugosus*, a small domal colony with a flat base: Silurian, Much Wenlock Limestone, Shropshire, UK; ×0.75. (E) *Thamnopora* sp., ramose growth form, the Palaeozoic coral that is structurally closest to the highly successful *Acropora* of living reefs (see Figure 11); Devonian, Portilla Formation, Bernesga Valley, northern Spain; ×0.5. (F) *Aulopora* sp., encrusting colonial mesh work of horn shaped coralla; Silurian, Klinteberg Formation, Gotland, Sweden; ×2. (G, H) *Halysites catenularius*, cateniform colony with corallites separated by thin tubules of colonial tissue along the ranks; Silurian, Much Wenlock Limestone, Shropshire, UK; ×2: (G) cross section and (H) longitudinal section. (I, J) *Propora* sp., coenenchymal colony with vesicular colonial tissue, from which corallites arise at the base of the colony; Silurian, Llandovery Series, Petalocrinus Limestone, Woolhope, Herefordshire, UK; ×3: (I) cross section and (J) longitudinal section. (K) *Subalveolites panderi*, longitudinal section of a colony formed by encrusting layers of oblique corallites; Silurian, Llandovery Series, Hughley Shales, Devil's Dingle, Shropshire, UK; ×2.5.

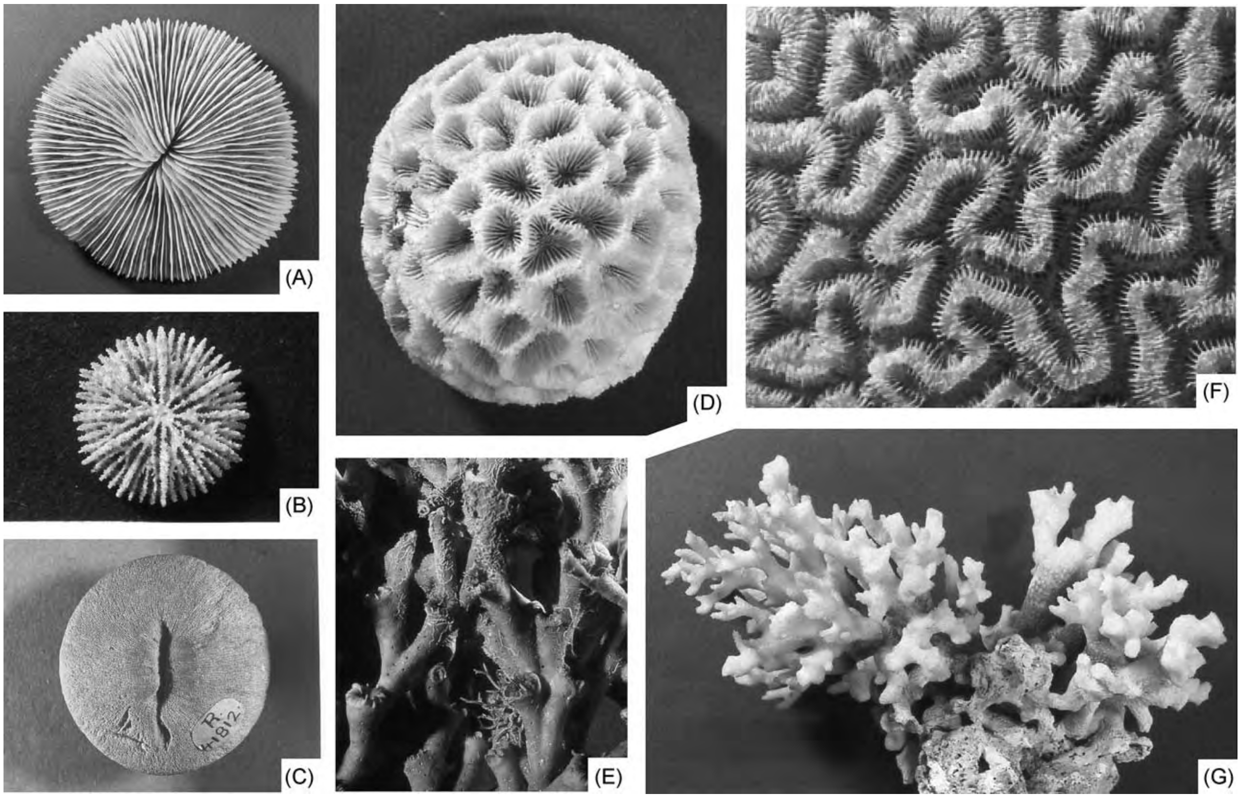


Figure 10 Representative Scleractinian corals. (A) *Fungia fungites*, calical view of discoidal zooxanthellate corallum; Recent, Great Barrier Reef, Queensland, Australia; ×0.25. (B) *Fungiacyathus symmetricus*, discoidal azooxanthellate corallum; Recent, locality unknown; ×3.5. (C) *Cyclolites* sp., calical view of zooxanthellate cupolate corallum with large numbers of closely spaced perforate septa; Cretaceous, Spain; ×0.8. (D) *Favites* sp., cerioid (scleractinian usage) colony in which intercorallite walls are formed by the interlocking of the peripheral ends of the septa (a septotheca); Recent, Great Barrier Reef, Queensland, Australia; ×0.7. (E) *Lophelia prolifera*, azooxanthellate dendroid colony with well spaced corallites; Recent, coral bank on continental shelf, North West Atlantic; ×2.5. (F) *Diploria* sp., typical zooxanthellate meandroid colony (brain coral); Recent, West Indies; ×1. (G) *Stylophora pistillata*, ramose zooxanthellate colony in which each branch is made up of many plocoid corallites; Recent, Great Barrier Reef, Queensland, Australia; ×0.5.

The role played by zooxanthellae is not entirely clear, but they make an important contribution to the energy budget of the coral, contributing up to 98% of its fixed carbon requirement. They appear to accelerate calcification, which in the extreme case of the Stag's horn coral *Acropora* can achieve a linear growth rate of 270 mm yr^{-1}. However, many massive Scleractinia appear not to have enhanced growth rates, and recent evidence suggests that, in this case, zooxanthellae may act to switch calcification on and off, rather than accelerate it, thus conserving energy. The algae also remove waste from the coral and may thereby allow very large colonies (in excess of 10 m) of tiny individuals (approximately 1 mm in diameter) to develop. There is a striking contrast in coloniality between zooxanthellate and azooxanthellate corals. The former are 95% modular by genus, whereas the latter are about 30%. The symbiosis with algae appears to have developed in the early Jurassic.

A major factor in the success of scleractinian corals in building rigid structures is their ability to cement their skeletons securely to other skeletal material and to encrust hard substrates (Figure 11). Shallow-water scleractinian-dominated reefs became important from the Jurassic onwards, although deep-water banks built by azooxanthellate scleractinians extend back into the Triassic.

Palaeozoic Corals

Rugose and tabulate corals enjoyed neither the advantages of algal symbiosis nor the ability to encrust securely to the extent seen in the scleractinians. The first point has been contentious. However, scleractinian zooxanthellate colonial corals show flattening of the skeleton and a significant reduction in growth rate with depth, as light levels decrease. A recent study of a widely distributed tabulate coral, among those considered most likely to have harboured algal symbionts, showed no signs of these effects. In addition, only 35% of rugosans (but all tabulates) are colonial, and massive growth forms of any Palaeozoic coral appear not to exceed about 2 m in diameter. Apart from some

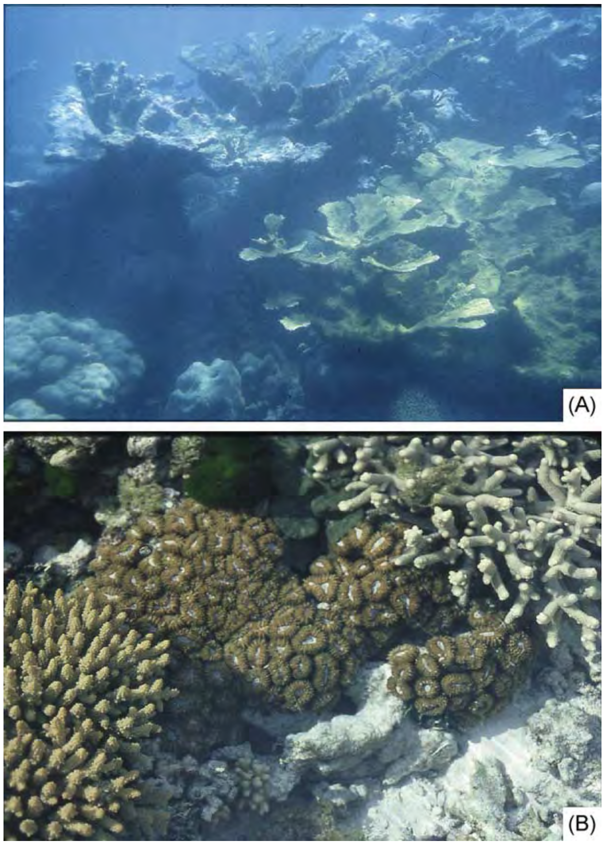


Figure 11 Reef forming scleractinian corals. (A) Reef edge dominated by *Acropora palmata*; open ocean to the right; East Rio Bueno, Jamaica. (B) Reef flat with phaceloid, partly polycentric colony of *Lobophyllia hemprichii* in the centre, flanked by ramose, plocoid colonies of *Acropora* spp; Heron Island, Great Barrier Reef, Queensland, Australia.

specialized forms, rugose and tabulate corals showing secure encrusting attachment or overgrowth as mature coralla are rare. This reflects the presence of an external wall (epitheca or holotheca) around the corallum and the lack of an extensive edge zone of soft tissue, the reverse of the situation in most scleractinians. Instead, the Palaeozoic corals were predominantly adapted to life on soft substrates. They were subsidiary contributors to reef cores, although common in peri-reefal communities, and reached their greatest importance in bioherms and biostromes, with little evidence of frame building and were often associated with mud-grade carbonate environments. Many show evidence of instability and smothering as soft substrates were disturbed by storm activity and sediment movement. They reached their maximum diversity in warm shallow waters, with diversity decreasing with depth and towards the contemporary poles. In addition, there was a distinctive low-diversity community of small, thick-walled, tabulate and solitary rugose corals characteristic of dysaerobic, dysphotic and other marginal environments that persisted from the Silurian to the

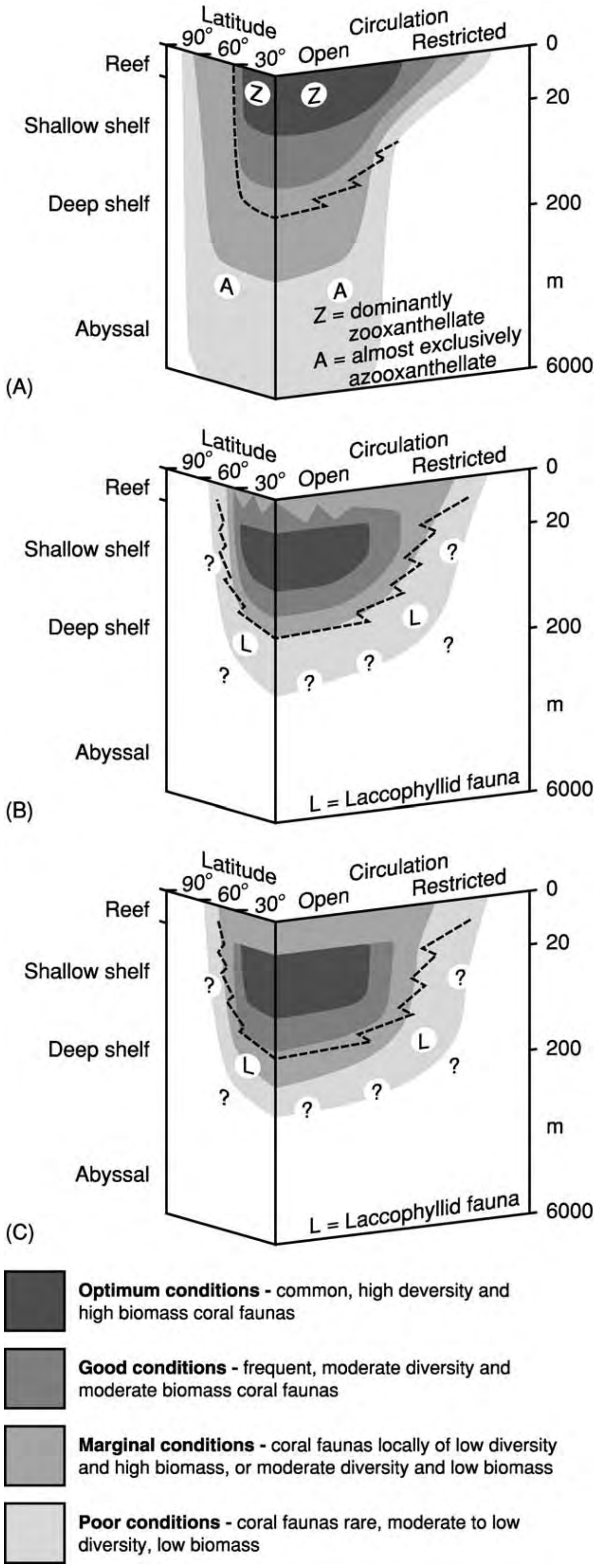


Figure 12 Generalized ecological and palaeoecological ranges of (A) the Scleractinia, (B) the Tabulata, and (C) the Rugosa. Depth not to scale, and divisions diagrammatic. For further discussion see text. Modified from Scrutton (1998, Figure 30) The Palaeozoic corals, II: structure, variation and palaeo ecology. *Proceedings of the Yorkshire Geological Society*, 52: 1 57 with the permission of the Yorkshire Geological Society.

Table 1 Comparative features of the two ecological groups of the Scleractinia, the Rugosa, and the Tabulata

	Scleractinia			
	Zooxanthellate	*Azooxanthellate*	*Rugosa*	*Tabulata*
Algal symbiosis	Present	Lacking	Lacking	Lacking
Epitheca/ holotheca	Mainly absent	Present in some	Almost universal	Present in most
Solitary forms	<1 m+ in diameter	Small	*c.* 4 140 mm in diameter	None
Colonial forms[a]	95% colonial; colonies may be large up to 10 m+ in diameter; corallites often very small, *c.* 1 50 mm in diameter	30% colonial; mainly dendroid; corallites well spaced, medium sized	36% colonial; colonies mainly small 30 400 mm, some up to 1 m or rarely 2 m in diameter; fasciculate rarely larger	All colonial; colonies mainly small 30 600 mm, but up to 2 m; corallites 1 20 mm in diameter
Growth rates	*Acropora* 100 150 mm yr^{1}, up to 270 mm yr^{1}; massive forms 5 15 mm yr^{1}	5 15 mm yr^{1}, up to 40 mm yr^{1} in some shallow branching forms	3 10 mm yr^{1} (massive), up to 27 mm yr^{1} (solitary),? rarely up to 70 mm yr^{1}	2 9 mm yr^{1} in most but up to 20 mm yr^{1}
Skeleton	Often light, porous	Usually solid, non porous	Usually solid, rarely porous	Usually solid, rarely porous
Depth	Most 0 190 m, abundant at less than 70 m, optimum 1 25 m	0 6200 m, most less than 800 m	Usually less than 60 m, recorded down to 120 m, probably extending much deeper	Common down to about 60 m, probably extending much deeper
Temperature	16 40°C, 25 29°C optimum	1°C 29°C	Most tropical to temperate	Most tropical to temperate
Ecology	Mostly hermatypic[b]; major component of shallow water reefs	Mostly ahermatypic[b] but fasciculate colonies form banks at depths of 100 1500 m	Common particularly in shelf bioherms and biostromes; minor component of reefs	Common particularly in shelf bioherms and biostromes; locally significant in reefs

[a] Percentage coloniality is based on numbers of genera.

[b] Hermatypic forms are those associated with shallow water reefs.

Modified from Scrutton (1998, Tables 3 and 4) The Palaeozoic corals, II: structure, variation and palaeoecology. *Proceedings of the Yorkshire Geological Society*, 52:1 57 with the permission of the Yorkshire Geological Society.

Late Permian (the Laccophyllid fauna; Figures 12B and 12C). It was characteristically cosmopolitan compared with the provincialism of other coral faunas.

Glossary

Amural Colonial rugose corals in which the walls separating individual corallites are lost

Ahermatypic Non-shallow-water reef-building scleractinian corals

Azooxanthellate Scleractinian corals lacking symbiotic algae (zooxanthellae)

Calceoloid Growth form of solitary rugose corals in which the corallite is semicircular in section and rests on its flattened face. Sometimes called 'slipper-shaped'

Calice Upper surface of the corallite on which the polyp rests. Calices may vary from convex to concave but most are shallow to deep and funnel or cup shaped

Cateniform Colonial corals in which corallites are in contact along their opposite sides like posts in a fence. This arrangement is called a rank, and ranks join at various points in the colony to enclose more or less irregularly shaped lacunae. Ranks in cross-section look like chains, hence 'chain corals'

Cerioid Structural arrangement of corallites in colonial corals such that each corallite is polygonal in section and in close contact with its neighbours. Usage differs in the Rugosa and Scleractinia in that in the former each corallite retains its external wall (epitheca), whereas in the latter the corallites are separated by the merging of peripheral septal ends from adjacent corallites to form a septotheca

Coenenchymal Structural arrangement of corallites in colonial corals such that corallite centres are set in common colonial tissue (coenenchyme). Coenenchyme may have various forms, but is usually tubular or vesicular

Corallite, corallum The skeleton secreted beneath a single polyp. Corallum is also used for the skeleton of a colonial coral consisting of many corallites

Cuneiform Growth form of solitary scleractinian corals in which the corallite is wedge shaped

Cupolate Growth form of solitary scleractinian corals in which the corallite has a flat base and a highly domed upper (calicular) surface

Dendroid Structural arrangement of corallites in a branching (fasciculate) coral in which each branch is a single corallite and the branches are irregularly divergent

Diploblastic Level of organization in primitive invertebrates in which the body wall consists of external and internal cellular layers (ectoderm and endoderm), which may be separated by a layer of noncellular connective tissue

Discoidal Growth form of solitary corals in which the corallite has a flat base and a flat to gently concave upper (calicular) surface

Dissepiment A vesicular skeletal plate, subhorizontal to inclined zones of which occupy the peripheral areas of many corallites (dissepimentarium). In scleractinian corals, vesicular and tabular dissepiments are distinguished

Dysaerobic Environment with 0.1–1 ml O_2 per litre of water

Dysphotic Environment with low light levels

Enteron Cavity in the body of a cnidarian polyp or medusa; the gut

Epitheca The external wall of a corallite

Flabellate Growth form of scleractinian corals, either a fan-shaped solitary coral with an elliptical cross-section or a meandroid corallum consisting of a single, linear, polycentric corallite

Foliose Growth form of colonial corals in which the branches are broad and more or less flat laminae

Hermatypic Scleractinian corals that build shallow-water reefs

Holotheca The external wall of a colonial coral in which the individual corallites have lost their own walls (epithecae)

Homoeomorphy Similarity of form resulting from convergence rather than genetic relationship

Increase The process of new corallite formation in a colonial coral

Lagerstätte Fossil deposit in which organisms not normally preserved, such as those lacking mineralized skeletons, are found

Meandroid Structural arrangement of corallites in scleractinian colonial corals in which a number of linear polycentric corallites are intertwined (a brain coral)

Medusa The free-floating stage of the cnidarian body plan in which the mouth faces downwards surrounded by tentacles

Mesentery A radially disposed fleshy infold of the endoderm of a cnidarian polyp, attached to the inner surfaces of the oral and basal discs. The inner end of the lamina may be attached to the stomodaeum, in which case the mesentery is said to be complete

Nematocysts Stinging or adhesive structures housed in specialized cells of the ectoderm, characteristic of cnidarians

Operculum Skeletal plate or plates covering the calice of a corallite in some corals

Periderm Outer stiff organic (scleroprotein) skeleton of some polypoid hydrozoans

Phaceloid Structural arrangement of corallites in a branching (fasciculate) coral in which each branch is a single corallite and the branches are relatively closely spaced and subparallel

Planula Larval stage of a cnidarian

Plocoid Structural arrangement of corallites in colonial scleractinian corals such that corallites are separated by septal costae, dissepiments, or other common colonial tissue

Polyp The fixed benthonic stage of the cnidarian body plan in which the mouth faces upwards surrounded by tentacles. In the corals and some other cnidarians the polyp secretes a calcium carbonate skeleton at its base

Ramose Branching growth form in which each branch consists of many corallites. The structural arrangement of the corallites may be cerioid, amural, coenenchymal, plocoid, or meandroid

Scyphistoma Benthonic polypoid stage in scyphozoans specialized for budding off medusae

Septum In corals, a radial skeletal element, which may be a solid or perforate vertical plate or be reduced to a series of spines. In Scyphozoa, one of four radial fleshy partitions of the gut

Stomodaeum Cylindrical collar extending down from the mouth into the gut of anthozoan polyps

Tabula A horizontally disposed skeletal plate, which may be flat, dished, arched, domed, or vesicular, stacks of which occupy the axial area of most rugosan corallites (tabularium) or the whole width of the corallite in some rugosans and most other Palaeozoic corals

Zooxanthellate Scleractinian corals with symbiotic algae (zooxanthellae) present in the endoderm

See Also

Fossil Invertebrates: Porifera. **Lagerstätten**. **Palaeoclimates**. **Palaeoecology**. **Precambrian:** Vendian and Ediacaran. **Sedimentary Environments:** Carbonate Shorelines and Shelves; Reefs ('Build-Ups').

Further Reading

Babcock LE and Feldmann RM (1986) The phylum Con ulariida. In: Hoffman A and Nitecki MH (eds.) *Problem atic Fossil Taxa*, pp. 135 147. New York: Oxford University Press.

Coates AG and Oliver WA Jr (1973) Coloniality in zoantharian corals. In: Boardman RS, Cheetham AH, and Oliver WA Jr (eds.) *Animal Colonies, Development and Function Through Time,* pp. 3–27. Stroudsburg: Dowden, Hutchinson and Ross.

Hill D (1981) Rugosa and Tabulata. In: Teichert C (ed.) *Treatise on Invertebrate Paleontology, Part F. Coelenterata (Supplement 1).* pp. xl + 762. Boulder, Colorado and Lawrence, Kansas: Geological Society of America and University of Kansas Press.

Jenkins RJF (1992) Functional and ecological aspects of Ediacaran assemblages. In: Lipps JH and Signor PW (eds.) *Origin and Early Evolution of the Metazoa,* pp. 131–176. New York: Plenum Press.

Moore RC (ed.) (1956) *Treatise on Invertebrate Paleontology, Part F. Coelenterata*. Boulder, Colorado and Laurence, Kansas: Geological Society of America and University of Kansas Press.

Nudds JR and Sepkoski JJ Jr (1993) Coelenterata. In: Benton MJ (ed.) *The Fossil Record,* 2, pp. 101–124. London: Chapman & Hall.

Oliver WA Jr and Coates AG (1987) Phylum Cnidaria. In: Boardman RS, Cheetham AH, and Rowell AJ (eds.) *Fossil Invertebrates,* pp. 140–193. Oxford: Blackwell.

Scrutton CT (1979) Early fossil cnidarians. In: House MR (ed.) *The Origin of Major Invertebrate Groups,* pp. 161–207. London: Academic Press.

Scrutton CT (1997) The Palaeozoic corals, I: origins and relationships. *Proceedings of the Yorkshire Geological Society* 51: 177–208.

Scrutton CT (1998) The Palaeozoic corals, II: structure, variation and palaeoecology. *Proceedings of the Yorkshire Geological Society* 52: 1–57.

Scrutton CT (1999) Palaeozoic corals: their evolution and palaeoecology. *Geology Today* 15: 184–193.

Scrutton CT and Rosen BR (1985) Cnidaria. In: Murray JW (ed.) *Atlas of Invertebrate Macrofossils,* pp. 11–46. Harlow: Longman.

Seilacher A (1989) Vendozoa: organismic constructions in the Proterozoic biosphere. *Lethaia* 22: 229–239.

Stanley GD Jr (1986) Chondrophorine hydrozoans as problematic fossils. In: Hoffman A and Nitecki MH (eds.) *Problematic Fossil Taxa,* pp. 68–86. New York: Oxford University Press.

Stanley GD Jr (ed.) (1996) *Paleobiology and Biology of Corals*. Lawrence, Kansas: The Paleontological Society.

van Iten H (1992) Microstructure and growth of the conulariid test: implications for conulariid affinities. *Palaeontology* 35: 359–372.

Veron JEN (1995) *Corals in Space and Time: Biogeography and Evolution of the Scleractinia*. Ithaca: Cornell University Press.

Echinoderms (Other Than Echinoids)

A B Smith, The Natural History Museum, London, UK

Introduction

Echinoderms are an important group of marine invertebrates with multiplated calcite skeletons and a long and rich fossil record. Echinoderms first appeared in the Early Cambrian and today are represented by approximately 6000 species distributed throughout all latitudes and from the intertidal zone to the deep ocean trenches. They have one limitation, however: they can tolerate only slightly reduced salinities and are absent from brackish and estuarine habitats. Sessile forms such as crinoids and blastoids are filtration feeders, using their network of arms or tentacles to capture organic particles from the water. Vagile forms are more diverse in their feeding strategies, with scavengers, algivores, carnivores, deposit feeders, and filter feeders all represented. Almost all are benthonic, living on the seafloor or buried just beneath it, although there are rare pseudoplanktonic and true nektonic species.

Key Attributes of Echinoderms

Crown-group echinoderms share several characteristics that mark them apart from other invertebrates and show them to be a monophyletic group.

- They possess multiplated skeletons whose elements are composed of high-magnesium calcite and under magnification have a highly distinctive mesh-like structure termed stereom (Figure 1). Because of this, even isolated plates in petrographic thin section are instantly recognizable as being derived from echinoderms. Stereom can have a variety of three-dimensional arrangements, largely depending on the nature of the investing soft tissue.
- Their body plan is basically pentaradiate. This is clearly expressed in all living groups, although there are exceptions, especially amongst some of the more primitive Early Palaeozoic blastozoan groups.
- All crown-group echinoderms possess a water vascular system built to a common plan. The water vascular system is a hydraulically operated system of tube feet and linking canals that is derived from the larval left mesocoel. It is composed of a central

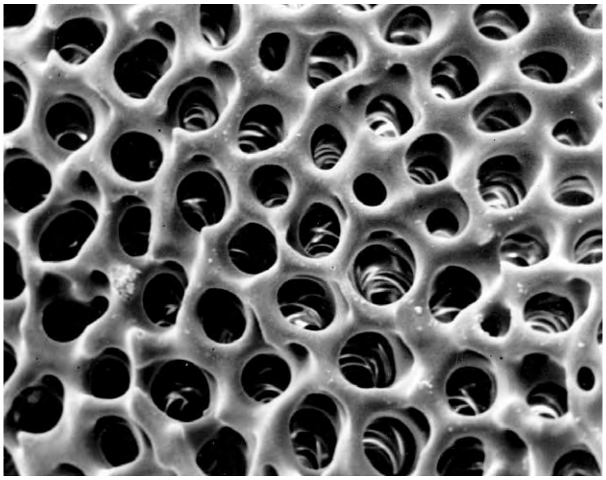


Figure 1 Scanning electron micrograph (magnification ×250) of an echinoid plate showing the three dimensional mesh (stereom) of which the echinoderm skeleton is composed.

ring canal, which surrounds the start of the digestive system just internal to the mouth. Five primary branches bud off from this ring forming long radial water vessels, which end blindly and which give rise on either side to a series of external tentacles termed tube feet. An additional short branch arising from the circumoral ring leads to the exterior via the hydropore and bears no tentacles. The skeletal elements associated with the radial canals and their tube feet comprise the ambulacral series. Ambulacral-plate series are thus always arranged radially around the mouth.

Of these characteristics, only possession of stereom is ubiquitous to all stem-group echinoderms. Pentameral symmetry is confined to crown-group echinoderms, stem-group members either being completely asymmetrical or showing bilateral or triradiate symmetry. Similarly, only some stem-group echinoderms show evidence of an ambulacral system and ambulacral plating.

Phylogenetic Relationships

Echinoderms are triploblastic Metazoa, which, despite their unique pentaradiate adult body plan, belong to the Bilateria. More specifically, echinoderms are members of the group Deuterostoma, i.e. bilaterians in which the embryonic blastopore develops into the anus rather than the mouth. This aligns echinoderms with chordates and hemichordates, the latter comprising acorn worms, pterobranchs, and the extinct graptolites (*see* **Fossil Invertebrates:** Graptolites). Recent molecular and morphological analyses unambiguously indicate that echinoderms and hemichordates are the more closely related, and together these form the taxon Ambulararia. Both hemichordates and echinoderms have a vascular system of tubes (the water vascular system of echinoderms and the tentacular system of pterobranchs) that originates from the homologous primary body coeloms. In hemichordates the tentacular system is symmetrically paired and forms from both the left and the right mesocoel. In echinoderms only the left mesocoel forms the water vascular system, and hence there is only a single hydropore. As both hemichordates and chordates are bilaterally symmetrical as adults and bear gill slits (pharyngeal openings), the latest common ancestor of echinoderms and hemichordates must also have been bilaterally symmetrical and have possessed gill slits.

Echinodermata are divided into five extant classes, which together form the crown group of the phylum. These are the Crinoidea (feather stars and sea lilies) (*see* **Fossil Invertebrates:** Crinoids), Asteroidea (starfishes), Ophiuroidea (brittle stars), Echinoidea (sea urchins) (*see* **Fossil Invertebrates:** Echinoids) and Holothuroidea (sea cucumbers). Major differences separate the body plan of crinoids from those characterizing the other four classes, and molecular and comparative data all point to crinoids as the most primitive of the extant echinoderm classes. The crinoids and their extinct stalked relatives the blastozoans together form the group Pelmatozoa. Asteroidea, Ophiuroidea, and the Echinozoa form a sister clade, the Eleutherozoa, and are free-living. Molecular evidence suggests that amongst eleutherozoans the Echinoidea and Holothuroidea are the most closely related, and together they form the group Echinozoa. Echinozoa have a globular to cylindrical body form with the aboral surface highly reduced. By contrast, asteroids and ophiuroids have a stellate body plan and their oral and aboral surfaces are equally developed. The extinct edrioasteroids are an important early eleutherozoan group.

There are some numerically small but interesting stem-group echinoderms; most important amongst these are the carpoids (see below). Carpoids all lack pentaradiate symmetry, and most lack an ambulacral system while retaining pharyngeal gill slits.

Geological History

The geological history of echinoderms is summarized in Figure 2. The earliest records we have are of isolated plates with the distinctive stereom structure from the Botomian (Early Lower Cambrian) of America. By the end of the Lower Cambrian we have evidence of carpoids, helicoplacoids, blastozoans, and edrioasteroids, suggesting that the crown-group split between pelmatozoans and eleutherozoans had

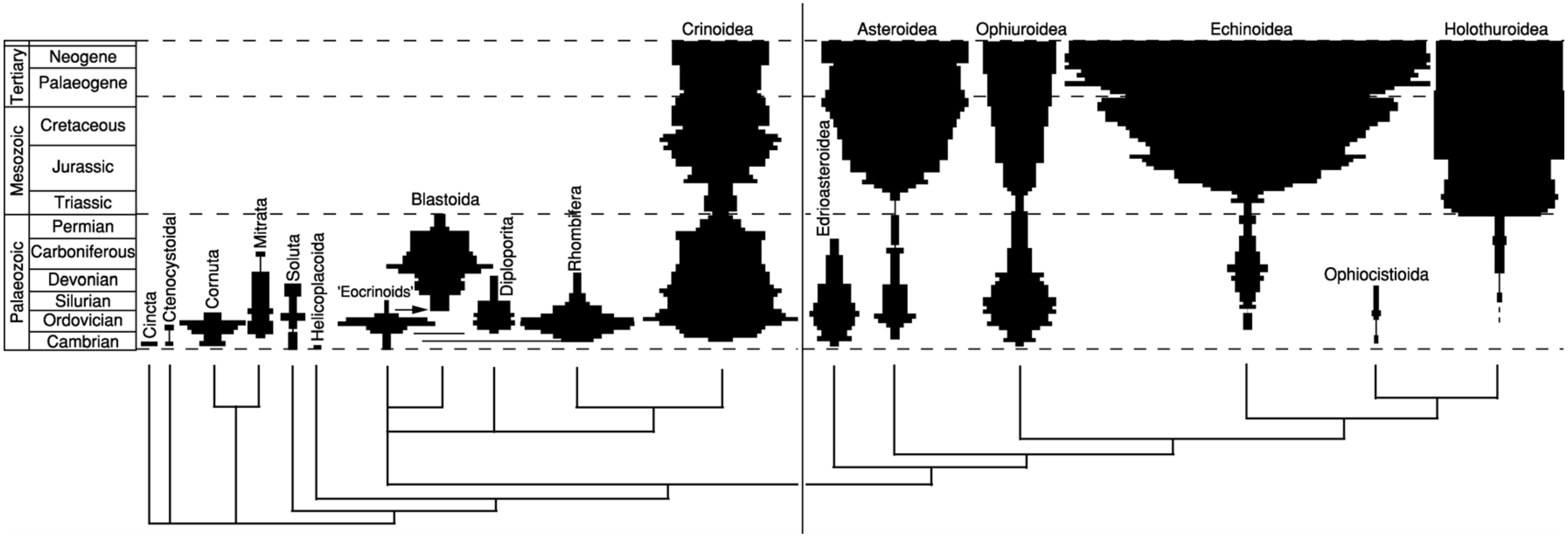


Figure 2 Stratigraphic ranges of the major groups of echinoderm. The widths of the bars are proportional to the taxic (family level) diversities. The cladogram shows their inferred phylogenetic relationships.

already taken place. However, it is not until the Early Ordovician that definite members of the five extant classes are found. The earliest crinoids, asteroids, and ophiuroids all appear close to the base of the Ordovician, while echinoids appear a little later, in the Middle Ordovician. Holothurians are not definitely known until the Late Silurian. During the Palaeozoic stemmed filtration feeders dominated benthic communities both numerically and taxonomically. Although isolated echinoderm ossicles can be present in such abundance as to be the major bioclastic constituent of some limestones, articulated echinoderm skeletons are for the most part confined to Lagerstätte deposits (*see* **Lagerstätten**) and are relatively rare.

All echinoderm groups went into decline towards the end of the Palaeozoic, and several groups became extinct. Edrioasteroids last occur in the Upper Carboniferous and blastozoans in the Permian. The groups that did survive into the Mesozoic did so only in small numbers, and indeed the modern fauna of echinoids, asteroids, crinoids, and ophiuroids were all initiated at around this time. The evolution in echinoids of a solid and robust test in the Triassic greatly enhanced their preservation potential. The history of echinoderms since the Triassic has been one of unbridled success, and the five classes have continued to diversify and expand their ecological ranges through the Mesozoic and Tertiary. Echinoderms are probably as diverse now as they have ever been.

Major Taxonomic Groups

Carpoids

Carpoids are the most primitive of the stem-group echinoderms. There are four major groups: Soluta, Cincta, Ctenocystoida, and Stylophora; Stylophora is further subdivided into the cornutes and the mitrates.The Stylophora (Figures 3C and 3D) are the most primitive and conform most closely in body plan to a basal deuterostome. Stylophora have a large anterior body with pharyngeal openings and a single posterior appendage (stele), which is muscular and bilaterally symmetrical. Cornutes (Figure 3D) are strongly asymmetric in outline, often rather boot-shaped, and have a well-developed marginal frame and a serially repeated set of external atrial openings. Mitrates (Figure 3C) on the other hand are more nearly bilaterally symmetrical with no differentiated marginal frame. Paired internal gill bars are present in at least some species. Both have a calcitic skeleton composed of stereom, and neither show evidence of possessing a water vascular system. The presence of gill openings and a motile muscular stele or tail is primitive for deuterostomes as a whole. Stylophora occur from the Middle Cambrian through to the Carboniferous and were recumbent suspension feeders that are thought to resemble tunicates in their feeding strategy (i.e. drawing water into the pharynx where it could be filtered through gill filaments).

Solutes are the most echinoderm-like of the carpoid groups. They have a sac-like body and two appendages asymmetrically arranged at opposite poles (Figure 3A). One of these appendages is an ambulacrum in the form of a single arm, so clearly this group had a tentacular water vascular system. Furthermore, the presence of a single rather than paired hydropore indicates that it was built on the echinoderm plan. The other appendage is a stalk, which is muscular near the theca and more rigid distally with a basal attachment pad. Solutes appear in the Lower Cambrian and survived until the Early Devonian. They were suspension feeders, using their ambulacrum to capture particles.

Cinctans have an ovate body with a well-developed marginal frame and a single rather rigid bilaterally symmetric appendage (Figure 3E). There is a small mouth opening through the marginal frame and to either side there is a groove that is roofed by a flexible sheet of cover plates. The right-hand groove is always the less well developed and may be lacking in some species. It seems likely therefore that cinctans possessed a pair of hydrocoel tentacles, with the left being better developed than the right. Whether these are both from the left hydrocoel or represent a hemichordate-like paired system is unknown, since hydropores have not as yet been definitely identified. At the anterior there is a large opening covered by an opercular plate, which acted as an outlet valve. This is best interpreted as an atrial opening and suggests that cinctans, like stylophorans, were active suspension feeders with some form of pharyngeal filtration basket. Like stylophorans they lived recumbent on the seafloor. Cinctans are restricted to the Cambrian.

Finally, the ctenocystoids are a small group of carpoids without a stem or tail, external gill slits, or ambulacra (Figure 3B). Their precise phylogenetic position remains uncertain, but they were free-living and possibly pharyngeal basket feeders like the Stylophora. They are found from the Middle Cambrian to the Upper Ordovician.

Helicoplacoids

The stratigraphically oldest fossil echinoderm assemblages known are dominated by helicoplacoids. Helicoplacoids are cigar-shaped echinoderms with a basal attachment at one end and a terminal anus at the other (Figure 3F). They have a lateral mouth, about two-thirds of the way up from which three ambulacra

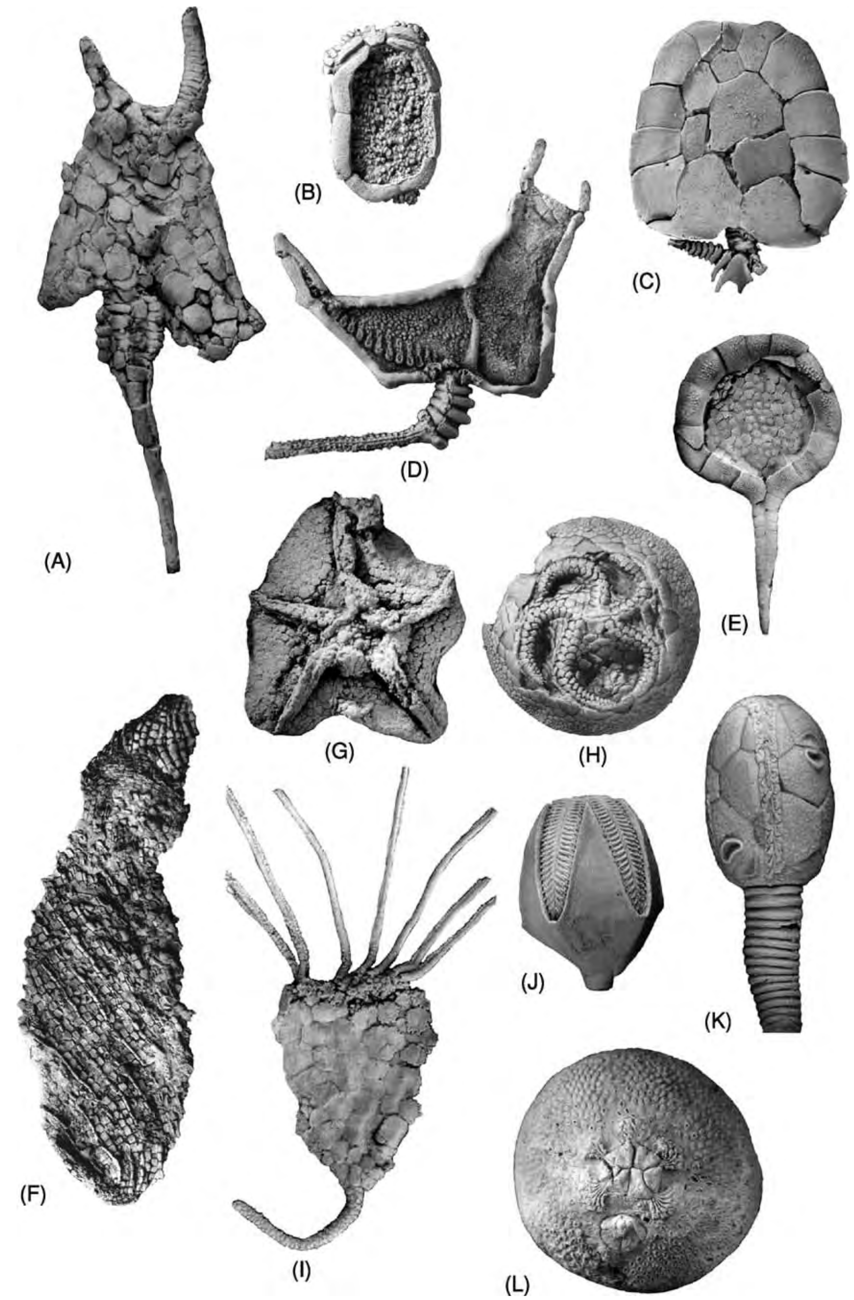


Figure 3 Representative fossil echinoderms. (A) The solute *Dendrocystoides* from the Upper Ordovician (magnification ×0.5). (B) The ctenocystoid *Ctenocystoides* from the Middle Cambrian (magnification ×3). (C) The mitrate stylophoran *Mitrocystites* from the Middle Ordovician (magnification ×1.4). (D) The cornute stylophoran *Cothurnocystis* from the Upper Ordovician (magnification ×1.4). (E) The cinctan *Trochocystites* from the Middle Cambrian (magnification ×1.4). (F) The helicoplacoid *Helicoplacus* from the Lower Cambrian (magnification ×1.4). (G) The stromatocystitid edrioasteroid *Stromatocystites* from the Middle Cambrian (magnification ×1). (H) The isorophid edrioasteroid *Carneyella* from the Middle Ordovician (magnification ×1.7). (I) The eocrinoid *Gogia* from the Middle Cambrian (magnification ×1.4). (J) The blastoid *Pentremites* from the Carboniferous (magnification ×1.4). (K) The rhombiferan *Apiocystis* from the Middle Silurian (magnification ×1.4). (L) The diploporite *Haplosphaeronis* from the Upper Ordovician (magnification ×1.4).

radiate, two spiralling upwards and one spiralling downwards. Each ambulacral series is composed of a double column of flooring plates and two flaps of cover plates. Pores indicate the positions of tube feet. The remainder of the body is covered in spirally arranged columns of rectangular elements. There is some evidence that helicoplacoids could extend and contract their body by changing the angle of spiralling. Helicoplacoids are confined to the Lower Cambrian and were suspension feeders living on algal-bound sediments. They used their tube feet to capture particles in the water column.

Blastozoans

Blastozoans are a large and morphologically diverse group of primitive stemmed echinoderms that mostly resemble crinoids in appearance and lifestyle. The plated body or theca bears a mouth, anus, and usually some form of specialized respiratory structure. The stem takes the form of a polyplated holdfast, a more regular stem comprising five series of vertically aligned ossicles, or a holomeric stem composed of disc-shaped columnals. A few blastozoans are secondarily stemless and either cemented directly to the seafloor or were free-lying. The mouth faces upwards, away from the seafloor. All blastozoans have some form of filtration fan composed of unbranched brachioles, which presumably bore tube feet and a water vascular system. However, in contrast to crinoid arms, which are formed as direct outgrowths of the body wall and carry extensions of the primary body coelom, blastozoan brachioles are side branches of the ambulacra and are not directly connected to the body coelom. Although many blastozoans show an obvious pentaradiality, diverse body forms, symmetries, and ambulacral architectures are to be found in this group. As currently constituted they are probably a polyphyletic group.

Eocrinoids are a paraphyletic assemblage of basal pelmatozoans that have an irregularly plated theca (Figure 3I). Simple sutural gaps termed epispires that notch the thecal plates are the only respiratory structures ever developed. Eocrinoids include the ancestors of all other blastozoan groups and probably of the crinoids also. Diploporites (Figure 3L) have larger and more regular plating and lack any form of ambulacral flooring plates; the brachioles arise directly from the thecal plates. Thecal plates are pierced by numerous pairs of pores (diplopores) that have a respiratory role. Rhombiferans (Figure 3K) have stout arms arising directly from around the mouth. Respiratory structures are always present and take the form of thin folded rhomboidal structures or tubes that run close to the inner or outer surfaces of the plates. Blastoids (Figure 3J) have the most regular thecal plating of any blastozoan, always constructed of three basals, five radials and five lancet plates. Pentameral symmetry is extremely well expressed. Ambulacra are well developed and form an integral part of the theca, giving rise to a dense fan of brachioles. Small openings along the margins of the ambulacra lead into an internal hydrospire system – a highly convoluted and thin-walled respiratory organ. The oldest blastozoans are eocrinoids of Lower Cambrian age. Diploporites and rhombiferans appear at the start of the Ordovician, while blastoids do not appear until the Silurian. The group finally went extinct at the end of the Permian.

Edrioasteroids

Edrioasteroids are an extinct group of sessile stem-group eleutherozoans that were discoidal, clavate, or subglobular in form (Figure 3G and 3H). They have a central mouth from which five ambulacra radiate. The ambulacra can be straight, but are commonly spiralled to provide a larger surface area, and extend to a marginal ring of plates marking the edge of the oral area. A single hydropore and gonopore open close to the mouth. Edrioasteroids lived with the mouth facing upwards and used their five ambulacra to capture organic material suspended in the water.

The earliest edrioasteroids (stromatocystitids: Figure 3G) were fully plated and appear not to have been attached to the substrate. These may represent basal eleutherozoans, ancestral to all later forms. The great majority of edrioasteroids, however, fall into a derived clade, the Isorophida (Figure 3H), that specialized as hard-ground colonizers. Isorophids lived attached to hard substrates, such as the shells of brachiopods, by their lower uncalcified surface. They were not permanently fixed but were able to realign themselves during life. Some isorophids developed the outer part of the disc as an extendible stalk, which allowed them to lower or raise the oral zone with the ambulacra, presumably in response to water flow conditions.

Stromatocystitids first appear in the Lower Cambrian, while isorophids appear a little later, in the Upper Cambrian, and survived through to the Upper Carboniferous.

Asteroids

Asteroids are stellate echinoderms with arms that merge without a break into a central body (Figure 4A). The mouth is central and faces down towards the substrate, and an ambulacral groove runs from the mouth along the lower surface of each arm. In asteroids the arms are hollow and extensions of the digestive system and gonads extend along each. Asteroids lack a jaw apparatus but are active predators, ingesting small prey whole. Some burrow after infaunal animals, while others are epifaunal hunters. One group has evolved an ingenious method of preying on bivalves. They clasp the bivalve using their tube feet and prise the two valves apart. They need to create only a small slit-like opening because they then evert their stomach into the bivalve and digest the animal. The asteroid skeleton is rarely robust and consists of a series of small elements embedded in a collagenous membrane. Consequently, they disarticulate rapidly after death and have left a rather poor fossil record. Asteroids first appear in the Early Ordovician and are never common or abundant.

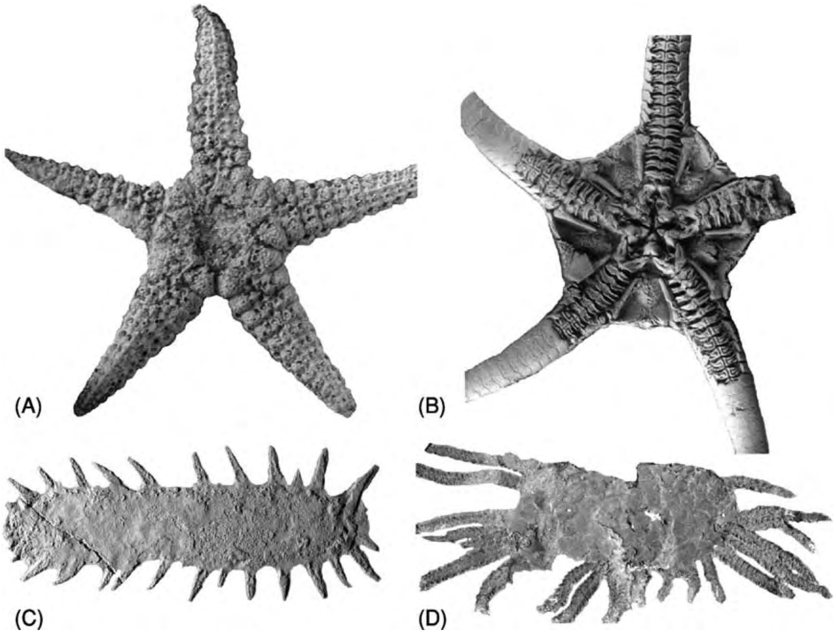


Figure 4 Representative fossil echinoderms. (A) The asteroid *Siluraster* from the Upper Ordovician (magnification ×2). (B) The ophiuroid *Palaeocoma* from the Lower Jurassic (the upper surface of the disc and the proximal parts of the arms have been lost, revealing the solid vertebrae filling the core of the arms and the central jaw) (magnification ×1). (C) The holothurian *Oneirophantites* from the Middle Triassic (magnification ×0.3). (D) The ophiocistiod *Sollasina* from the Upper Silurian (magnification ×0.7).

Ophiuroids

Ophiuroids resemble asteroids in having a stellate body plan, with five or more arms radiating from a small circular disc (Figure 4B). Unlike asteroids, however, their arms are solid, being supported by a series of internal disc-like ossicles termed vertebrae. Consequently, there are no extensions of the gonads or digestive system into the arms. The mouth is central and faces downwards. The digestive system is a simple blind sac (there is no anus), and the lower surfaces of the arms carry the water vascular system and the tube feet. All ophiuroids are carnivorous and possess a formidable jaw apparatus with strong musculature and batteries of tooth-like spines.

There are two major post-Palaeozoic groups, the euryalids and the Ophiuroida. In euryalids the arms are typically branched many times to form a filtration fan and bear short spike-like spines. Euryalids are suspension feeders that use their network of arms to ensnare small nektonic prey such as arrow worms, which are then passed to the mouth. The great majority of Ophiuroida are active predators and are able to move rapidly over the seafloor by using their highly motile arms. A few are also able to suspension feed by extending one or more arms into the water column and using their tube feet to secrete ribbons of sticky mucous to ensnare small prey.

The ophiuroid skeleton is rather fragile and readily disintegrates upon death. Consequently, like asteroids they have left a relatively sparse fossil record, starting in the Lower Ordovician.

Ophiocistioids

Ophiocistioids are a small but interesting extinct group of globular eleutherozoans with large plated tube feet (Figure 4D) that are related to both echinoids and holothurians. The mouth in ophiocistioids is central and downward facing, and there is a complex jaw apparatus identical in all important respects to the Aristotle's lantern of echinoids. They were therefore presumably active predators, like early echinoids, using their jaws to capture small benthic prey. The large plated tube feet were locomotory in function, and their water vascular system lay beneath the ambulacral plates, as in echinoids and holothurians, rather than externally, as in asteroids and crinoids. In early members the body is covered in a series of small plates. In at least one taxon, however, the body wall is reduced to microscopic spicules, which are wheel-shaped and identical to those seen in apodid holothurians. Ophiocistioids first appear in the Middle Ordovician and continue to the Early Carboniferous. They are always rare.

Holothurians

Holothurians have a cylindrical body plan with the mouth at one pole and the anus at the other (Figure 4C). The mouth is surrounded by a ring of

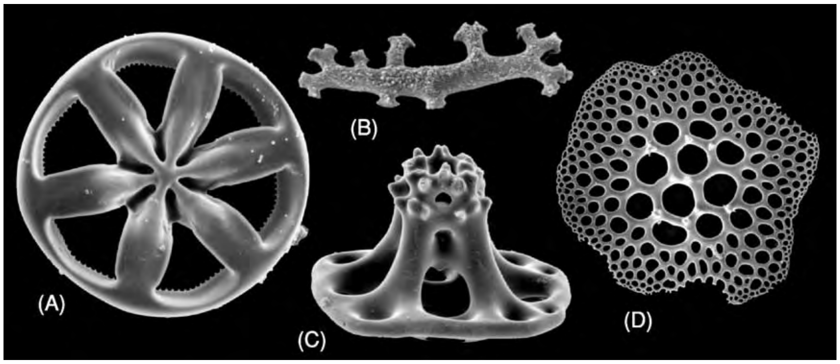


Figure 5 Characteristic spicules from the body walls of holothurians. (A) Wheel (magnification ×600). (B) Rod (magnification ×700). (C) Button (magnification ×800). (D) Plate (magnification ×50).

tentacles, which are often highly branched and are used in feeding. There is no jaw apparatus. Some holothurians have long highly branched tentacles that are extended into the water column and used for suspension feeding. Others have short finger-like tentacles that are used in deposit feeding. Today holothurians are the major deposit feeders in back-reef habitats. They live either on the surface or in shallow burrows.

Holothurians are the most diverse of the five extant classes of echinoderm, with over 2000 extant species, but they have the poorest of fossil records. This is because their skeleton is reduced to microscopic spicules. The only elements of any size are 10 ossicles that surround the mouth and provide an anchorage for the oral tentacles; these form the circumoral ring. Each holothurian has many thousands of microscopic spicules in its body wall. Many are distinctive, such as the anchor-like elements of apodians and the wheel-like elements of molpadiids (Figure 5). A few body fossils are known from Lagerstätte (Figure 4C) but almost all of our information on the history of holothurians comes from the study of their isolated spicules distributed in the sedimentary record. The first holothurian body fossil comes from the Late Silurian, but spicules probably attributable to holothurians are known from the Ordovician onwards.

See Also

Fossil Invertebrates: Crinoids; Echinoids; Graptolites. **Lagerstätten**. **Palaeoecology**.

Further Reading

Beaver HH, Caster KE, Durham JW, *et al.* (1978) *Treatise on Invertebrate Paleontology. Part S, Echinodermata 1.* Boulder, Colorado and Lawrence, Kansas: The Geological Society of America and the University of Kansas Press.

Bell BM (1976) A study of North American Edrioasteroidea. *New York State Museum, Memoirs* 21: 1 447.

Gilliland P (1993) The skeletal morphology, systematics and evolutionary history of holothurians. *Special Papers in Palaeontology* 47: 1 147.

Guensburg TE and Sprinkle J (1994) Revised phylogeny and functional interpretation of the Edrioasteroidea based on new taxa from the Early and Middle Ordovician of Western Utah. *Fieldiana Geology* 29: 1 41.

Hess H (1975) Die fossilen Echinodermen des Schweizer Juras. *Veroffentlichungen aus dem Naturhistorischen Museum Basel* 8: 1 130, pls 1 48.

Jagt JWA (2000) Late Cretaceous Early Palaeogene echinoderms and the K/T boundary in the southeast Netherlands and northeast Belgium. Part 3, Ophiuroids. *Scripta Geologica* 121: 1 179.

Jagt JWA (2000) Late Cretaceous Early Palaeogene echinoderms and the K/T boundary in the southeast Netherlands and northeast Belgium. Part 5, Asteroids. *Scripta Geologica* 121: 377 503.

Littlewood DTJ, Smith AB, Clough KA, and Ensom RH (1997) The interrelationships of the echinoderm classes: morphological and molecular evidence. *Biological Journal of the Linnean Society* 61: 409 438.

Paul CRC (1973 1997) British Ordovician cystoids. *Monograph of the Palaeontographical Society,* 1 213, pls 1 33.

Smith AB, Peterson K, Littlewood DTJ, and Wray GA (2004) From bilateral symmetry to pentaradiality: the phylogeny of hemichordates and Echinodermata. In: Cracraft J and Donoghue M (eds.) *Assembling the Tree of Life.* Oxford: Oxford University Press.

Sprinkle J (1982) Echinoderm faunas from the Bromide Formation (Middle Ordovician) of Oklahoma. *The University of Kansas Paleontological Contributions Monograph* 1: 1 369.

Waters JA and Maples CG (1997) Geobiology of echinoderms. *The Paleontological Society Papers* 3: 1 355.

Crinoids

M J Simms, Ulster Museum, Belfast, UK

Introduction

Crinoids are one of five extant classes of echinoderm (the others being echinoids, asteroids, ophiuroids and holothuroids (*see* **Fossil Invertebrates:** Echinoderms (Other Than Echinoids))) with a rich fossil record extending back nearly 500 Ma to the Early Ordovician. They are exclusively marine suspension feeders and, with a few exceptions, sessile benthos. Crinoids reached a peak of taxonomic diversity in the Late Palaeozoic but experienced a catastrophic decline around the Permo-Triassic boundary, from which they subsequently rediversified. More than 700 extant species and over 15 000 fossil species have been described.

Like other echinoderms, crinoids have a multielement endoskeleton of high-magnesium calcite ossicles connected by soft tissue and enclosed by a thin veneer of living tissue. Each ossicle is a single optically continuous calcite crystal with a stereom structure – a labyrinthine network of cavities permeated with soft tissue called stroma.

In the basic crinoid design the endoskeleton can be divided into three morphological sections. The stem or column is a usually slender flexible elongate structure for attachment to the substrate and elevation above it. The stem is surmounted by the cup, which is a fairly rigid structure, usually constructed of two or three circlets of plates, containing organs of digestion, movement, and nervous control. Arising from the upper circlet of the cup are elongate flexible arms involved in food gathering. Typical crinoids have a superficial similarity to flowers, giving rise to their popular name of ‘sea lilies’ and indeed to the name ‘crinoid’ itself (from the Greek *krinos* = a lily). Stalked crinoids are very much in a minority among extant taxa, with fewer than 100 species. In the remaining more than 600 species a stem is present only at the larval stage; the stemless adults are often termed ‘feather stars.’

The soft parts of crinoids comprise a small and inconspicuous proportion of the total mass. They are virtually unknown in fossil crinoids, although their presence and functions can be inferred from the structure of the preserved hard parts and by analogy with extant crinoids. The digestive system is contained entirely within the cup, with both mouth and anus opening onto the upper surface. In common with all echinoderms, and unique to them, crinoids possess a network of fluid-filled tubes called the water vascular system. Branching canals extend radially into the arms from a central ring canal housed in the cup and terminate in the tube feet – small tentacle-like structures that detect and capture food particles before passing them to ciliated food grooves, which run down the arms to converge on the mouth. A further unique echinoderm character, almost certainly present in fossil taxa, is catch connective tissue, which is a type of ligament that can change its properties from pliable to rigid, enabling parts of the skeleton, such as the stem or arms, to ‘lock’ into position for prolonged periods with minimal energy expenditure. Catch connective tissue is noncontractile and cannot contribute to active movement. Only muscles are able to do this; they are restricted to the arms of only some taxa and are entirely absent from the stem.

Morphology and Functional Interpretations

A crinoid’s morphology is determined by the interaction of three distinct factors: inherited characteristics, i.e. the morphology of its immediate ancestor; architecture, i.e. physical properties of the materials; and evolutionary selection pressure from the external environment, i.e. ecological factors.

The skeleton of most crinoids is readily divisible into three distinct sections – the stem, cup, and arms (**Figure 1**) – a basic morphology that was inherited from a pre-crinoid ancestor and subsequently modified within limits imposed by ancestry, architecture, and ecology. In a few specialist taxa the stem or arms may be greatly reduced or absent.

The stem’s main functions are attachment to the substrate and, especially, elevation above it. Most fossil crinoids have a stem ranging from a few centimetres to perhaps a metre or so in length, similar to that of extant stalked crinoids, although in some taxa it is very much longer while in others it may be reduced or absent (particularly in free-living crinoids such as the comatulids). Typically, the stem comprises a stacked series of ossicles, or columnals, pierced by a central canal containing extensions of the coelom and nervous system. Each columnal is typically a single ossicle, but a few Early Palaeozoic taxa have what are known as meric columnals constructed from several ossicles or meres. Each columnal is connected to adjacent columnals by catch connective tissue. Short

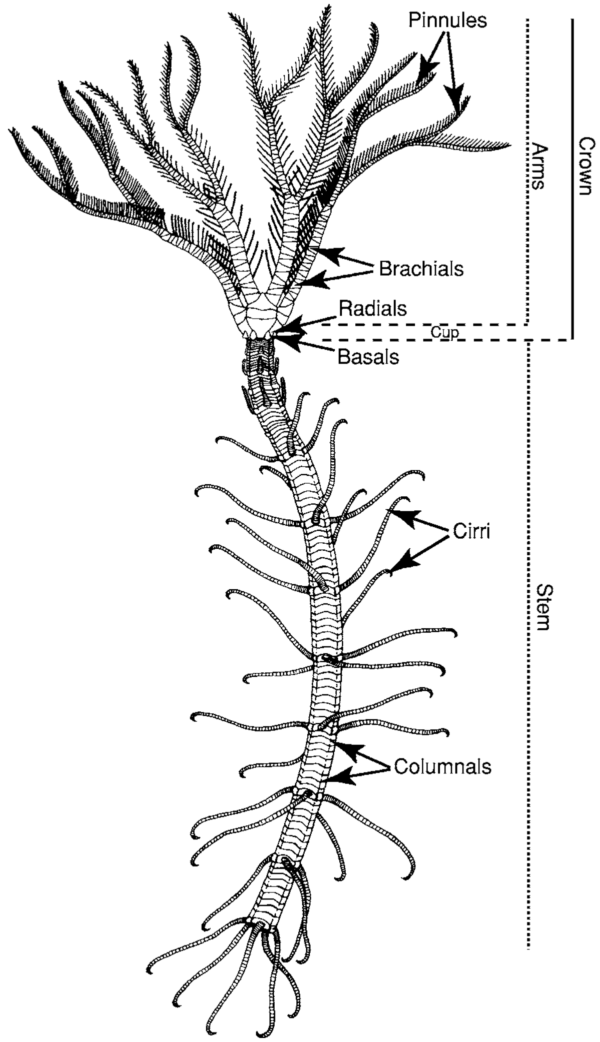


Figure 1 The main morphological features of an extant stalked crinoid (isocrinid). Some of the arms have been omitted for clarity.

ligaments connect opposing faces of adjacent columnals, with longer through-going ligaments linking groups of columnals together. Autotomy – the voluntary shedding of distal parts of the stem to allow the animal to drift to a new site – can occur at a junction between adjacent linked groups of columnals through a change in the properties of the catch connective tissue. Hence, discarded columnals can contribute to the future fossil record long before the animal itself has actually died! Articulating surfaces of adjacent columnals show a range of morphologies, most typically a radial pattern of crenulations (Figure 2A), which interlock with opposing crenulations on the adjacent columnal to resist shearing and torsional stresses. These crenulations may be arranged in more complex, often pentaradiate, configurations (Figure 2B). Another distinctive type of columnal articulation developed independently in the Silurian to Permian

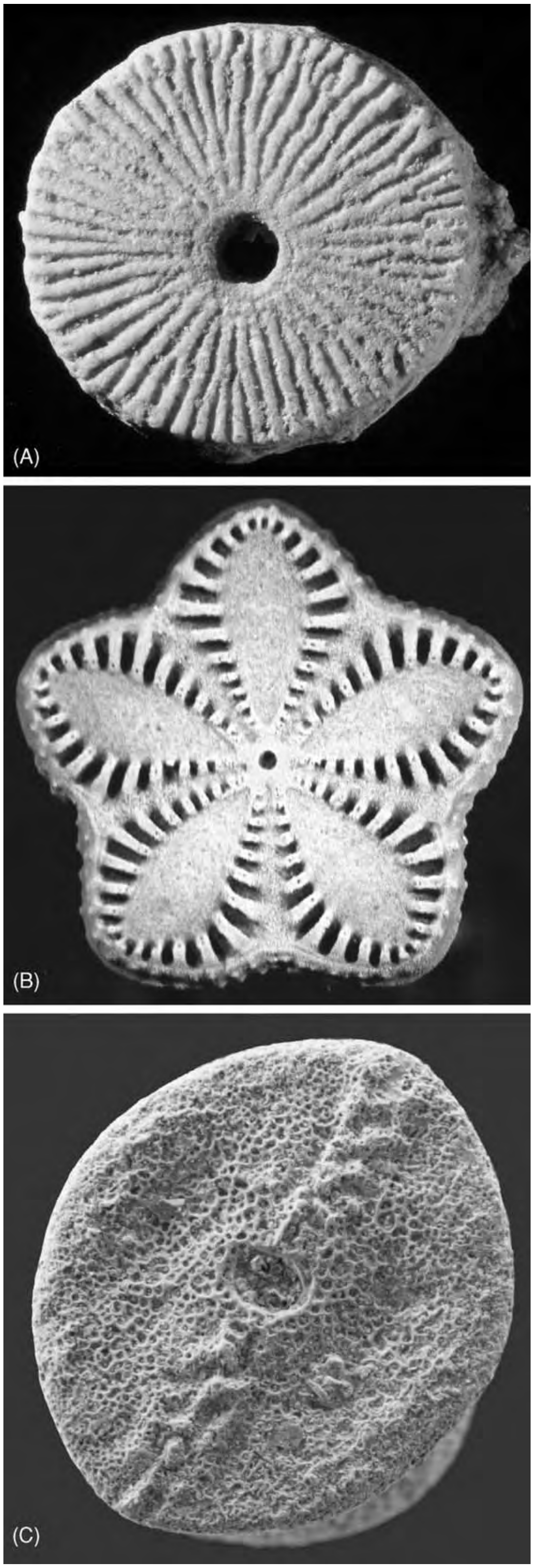


Figure 2 The main types of columnal articulations encountered in crinoids: (A) multiradiate is 8 mm across; (B) pentaradiate is 12 mm across; and (C) synarthrial. The columnal in (C) is from a larval isocrinid and is less than 0.7 mm across; the stereom is clearly visible.

platycrinitid camerates and the Cretaceous to Holocene bourgueticrinids and their allies. In these crinoids, the articulations (Figure 2C) are traversed by a prominent fulcral ridge flanked by two ligament pits, with the fulcral ridges on the proximal and distal faces of each columnal typically at an angle to each other.

The crinoid stem may be more or less permanently attached to a hard substrate by a cemented holdfast, may be rooted in sediment by irregular branches or outgrowths from the stem or may attach on a more temporary basis to hard or soft substrates using regular offshoots of the stem, known as cirri. Cirri are the most versatile form of attachment and were adopted independently by several major crinoid groups, including the dominant extant group – the stemless comatulids.

The cup or calyx, located between the stem and the arms, houses the vital organs of the digestive, nervous, and water vascular systems. Typically the cup consists of a regular series of interlocked circlets, each usually of five plates, often with one or two additional plates intercalated towards the top. In the earliest known crinoid (the Early Ordovician *Aethocrinus*) there were four circlets, but in virtually all other crinoids the plates of the lowest circlet have been lost and the cup comprises either two or three circlets (see below). Typically the plates of each circlet are offset by 36° relative to the plates of adjacent circlets, imparting rigidity even with only two circlets. The shape of the cup varies, from shallow and bowl shaped to almost globular, with the same basic designs having evolved repeatedly in different groups. The calyx incorporates all of the plates between the top of the stem and the base of the free arms, including the cup, the tegmen (which covers the oral surface), and any parts of the arms that are incorporated into the tegmen. In many crinoids the cup plates are connected only by ligaments and the tegmen is weakly constructed; hence, it disarticulates rapidly after death. However, in the Palaeozoic subclass Camerata the calyx forms a relatively rigid and often globose structure, which may remain intact for some time after death. Consequently camerate calyces are better represented as fossils than are those of the other crinoid subclasses (see below).

The food grooves of the arms converge on the mouth, which is located on (or beneath in the case of camerates) the surface of the tegmen along with the anus. The anal opening is commonly elevated on an anal tube or anal sac, forming a large and complex structure in some fossil taxa, presumably to avoid faeces entering the mouth.

The arms are the food-gathering parts of crinoids, although ultimately it is the tube feet, lining a groove on the oral side of the arms, that are directly involved in capturing food and moving it towards the mouth. The arms increase the support area for tube feet but are not essential, as evidenced by a few crinoid taxa in which they are reduced in number or even absent. Secondary functions of the arms include respiration and locomotion, but generally selection pressure is towards improvement of the food-gathering mechanism within the constraints imposed by other factors.

Flexibility of the arms is achieved in the same manner as flexibility of the stem, with brachial plates connected by ligaments and/or muscles. Muscle tissue is present in all echinoderm tube feet but appeared in brachial articulations in only one mid-Palaeozoic group, which ultimately evolved into the post-Palaeozoic subclass Articulata. As in the stem, specialised ligamentary articulations can be used to autotomize parts of the arms.

At their simplest, crinoids have five unbranched arms, each arising from one of the plates in the uppermost circlet of the cup. Increasing the effective arm length, by branching, increases the number of tube feet that can be supported and hence increases the filtration efficiency. The development of numerous small side branches, or pinnules, is one strategy that has evolved independently in several major groups. The arms themselves show a wide range of branching patterns, from simple dichotomous (isotomous) (Figure 1) through to strongly endotomous (Figure 3). Endotomous branching appears to be the most efficient pattern, in terms of expenditure of materials versus food-gathering capabilities, and an analogy has been drawn with the arrangement of roads on banana plantations. Intriguingly, only a few fossil crinoid taxa developed this pattern to a significant extent.

Phylogeny, Systematics, and Geological History

With their morphologically complex multielement skeletons, crinoids are ideal subjects for phylogenetic analysis. Each part of the skeleton – stem, cup, and arms - can provide a wealth of morphological data for use in descriptive and phylogenetic investigations. However, all too often primary descriptions of genera and species are lacking in detail, particularly for the stem, while interpretations at higher taxonomic levels have been hindered both by this primary deficiency of data and, until the mid-1990s, by a rigid adherence to traditional interpretations of skeletal homologies.

At high taxonomic levels the structure of the calyx is considered to be of major phylogenetic significance.

Figure 3 Crown, stem, and cirri of *Pentacrinites fossilis*, a pseudoplanktic crinoid from the Early Jurassic. The arms are strongly pinnulate and branch endotomously, with each of the main arms bifurcating twice and subsequent side branches arising only from the inner sides of the main branches. Crowded cirri can be seen arising from the stem in the lower part of the picture.

A few of the earliest crinoids, grouped together in the subclass Aethocrinea (named after the genus *Aethocrinus*), have four circlets of plates in the cup: lintels (at the base), infrabasals, basals, and radials (top) (Figure 4). This number is reduced in all other crinoids, although the exact circlet(s) missing may vary, with potentially any one or more of the four circlets being lost to produce a three-, two- or, in at least one species, one-circlet cup. In the initial dichotomy after the four-circlet ancestor, one group (the subclass Disparida) lost the basal plates to leave a cup composed of lintels, infrabasals, and radials. Disparid cups may display some bilateral symmetry, with the constituent plates of each circlet being of unequal, or disparate, size, giving the subclass its name. Later disparids lost some or all of the radials and, in a few taxa, even the infrabasals, reducing the cup to a single circlet of lintels. The other branch of the dichotomy lost the lintels to form a cup of infrabasals, basals, and radials. This branch comprises two distinct subclasses, the Cladida and the Camerata, each of which contributed significantly to Palaeozoic crinoid diversity alongside the disparids. In both cladids and disparids the free arms arise directly from the top of the cup, but in camerates the calyx incorporates fixed brachials and interradials in addition to the cup, so that the free arms begin above the top of the cup. Camerates also have the tegmen firmly fixed to the cup to form a rigid chamber, from which their name derives. Camerates experienced a further important dichotomy early in their history, giving rise to the three-circlet diplobathrids and the two-circlet monobathrids in which the infrabasal circlet was lost. Cladids, which derive their name from the often multiple branching of the arms, also gave rise to two subclasses – the Flexibilia and the Articulata. The morphological differences that separate articulates or flexibles from the cladids are less profound than those that separate the other four subclasses from each other. Flexibles have a three-circlet cup like that of cladids, although the infrabasal circlet is typically reduced to three plates; their name derives from the weakly sutured plates of the cup and the inrolled tips of the arms. The articulates are the only post-Palaeozoic subclass, although their origins can be traced back well into the Palaeozoic. With a few exceptions the articulate cup is reduced to two circlets – basals and radials – with the basals also greatly reduced in some taxa. The brachial articulations are predominantly muscular, a characteristic after which the subclass is named.

The precise group from which crinoids originated is unclear. A considerable diversity of noncrinoid pelmatozoan (stalked) echinoderms characterized the Early Palaeozoic, and crinoids evolved from one of these through a reduction in the complexity of the cup and the evolution of true arms carrying extensions of the coeloms and of the nervous and water

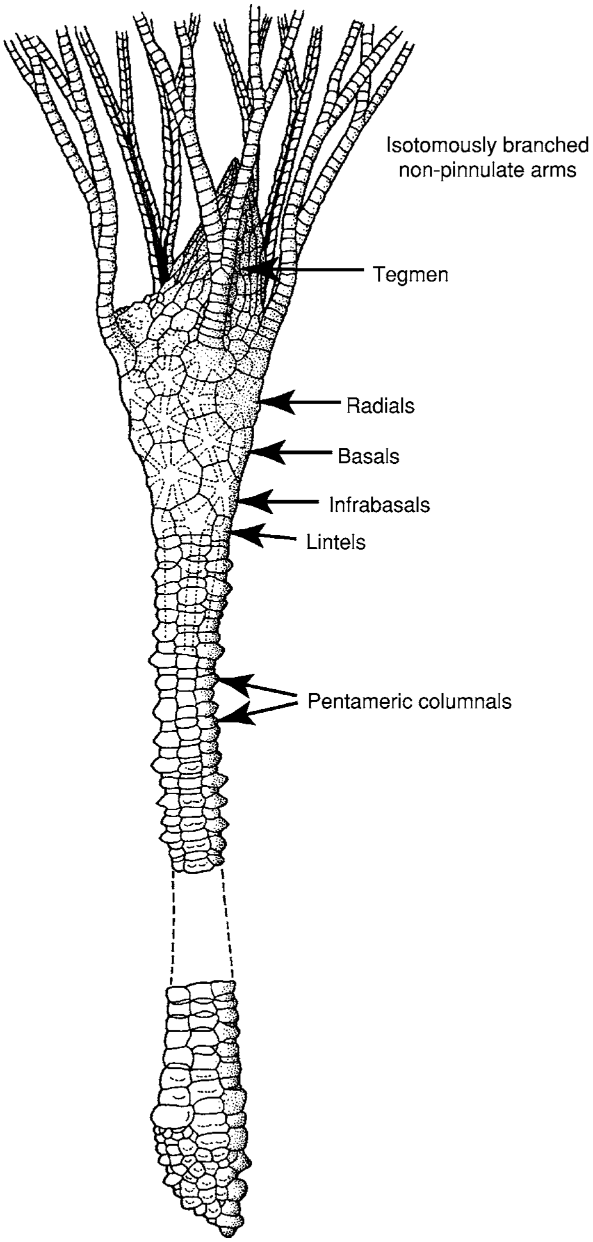


Figure 4 Reconstruction of *Aethocrinus moorei* from the Early Ordovician the earliest undisputed crinoid known. This has the primitive four circlet arrangement of plates in the cup, simple isotomous arms, and a pentameric stem. Annotated and reproduced with permission from Hess H, Ausich WI, Brett CE, and Simms MJ (1999) *Fossil Crinoids*. Cambridge: Cambridge University Press.

vascular systems. Some authorities favour an origin among the Late Cambrian or earliest Ordovician rhombiferan ‘cystoids’; others have suggested, based on several crinoid-like pelmatozoans discovered in the earliest Ordovician (Tremadoc), that they are descended from an edrioasteroid ancestor.

The earliest undisputed crinoid yet discovered is *Aethocrinus moorei*, from the Early Ordovician (Tremadoc or Arenig) (Figure 4). *Echmatocrinus brachiatus*, from the Middle Cambrian Burgess Shale, was for nearly two decades regarded as a possible ancestor of all later crinoids on account of its supposedly irregularly multiplated cup, but its identity as an echinoderm is now disputed.

There was considerable diversification in the Early Ordovician such that by the start of the Llanvirn Stage nearly all of the major Palaeozoic clades (aethocrinids, disparids, cladids, diplobathrid camerates and monobathrid camerates) were represented (Figure 5). The remaining subclass, the Flexibilia, arose from among the cladids during the Caradoc Stage of the Late Ordovician. This diversification continued into the Late Ordovician, and crinoids as a whole suffered only a minor drop in diversity during the mass extinction near the close of the Ordovician, although thereafter disparids ceased to be a dominant element of crinoid faunas. The fossil record of crinoids is poor for the Llandovery Stage but those from later in the Silurian and into the Devonian show clearly that the overall diversification continued unabated, though several important families declined or disappeared as a result of the Late Devonian (Frasnian–Famennian) extinction event. Crinoid diversity reached its acme during the Late Palaeozoic; several thousand species have been described from the Lower Carboniferous, and an apparent reduction in diversity into the Permian may well be merely an artefact of collection or description failure. Crinoid debris is often abundant and diverse in late-Permian marine successions, even in the very highest strata, but this disarticulated material has not been thoroughly investigated. It is clear that, along with so many other groups, crinoids experienced a profound drop in diversity from the Permian into the Triassic, but whether crinoid diversity remained high right up to the Permo-Triassic boundary, or was declining prior to this, remains unclear.

In Early Triassic marine sediments crinoid remains are rare and form a very low-diversity or even monospecific assemblage of small crinoids, assigned to the genus *Holocrinus*, which is the common ancestor of the subclass Articulata and all post-Palaeozoic crinoids. The articulates arose from a group of morphologically advanced cladids with muscular arm articulations; all of the other cladids, along with the flexibles, disparids and camerates, appear to have been eradicated entirely by the end-Permian extinction.

The articulates diversified rapidly in the mid- to Late Triassic but experienced a significant extinction early in the Late Triassic (mid-Carnian Stage), marked particularly by the disappearance of the distinctive and diverse encrinids. Jurassic faunas were

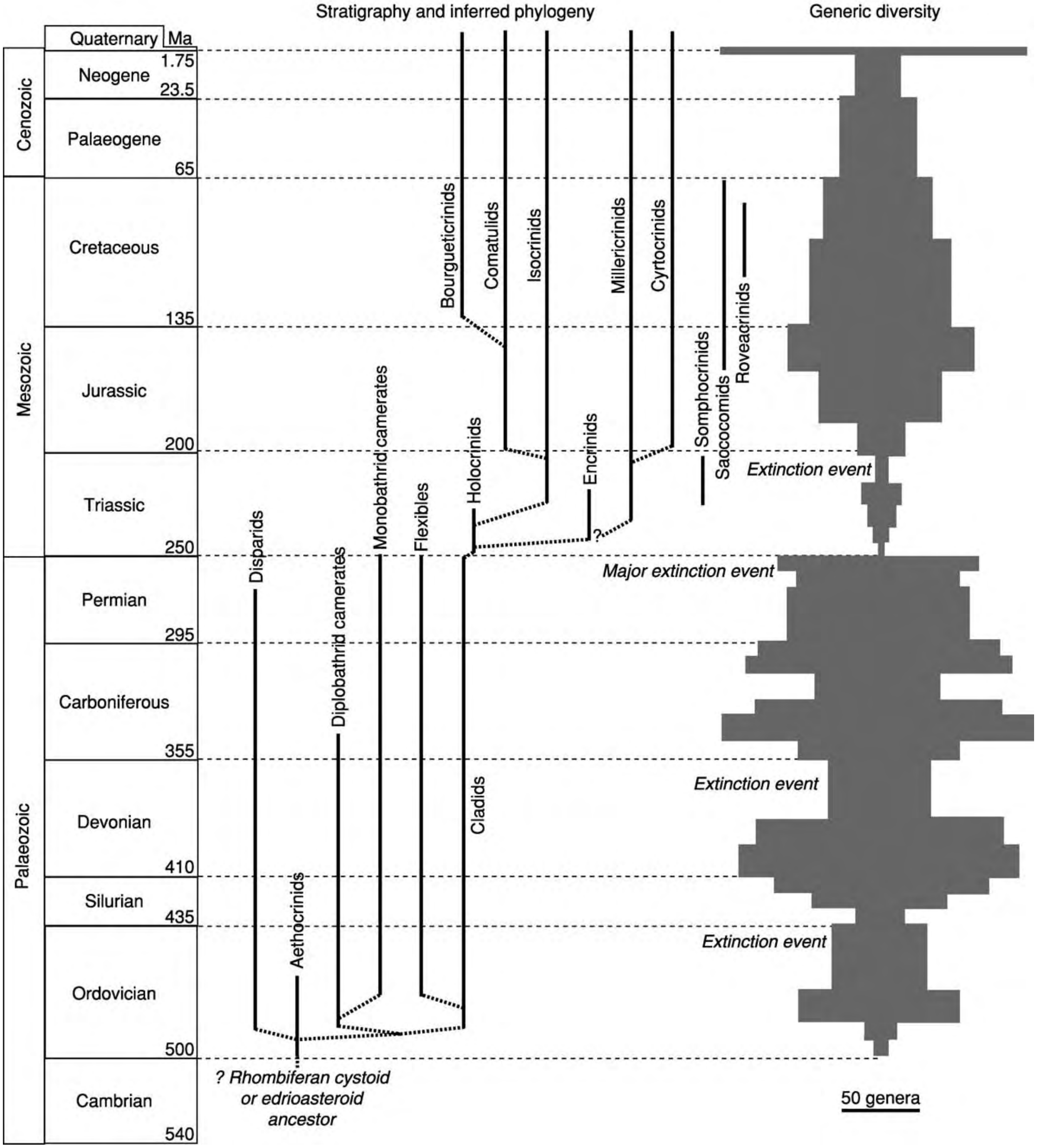


Figure 5 Stratigraphic distribution and phylogenetic relationships of the main crinoid groups, with total crinoid diversity indicated on the right.

dominated by two distinct clades – the isocrinids and the millericrinids. A stemless offshoot of the former, the comatulids, became increasingly dominant from the mid-Jurassic and now accounts for some 85% of the total crinoid taxonomic diversity.

A recurrent theme in the history of the Crinoidea is the appearance of microcrinoids, which may be broadly defined as those with an adult cup diameter of about 2 mm or less. Palaeozoic microcrinoids are known from the Silurian to the Permian and were exclusively sessile stalked benthos; most are assigned to a small number of cladid and disparid families. Mesozoic microcrinoids were quite different, lacking a stem during all known growth stages, leading to suggestions that they were planktic in habit. Three distinct groups are recognized: the mid- to Late Triassic Somphocrinidae, the Late Jurassic Saccocomidae, and the Cretaceous Roveacrinidae. They are often grouped together as 'roveacrinids,' but each group probably evolved independently from a more 'normal' ancestor. No microcrinoids are known after the end of the Cretaceous.

Ecology and Taphonomy

All extant crinoids are fully marine, intolerant of brackish or hypersaline conditions, and have relatively high oxygen demands. Evidence from the fossil record suggests similar ecological limits for extinct taxa, and hence they are useful palaeoenvironmental indicators. Extant crinoids are exclusively passive suspension feeders, and this was probably the case for most fossil taxa, although there have been suggestions of active filtration mechanisms in some highly specialized taxa, while among the microcrinoids there may have been an element of active podial capture of plankton. For many decades it was assumed that typical crinoids adopted a 'feeding bowl' arm configuration to intercept detritus sinking through the water column. However, observations by deep submersible in the 1970s indicated that many stalked crinoids adopt a quite different strategy, that of a 'parabolic filtration fan' in which the crown is held roughly at right angles to the current to intercept material carried along by it (Figure 6). Crinoids in turbulent environments where there is no single, or tidally reversing, dominant current adopt a range of feeding configurations depending on the degree of turbulence and the number of arms, but all appear to exploit currents rather than depending on passive settling of detritus.

Elevating the feeding arms helps to place them into faster-moving currents above the benthic boundary layer (Figure 6). The development among Early Ordovician crinoids of columns substantially longer than those of contemporaneous non-crinoid pelmatozoans may well have been a major factor in the overwhelming success of crinoids, compared with other echinoderms, in the Palaeozoic. Elevation of the arms also reduced competition with other benthic suspension feeders, including other species of crinoid. During periods of high crinoid diversity, such as the Early Carboniferous, crinoid crowns occupied several distinct tiers, up to a metre or so above the sea floor, thereby partitioning the food resource brought along by horizontal currents. In some environments swept by particularly strong or turbulent currents the entire crinoid morphology may show profound ecological adaptations, sometimes with very short and robust stem and arms, and occasionally with the cup cemented directly to the substrate. The Jurassic to Holocene cyrtocrinids exemplify a range of such morphologies.

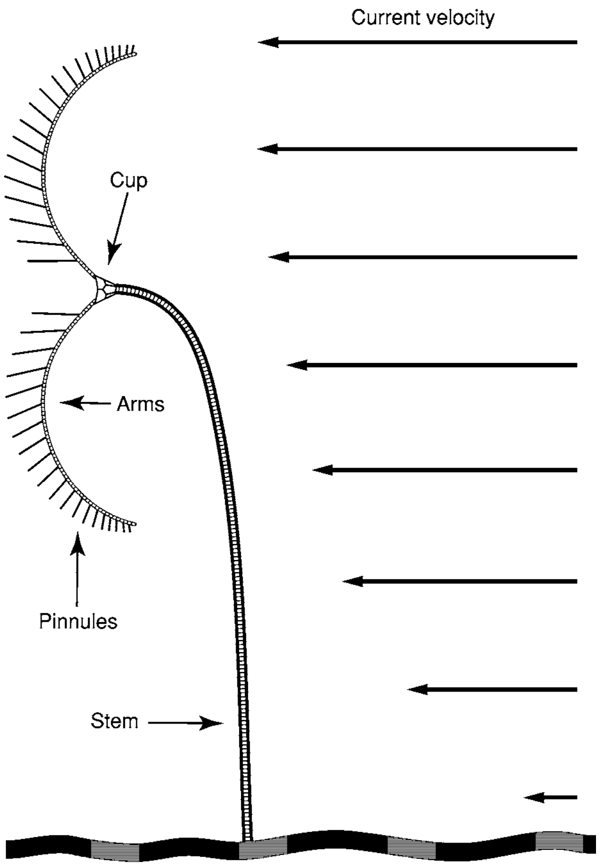


Figure 6 Diagrammatic reconstruction of a stalked crinoid in a rheophilic feeding posture, showing the position of the arms and the reduction of current flow velocity near to the benthic boundary layer.

Despite the relatively low proportion of soft tissue in crinoids, examples of predation are well documented among extant taxa, and the spinose arms and/or calyces of some fossil taxa may have evolved as predator deterrents. However, the regenerative powers of crinoids are formidable, and entire crinoid crowns may regenerate from just a surviving basal circlet. The ability of crinoids to autotomize arms and regenerate new ones is important in this respect, and there are many fossil examples known where this has clearly happened.

The construction of the crinoid endoskeleton, with a multi-element skeleton of calcite ossicles held together by soft tissues, confers on them poor preservation potential as intact specimens. The soft tissues decay quickly after death, often within days, and disarticulation is rapid in the often turbulent and well-oxygenated environments favoured by crinoids.

Muscles decay most rapidly, with the short ligaments connecting adjacent ossicles decaying faster than the longer through-going ligaments found in the stems. Hence the postmortem disarticulation of crinoids often follows a definite sequence, with lengths of stem (pluricolumnals) held together by the through-going ligaments being the last to disarticulate into individual ossicles. This disarticulated material is common in the fossil record, in some instances attaining rock-forming abundance. Such crinoidal limestones, often termed 'encrinites' (Figure 7), occasionally reach thicknesses of tens of metres over many thousands of square kilometres. Many examples are known from the Upper Palaeozoic, but they are rare thereafter, with only a few known from the Triassic and Jurassic.

Preservation of intact crinoids indicates an unusual event. In most instances it can be attributed to obrution – the smothering of the living animals by a sudden influx of sediment sufficient both to cause death and to protect the remains against subsequent disruption by burrowing scavengers. Storm resuspension of sediment is probably the most common cause of obrution, and hence some successions, deposited in environments prone to obrution, may include numerous horizons with intact crinoids. Crinoids are also sometimes found intact in dark laminated mudstones that were clearly deposited in anoxic conditions inimical to their survival but greatly enhancing their preservation potential through the exclusion of scavengers and the slowing of decay rates. Some of these crinoids can be interpreted as allochthonous faunal elements catastrophically swept in from a more favourable adjacent environment. However, a few crinoid taxa are regularly found in such anoxic facies and appear to be poorly represented elsewhere, indicating an unusual and specialized ecology. Four such genera – the Early Jurassic *Pentacrinites* (Figure 3) and *Seirocrinus*, the Late Triassic *Traumatocrinus*, and the Late Devonian *Melocrinites* – are often intimately associated with fossil driftwood, and there is considerable evidence, such as the crinoids almost invariably being preserved beneath the wood, that in life these crinoids were pseudoplanktic. They were introduced into these anoxic environments by accident when the floating driftwood to which they were attached finally sank, perhaps under the burden of its rapidly growing crinoid passengers. Pseudoplanktic crinoids show remarkable morphological convergence, particularly in their long stems (examples of *Seirocrinus* are known with stems over 20 m long!) and strongly endotomous arm branching, suggesting intense evolutionary selection pressure in this highly specialized niche.

There have been suggestions of a planktic mode of life for several taxonomically, and morphologically, disparate groups of fossil crinoids with near cosmopolitan distributions. The small and delicate construction of the Mesozoic 'roveacrinids' has been considered to be an adaptation to a planktic or nektic habit, although this is still subject to debate. Similarly, a planktic mode of life has often been proposed for

Figure 7 A typical encrinite, or crinoidal limestone, from the Early Carboniferous of North Wales, UK. Field of view 30 cm.

two large stemless crinoid genera from the Cretaceous – *Marsupites* and *Uintacrinus* – although recent interpretations favour a benthic habit. In the Late Silurian to Early Devonian a bizarre group of large camerate crinoids – the Scyphocrinitidae – modified the distal end of the stem into a large chambered 'bulb,' which has been interpreted as a flotation structure.

See Also

Fossil Invertebrates: Echinoderms (Other Than Echinoids); Echinoids. **Mesozoic:** Triassic. **Palaeoecology**. **Palaeozoic:** Devonian; End Permian Extinctions. **Sedimentary Environments:** Carbonate Shorelines and Shelves.

Further Reading

Ausich WI (1998) Early phylogeny and subclass division of the Crinoidea (Phylum Echinodermata). *Journal of Paleontology* 72: 499–510.

Cowen R (1981) Crinoid arms and banana plantations: an economic harvesting analogy. *Paleobiology* 7: 332–343.

Guensburg TE and Sprinkle J (2003) The oldest known crinoids (Early Ordovician, Utah) and a new crinoid plate homology system. *Bulletins of American Palaeontology* 364: 1–43.

Hess H, Ausich WI, Brett CE, and Simms MJ (1999) *Fossil Crinoids*. Cambridge: Cambridge University Press.

Meyer DL and Ausich WI (1983) Biotic interactions among Recent and among fossil crinoids. In: Tevesz MJS and McCall PL (eds.) *Biotic Interactions in Recent and Fossil Benthic Communities*, pp. 377–427. New York: Plenum Publishing.

Moore RC and Teichert C (eds.) (1978) *Treatise on Invertebrate Paleontology, Part T, Echinodermata 2*, vols 1–3. Boulder, Colorado and Lawrence, Kansas: Geological Society of America and University of Kansas.

Seilacher A, Reif W E, and Westphal F (1985) Extraordinary fossil biotas: their ecological and evolutionary significance. *Philosophical Transactions of the Royal Society of London B* 311: 5–23.

Simms MJ (1986) Contrasting lifestyles in Lower Jurassic crinoids: a comparison of benthic and pseudopelagic Isocrinida. *Palaeontology* 29: 475–493.

Simms MJ (1994) Reinterpretation of thecal plate homology and phylogeny in the Class Crinoidea. *Lethaia* 26: 303–312.

Echinoids

A B Smith, The Natural History Museum, London, UK

Introduction

Echinoids are a diverse and widespread group of marine macroinvertebrates and one of the five classes of echinoderm (*see* **Fossil Invertebrates:** Echinoderms (Other Than Echinoids)). They are exclusively marine and benthic, living either on or within the sea floor, and with their mouth downward-facing. The group includes such well-known animals as the sand dollar, heart urchin and sea urchin. Like all echinoderms, they have a calcitic skeleton constructed of a mosaic of plates, and these are usually firmly bound together to form a rigid and robust test. This skeleton is architecturally complex and, being mesodermal, is able to provide a wealth of information about the musculature and soft-tissue anatomy. Because echinoids have a rigid skeleton, they have left the best fossil record of any echinoderm class. They are also the best understood echinoderm group from a phylogenetic perspective. This combination of a complex skeleton, good fossil record and extensively studied phylogeny has made echinoids the focus of many important palaeobiological studies in recent years.

There are about 900 extant species equally divided between regular forms ('regulars') whose anus opens in the aboral plated surface and that live epifaunally, and irregular forms ('irregulars') whose anus is displaced away from the aboral plates into the posterior interambulacral zone and that live predominantly infaunally (Figure 1). All echinoids have a well-developed system of hydraulically operated tube-feet that are important for respiration, feeding and locomotion. Echinoids also have a dense and often highly differentiated covering of calcitic spines used for locomotion and defence.

Morphology of Post-Palaeozoic Echinoids

Echinoids range in size from just a few millimetres in diameter to over 350 mm and come in a variety of body forms: globular, heart-shaped, cylindrical, hemispherical or even flattened discoidal (Figures 1 and 2). However, irrespective of their overall shape, the skeleton (also known as the 'test') in post-Palaeozoic echinoids is always constructed along the same standardized plan. At the apex of the test is a small cluster of up to 10 plates known as the apical disc (Figure 1A). In regular echinoids this series of plates surrounds the

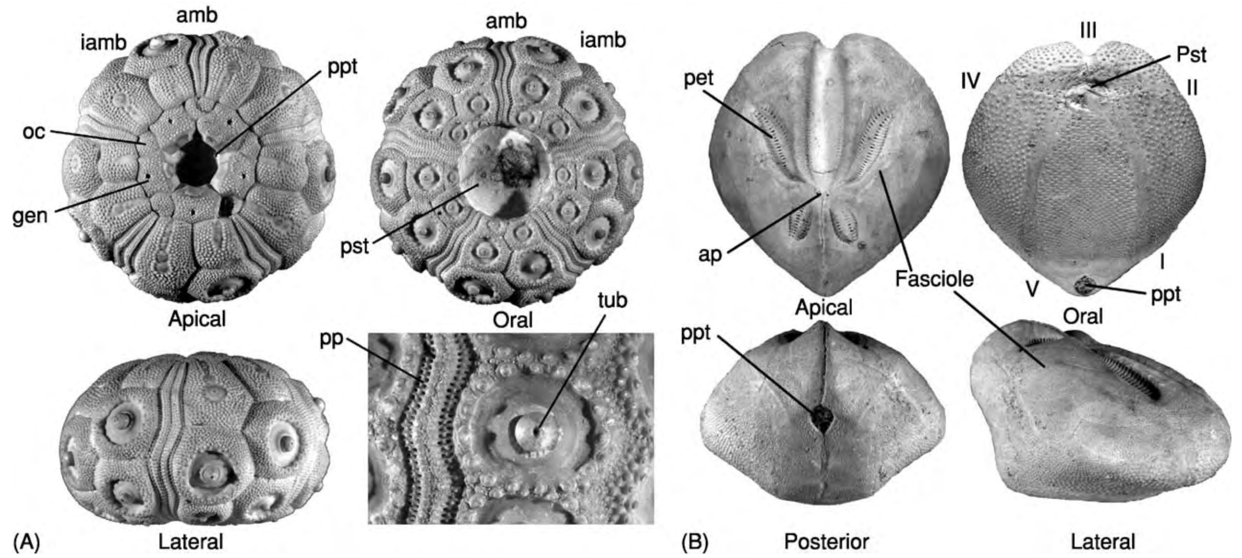


Figure 1 Morphology of (A) a regular echinoid *Stereocidaris,* an Upper Cretaceous cidaroid from England, and (B) an irregular echinoid *Schizaster*, an Eocene spatangoid from Italy. amb ambulacrum; ap apical disc; gen genital plate; iamb interambulacrum; oc ocular plate; pet petal; pp pore pair; ppt periproct; pst peristome; tub tubercle. I V numbered ambulacral zones.

anus but in irregular echinoids the anus is usually outside and to the posterior (Figure 1B). The larger genital plates are in part derived from larval skeleton and are all that remains of the aboral plated surface that is so well developed in other echinoderms. Genital plates are each perforated by a gonopore, which acts as the outlet for the gonads. One of the five genital plates is also perforated like a pepper pot, which forms the opening to the water vascular system. The smaller plates are also perforate, with their tiny opening marking where the tip of the radial water vessel reaches the exterior: these are the ocular or terminal plates. Plates are added at the distal edge of ocular plates during growth.

The remainder of the test is composed of 20 columns of plates arranged into alternate pairs of ambulacral and interambulacral elements (Figure 1). Ambulacral plates overlie the radial water vessel and are pierced by single or double pores where tube-feet connect to the interior. Interambulacral plates by contrast are generally larger and are imperforate. They arise on either side of the ocular plate. The external surfaces of both ambulacral and interambulacral plates are covered in tubercles of various sizes and these are the articulation points for spines. Regular echinoids have only a small number of large tubercles whereas irregular echinoids typically have uniform, fine, dense tuberculation (Figure 1A, B). Ambulacral and interambulacral columns of plates converge around the mouth, which lies at the opposite pole from the apical disc.

The mouth is situated in a large opening termed the peristome. This is covered in life by a flexible membrane with embedded plates that is rarely fossilized. All regular echinoids and many irregular echinoids have an internal jaw apparatus that largely fills the peristome. The lantern is a pentaradially symmetrical and highly complex in structure (Figure 3), being composed of 50 skeletal elements and 60 individual muscles. There are five teeth, each braced in a hemipyramid, and the entire structure can open and close like a grab, as well as moving in and out of the test. Because echinoids have a rigid test of fixed internal volume, having a large and active lantern that moves in and out creates problems. To compensate for changes in internal volume there are 10 expandable soft-tissue sacs around the edge of the peristome that connect directly to the interior and that accommodate any displaced body cavity fluids. Each sac passes to the exterior at a small notch in the peristome (buccal notch). Echinoids that either lack a lantern or have an entirely internal lantern lack buccal notches.

The primary muscles that move the lantern attach to the interior of the test around the rim of the peristome. Enlarged skeletal flanges mark the site of muscle attachment (Figure 3). Flanges developed from the adoral ambulacral plates are termed auricles while those developed from interambulacral plates are termed apophyses. These lantern muscle attachments together form the perignathic girdle.

Appendages that attach to the test include both spines and pedicellariae. Spines vary tremendously in size and shape according to their function (Figure 4). They can be either hollow or solid and either smooth or ornamented externally. There may be a central ligament binding the spine to its articulation ball,

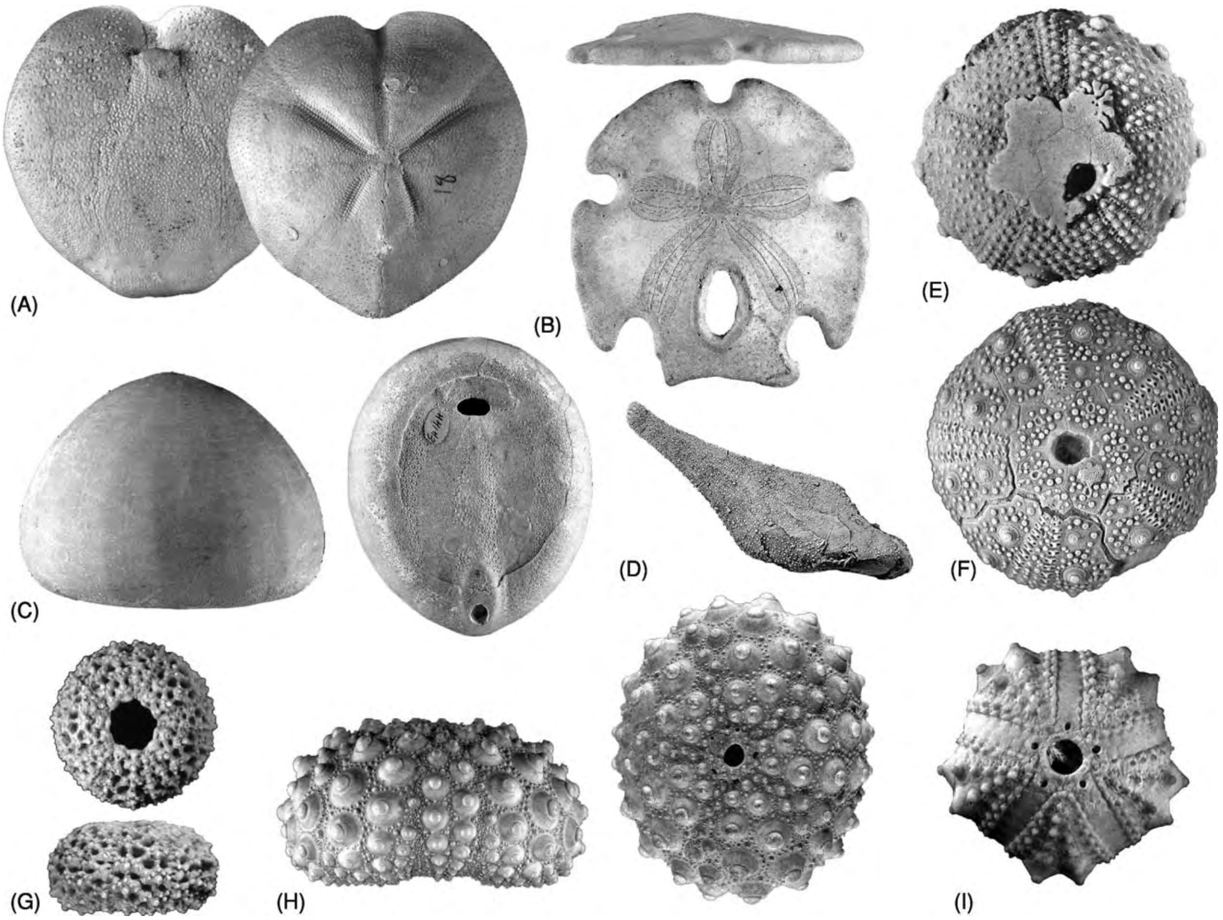


Figure 2 Representative fossil post Palaeozoic echinoids. (A) the spatangoid *Micraster* in oral and apical views, from the Upper Cretaceous of England (×0.7). (B) the clypeasteroid *Encope* in lateral and apical views, from the Pliocene of the U.S.A. (×0.4). (C) the holasteroid *Echinocorys* in lateral and oral views, from the Upper Cretaceous of England (×0.4). (D) the holasteroid *Hagenowia* in lateral view from the Upper Cretaceous of England (×2). (E) the salenioid *Novosalenia* in apical view, from the Cretaceous of the Czech Republic (×2). (F) the pedinoid *Hemipedina* in apical view, from the Lower Jurassic of England (×3). (G) the temnopleuroid *Viaudechinus* in apical and lateral views, from the Miocene of France (×2). (H) the echinoid *Heterocentrotus* in lateral and apical views from the Recent of the Indo West Pacific (×0.7). (I) the arbacioid *Murravechinus* in apical view from the Miocene of Australia (×1.5).

in which case the associated tubercle is perforate (Figure 1A) and the surrounding platform may be smooth or crenulated. Greatest functional differentiation is found in irregular echinoids where spines perform a variety of different roles. Pedicellariae are microscopic stalked, jawed appendages that resemble tulips in shape and that are used to deter small ectoparasites. They evolved from clusters of spines and are present in all echinoids from the Silurian onwards. There are many different forms of pedicellaria making them very useful species-level indicators but, unfortunately, they are rarely preserved in fossils.

Morphology of Palaeozoic Echinoids

Although Palaeozoic echinoids mostly retain a pentaradiate pattern of alternating ambulacral and interambulacral zones, the precise way their test

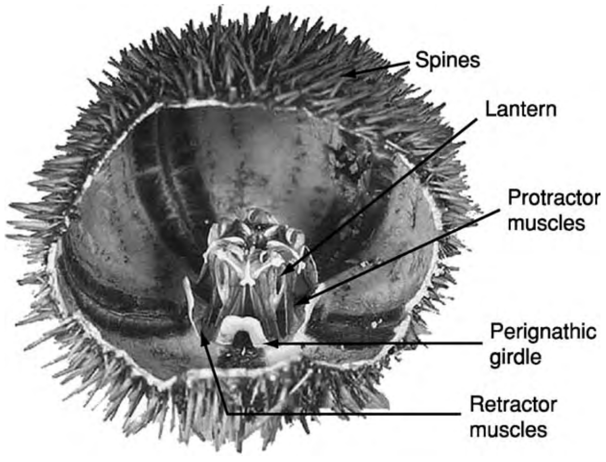


Figure 3 Test of a modern echinoid (*Echinus*) broken open to show the internal lantern and perignathic girdle.

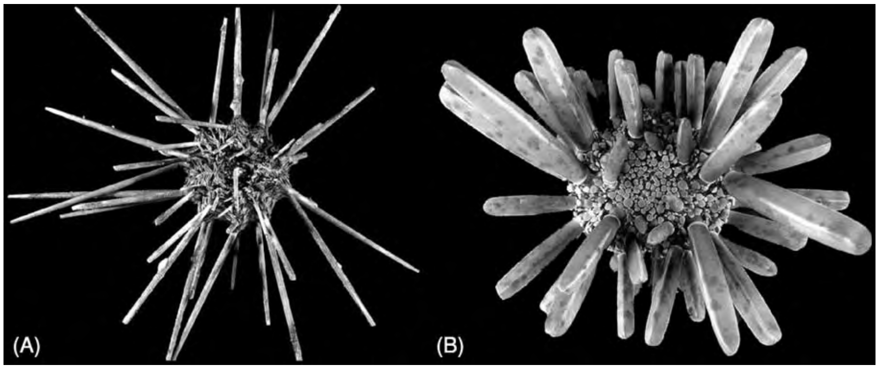


Figure 4 Modern echinoids with their complement of spines attached. (A) the cidaroid *Cidaris*; (B) the echinoid *Heterocentrotus*.

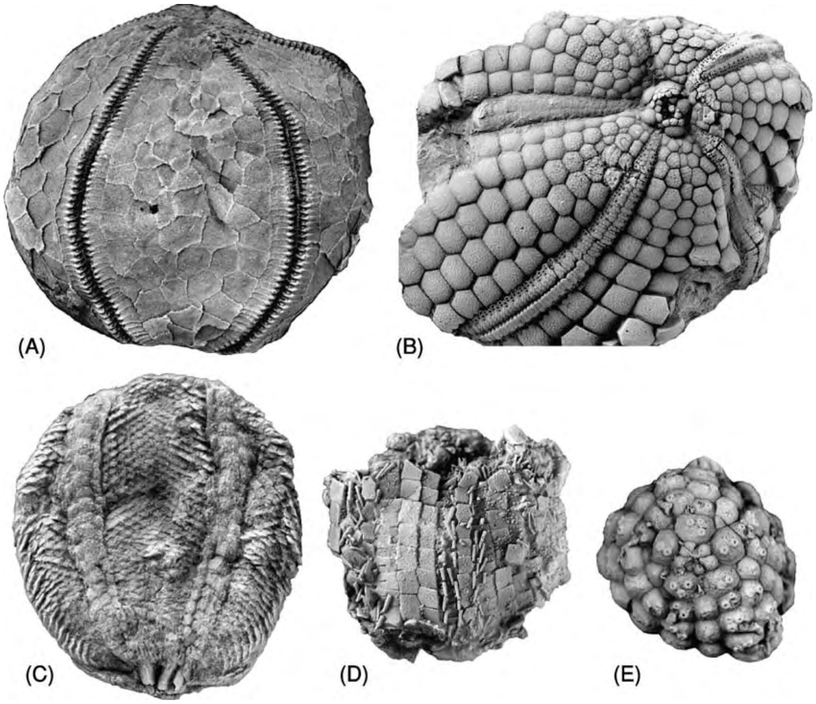


Figure 5 Representative fossil Palaeozoic echinoids. (A) *Aulechinus* (internal mould) from the Upper Ordovician of Scotland (×1). (B) *Maccoya* from the Carboniferous of England (×0.5). (C) *Lepidesthes* from the Carboniferous of the USA (×1). (D) *Bromidechinus* from the Upper Ordovician of the USA (×1). (E) *Bothriocidaris* from the Upper Ordovician of Estonia (×3).

is constructed shows a huge amount of variation (Figure 5). Only the Permian genus *Miocidaris* has a test constructed of 10 ambulacral and 10 interambulacral columns like that of post-Palaeozoic echinoids. In all others the number of columns of plates, both ambulacral and interambulacral, varies tremendously. Some, such as *Palaechinus* and *Aulechinus* (Figure 5A) have biserial ambulacra and multiple columns of non-ambulacral plates. Others, such as *Proterocidaris* and *Lepidesthes* (Figure 5C), have very wide ambulacra of up to 20 columns of plates, and narrow interambulacra. *Bothriocidaris* (Figure 5E) has no interambulacral plates at all, but there is an imperforate series of ambulacral plates. Primitive echinoids also, interestingly, have the radial water vessel enclosed within the ambulacral plates rather than being fully internal, as it is in all post-Palaeozoic forms. A lantern is present, but the teeth are very primitive in structure and no Palaeozoic echinoid has a perignathic girdle. Finally, plating is imbricate (Figure 5D) and so the test rapidly fell apart at death so that even partially articulated specimens are rare except in Lagerstätte deposits.

Echinoid Palaeobiology

Locomotion and Burrowing

Regular echinoids are vagile benthos, moving by means of their oral spines over the sea floor and using their tube-feet to climb and grip hard substrata. In forms living in high-energy, rocky, shore environments, the oral tube-feet that provide grip are highly muscular and also densely packed, an adaptation that is easily recognizable from fossil tests. Only a few regular echinoids, notably cidaroids and arbacioids, evolved specialized aboral respiratory tube-feet: for the most part tube-feet are undifferentiated. Spines would have been the primary means of locomotion in Palaeozoic echinoids, although some of the most primitive presumably used their tube-feet.

Irregular echinoids rely entirely on their spines for locomotion, and most burrow into unconsolidated sediment. This raises a number of problems: locomotion is harder, and circulation of oxygenated water and removal of waste products more difficult, particularly in finer-grained, more impermeable sediments. Infaunal echinoids have highly specialized aboral tube-feet for gaseous exchange and these regions form characteristic petals on the surface of the test (Figure 1B). However, only heart urchins have successfully adapted for life in fine-grained sediments. In order to do this they have specialized aboral tube-feet that are used to build and maintain an open shaft to the surface down which fresh water can be drawn. Water is drawn into the burrow by bands of highly ciliated spines, termed fascioles. These fascioles also generate a mucous coat that is held over the tips of the spines and prevents fine-grained sediment from falling between the spines. The mucous coat thus allows heart urchins to maintain a water-filled space within the burrow.

The traces of burrowing echinoids have been described in rocks from the late Lower Cretaceous onwards.

Feeding

Regular echinoids feed using their lantern. Palaeozoic echinoids had relatively weak lantern musculature and feeding must have largely involved scooping material from the sea floor. With stronger lanterns echinoids were able to utilize new food sources and could rasp encrusting organisms from hard substrata. Echinoid tooth bite marks first appear as trace fossils in the late Triassic, at about the same time that well-developed perignathic girdles and stronger lanterns evolved. During the Mesozoic the biting force that lanterns could exert was further increased through the development of teeth with a stronger T-shaped cross-section, and the stronger and more extensive bracing of hemipyramids. Mesozoic and Tertiary echinoids were mostly either algivores or predatory on sessile organisms.

Deposit feeding may have evolved in the Carboniferous, when echinoids such as *Proterocidaris* evolved large numbers of massive oral tube-feet whose most likely function was in detritus gathering. However, this clade became extinct towards the end of the Permian and deposit feeding was not adopted again by echinoids until the Early Jurassic. The evolution of irregular echinoids in the Jurassic is really the story of adaptation towards improved deposit feeding. Some irregular echinoids (cassiduloids) are bulk sediment swallowers, passing a continuous stream of sediment through the gut in order to extract the small quantities of organic mater it might contain. Others (spatangoids and holasteroids) evolved as selective deposit feeders using specialised tube-feet to pick out organic-rich detritus. Sand dollars are rather special in that they appear to be using their aboral spine canopy as a sort of sieve to separate fine organic material from amongst the sand grains. The finer particles fall between the spines and are swept towards the mouth and entrapped in mucous strings that run towards the mouth. These food grooves are easily visible even on a denuded test.

Predation and Defence

Echinoids today are preyed upon by a variety of animals (including humans), and predation has, presumably, been a major factor since the Devonian, when the first echinoids with long, lance-like spines evolved. Long, sharp spines (Figure 4A) are effective deterrents to many predators, but alternative strategies have also been evolved. Cidaroids have spines that are not covered in a living tissue, and that, consequently, can become heavily encrusted with algae and sessile organisms. This encrustation on flattened, paddle-shaped spines provides very effective camouflage. An alternative ploy has been to develop very thick club-shaped spines that cannot easily be snapped or broken off (Figure 4B). A few regular echinoids have abandoned spines in favour of poisonous pedicellariae as their first line of defence. *Toxopneustes* has a very short, uniform coating of spines and their highly venomous pedicellariae are almost as long. The aboral surface of this animal thus presents a formidable battery of dense, poisonous pedicellarial jaws.

Spines can also provide a first level of protection against impact in high-energy environments, and both in the Jurassic and in the Tertiary echinoids evolved that were short, stout and umbrella-like, forming a tessellate pavement over the surface of the test.

Reproduction

In echinoids the sexes are separate, but it is usually impossible to tell the sexes apart visually. Reproduction occurs by broadcast spawning: eggs and sperm are released into the water column where fertilization takes place. The fertilized egg then develops into a planktonic larva termed pluteus, which feeds in surface waters for a few weeks before metamorphosing into the adult form and settling to the bottom. A small number of species, most from high-latitude, do not pass through a planktonic larval stage, but develop directly. In such cases the female produces large yolk-rich eggs that are brooded on the test. Females then have special brood pouches termed marsupia where the eggs are protected, and these form sunken zones around the gonopore openings. Sexual dimorphism first evolved in the Late Cretaceous and was particularly prevalent in Australian faunas in the Early Tertiary.

Geological History

The closest living relative of the echinoids are holothurians (*see* **Fossil Invertebrates:** Echinoderms (Other Than Echinoids)) and their ancestry presumably lies amongst the 'asterozoan' taxa of the Early Ordovician. When echinoids first appear in the Upper Ordovician, there were already three very different body forms. Bothriocidarids have a small globular test composed of thick plates of the ambulacral series only (Figure 5E), lepidocentrids have narrow biserial ambulacra and wide zones of irregularly arranged imbricate interambulacral plates (Figure 5A), and *Bromidechinus* has biserial interambulacra and triserial ambulacra (Figure 5D). Clearly the earliest phase in the history of echinoids has yet to be uncovered.

During the Lower Palaeozoic, echinoids remained a minor component of marine benthic communties, never becoming diverse or abundant. By the Devonian one lineage had developed enlarged adoral tube-feet and appears to have specialized as a deposit feeder, and this group thrived through to the Permian before going extinct. In the Carboniferous another important group, the archaeocidarids, made their appearance. Archaeocidarids have a single large tubercle on each interambulacral plate and possessed long highly muscular spines. On functional grounds they were probably the first active predators. *Miocidaris*, the only Palaeozoic echinoid with the test architecture of post-Palaeozoic forms, evolved directly from *Archaeocidaris* by reduction of plating columns in each interambulacral zone.

The modern crown group was initiated and had already begun to diversify before the end of the Permian, with both a cidaroid and a non-cidaroid lineage passing into the Triassic. The cidaroids (Cidaroidea) developed rigid test plating in the Triassic and have been an important constituent of deeper-water, shelf communities ever since. Despite attaining a modest diversity, cidaroids have always remained very conservative in their morphology.

Small non-cidaroid forms flourished in reefal settings in the Late Triassic, but it was not until the early Jurassic that the other lineage (Euechinoidea) started to diversify in a major way. Irregular echinoids make their first appearance in the Lower Jurassic and rapidly evolved as deposit feeders. The earliest irregular echinoids still possessed a lantern like that of regular echinoids, but by the Middle Jurassic both cassiduloids and spatangoids had evolved. Cassiduloids specialized as bulk sediment feeders using dense tube-feet for handling sediment particles. Sand dollars arose in the Early Tertiary from cassiduloid ancestors and specialized for life in the near-shore. Unique amongst echinoids, they developed large numbers of microscopic tube-feet to each ambulacral plate and adopted a very singular method of sediment sifting to obtain their food. Spatangoids and holasteroids on the other hand evolved a special kind of tube-foot around the mouth for selective deposit feeding. Modern heart urchins first appear at the base of the Cretaceous and have diversified more or less constantly since then. The end-Cretaceous extinction (*see* **Mesozoic:** End Cretaceous Extinctions) affected echinoids selectively, with deposit feeders being hardest hit. Regular echinoids also thrived throughout the Mesozoic, although today's major modern groups did not become established until the Late Cretaceous.

Phylogeny and Classification

Major groups and their characteristic features are listed below and their geological ranges are shown in Figure 6:

- Cidaroida (Figure 1A) (Permian – Recent): regular echinoids with simple ambulacral plating and interambulacra with a single large primary tubercle; teeth U-shaped in cross-section.
- Echinothurioida (Jurassic – Recent): regular echinoids with hollow spines and an imbricate plated test; tubercles perforate and non-crenulated; teeth U-shaped in cross-section.
- Diadematoida (Jurassic – Recent): regular echinoids with rigid, thin-plated test, compound ambulacra and perforate and crenulated tuberculation; teeth U-shaped in cross-section.
- Pedinoida (Figure 2F) (Triassic – Recent): regular echinoids with perforate non-crenulate tubercles,

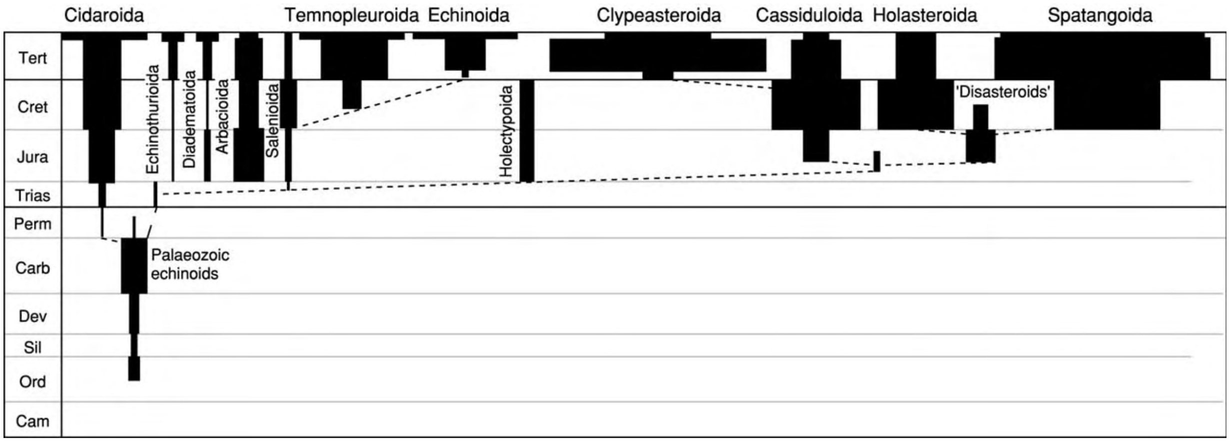


Figure 6 Stratigraphic range charts for the major groups of echinoid. Width of bars is proportional to taxic (genus level) diversity. The dashed lines show their phylogenetic relationships.

compound ambulacral plating and teeth U-shaped in cross-section.

- Arbacioida (Figure 2I) (Jurassic – Recent): regular echinoids with imperforate, non-crenulate tubercles, compound ambulacral plating and teeth T-shaped in cross-section.
- Salenioida (Figure 2E) (Jurassic – Recent): regular echinoids with imperforate tubercles, compound ambulacral plating and teeth T-shaped in cross-section. The apical disc is large and cap-like and has extra plates incorporated into it.
- Temnopleuroida (Figure 2G) (Upper Cretaceous – Recent): regular echinoids with compound ambulacral plating, imperforate tuberculation that is either crenulated or non-crenulate, and teeth T-shaped in cross-section. The test surface is often highly ornamented with pits.
- Echinoida (Figure 2H) (Upper Cretaceous – Recent): regular echinoids with compound ambulacral plating, imperforate, non-crenulate tuberculation and teeth T-shaped in cross-section.
- Holectypoida (Jurassic – Cretaceous): irregular echinoids with a large functional lantern, no petals, and peristome with buccal notches. The periproct is large and tuberculation relatively coarse.
- Cassiduloida (Jurassic – Recent): irregular echinoids with no lantern and small peristome; tube-feet concentrated around the mouth to form phyllodes; aboral ambulacra petaloid; spines fine and dense.
- Clypeasteroida (Figure 2B) (Palaeocene – Recent): irregular echinoids, often flattened and disc-like, characterized by having petals, a fully internal lantern and small peristome and very many microscopic tube-feet to each ambulacral plate.
- Holasteroida (Figure 2C, D) (Cretaceous – Recent): ovate to heart-shaped irregulars, but may be more elongate; no lantern; apical disc elongate with ocular plates II and IV meeting behind anterior genital plates.
- Spatangoida (Figures 1B, 2A) (Cretaceous – Recent): ovate to heart-shaped irregulars; no lantern; apical disc compact.

See Also

Fossil Invertebrates: Echinoderms (Other Than Echinoids); Crinoids. **Lagerstätten**. **Mesozoic:** End Cretaceous Extinctions. **Sedimentary Environments:** Reefs ('Build-Ups'). **Trace Fossils**.

Further Reading

Kier PM (1965) Evolutionary trends in Paleozoic echinoids. *Journal of Paleontology* 39: 43 465.

Kier PM (1974) Evolutionary trends and their functional significance in the post Paleozoic echinoids. *Journal of Paleontology* 48(supplement 5): 1 95.

Lewis DN and Donovan SK (1998) Fossils explained 23: Palaeozoic echinoids. *Geology Today November/December:* 235 240.

Smith AB (1984) *Echinoid Palaeobiology,* p. 199. London: George Allen & Unwin.

Smith AB and Savill JJ (2002) *Bromidechinus*, a new Middle Ordovician Echinozoa (Echinodermata), and its bearing on the early history of echinoids. *Transactions of the Royal Society of Edinburgh* 91: 137 147.

Smith AB, Littlewood DTJ, and Wray GA (1996) Comparative evolution of larval and adult life history stages and small subunit ribosomal RNA amongst post Palaeozoic echinoids. In: Harvey PH, Leigh Brown AJ, Smith JM, and Nee S (eds.) *New Uses for New Phylogenies,* pp. 234 254. Oxford: Oxford University Press.

Smith AB (2004) *The Echinoid Directory.* World Wide Web electronic publication.www.nhm.ac.uk.

Graptolites

R B Rickards, University of Cambridge,
Cambridge, UK

Introduction

The graptolites were a group of colonial animals that inhabited the Palaeozoic seas, often in very great abundance. Their remains can be found as small elongate (3 mm–1 m) skeletons of thin collagenous periderm forming simple linear or complexly branched colonies. Each branch is composed of a linear succession of interconnecting tubes (thecae) originating from a single conical individual called a sicula (Figure 1). Graptolites are of value in the study of the stratigraphy of many Ordovician and Silurian deposits the world over, because of their abundance, wide geographical distribution, and rapid evolution. Benthic graptolites appeared in the Middle Cambrian, becoming extinct in the Upper Carboniferous; planktonic graptolites ranged from the earliest Ordovician to the Middle Devonian.

The Chequered History of Graptolite Studies

Although graptolites were originally observed and illustrated as long ago as the late eighteenth century, it was not until the mid-nineteenth century that much attention was paid to them as fossils and their animal nature was fully appreciated. Even then there was considerable doubt about their zoological affinities, and they were variously allied to the plants, molluscs, corals, protozoa, and bryozoans. Modern interpretation of their morphology places them in the Hemichordata based on comparisons with living forms. Thus, their zoological position has been elevated over the years from plants, through the lower invertebrates, to the lower vertebrates.

Understanding of their morphology has progressed greatly, aided by technical developments in microscopy, especially since the late 1950s with the use of scanning electron microscopes. Consequently, there is now a tremendous amount of detailed information available on this extinct group of strange little animals that might at first appear to be, as their name (graptos, written or marked; lithos, stone) implies, little more than markings on rock.

Classification, Broad Evolution, Stratigraphy, and Mode of Life

The ancestors *Eorhabdopleura* and *Eocephalodiscus* of the living Hemichordata probably arose with the order Dendroidea from a common ancestor in the early Middle Cambrian (Figure 2). The common ancestor may have been a phoronid-like worm that evolved through one or more of the following stages of evolution: infaunal, in dense association; infaunal with loose coloniality; and finally sessile with several stolons inside one upright peridermal tube. This last stage would have been equivalent to a primitive *Mastigograptus*-like graptolite. The evolutionary relationships of the orders other than dendroids and graptoloids are unknown, and the stratigraphic record of these other orders is poor. The evolution within the Order Graptoloidea is discussed in further detail later in this article, and the distinctions between dendroids and graptoloids are discussed below.

The Sicula (or Larval) Stage

Eggs and embryos are known but are extremely rare in graptolites. However, the peridermal sheath of the larval stage, called the sicula, is known in almost all graptolites and is very common in the rock record. Eggs, embryos, and larval stages were produced by sexual reproduction, but the remainder of the colony grew by asexual budding from the sicula. In dendroid siculae (Figure 3A) a resorption foramen (or hole) formed in the prosicula, and from this grew the first asexual bud of the colony. In most graptoloids (Figure 3B) the resorption foramina were in the metasicula, but in some, the Monograptina (Figure 3C), the growth lines construct a notch, then a hole (lacuna), through which the first theca grew. The dendroid sicula shown here (Figure 3A) is of a benthonic type: a planktonic type would have a nema instead of a basal disc and may have longitudinal threads.

The Stolon System

In dendroids and tuboids the thecal tubes are attached at their bases to a black (sclerotized) stolon system; in dendroids (Figure 4) each node divides into three stolons, but in tuboids (Figure 5) the division is always into two stolons and some species have only partly sclerotized stolons (i.e. only partly hardened). In graptoloids (Figure 6) there are no sclerotized stolons,

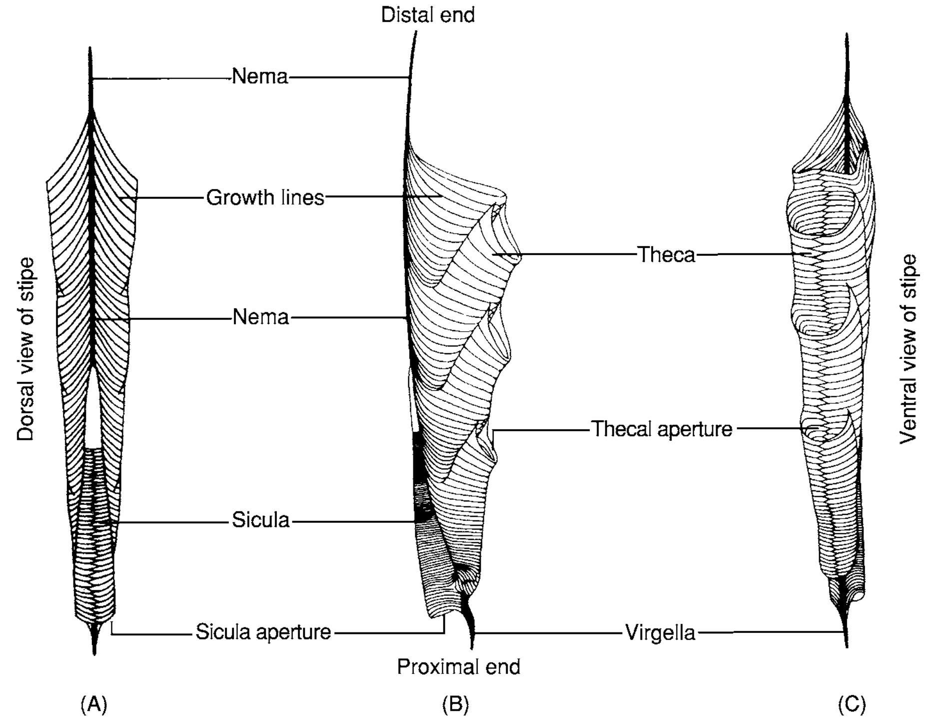


Figure 1 (A) Dorsal, (B) lateral and (C) ventral views of a graptolite stipe, showing the arrangement of the thecae.

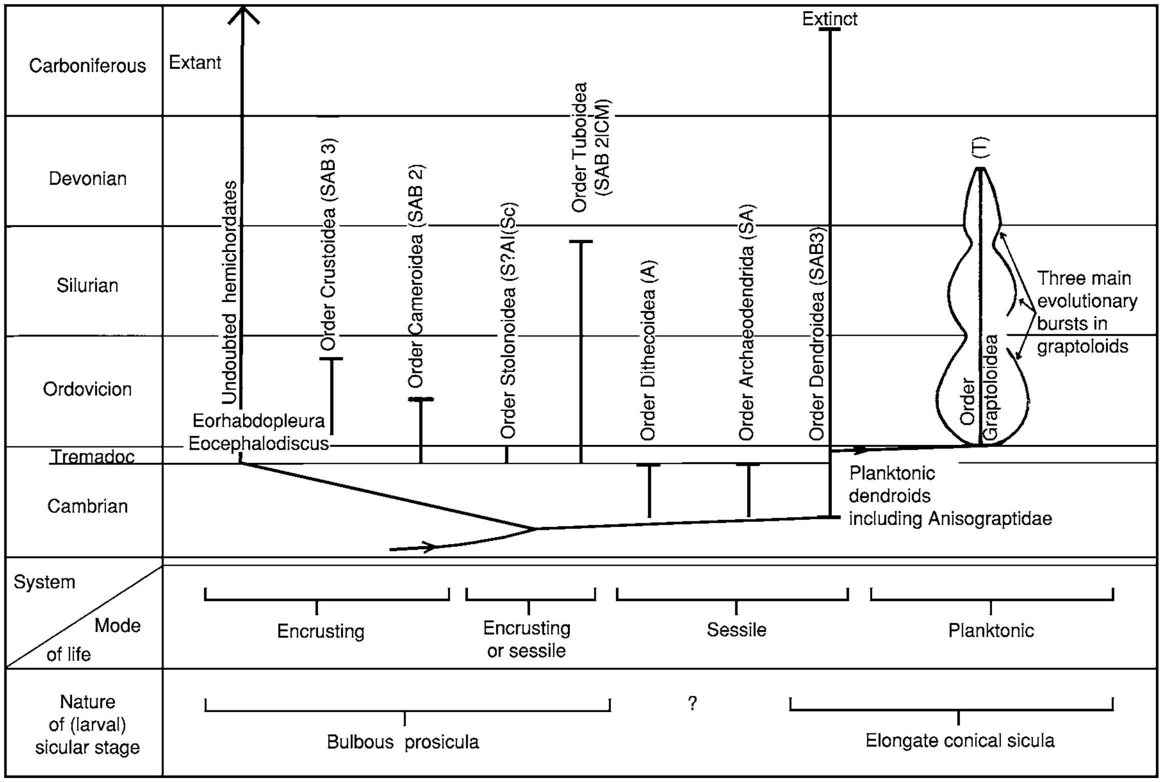


Figure 2 The evolutionary relationships of the graptolites to other early Hemichordata. The letters in parentheses following the ordinal name indicate briefly the defining characters of the order. Abbreviations are as follows: S, stolotheca; A, autotheca; B, bitheca; 2 and 3, diad and triad divisions of the stolon system, respectively; I, irregular stolons; C, conotheca; M, microthecae; I(Sc), irregular stolons, which may be partially sclerotized; T, theca.

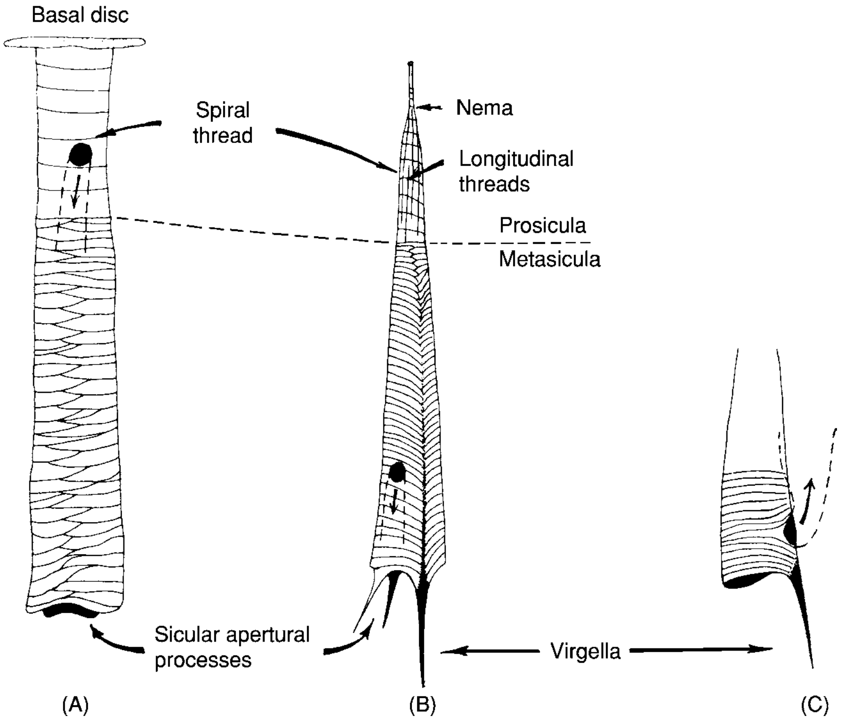


Figure 3 The sicula or larval stage, showing the resorption foramen, from which the first theca forms.

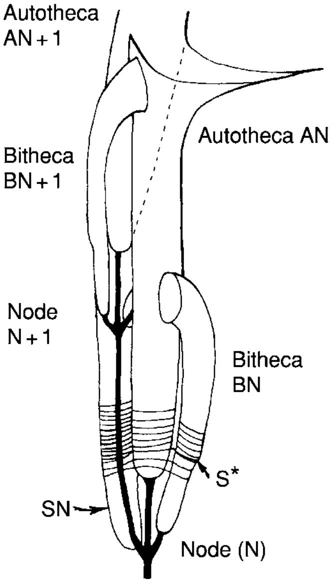


Figure 4 Schematic illustration of a dendroid, showing each node dividing into three stolons. The autotheca AN 1 is omitted for simplicity; bitheca BN would open into it, as bitheca BN + 1 opens into autotheca AN. S* marks the distal limit of growth lines that completely encompass tubes AN, BN, and SN. Distal to this, the tubes SN and BN have their own independent fuselli. Proximal to S* it is presumably the autothecal zooid in AN that secreted the fusellar periderm, because the growth lines below S* are continuous with those of AN but have an uncomfortable relationship with those of SN and BN. All tubes, whether developed at one node or at an earlier node, grow together, so that a growing end of the stipe just below the aperture of AN would have developing tubes, and hence zooids, of AN, BN + 1, AN + 1, and SN + 1.

and the individuals of the colony were probably connected by unsclerotized homologues of stolons.

The Structure of the Periderm

A modification of Kozlowski's classic interpretation is shown in Figure 7. The graptolite periderm is composed of two layers (Figure 8) – an inner fusellar layer constructed of fuselli in the form of half rings with a double zig-zag contact and an outer cortical layer comprising a bandage-like structure laid down by the zooids (Figure 9) after the tube has been built of fuselli. The bandages become more thinly layered towards the thecal apertures, at least during the early stages of thecal development. A number of graptolites have an inner, thinner, layer of cortex, which is also bandaged (Figure 10B–G, I, J).

Ultrastructure

In recent years the use of electron microscopes has increased our knowledge of detailed graptolite structure. Each fusellus has an envelope of largely granular sheet fabric within which is a sponge-like mesh of anastomosing fibrils (probably related to collagen) termed 'fusellar fabric'. Each cortical bandage, seen in transverse section in Figure 8, is also bounded by sheet fabric and comprises bundles of fibrils that are certainly fossil collagen. Within the sheet fabric are vesicular bodies, which may have been gas or

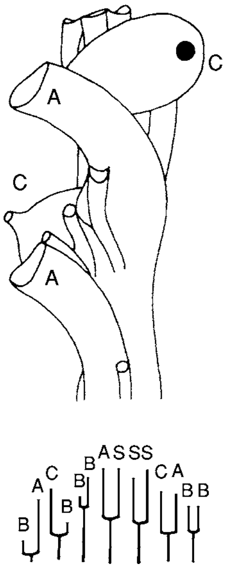


Figure 5 Schematic illustration of a tuboid, showing the division into two stolons and the tangle of tubes typical of some tuboids. The lower panel illustrates the diad stolon divisions that could give rise to such stripe complexity: A, autothecal stolon; B, bithecal stolon; S, stolothecal stolon; C, conothecal stolon. Note the great irregularity of stolon length compared with that of the dendroids (Figure 4).

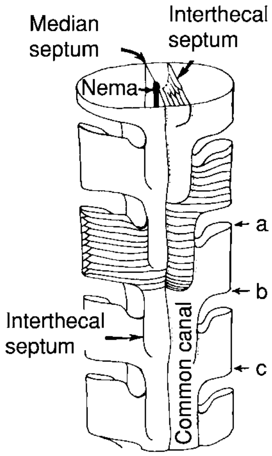


Figure 6 Schematic illustration of the graptoloid *Climacograptus*, showing the individuals of the colony connected by unsclerotized homologues of stolons. The thecal overlap is ab/ac. The distal end of the stipe is cut to show the internal structure, and growth lines are shown on two of the thecae.

fat bodies. The collagen fibrils are parallel to the long edges of the bandages (shown schematically in Figure 9) and in Figure 8 are shown cut transversely in layers (or bandages) 1, 2, and 4, and obliquely in layer 3. The oblique section of layer 3 shows the typical collagen annulations, seen here as short wavy bars, which in reality are ring-like thickenings around each fibril at specific intervals.

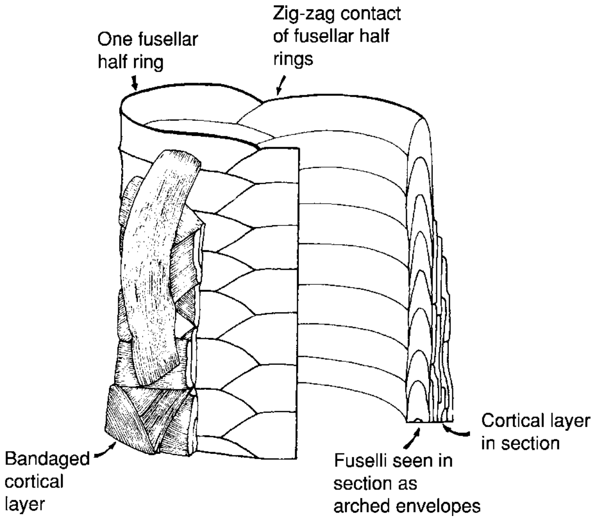


Figure 7 A modification of Kozlowski's classic interpretation of the structure of the periderm. Part of the thecal tube is cut away to show both the internal arch and overlap structure of the fuselli and the piles of bandages. The fibrous structure of the bandages is diagrammatic.

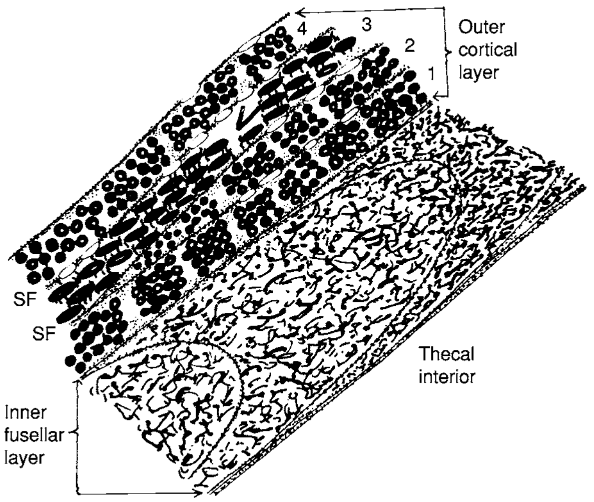


Figure 8 The graptolite periderm, showing the two layers. The minute fibrils in the outer cortical layer are parallel to the lengths of the bandages and there is a characteristic bundling of fibrils within each layer. Layer 2 shows the thinly tapering edge of one bandage. SF, sheet fabric; V, vesicular body; the magnification is about ×13 000.

Bandaging was almost certainly a strengthening feature and was thin or absent in forms that lived encrusted in sheltered environments, such as the insides of shells, or which needed to be light, in order to live in a planktonic environment. The graptolite zooid has only rarely been found preserved, and its structure has been deduced from living zooids known in other hemichordates. Figure 10 summarizes this similarity in transverse section, although there

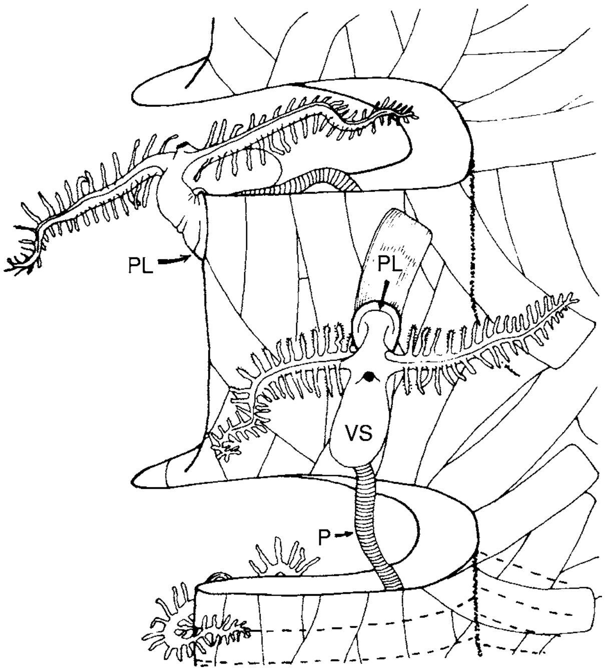


Figure 9 A zooid laying down bandages on the outside of a *Climacograptus* theca, using the pre oral lobe or disc (PL). The anus is shown as a small black foramen on top of the zooid. The mouth is underneath the visceral sack (VS), between it and the pre oral lobe. The ciliated tentacles have food gathering grooves on the upper surface, which lead down to the mouth. The zooids were probably connected by a soft pectocaulus (P). The positions of some fusellar growth lines are shown by the dashed lines, although these are not always visible if the bandaging is heavy.

are many other similarities between graptolites and living hemichordates.

Rhabdosomes and Synrhabdosomes

A complete rhabdosome of a *Monograptus* species (Figure 11A) consists of a single scandent stipe with one series of thecae (uniserial) growing upwards from the sicula and along the nema. Occasionally graptolite rhabdosomes are found in the association shown in Figure 11B, termed synrhabdosomes, which are preserved when the nemata remained fortuitously tangled. The reasons for such associations are unknown. The nemata could have been embedded in vacuolated (gas-filled) tissue and the unit as a whole could have assisted the buoyancy of the colonies. Alternatively, the association could have been related to sexual reproduction. In any event there must have been soft tissue in the centre of the synrhabdosome.

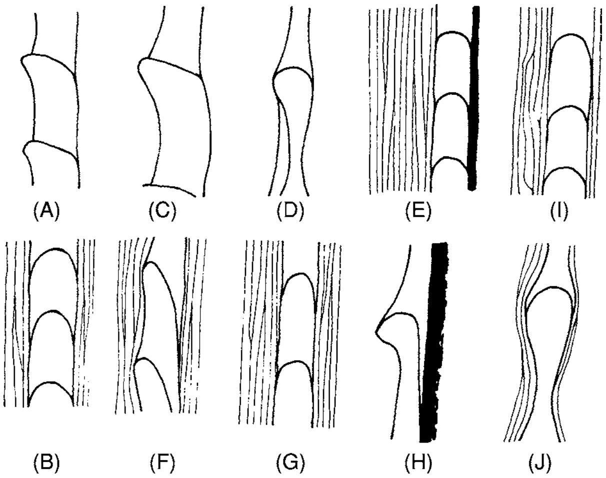


Figure 10 The similarity in transverse section between graptolites and living hemichordates. (A, B) Distal and proximal, respectively, fuselli and cortex of a living hemichordate. (C E) Distal, proximal, and support stem, respectively, fuselli and cortex of *Mastigograptus*, a dendroid approximately intermediate between graptolites and other hemichordates. (F) A tuboid (see Figure 5). (G) A dendroid. (H) A graptoblast (enigmatic encysted graptolite). (I) *Diplograptus*, showing scarped edges to the bandages. (J) *Monograptus*, showing a relatively thin cortex. Fusellar like fabric is shown in white and cortical fabric with a stipple.

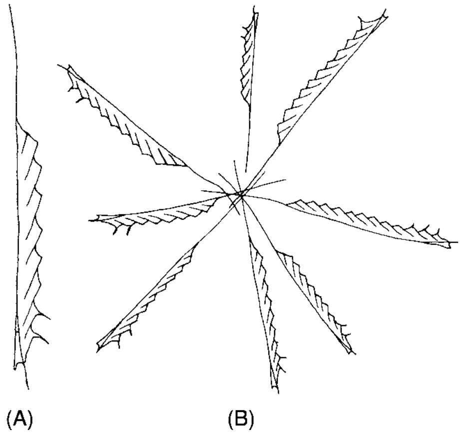


Figure 11 (A) A complete rhabdosome of a *Monograptus* species. (B) Occasionally, graptolite rhabdosomes are found in associations that are termed synrhabdosomes; these are preserved when the nemata remained fortuitously tangled.

Distinctions Between the Dendroidea and the Graptoloidea

The Dendroidea are characterized by two types of theca (autothecae and bithecae; Figure 12) linked by a black stolon system and a thick outer peridermal layer (cortex). Strengthening rods connect the numerous stipes, which each have several hundred thecae. A typical example of a dendroid is *Dictyonema* (Figures 13 and 14). In evolving from the Dendroidea, the Graptoloidea gradually lost the bithecae and stolon system, the result being colonies with one type of individual (simply termed 'thecae');

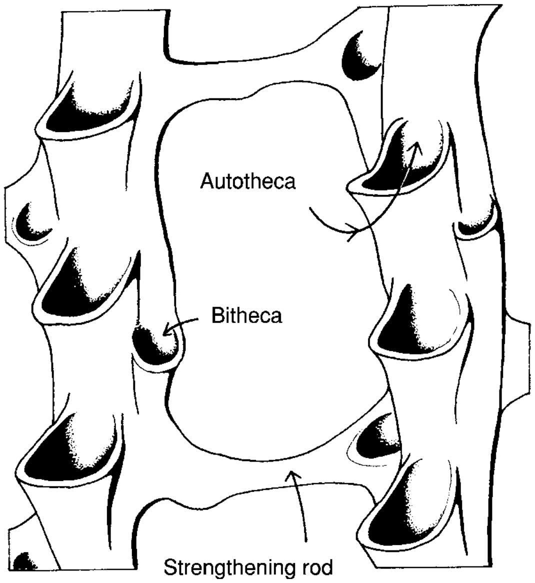


Figure 12 The structure of the Dendroidea, showing the two characteristic types of theca.

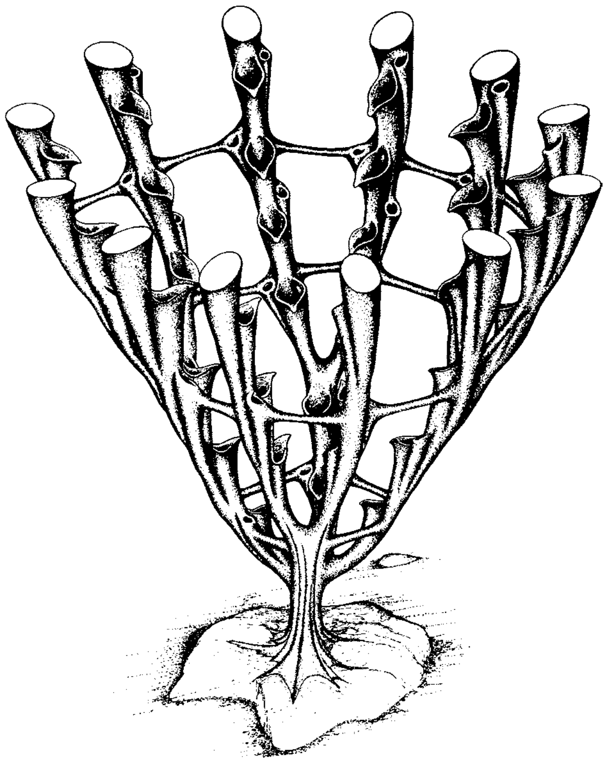


Figure 13 A typical dendroid, *Dictyonema*, which lived attached to the substrate by a basal holdfast.

the outer cortex became thinner and the colonies became lighter in weight; the stipes became fewer, but longer and more widely spread, without connecting rods; in the diplograptids the stipes came to lie

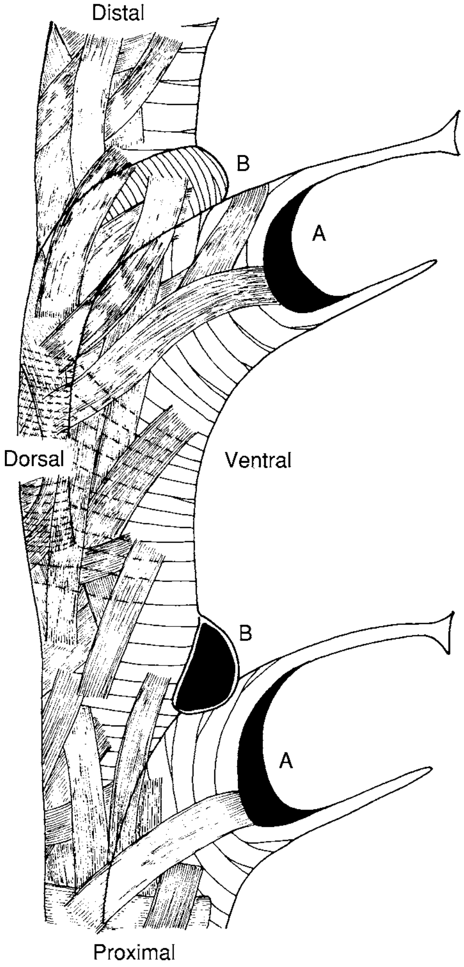


Figure 14 A section of the dendroid *Dictyonema rhinanthiforme*, showing two autothecae (A) and two bithecae (B) to illustrate the relationship of the two types of dendroid thecae. Magnification is approximately ×133. Fusellar growth lines and cortical bandages (Figure 7) are shown; the latter are less thickly deposited in the apertural regions of the autothecae. In most dendroids the cortical bandages are not easily seen, but the growth lines are commonly well preserved. The lower of the two bithecae grows around the far side of the stipe and opens in the angle between the two autothecae, facing the reader. The upper bitheca does exactly the opposite. The internal relationships of the bithecae and autothecae are shown in Figure 4. Some bandages are omitted near the ventral side of the stipe so that the fusellar growth lines can be shown. In actual specimens the growth lines can often be seen through the bandages. The positions of the fuselli are shown by dashed lines in the centre of the diagram.

back-to-back. *Rastrites* (Figure 15) is an extreme evolutionary product of *Monograptus* (Figure 16) in which the protheca is reduced to a fine thread-like tube and the metatheca is completely isolated from neighbouring thecae; there is thus no thecal overlap and no interthecal septum. The nema, virgella, and sicula are quite normal. The thecal apertures have hooks.

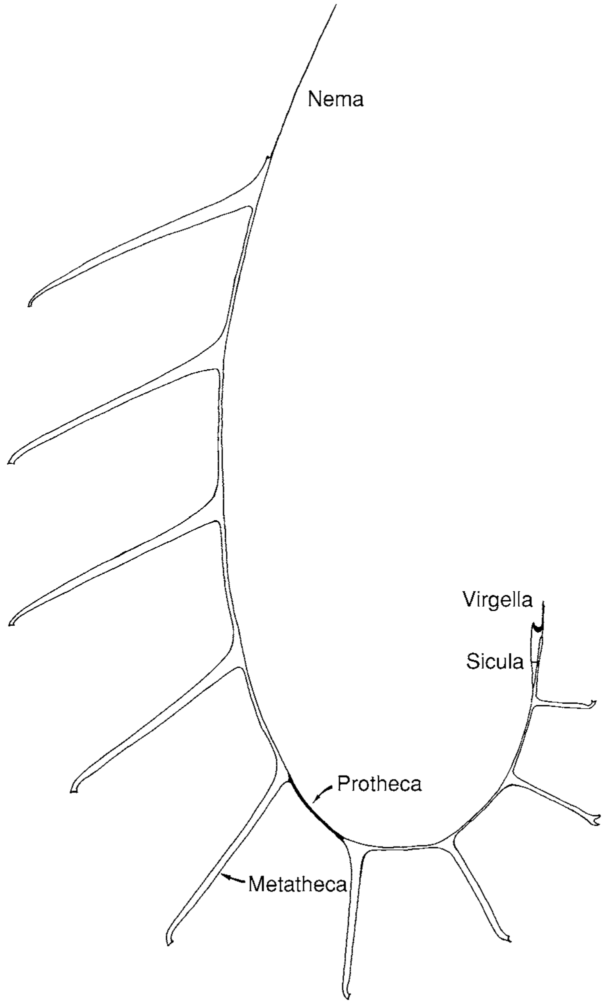


Figure 15 *Rastrites*, showing the protheca reduced to a fine thread like tube (shown in black on protheca 5) and the metatheca completely isolated from neighbouring thecae. The nema, virgella, and sicula are quite normal. The thecal apertures have hooks; theca 2 is deliberately twisted so that the two tubes at the apertures are visible; the remainder are shown in profile so that only one tube is easily visible.

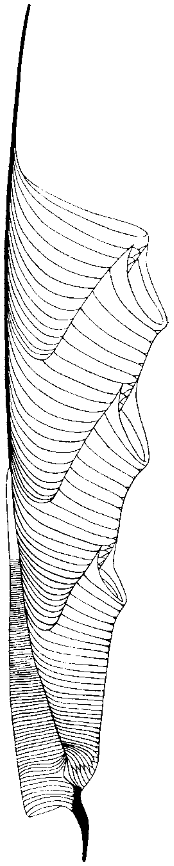


Figure 16 A typical graptoloid, *Monograptus*.

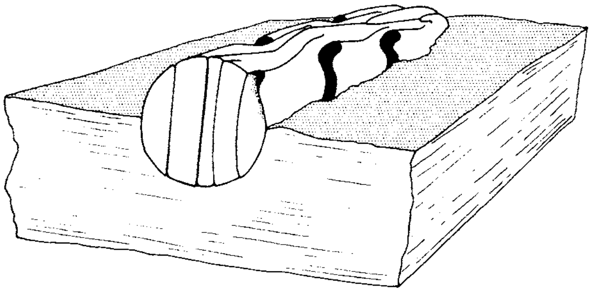


Figure 17 Preservation of a dendroid stipe in three dimensions.

Occurrence in the Rocks and Preservation

Graptolites usually occur in one of three ways in the rock record.

1. The periderm may be more or less intact and the stipes preserved in three dimensions (Figure 17). This form of preservation is most common in chemically precipitated limestones, clastic limestones, and sandstones. Dendroidea tend to be preserved most commonly in this way.
2. Much more common than (1) is the preservation of graptolites in three dimensions with an infilling of iron pyrite; the entombing rock is usually a black shale rich in organic debris. Such preservation probably occurred in oxygen-depleted conditions on the seafloor, where large amounts of decaying organic matter provided a suitable environment for bacteria that were capable of extracting iron from the seawater and precipitating it as iron pyrite.
3. Most commonly of all, graptolites may have been quickly flattened by the weight of sediment (Figure 18) such that they are now seen by the collector as films upon the bedding plane, usually white or silver on a dark background in a shale environment or as rather dark films in siltstones.

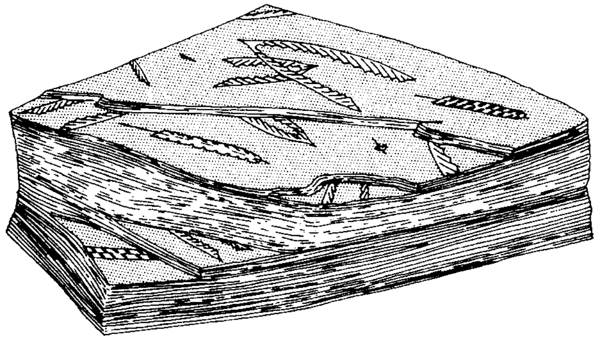


Figure 18 Graptolites flattened by the weight of the overlying sediment, to be preserved as films on the bedding plane. This is the most common mode of preservation.

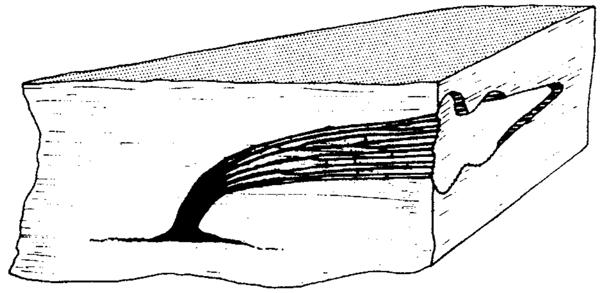


Figure 19 A dendroid colony preserved in growth position.

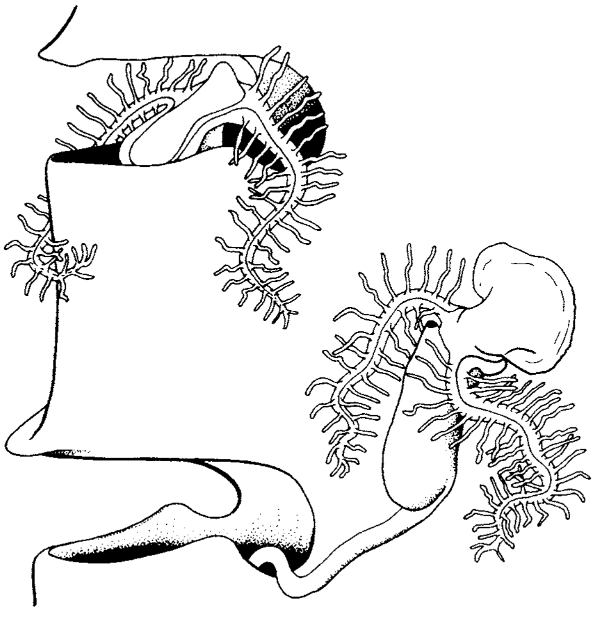


Figure 20 A hypothetical illustration depicting one zooid in the thecal tube and another extended from the thecal tube. This diagram is based upon an understanding of the detailed micro structure of the periderm or skeleton.

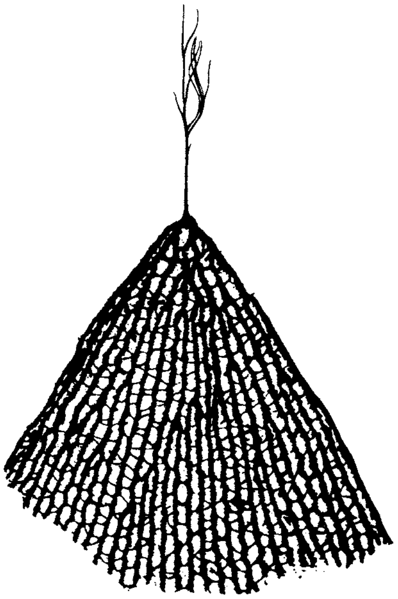


Figure 21 A planktonic species of *Dictyonema*, showing how it probably hung suspended in the surface layer of the ocean.

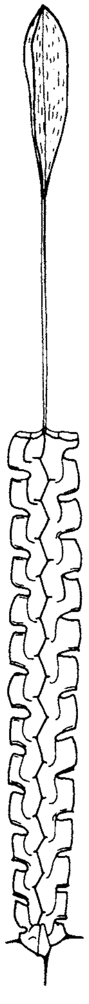


Figure 22 The diplograptid *Pseudoclimacograptus*, which probably hung from a dart like structure and rotated in turbulent currents.

All Graptoloidea, and the planktonic Dendroidea, occur lying flat on the bedding planes, reflecting the fact that they sank upon death from the overlying water. Sessile Dendroidea are occasionally found in positions of growth, with the basal discs or holdfasts downwards and the colony itself cross-cutting the bedding (Figure 19).

The Living Graptolite Colony

Preservation of graptolite soft parts is exceedingly rare, so that Figure 20 is largely hypothetical and is based on an understanding of the detailed microstructure of the periderm or skeleton. It is known, however, that the colonies themselves were either sessile or broadly planktonic. Figure 13 shows a species of the genus *Dictyonema* that had a basal disc or holdfast attached to the substrate of mud, sand, algal fronds, or rock or shell debris; such graptolites occurred in relatively shallow inshore environments along with a normal benthos of brachiopods, trilobites, and molluscs.

Figure 21 illustrates a planktonic species of *Dictyonema*, which probably hung suspended in the surface layer of the ocean from a termination of the nema consisting of a bundle of fibres enclosing soft tissue and tiny gas bubbles. The diplograptid *Pseudoclimacograptus* (Figure 22) was almost certainly pendant from a three-vaned dart-like structure that caused the colony to rotate in turbulent currents. Sessile graptolites have a limited geographical distribution, sometimes of only a few hundred miles, whereas the free-floating planktonic species often had worldwide distributions.

Broad Evolution of the Graptolites

Four successive, evolving, and overlapping faunas are known, leading from the Dendroidea (through the family Anisograptidae) to the Graptoloidea (Dichograptina, Diplograptina, and Monograptina). Many evolutionary lineages are known and some typical faunal elements are depicted in Figure 23. In the change from anisograptid to dichograptid we see a reduction in the number of stripes (and hence in the

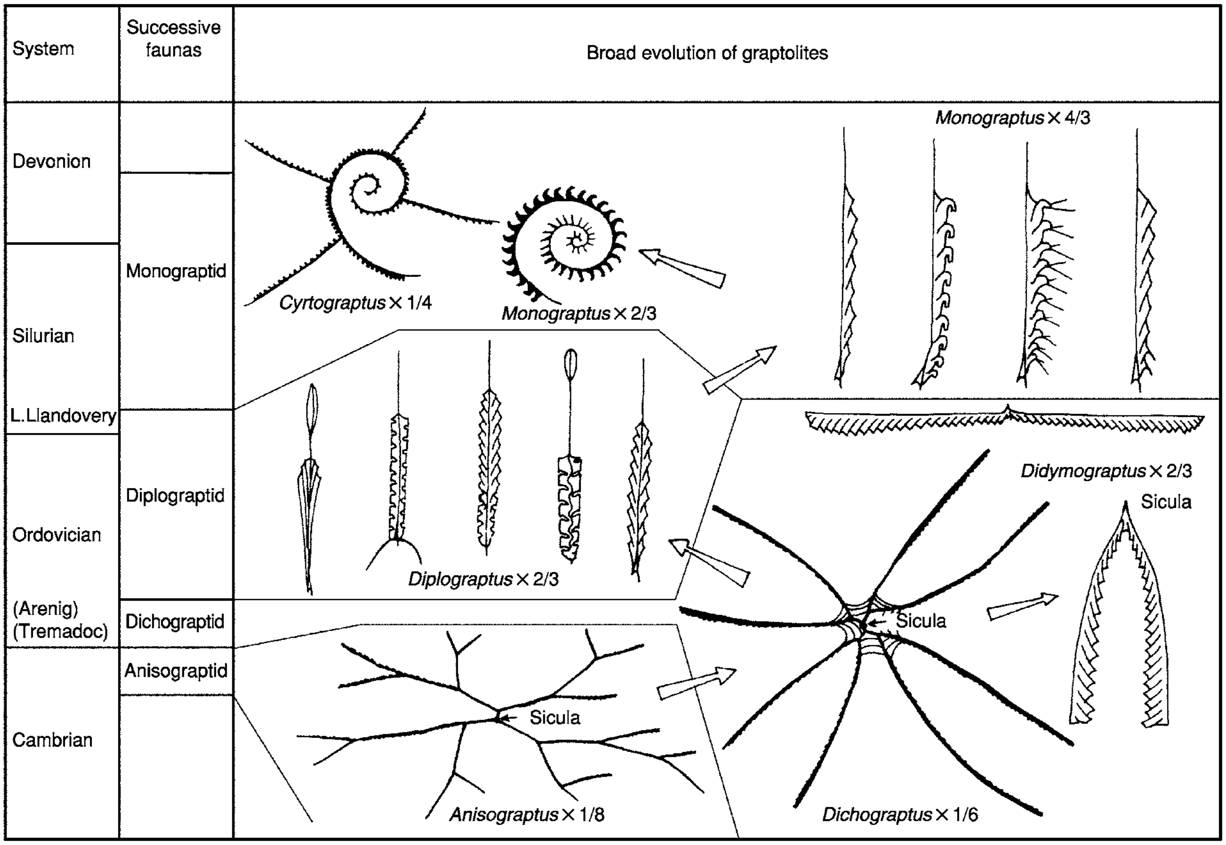


Figure 23 The broad evolution of the graptolites, from the anisograptids of the Upper Cambrian, through the dichograptids and diplograptids, to the monograptids of the Silurian and Devonian.

number of thecae), a reduction in the thickness of the cortical layer of the periderm, a loss of bithecae (typical of the dendroids; Figures 12 and 14), and the beginnings of a tendency towards a scandent position of the stipes (i.e. they tend to come upwards in later species, rather than hanging down as in earlier species). In the diplograptid fauna this last tendency reaches a peak with two stipes back-to-back, as in *Diplograptus* and *Climacograptus*, enclosing the nema for much of the rhabdosome length. In the monograptid fauna only one stipe remained in most forms, which was uniserial scandent (Figure 11A), although some did develop branches or become coiled (Figure 24).

Speciation in *Monograptus*

Further examples of speciation in the Monograptina are illustrated in Figure 24. All monograptids probably developed from a form with relatively simple thecae, which occurs first in the latest Ordovician. This row of simple thecae, uniserially arranged on a single scandent stipe, had the potential to develop thecal hooks, spines, enrolled tubes, and isolated thecae (Figure 15), whereas the rhabdosome as a whole had less potential, other than to curve, coil, or develop branches.

Stratigraphic Use of Graptolites

In the Ordovician and Silurian, and to a lesser extent in the Devonian, graptolites are perhaps the most important fossils for global correlation. In the Silurian, especially, they are particularly useful, and it has been calculated that the mean duration of a graptolite biozone is less than half a million years; in work done around the Wenlock–Ludlow boundary (mid-Silurian) the biozone duration is closer to one-quarter of a million years.

The reason for this great utility is that graptolites evolved very quickly yet often had a worldwide distribution in the tropical plankton realms. The mean duration of a graptolite species in the Silurian is under half a million years; only hominids are comparable.

Graptolites are used in several different ways in the relative dating of Lower Palaeozoic strata. The assemblage biozone depends for its recognition upon a unique association of species. The total-range biozone (acrozone) depends upon the known total temporal range of an individual species. Evolutionary lineages are also used where they are understood with some certainty: the change from one species to another occurred at a unique point in time. Finally, mathematical methods can be applied to the first and last appearances of species in different parts of the

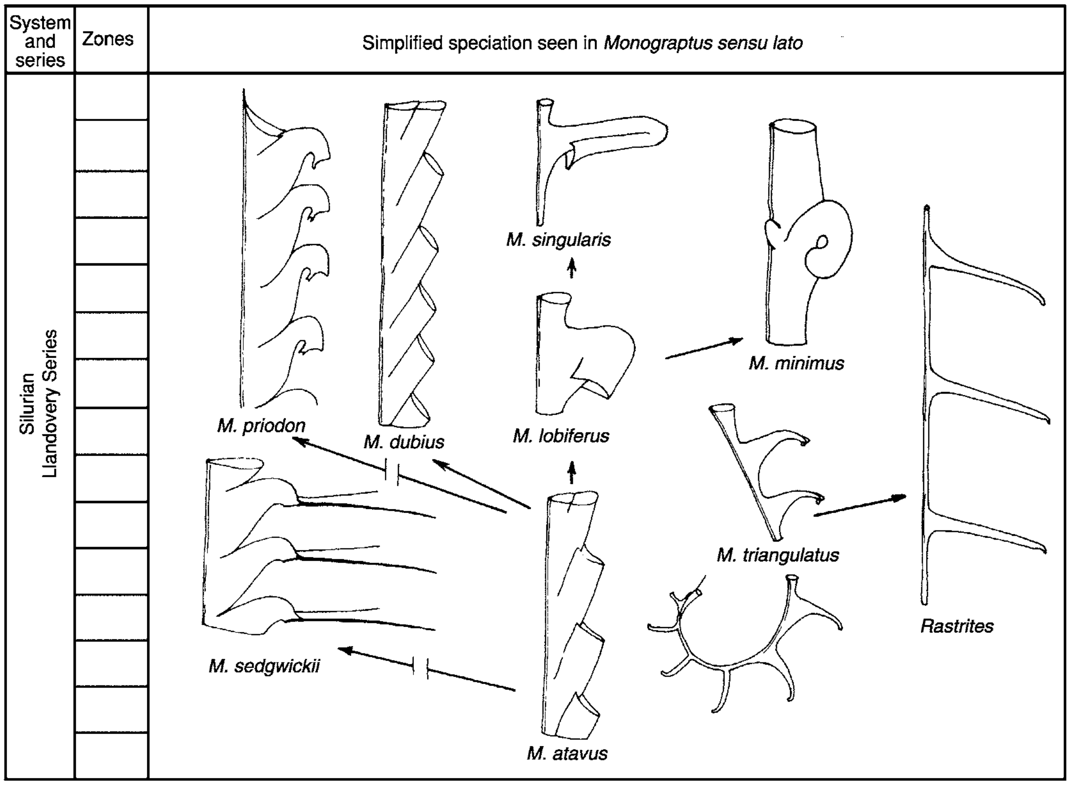


Figure 24 Examples of speciation in the Monograptina, showing very simplified sketches of some of the many known evolutionary lineages. For the purposes of simplicity, numerous morphological and stratigraphical intermediates have been omitted.

world. When these approaches are used in concert, or in conjunction with information from other groups, such as conodonts, acritarchs, and chitinozoans, an extremely powerful corrective tool results.

See Also

Biozones. **Fossil Invertebrates:** Bryozoans; Corals and Other Cnidaria. **Palaeozoic:** Ordovician; Silurian; Devonian.

Further Reading

Bulman OMB (1970) Graptolithina with sections on Enteropneusta and Pterobranchia. In: Teichert C (ed.) *Treatise on Invertebrate Paleontology,* Part V, 2nd edn, pp. 1 163. Geological Society of America and University of Kansas Press, Kansas, USA.

Elles GL and Wood EMR (1901 1918) *Monograph of British Graptolites*. Palaeontographical Society, London.

Fortey RA and Cooper RA (1986) A phylogenetic classification of the graptolids. *Palaeontology* 29: 631 654.

Kozlowski R (1949) Les graptolithes et quelques nouveaux groups d'animaux du Tremadoc de las Pologne. *Palaeon tooigca Polonica* 3: 1 235.

Palmer DC and Rickards RB (1991) *Graptolites: Writing in the Rocks*. Woodbridge: The Boydell Press.

Rickards RB (1979) Early evolution of graptolites and related groups. In: House MR (ed.) *The Origin of Major Inverte brate Groups,* pp. 425 441. Special volume 12. Systemat ics Association, London, New York, San Francisco.

Rickards RB (1995) Utility and precision of Silurian graptolite biozones. *Lethaia* 28: 129 137.

Ruedemann R (1947) *Graptolites of North America*. Memoir 19. Geological Society of America, Baltimore USA.

Urbanek A (1978) Significance of ultrastructural studies for graptolite research. *Acta Palaeontologica Polonica* 23: 595 629.

Molluscs Overview

N J Morris, The Natural History Museum, London, UK

The Mollusca is a phylum of invertebrate Metazoa that is semi-coelomate, protostomatous, and triploblastic. Molluscs inhabit marine, fresh water, and land habitats. The Mollusca are recorded from the beginning of the Cambrian Period, spreading into non-marine aqueous habitats during the latest Silurian or Early Devonian and apparently onto the land during the Carboniferous. Today they are only outnumbered, in terms of species, by the Arthropods (*see* **Fossil Invertebrates:** Arthropods).

In most modern classifications nine classes of molluscs are generally recognised, of which eight are extant. These are the Monoplacophora, the Cephalopoda (*see* **Fossil Invertebrates:** Ammonites; Cephalopods (Other Than Ammonites)), the Gastropoda (*see* **Fossil Invertebrates:** Gastropods), the Rostroconchia (now extinct), the Bivalvia (*see* **Fossil Invertebrates:** Bivalves) and the Scaphopoda making up the subphyllum Conchifera; and the Amphineura, the Solenogasters and the Caudofoveata making up the subphyllum Aculifera.

The molluscs are considered to be primitively bilaterally symmetrical, with an anterior head, a posterior anus, and a ventral foot. They have a mucus-producing skin which needs to avoid dessication. The dorsal surface of the body usually extends laterally to form an external cover (the mantle) which protects lateral pallial or mantle cavities. Primitively the mantle encloses laterally placed appendages such as the labial palps and the gills. The Mollusca need oxygen for their metabolism and usually use the blood pigment haemocyanin. However, haemoglobin is sometimes used, particularly in areas of the oceans where the oxygen concentration is low.

A characteristic feature of the Mollusca is the presence of a calcareous shell. The dorsal surface of the body and the mantle commonly secrete the shell, which has an outer organic layer, a periostracum, and an inner part made up of one or more layers consisting of complex microstructures of calcium carbonate with an organic matrix. The periostracum is secreted at the mantle margins except in the earliest growth stages and at times of shell damage when repairs may be facilitated by other parts of the mantle surface. The calcified shell is secreted across a space filled with extra-pallial fluid, by the dorsal surface and mantle margin on to the inner surface of the periostracum and or previous shell layers. The calcium carbonate of the shell may be in the form of crystals of aragonite or calcite but which are modified by their organic matrix.

The Mollusca primitively have an anterior mouth and a posterior anus with a gut that passes through the heart. They primitively have bilaterally paired kidneys and gonads. The gut is commonly coiled making it longer than the total body length in many living taxa, including the Monoplacophora and Amphineura. The gut is not coiled in the Scaphopoda and only slightly coiled in the Cephalopoda. Natural casts of stomachs and intestines are sometimes fossilised in detritus feeding taxa, notably in some examples of the nuculoid bivalves and trochid gastropods. In all of the living classes, except the Bivalvia, feeding is assisted by a tongue like apparatus (the radula) just within the mouth, which continuously produces rows of chitinous teeth. It is generally accepted that organisms ancestral to the bivalves did have a radula. Cephalopods had developed a horny beak (superficially parrot-like) which may also be preserved as a fossil, while certain opisthobrach gastropods developed chitinous gizzard plates which have also been preserved in Cenozoic sediments.

The soft tissues of the body of Mollusca are attached to the shell by muscles which in most Bivalvia, Amphineura, Monoplacophora, and Cephalopoda are arranged in a bilaterally symmetrical pattern. In the Amphineura and the Monoplacophora this musculature is arranged in eight serially arranged blocks (metameres), giving the body a repetitive structural pattern, known as pseudometamery. In the Amphineura the musculature of the anterior two metameres is attached to the head, while the poserior six metameres are attached to the more posterior part of the body (foot). In the tryblidioidean Monoplacophora, the head makes up the three anterior metameres. This pseudometamery differs from true metamery of the Annellida in not having a discrete coelom associated with each metamere. The attachment of the muscles to the shell is by way of a particular shell microstructure, low aragonitic prisms forming the myostracum which is virtually universal in the Conchifera but not obvious in the Amphineura.

Reproduction in the Mollusca occurs primitively by fertilization of gametes released into the sea. The fertilised ovum develops into a trochophore larva similar to those of the annellids. After a short period of time, sometimes two days, the larva settles and metamorphoses into a benthonic adult which develops a shell. The larvae of many groups may extend their life by the development of a shell (secreted by a shell gland) invariably consisting of aragonite and organic material. In the larvae many bivalves and gastropods, paired lateral processes (the velar lobes) help the growing animal to support itself in seawater and provide a certain amount of locomotion. Where this stage occurs it is known as the veliger larva. Any of these stages may be modified, often within the protection of an egg capsule, which is clearly of importance in the invasion of the land habitat. Most larval shells are external, but in the scaphopods and nuculoid bivalves, the larval shell develops within the expanded skin of the trochophore, and is known as the test cell larva.

The history of larval development in the Mollusca is not firmly established. During the nineteenth century larvae with a seven-part bilaterally symmetrical chitinous shell, resembling the early development of the Amphineura, were described for the Solenogasters. These observations need to be confirmed, but if correct might indicate that the Solenogasters and Caudofoveata are secondarily without a shell. This view is contrary to that of some modern workers that the non-shelled chitons are more primitive than those with a shell and are the living representatives of a ‘pre-shelled’ molluscan stock.

Living Cephalopoda emerge from their eggs as (free swimming) miniatures, resembling their adults.

In many bivalves and gastropods, extension of larval life is achieved by increasing the size of the larval shell by growth at the shell margins, which often produces daily growth lines. This extension of larval life enables some achitectonicid gastropods to cross the Atlantic Ocean and may have allowed some Mesozoic inoceramid bivalves to exist in the plankton for more than sixty days, a feature coincident with the almost global distribution of some of their species.

The periostracum consists of proteinaceous material known as conchyolin and may have as many as three layers. It may be smooth or hairy and can include calcareous spicules which are sometimes fossilised. There are up to three layers of calcareous shell. The primitive conchiferan shell is considered to be made up of an external layer of aragonite prisms together with an internal layer of nacre. This shell microstructure is known to occur in the types of bivalves, gastropods, cephalopods, and monoplacophora which have other features thought to be primitive. This view is supported by evidence of shell microstructures preserved from the Ordovician. However, it has been suggested that some shells from the early Cambrian, whose inner surface structure has apparently been replicated by phosphatic material, may have consisted of calcite.

Cross-lamellar aragonite and complex cross-lamellar aragonite may develop in any of the layers and at least one of these microstructures occurs in some scaphopods, gastropods and bivalves.

‘Amorphous aragonite’, probably very fine-grained cross-lamellar structure, is present in some bivalves.

Outer calcite layers have evolved in some epifaunal benthonic molluscs, including some bivalves (Ostreoida, Pterioida, etc.) and gastropods (Euomphalidae, Patellidae, Buccinidae). This development may be related to the greater resistance of calcite to solution in seawater. In the oysters, bivalves in which all the shell layers have become calcitic, the muscle attachment layers (myostracum) alone remain as typical aragonitic prisms which are joined to the muscular tissue by microvilli on the basal cell membrane.

Amphineura have a dorsal integument of conchiolin, known as the girdle, which is sometimes pustulose and rarely with calcified pustules. The shell consists of eight imbricating plates, symmetrically arranged in sequence along the centre of the girdle. The microstructure of the calcareous shellis aragonitic, and has been likened to the cross-lamellar structure of the conchifera, but the general pattern of the structure differs and is traversed by tubules which house the sensory aesthetes. The muscle insertions closely resemble those of the conchifera but they seem to have no substantial myostracum. The occurrence of fossil chitons is sporadic but their record is good back to the Early Ordovician. A number of probable chitons have been described from the earliest Cambrian (Meisuchun Stage) from Southern China.

A number of fossils of uncertain affinities may belong to the Mollusca. The hyolithids have comparable shell microstructure to some Mollusca but their appendages (helens) and the arrangement of the gut, as occasionally preserved, are quite different. Some Machaeridia may be forms intermediate between shelled and non-shelled chitons, as may be the Cambrian Halkeriids and *Wiwaxia*.

See Also

Fossil Invertebrates: Arthropods; Bivalves; Gastropods; Cephalopods (Other Than Ammonites); Ammonites.

Further Reading

Fretter V and Graham A (1994) *British prosobranch mol luscs: their functional anatomy and ecology.* (Revised Edition), vol. 144. London: Ray Society.

Moore RC (ed.) (1952 1996) *Treatise on Invertebrate Paleontology.* Parts I,K,L,N, Mollusca. 7 Vols. University of Kansas Press, Lawrence: Geological Society of America.

Natural History Museum (London) (1996) *British Palaeo zoic fossils,* edn. 4. London: HMSO.

Natural History Museum (London) (1993) *British Meso zoic fossils,* edn. 6. London: HMSO.

Natural History Museum (London) (1993) *British Caeno zoic fossils. Tertiary and Quaternary,* edn. 5. London: HMSO.

Pojeta J and Runnegar B (1976) *The paleontology of ros troconch mollusks and the early history of the phylum Mollusca.* Washington, DC: United States Geological Survey Professional Paper No. 968.

Stanley SM (1970) *Relation of shell form to life habits of the Bivalvia (Mollusca).* Boulder, Colorado: Geological Society of America. Memoir 125.

Wenz W and Zilch A (1939 1962) Gastropoda. In: Schindewolf OH (ed.) *Handbuch der Paläozoologie.* Berlin: Borntraeger.

Wilbur KM and Yonge CM (1964, 1966) *Physiology of mollusca.* 3 Vols. New York, London: Academic Press.

Bivalves

E M Harper, University of Cambridge, Cambridge, UK

Introduction

The Bivalvia, the second largest class within the phylum Mollusca, is one of the most familiar of all invertebrate taxa. Modern representatives, such as mussels, cockles, oysters, and scallops, are well known from excursions to the coast, and in many parts of the world they are important commercial species. Their generally excellent fossil record has allowed their evolutionary history to be traced back to the Early Palaeozoic and, for much of this time, they have been important components of many faunas. From rather modest beginnings, they have conquered a range of habitats from the deep sea to freshwater, exploited a wide range of life habits (from deep burrowing to swimming), and undergone a near-exponential taxonomic proliferation, a spectacular example of an adaptive radiation.

Bivalves come in all manner of shapes and sizes, from tiny, thin-shelled commensals that live in association with sea anemones, to giant clams and the extinct rudists and inoceramids which reach(ed) sizes well over 1 m. Shell morphology is extremely plastic, but all are modifications of the same basic theme. The intimacy of the shell morphology to life habit has been a great benefit in reconstructing the life habits of extinct bivalves, but has also frustrated many attempts to establish the relationships between different groups within the class. Bivalves have been proven to be good palaeoenvironmental indicators, but they have only limited use in biostratigraphy. Freshwater 'mussels' have been used to date fluvial deposits in the Carboniferous Coal Measures of Western Europe, and inoceramids have been used for Late Cretaceous deep marine settings (e.g., in New Zealand). In general, however, species of the class are too long lived and too facies specific to be of any great value.

General Morphology

As the name implies, bivalves comprise two calcareous valves. These are arranged laterally (left and right), are joined dorsally by a partially calcified elastic ligament, and enclose the soft tissue. Each valve has clearly differentiated posterior and anterior features, i.e., inequilateral. The primitive arrangement, retained by most bivalves, was to have a plane of symmetry parallel to the commissure (the join between the two valves), resulting in valves which are mirror images of one another (i.e., equivalve). Although this symmetry is found in virtually all bivalves which live with the commissural valve perpendicular to the substrate surface (orthothetic), it has been lost in those which have adopted a pleurothetic habit where they lie on one valve (e.g., oysters, scallops). In these cases, there is a tendency for the two valves to become dissimilar (i.e., inequivalve), typically with the underlying valve becoming more bowl-like and the 'upper' one more reduced like a lid.

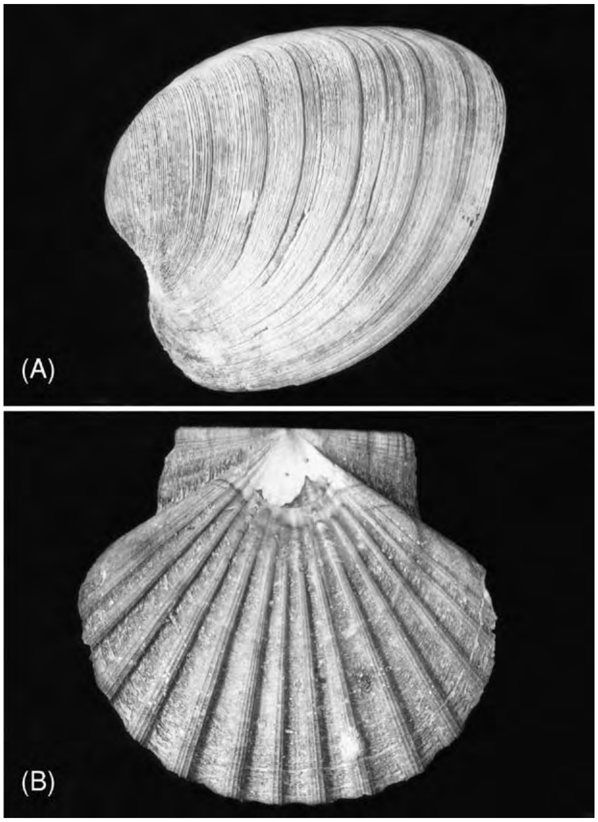


Figure 1 (A) *Mercenaria mercenaria*, a shallow burrowing bivalve from the Pliocene of Florida. Note the prominent annual growth bands. (B) *Pecten maxima*, a free living epifaunal scallop from the Holocene Atlantic.

Shell Morphology

All bivalves possess a pair of shells which may be shut to provide protection from both environmental stresses (e.g., desiccation in the intertidal habitat) and the threat of predation. Most shells are reasonably robust, which has provided the class with a generally excellent fossil record. Although shell morphology in bivalves is very variable and intimately linked to their life habits (see below), all shells are simple modifications of the basic shell secretion model used by all shelled molluscs. The shell is secreted by the mantle lobes and grows by marginal accretion, as evidenced by the growth lines on the surface of the valve (Figures 1A and 2A). These growth lines are particularly marked in bivalves from intertidal and shallow temperate habitats, where the animals experience pronounced seasonality and largely stop growing during the winter months. Bivalves which experience more equable conditions do not show such obvious or regular patterns. Inspection of the growth lines in sectioned valves shows that, although most shell material is added ventrally, the shell is also thickened during growth (Figure 2B), demonstrating that the entire mantle surface is responsible for adding material. The outermost part of the shell is an organic layer called the periostracum secreted at the mantle edge (Figure 2C). The thickness of the periostracum varies between taxa, from less than 1 μm in oysters and scallops to several hundred micrometres in some mussels. In many cases, the periostracum is lost by abrasion and decay during the life of the animal, particularly on the older parts of the shell, and there is no real prospect of it being preserved in any but the most exceptional circumstances. The primary function of the periostracum is to act as the template on which the calcareous part of the shell is deposited, but it may also provide protection from both corrosive waters and predators that dissolve the shell. It is particularly noticeable that freshwater bivalves have very thick periostraca.

The main part of the shell, however, is calcareous. It is in effect a ceramic made up of calcium carbonate crystals in an organic matrix (the latter accounting for <5% of the dry weight of the shell). The proteinaceous matrix controls both the polymorph of calcium carbonate used and the arrangement of the crystals. All bivalves contain aragonite in their shells and the vast majority are wholly so. Some taxa, however, chiefly those exploiting epifaunal life habits, also secrete calcite in their outer layers. The oysters have taken this to its extreme and the bulk of the shell is

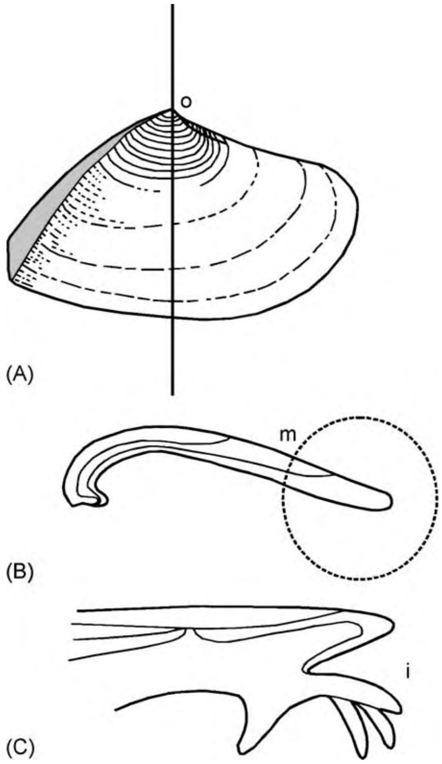


Figure 2 Details of shell formation in a generalized bivalve. (A) Marked comarginal growth lines on the shell surface. (B) Section through the shell along the line indicated in (A) showing the arrangement of growth lines within the shell. (C) The relationship between the shell and the underlying mantle edge (close up details of the circled area in (B)). i, inner mantle fold; m, middle mantle fold; o, outer mantle fold.

calcitic, with aragonite being confined to the sites of muscle attachment and the ligament.

Molluscan shell is immensely strong, in fact often much stronger than vertebrate bone. There are a number of different microstructures (Figure 3), each with different mechanical properties, and most shells are made up of two or three arranged in different layers. Different taxa show different arrangements and these are considered to be of phylogenetic significance. It is apparent that the earliest bivalves were wholly aragonitic and chiefly composed of nacre (Figure 3A), and that subsequent evolution has produced the wide array of microstructural arrangements seen today. The effect of differing crystal sizes, amount of organic material, and polymorph used has affected the preservation potential of different taxa; many of the Palaeozoic and Mesozoic taxa that were originally aragonitic are either preserved as internal moulds or are replaced by calcite.

Details of the internal features of the shell are shown in Figure 4. The hinge plate is situated dorsally and houses the ligament and teeth. The ligament is an elastic, partially calcified layer that provides a very energy-efficient opening mechanism. During valve closure, energy is stored in the ligament as it is flexed by the contracted adductor muscle(s) (Figure 4C). When the muscle is relaxed, the ligament springs the valves apart causing them to gape. This passive valve opening mechanism is the reason why many fossil bivalves are found in a disarticulated state. Although the ligament itself is seldom preserved, its position may be inferred from the presence of the ligament pits in which it is anchored (Figures 4 and 5). Most bivalves have teeth on the hinge plate which fit into corresponding sockets on the opposite valve and function to keep the valves in perfect alignment. Both ligamenture and dentition vary markedly amongst higher taxa of bivalves, and both are often used as informative characters in establishing phylogenies. Some of the range of hinge plate architectures is shown in Figure 5.

A number of attachment scars mark the locations where muscles are anchored to the shell. The most significant of these are the adductor scars (Figures 4A and 4B). If the adductor scars are paired (i.e., dimyarian), they occur posteriorly and anteriorly. If an animal is monomyarian, the single muscle (the posterior) occupies a more central position. In many taxa, there is a thin pallial line running around the shell a small distance from the ventral edges that marks the attachment of the mantle to the shell. In infaunal taxa, where the posterior mantle has been fused and elongated to form siphons, the pallial line is inflected forming the pallial sinus. The sinus represents the space into which the siphons are withdrawn when the valves are shut. Other muscle attachment scars may be more or less apparent, including the insertions of the pedal musculature (particularly in burrowers and byssate taxa).

Soft Part Anatomy

Bivalves are laterally compressed and, unlike most molluscs, there is no head or radula. The internal organs are enclosed by the two mantle lobes that are joined dorsally (Figure 4C). The chief function of the mantle is to secrete the shells, but the ventral edges of each mantle lobe are differentiated into three folds (Figure 2C), only the outermost of which is directly concerned with shell manufacture. The innermost fold controls water flow into and out of the mantle cavity, whilst the middle fold has sensory capability. In several bivalve groups (such as scallops), the middle fold is well developed with tentacles and eyes. In some taxa, the mantle is extended posteriorly and fused to form a pair of siphons through which water is directed into (inhalant) and out of (exhalant) the mantle cavity.

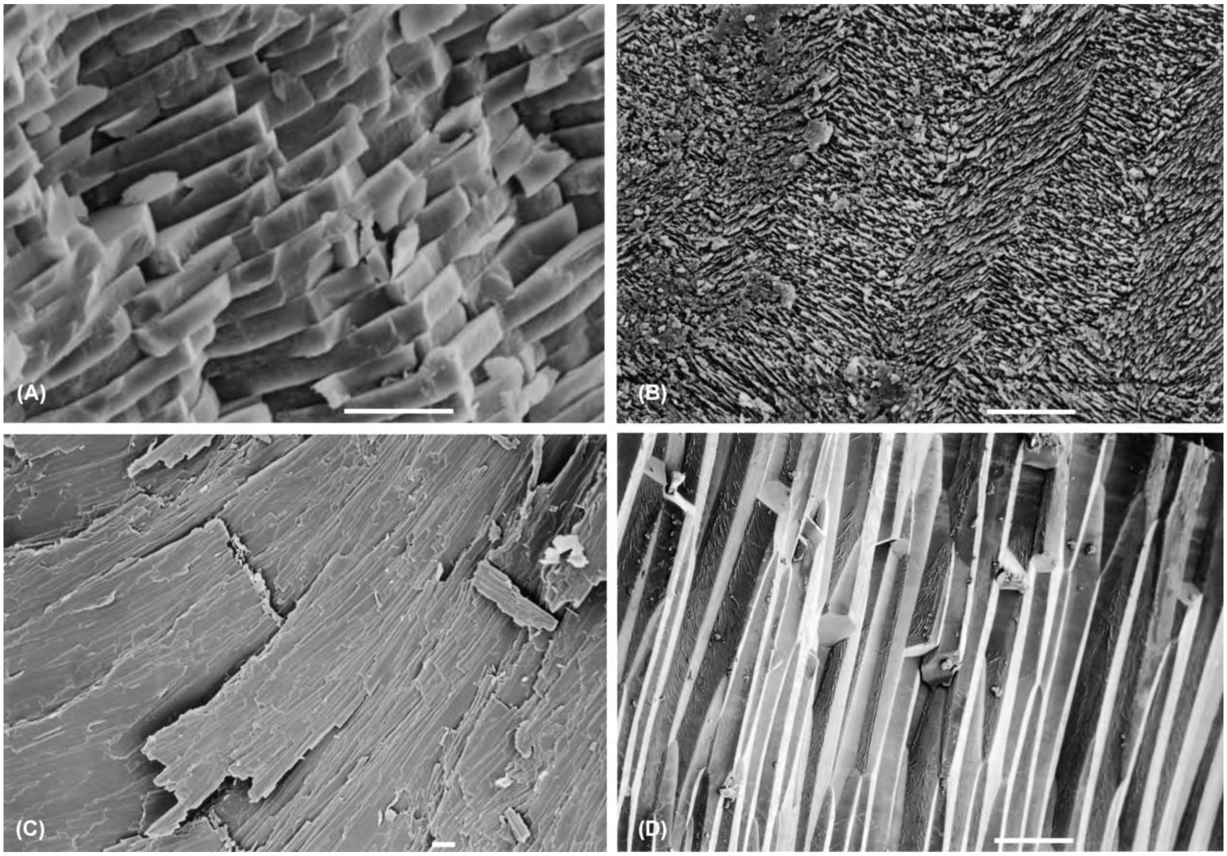


Figure 3 Scanning electron micrographs of four of the most common bivalve shell microstructures. (A) Aragonitic nacre (from the inner shell layer of *Pinna nobilis*). (B) Aragonitic crossed lamellar structure (from the outer layers of *Corbula gibba*). (C) Foliated calcite (from *Ostrea edulis*). (D) Calcitic prisms (from the outer shell layer of *Pinna nobilis*). Scale bars for (A) (C) represent 10 μm and for (D) represents 100 μm.

The mantle cavity is spanned by one or two adductor muscles that attach to the shell and act antagonistically with the ligament to close the valves on contraction (Figure 4C). A significant part of the mantle cavity is occupied by a pair of gills (the ctenidia) lying on either side of the rest of the viscera. In most bivalves, the gills are involved with both respiration and ciliary suspension feeding (filtering small particles out of the water which are then transferred to the mouth by a pair of labial palps). Recent bivalves show a number of different gill morphologies depending largely on the feeding process employed. Deposit feeders, e.g., *Nucula*, have less well-developed (protobranch) gills, whilst members of the Lucinidae augment their filter feeding by energy gained from the activities of sulphide-oxidizing chemosymbiotic bacteria living within the modified gills. The carnivorous septibranch bivalves (e.g., *Cuspidaria*, *Poromya*) suck in small prey (such as amphipods) using their modified siphons. These extraordinary bivalves have 'lost' their gills and respire over the inner surface of the mantle. Other significant organs within the mantle cavity include the gut, heart, circulatory system, and the foot. The gut runs between the anteriorly positioned mouth and the posterior anus, and includes a complex stomach which, again, has a number of configurations depending on the feeding biology of the animal. Blood is circulated throughout the animal by a three-chambered heart. A muscular foot is present in all juvenile and most adult bivalves and occupies the centre of the mantle cavity.

Naturally, the soft part anatomy of bivalves is very seldom preserved, although preservation of gill and muscle material has been reported in exceptional circumstances. Various details, however, can be inferred from the study of the internal surface of the shells. Apart from adductor muscle scars, the practised eye may pick out the attachment points of more minor muscles and impressions of radial muscles and blood vessels within the mantle.

Ecology

Modern bivalves exploit a wide range of life habits. Many burrow to varying depths within soft sediments,

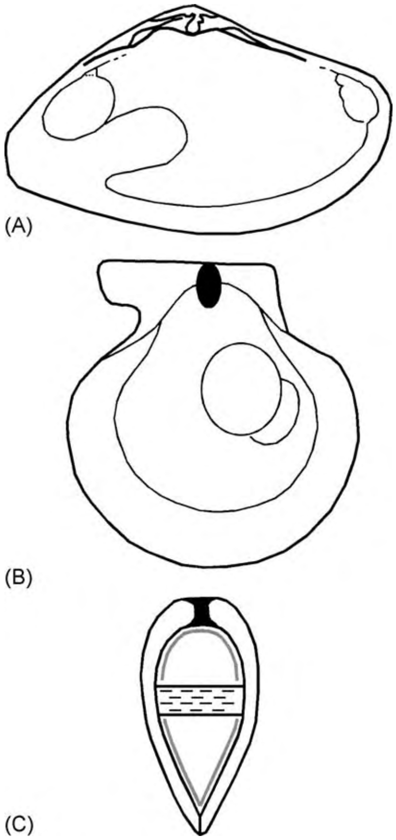


Figure 4 Various aspects of the internal morphology of bivalve shells. (A) The left valve of a generalized burrowing dimyarian. (B) The right valve of a generalized byssate scallop. (C) The relationship between the two valves and their associated mantle lobes, adductor musculature and ligament.

but others attach to or bore into hard substrates by a variety of means; others have become free living, some with the ability to swim. Most bivalves are marine, exploiting niches from the intertidal zone down into the abyssal depths, but successful groups (including the oysters) have invaded more brackish conditions and even freshwater, where modern unionid mussels cause enormous damage as biofoulers. It is clear that the most primitive bivalves were marine shallow burrowers and that other life styles evolved later. It is also apparent that many of the more specialized life habits have evolved separately in a number of different lineages (i.e., polyphyletically). Seminal work by S. M. Stanley firmly established how different aspects of the morphology of living bivalves could be related to their life habits, such that it is possible to use these characteristics of extinct taxa to reconstruct the life habits of fossil groups.

Burrowing

A large proportion of all bivalves (around 50% of all modern families) burrow into soft sediments using the foot. Most are equivalve and are isomyarian (i.e., the posterior and anterior adductor muscles are of equal size). The depth to which they burrow varies between taxa, from those which lie just under the surface with the edge of the shell virtually level with the sediment–water interface (e.g., *Cerastoderma*; Figure 6C), to depths of several centimetres (e.g., *Mya*; Figure 6B), with *Panopea* reaching spectacular depths of up to 1 m. The key to successful burrowing is maintaining contact with the seawater in order to continue both feeding and respiration. This is achieved by the siphons, snorkel-like extensions of the posterior mantle. The length of the siphons, and therefore the depth of burrowing, can be inferred from the shells by the size of the pallial sinus; deeper burrowers have more indented pallial sinuses, whereas very shallow burrowers have no sinus at all (see Figure 6). Very deep burrowers, such as *Mya*, have siphons so long that they are unable to withdraw them fully into the shell when it shuts, and have a permanent posterior siphonal gape through which they protrude. Shallow burrowers generally have strong, robust shells, often with a pronounced radial or concentric ornament that may assist the burrowing process or help the animal remain ‘locked’ into the sediment. Deeper burrowers tend to have thinner shells and are often smooth shelled. Although the foot is never preserved, its presence may be inferred from the pedal musculature on the inside of the valves and, in cases where the animal is a rapid and deep burrower (such as the razor shell *Ensis*), the foot may be so well developed as to require an anterior pedal gape.

It is clear from studies of the siphons of living bivalves that they are constructed in a number of different ways, suggesting that the deep burrowing habit has evolved independently in several clades.

Attachment

Almost all larval bivalves attach to the substrate, if only briefly, with tanned protein threads (the byssus) secreted by a gland at the base of the foot. In a large number of taxa, this habit has been neotenously retained into adulthood, and again it is clear that this has happened repeatedly in different groups. Byssate bivalves fall into two categories: those like *Pinna* (Figure 6A) and *Modiolus* that are orthothetic and live attached to clasts within the sediment in which they are partially buried (endobyssate), and those that are attached to the surface of hard substrates (epibyssate), either in an orthothetic (e.g., *Mytilus*; Figure 7D) or a pleurothetic (e.g., *Isognomon ephippium*; Figure 7C) orientation. Orthothetic byssate bivalves tend to be equivalve and have much reduced anteriors. This anterior reduction is reflected

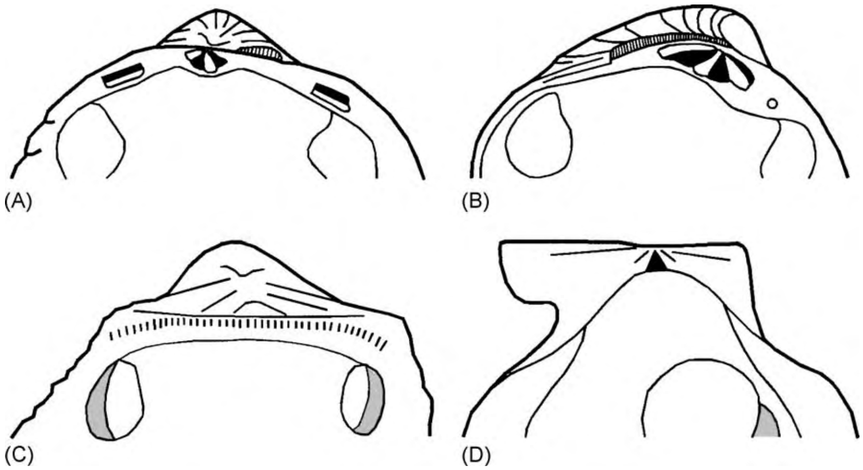


Figure 5 Selected hinge plates showing some of the variety of ligament insertion and arrangement of teeth. (A) *Cerastoderma edule*: heterodont dentition with two centrally placed cardinal teeth and two lateral teeth. (B) *Venus casina*: heterodont (similar to *Cerastoderma* but with no lateral teeth). (C) *Arca tetragona*: taxodont dentition with numerous teeth arranged in a row; the ligament forms a chevron pattern on the broad triangular area below the umbones. (D) *Chlamys varia*: two simple teeth with the internal ligament occupying a triangular pit below the umbones.

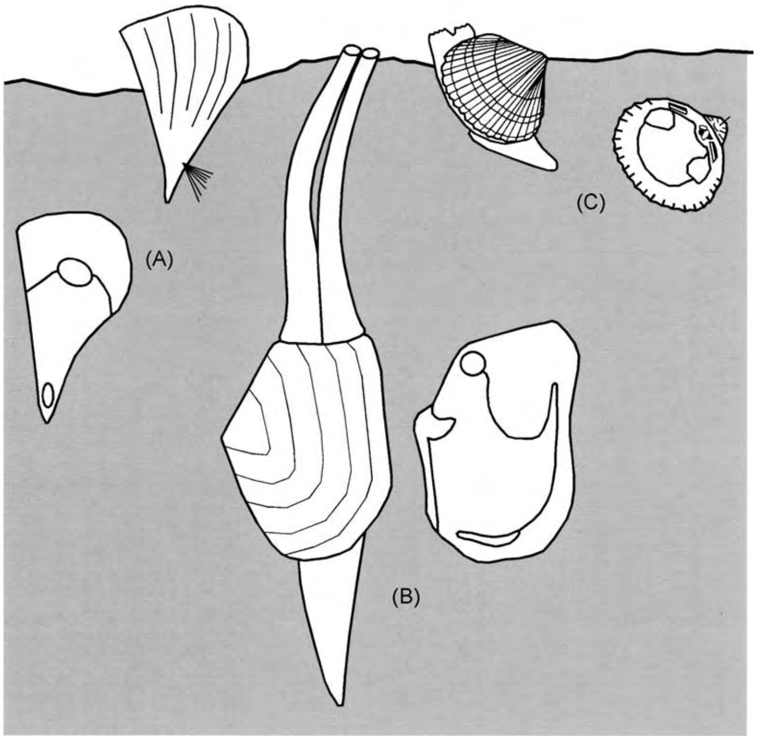


Figure 6 Morphology and mode of life of bivalves living in or partially within soft substrates. (A) *Pinna nobilis*. Not to scale. (B) *Mya truncata*. (C) *Cerastoderma edula*.

in the adductor musculature, which (although still dimyarian) is heteromyarian, with the anterior adductor much smaller than the posterior (Figure 7D). Pleurothetic byssate bivalves are often markedly inequivalve, with the 'lower' valve (which in the majority of cases is the right) often larger than the other. Although they are dimyarian early in ontogeny, the 'adults' are monomyarian, having lost the anterior muscle during ontogeny; the remaining posterior muscle is often large and centrally placed (Figure 7C). The presence of a byssus may be inferred from either a slight gape between the valves through which it passes (the byssal gape), or more obviously the byssal notch in scallops (Figure 4B).

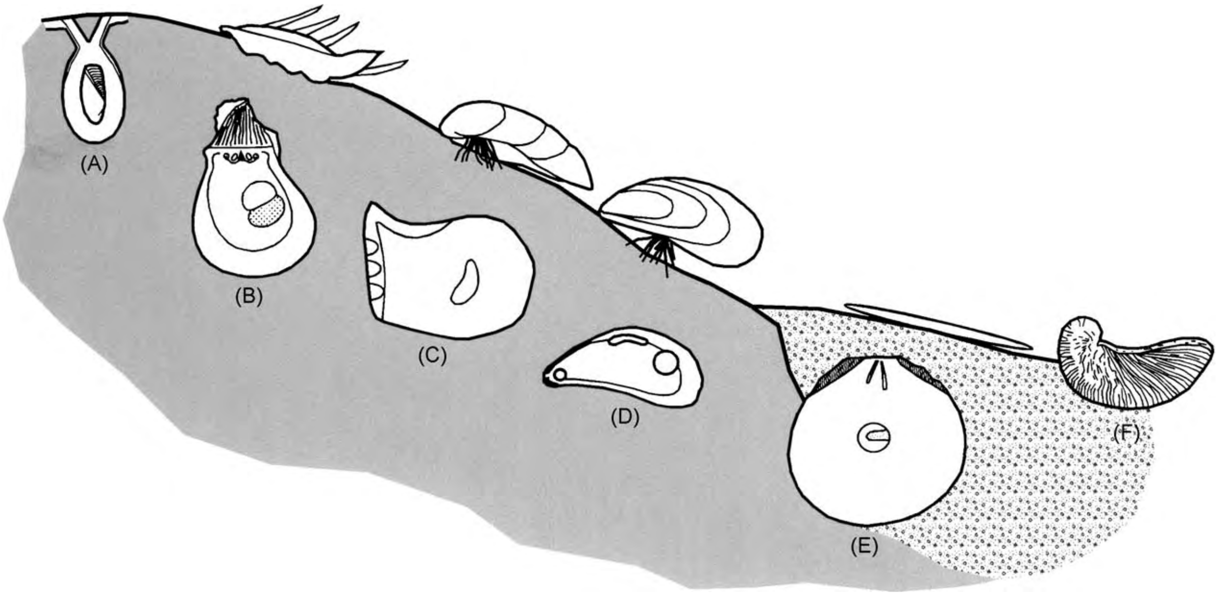


Figure 7 Morphology and mode of life of bivalves living on or boring into substrates. (A) *Spengleria rostrata*. (B) *Spondylus americanus*. (C) *Isognomon ephippium*. (D) *Mytilus edulis*. (E) *Placuna placenta*. (F) *Gryphaea arcuata*. Not to scale.

Whereas byssate attachment is flexible and also renewable, some bivalves permanently attach to hard substrata by a calcareous cement. The cemented habit always succeeds a byssate phase and it is clear that it has evolved independently in a number of different clades (e.g., oysters, rudists, and a number of pectinoids including *Spondylus*). Cemented bivalves are often easily recognizable from their irregular morphology, developed because of the requirement to conform to substratal irregularities, and are markedly inequivalve. As they tended to evolve from pleurothetic byssate stock, most are also monomyarian (Figure 7B). Most cement mainly by the right valve, but a major group, the oysters, do so by the left valve. The size of the attachment scar varies and the substrate may be instantly recognizable; for example, oysters were often attached to ammonite shells, even if the scar and substrate are no longer attached. Most cementers have thick, robust shells and may be extravagantly ornamented with spines or flanges (e.g., *Spondylus*; Figure 7B).

Free Living

A number of different taxa have independently abandoned attachment to become free living on softer sediments. Here, the challenge is not to sink into the substrate, and this has been solved in two ways. The first is by adopting a 'snow-shoe'-type morphology, i.e., resting on a large surface area, epitomized by the wafer-thin window pane shell *Placuna* (Figure 7E). Alternatively, they may be semi-submerged in the soft sediment like an iceberg. This strategy is inferred for the thick-shelled devil's toenail *Gryphaea*, common in Mesozoic clay facies (Figure 7F). These 'aberrant' oysters clearly had a cemented phase, marked by a small attachment scar at the umbo. A few free-living bivalves (notably several groups of scallops) also have the ability to swim short distances if they are threatened. These have smooth hydrodynamic shells and a well-developed posterior adductor muscle whose vigorous contraction provides the propulsion.

Boring

A number of groups, once again polyphyletically, have evolved to excavate 'burrows' in hard substrates by boring. The most successful of these, the mytilid lithophagids, do so principally by acidic secretions (which presents its own challenge of not dissolving its own shell), whilst others, e.g., *Pholas*, bore at least partly by physically rasping the substrate with small projections on the outside of the shell. Some of the most bizarre borers are the teredinids which excavate long cylindrical boreholes in wood and have the enzymatic capability of digesting the cellulose. These 'shipworms' are thought to have been part of the undoing of the ships of the Spanish Armada. Members of the boring group as a whole have a very varied morphology, but may be easily recognized because they are almost invariably fossilized within their characteristic boreholes (Figure 7A).

Taxonomy and Biological Relationships

There are around 100 families of living and a further 150 families of extinct bivalves. Traditionally, these are split into six subclasses which have been defined on the basis of a mosaic of characters, few of which are diagnostic in their own right. Table 1 gives the names and basic attributes of each of these subclasses, providing examples of each. Most modern bivalves belong to one of two subclasses: the infaunal Heterodonta and the largely epifaunal Pteriomorphia.

Despite the excellence of the bivalve fossil record, the wealth of Holocene material on which to base studies, and the general familiarity of the class, our understanding of the relationships between higher taxa is extremely unclear, and the monophyly of particular subclasses has been called into question. Even the relationship between the Bivalvia and the other molluscan classes is debatable. The principal difficulties stem from the high degree of convergent and parallel evolution. Some instances of convergence are clear; for example, the multiple evolution of byssate attachment has led to the heteromyarian ‘mussel’ form being adopted in unrelated taxa, such as the marine mytilids (e.g., *Mytilus*) and the freshwater driessenids. In these instances, close study of other anatomical characters clearly demonstrates that overall morphological similarity is superficial. The problem is more acute where convergence has affected much smaller groups, for example where the cemented habit has evolved repeatedly (five times) within the single family Pectinidae.

Most attempts to classify the bivalves have used either a single or very few characters. Zoologists and palaeontologists have invented different schemes because, not unnaturally, they have placed different importances on anatomical (e.g., gill or stomach architecture) and shell characters, such as hinge structure and shell microstructure, respectively. These are all subject to convergence. Modern computer-driven cladistic analyses which utilize large numbers of different characters offer a better hope of establishing phylogenetic relationships and of identifying convergent characters. Most informative of all are the new battery of molecular techniques; in particular, studies of the 18S rRNA gene have proved to be illuminating. Superimposing morphological data onto molecular phylogenies provides perhaps the ultimate means of establishing convergent characters. However, both the use of data sets combining shell and anatomical characters and the use of molecular methodologies are clearly of little help when dealing with extinct taxa. The fossil record has a critical role

Table 1 Six subclasses of bivalves

Subclass	*Examples*	*Characteristics*
Palaeotaxodonta	*Nucula*	Wholly aragonitic shells of nacre and prisms. Taxodont dentition. Equivalve. Deposit feeders with small protobranch gills. All infaunal
Cryptodonta	*Solemya*, *Cardiola*	Wholly aragonitic shells of nacre and prisms. Toothless. Infaunal. Protobranch gills, specialized to house symbiotic bacteria. Active burrowers
Pteriomorphia	*Mytilus*, *Modiolus*, *Lithophaga*, *Coralliodomus*, *Pinna*, *Ostrea*, *Gryphaea*, *Pecten*, *Spondylus*, *Isognomon*, *Imoceramus*, *Arca*	All include some aragonite within the shell (nacre, crossed lamellar microstructures), but many also have calcitic layers (prisms, foliae). A range of different hinge architectures and musculature (iso and heteromyarian). Eulamellibranch and filibranch gills. Many inequivalve. Pteriomorphs are primitively byssate, but have repeatedly evolved cementing, boring, and free living habits
Palaeoheterodonta	*Modiolopsis*, *Unio*, *Trigonia*	Wholly aragonitic nacre and prisms. Equivalve and a small number of heterodont teeth. Burrowers. Includes freshwater forms
Heterodonta	*Lucina*, *Mya*, *Corbula*, *Venus*, *Mercenaria*, *Ensis*, extinct rudists	Wholly aragonitic, mainly crossed lamellar and complex crossed lamellar microstructures. Mostly large heterodont teeth. Eulamellibranch gills. Equivalve and inequivalve forms. Majority are shallow and deep burrowers, but the extinct rudists were either cementing or free living
Anomalodesmata	*Cuspidaria*, *Poromya*	Wholly aragonitic (largely prisms and nacre). Largely toothless. Mainly isomyarian musculatures. Variety of morphologies reflecting very diverse life habits (burrowing, cementing, byssate)

to play in providing the temporal data that will help to discriminate between putative hypotheses, and provides a time-frame for establishing branching points on suggested trees.

Evolutionary History

Figure 8 shows the changing familial diversity of bivalves over the Phanerozoic. Aside from fluctuations, largely associated with mass extinction events, the overriding impression is one of almost exponential increase in diversity towards the present day. This taxonomic explosion has been accompanied by a movement away from the primitive shallow burrowing and byssate habits into more specialist habits (detailed above) in a series of adaptive radiations.

Bivalves are generally thought to have evolved from the extinct molluscan class Rostroconchia, although this is not universally accepted. The first fossils widely accepted (but again not universally) as bivalves are from Early Cambrian rocks. *Fordilla troyensis* is known from a large number of specimens from Tommotian localities in North America, Greenland, and Western Europe, whilst *Pojetaia runnegari* is known from Australia. Both *Fordilla* and *Pojetaia* are small, only a few millimetres in length, and are generally thought to have been shallow burrowers. Despite this early appearance, there is then a gap in the bivalve fossil record throughout the Middle and Upper Cambrian before their reappearance in the Early Ordovician (Tremadoc). This gap is both perplexing and frustrating; it spans some 4% of the evolutionary history of the class and covers an interval when bivalves apparently became larger and more diverse in terms of both taxa and the life habits employed. There have been several claims for Middle and Upper Cambrian bivalves, but most of these have been dismissed as misidentified inarticulate brachiopods. It is difficult to explain this gap in the bivalve record, but one plausible explanation is that the earliest bivalves may have lived in nearshore silty facies which are seldom preserved.

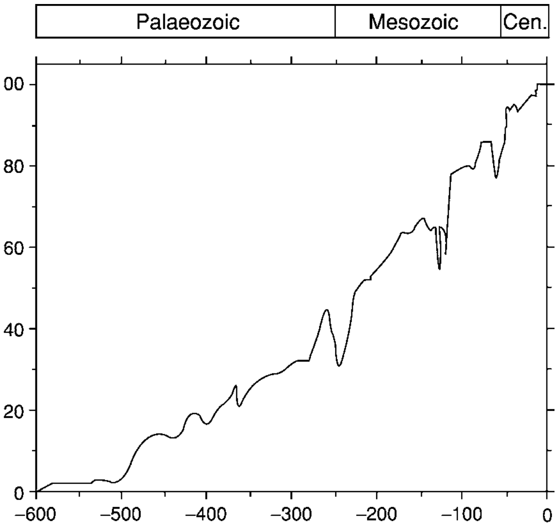


Figure 8 Familial diversity of bivalves over the Phanerozoic.

By the close of the Ordovician, bivalves were more common and abundant and all the subclasses recognized in Table 1 had appeared. There is evidence of shallow burrowers and both endo- and epibyssate forms and signs of different feeding strategies (deposit feeders, suspension feeders). Bivalve diversity continued to increase throughout the Palaeozoic, although there are small decreases associated with the Ordovician and the Devonian (Frasnian/Famennian) mass extinctions. Throughout the era most bivalves remained in 'primitive' life habit, with only one genus each of borers (*Coralliodomus*) and cementers (*Pseudomonotis*) being recognized (neither of which is widespread). Very few Palaeozoic bivalves (e.g., *Lyrodesma*) have pallial sinuses. Although the effect of the end-Permian mass extinction is clearly evident in Figure 8, losses were small in contrast with those sustained by other shelly benthos, for example the brachiopods.

The beginning of the Mesozoic saw a further combined increase in the number of taxa, with an ever increasing repertoire of life habits. Many shallow burrowers moved deeper, and large numbers of byssate taxa adopted more specialized habits, such as cementing, boring, and free living. The cementing habit is exemplified by the successful oysters and rudists. The first evidence of the adaptations for swimming in scallops are recognized in Jurassic taxa, and free-reclining gryphaeid oysters became abundant and widespread in this period. This major adaptive radiation is widely thought to have been linked to the Mesozoic Marine Revolution – the profound restructuring of shallow marine communities and, in particular, a massive increase in predation pressure at that time. Bivalves are extremely vulnerable to predators and are attacked by a wide range of important predatory groups (e.g., fish, gastropods, starfish, crustaceans, mammals, and birds) and by a variety of different methods (swallowing whole, crushing, drilling, or prising apart). It has been suggested that different groups of bivalves responded in a variety of different ways, for example by living deeper within the sediment, encasing themselves in hard substrates and becoming more robust, and, in some cases, growing formidable spines. Although the end-Cretaceous extinctions had a marked effect on the bivalves, with major groups such as the inoceramids and rudists being lost, their diversity continued to increase during the Cenozoic. The number of groups exploiting the more specialist boring and cementing habit has continued to grow and there has been a further taxonomic explosion of the shallow burrowers at lower taxonomic levels.

See Also

Biological Radiations and Speciation. **Fossil Invertebrates:** Molluscs Overview; Gastropods; Cephalopods (Other Than Ammonites); Ammonites. **Mesozoic:** End Cretaceous Extinctions. **Palaeoecology**.

Further Reading

Beesley PL, Ross GJB, and Wells A (1998) *Mollusca: the Southern Synthesis,* vol. A. Collingwood, Australia: CSIRO Publishing.

Harper EM (1998) The fossil record of bivalve molluscs. In: Donovan SK and Paul CRC (eds.) *The Adequacy of the Fossil Record,* pp. 243 267. Chichester: John Wiley and Sons.

Harper EM and Skelton PW (1993) The Mesozoic Marine Revolution and epifaunal bivalves. *Scripta Geologica, Special Issue* 2: 127 153.

Harper EM, Taylor JD, and Crame JA (2000) *The Evolutionary Biology of the Bivalvia. Geological Society of London Special Publication 177.* London: Geological Society of London.

Johnston PA and Haggart JW (1998) *Bivalves: an Eon of Evolution. Paleobiological Studies Honoring Norman D. Newell.* Calgary: Calgary University Press.

Skelton PW, Crame JA, Morris NJ, and Harper EM (1990) Adaptive divergence and taxonomic radiation in post Palaeozoic bivalves. In: Taylor PD and Larwood GP (eds.) *Major Evolutionary Radiations. The Systematics Association Special Volume 42,* pp. 91 117. Oxford: Clarendon Press.

Stanley SM (1970) Relation of shell form to life habits of the Bivalvia. *Geological Society of America Memoir* 125: 1 296.

Taylor JD (1996) *Origin and Evolutionary Radiation of the Mollusca.* Oxford: Oxford University Press.

Vermeij GJ (1987) *Evolution and Escalation. An Ecological History of Life.* Princeton, NJ: Princeton University Press.

Vermeij GJ (1993) *A Natural History of Shells.* Princeton, NJ: Princeton University Press.

Gastropods

J Frýda, Czech Geological Survey, Prague, Czech Republic

Introduction

Gastropods are well-known animals which have been associated with humans since the dawn of civilization. Their bodies were gathered for food and their shells were used as tools, ornaments, and later as money. Their widespread occurrence is clear evidence of their successful adaptation to different environments. During a long evolution, they are the only molluscan class to have colonized the majority of marine, freshwater, and terrestrial environments. Marine gastropods occur mostly in shallow-water benthic communities; however, some gastropod species have also lived in the deep sea (e.g., faunas associated with hydrothermal vents), and others, such as holoplanktic animals, have spent their whole lives as free-swimming gastropods. The terrestrial gastropods colonized most land environments, ranging from lowlands to high mountains, and including humid to arid biotopes of tropical to subarctic areas. Such adaptive radiation is quite exceptional amongst all animal phyla and is linked to the extraordinary morphological and functional diversity of their bodies and shells. The gastropods comprise one of the most diverse groups of living animals (the second after Insecta). All these facts, together with their long and rich fossil record, make gastropods a unique animal group for evolutionary, ecological, and biogeographical investigations. There follows a brief review of gastropod anatomy, shell morphology, classification, and more than 500 million years of evolution.

Definition and General Description

The Gastropoda forms one of eight molluscan classes, and is defined by several unique anatomical features which support its interpretation as a molluscan group derived from the same ancestor (i.e., monophyly). The most characteristic feature of gastropods is torsion of their soft bodies during early larval stages, producing a crossing of their nerve connectives, bending of the intestine, and twisting of the mantle cavity (together with associated structures, including the ctenidia, anus, kidney openings, etc.) anteriorly over the gastropod head.

Anatomical Features

The exceptional morphological and functional diversity of gastropod bodies is also reflected in their anatomy. Generally, the body consists of a large foot, a visceral mass, and a head with a mouth, tentacles, and eyes. The visceral mass is mostly enclosed, together with the mantle cavity, in a calcareous shell (Figure 1). Gastropods as soft-bodied animals use the pressure of their blood and muscles for movements of different organs. The circulatory system comprises a contractile heart with one (Caenogastropoda and Heterobranchia) or two (Patellogastropoda, Archaeogastropoda, and Neritimorpha) auricles, and a ventricle, as well as a system of arteries and veins. Gastropod blood transports oxygen using the copper-bearing pigment haemocyanin. In most marine gastropods, one or two gills (ctenidia) situated in the mantle cavity are used for respiration. Some marine and freshwater gastropods developed secondary gills after the loss of their ctenidia in previous evolution. In the terrestrial gastropods (e.g., Pulmonata), a highly vascularized internal wall of the mantle cavity (lung) is used for respiration. One or, rarely, two (Patellogastropoda and Vetigastropoda) kidneys serve for excretion through the mantle cavity. The digestive system starts with a mouth containing a tooth-bearing ribbon (radula). The organization of gastropod radulae and stomach, as well as additional parts of the digestive system, reflects their different feeding habits (herbivory, detritivory, carnivory, or parasitism). The anus opens into the mantle cavity. The gastropod nerve system includes paired ganglia which are linked with different sensory receptors by connectives and commissures. The morphology of the reproductive organs and the reproductive strategies are highly diverse. Generally, more ancient gastropod groups are gonochoristic with a simple reproductive system and external (Patellogastropoda and Archaeogastropoda) or internal (Neritimorpha) fertilization. Caenogastropoda also use internal fertilization with complex reproductive morphology, and some may be simultaneous hermaphrodites. The Heterobranchia have the most complex and variable reproductive system and are hermaphroditic.

Ontogeny

Gastropods, like all Mollusca, have a biphasic life cycle (i.e., larval and post-metamorphosis stages), and this feature is shared with closely related animal phyla (Kamptozoa, Sipunculida, Polychaeta, etc.). Like other molluscan groups, the embryonic development is characterized by a spiral cleavage, which differs slightly in the main gastropod groups. The subsequent larval stage is called the trochophore

Figure 1 Some variations in shell form of living gastropods. (A) High spired: *Mitra mitra* (Muricoidea). (B) Strombiform: *Lambis chiragra* (Stromboidea). (C) Turbiniform: *Liguus vittatus* (Orthalicoidea). (D) Convolute: *Cyprea tigris* (Cypraeoidea). (E) Spinose fusiform: *Chicoreus ramosus* (Muricoidea). (F) Fusiform: *Pleuroploca trapezium* (Muricoidea). (G) Conoidal: *Conus litteratus* (Conoidea). (H) Discoidal: *Architectonica perspectiva* (Architectonicoidea). (I) Turriculate: *Terebra* sp. (Conoidea). (J) Ovoid: *Olivancillaria gibbosa* (Olivoidea). (K) Involute: *Cypraecassis rufa* (Tonnoidea). (L) Irregularly coiled: *Siliquaria ponderosa* (Cerithioidea). (M) limpet: *Megathura crenulata* (Fissurelloidea). (A, E G, I, J) Neogastropoda (Caenogastropoda); (B, D, K) Littorinimorpha (Caenogastropoda); (C, H) Heterobranchia; (I) Sorbeoconcha (Caenogastropoda); (M) Vetigastropoda (Archaeogastropoda).

larva, and a similar larval type is developed in all molluscan groups. The trochophore larvae may be free swimming, as in the ancient gastropod groups (Patellogastropoda and Archaeogastropoda), or may occur in egg capsules, as in more advanced gastropods. The last larval stage is termed veliger, which typically bears two ciliate paddles (velum), sometimes subdivided into several lobes. If free-swimming gastropod larvae use planktic organisms for their nutrition, their development is termed planktotrophic. Marine gastropods with such development have small eggs, but numbering over half a million. Planktotrophic larvae may stay planktic for several months and thus can be carried for long distances by oceanic currents. The gastropods, however, developed another ontogenetic strategy in which their larvae were not dependent on an external food source, but on the yolk of their eggs. Gastropods with such a non-planktotrophic development (lecithotrophic) typically produce fewer eggs, which are relatively large. The larval stages end with a metamorphosis that involves anatomical and physiological reorganization of the larval body into the juvenile, post-larval body. Terrestrial and freshwater gastropods have simplified their development, and their embryonic and larval stages are fixed on egg capsules or the female body (direct development). Such ontogenetic changes considerably decreased their dispersal potential.

The Gastropod Shell

Gastropods are not only one of the most diverse animal groups, but the morphology of their shells is extremely varied (Figures 1 and 2). During more than 500 million years of evolution, they developed shells with various shapes and ornament, ranging in size from about 1 mm up to more than 1 m (Eocene Campaniloidea, Caenogastropoda). The shell and its ornament may be broadly linked to the mode of gastropod life (e.g., origin of limpet-shaped shells in unrelated gastropod groups). Generally, the most ornate shells occur in tropical marine environments, but freshwater and terrestrial gastropods are often less ornate.

Protoconch and Teleoconch

In shell-bearing gastropods, the shell grows during almost the whole of their ontogeny. The part of the shell formed during the embryonic and larval stages is called a protoconch (Figure 3), and that growing after metamorphosis is termed a teleoconch. The main gastropod groups differ in their early development, which is reflected in their protoconch morphology. The more ancient gastropod groups (Patellogastropoda and Archaeogastropoda) have the simplest shell ontogeny and their protoconchs have only an embryonic shell (protoconch I), which is followed by a teleoconch (Figures 3B and 3F–3H). On the other hand, the protoconchs of more advanced gastropods (Neritimorpha, Caenogastropoda, and Heterobranchia) consist of an embryonic shell (protoconch I) and a subsequent larval shell (protoconch II). In most caenogastropods, the larval shells have different ornament from the teleoconchs (Figures 3K and 3L), and both shells are coiled in the same direction (such a condition is termed homeostrophic; Figure 4). In contrast, in the Heterobranchia with planktotrophic development, the protoconchs are coiled in the opposite direction to the teleoconchs (Figures 3J and 4). Such shells are termed heterostrophic. The Neritimorpha form typical, strongly convolute protoconchs during planktotrophic development, which are homeostrophic (Figures 3A, 3N, and 4). Higher gastropods with non-planktotrophic development (some marine, freshwater, and terrestrial gastropods) have simplified their early ontogeny and thus also the morphology of their protoconchs. The latter strategy is documented from the Devonian (400 Ma).

Operculum

The majority of gastropods have a lid-like structure (operculum) to close their aperture. This operculum is present in all living gastropods during their larval stages, but is lost in some adults (e.g., limpets and the majority of terrestrial gastropods). The operculum is mostly horny (corneous) and may be tightly (multispiral) or loosely (paucispiral) coiled or concentric. Some gastropod groups have calcareous opercula, and the oldest operculum known is from the Ordovician (Macluritoidea).

Shell Structure

Most gastropod shells are composed of an outer organic layer (periostracum) and an inner, mostly much thicker, calcified layer. The colour pattern typical of many gastropod shells (Figure 1) is formed by different organic pigments which are limited to the periostracum and the uppermost calcified layer. This shell feature, sometimes reflecting the mode of life, has been known since the Palaeozoic (Figure 5). The inner layers of gastropod shells consist of minute calcium carbonate crystals (aragonite or calcite) in an organic matrix. There are over 20 structural types of gastropod shell and, in general, more ancient groups exhibit more diverse shell structures. The Patellogastropoda (Eogastropoda) had the most complex shell structure. On the other hand, the majority of the higher gastropods have developed simple

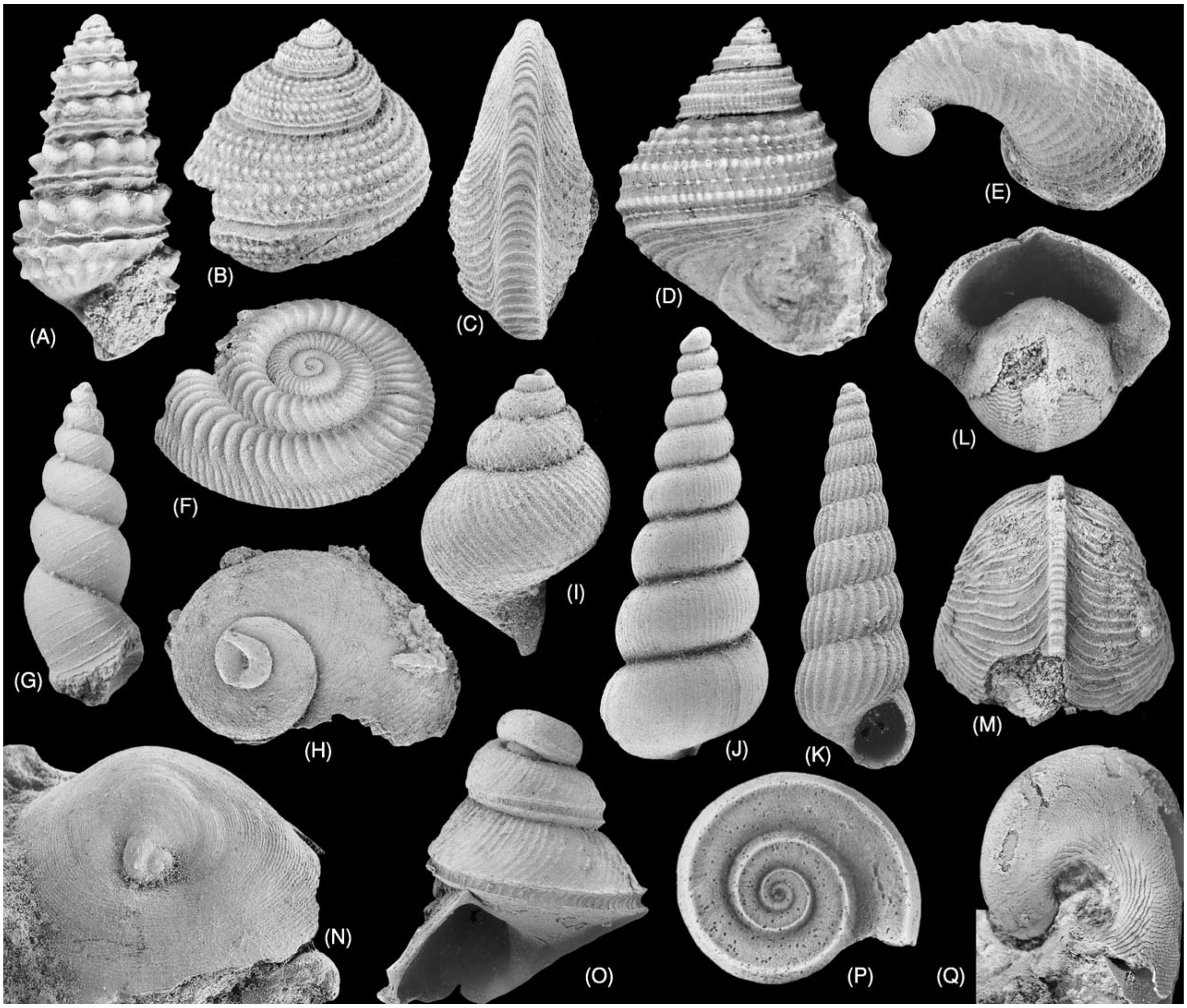


Figure 2 Some variations in shell form in the main groups of middle Palaeozoic gastropods. (A) High spired shell: *Murchisonia coronata* (Murchisonioidea). (B) Trochiform, slit bearing shell: *Devonorhineoderma orbignyana* (Eotomarioidea). (C) Bilaterally symmetrical shell with a prominent selenizone: *Kolihadiscus tureki* (Cyrtolitoidea). (D) Turbiniform shell: *Gyronema armata* (Gyronematidae). (E) Openly coiled shell: *Pragoserpulina tomasi* (Pragoserpulinidae). (F) Discoidal shell: *Stusakia pulchra*. (G) Sinistrally coiled shell: *Voskopiella barborae* (Onychochilidae). (H) Naticiform shell: *Eifelcyrtus blodgetti* (Vltavielidae). (I) Fusiform shell: *Havlicekiela parvula* (Peruneloidea). High spired shells: (J) *Pragozyga costata*; (K) *Palaeozyga bohemica* (Loxonematoidea). (L, M) Bilaterally symmetrical shell: *Bellerophon vasulites* (Bellerophontoidea). (N) Limpet: *Pragoscutula wareni* (Pragoscutulidae). (P) Sinistrally coiled shell: *Alaskiella medfraensis* (Porcellioidea). (P) Discoidal shell: *Nodeuomphalus labadyei* (Euomphaloidea); (Q) Bilaterally symmetrical shell covered by secondary shell deposits: *Branzovodiscus bajae* (Bellerophontoidea). (A, B, D, F, O) Archaeogastropoda; (C) Cyrtonellida; (G) Mimospirina; (H) Cyrtoneritimorpha; (L, M, Q) Bellerophontida; (I) Perunelomorpha; (E, J, K, N) Order uncertain; (P) Euomphalomorpha.

aragonitic shells with a crossed lamellar structure (Figures 6A and 6B). Some structural types are restricted to certain groups (e.g., nacre) and this may be used for their identification in fossils (Figure 6C). Nacreous and crossed lamellar structures have been known since the Palaeozoic.

Shell Coiling

The majority of the shell-bearing gastropods have right-handed (dextral) shells, but some have left-handed (sinistral) shells (Figures 1 and 2). Only a few gastropods have bilaterally symmetrical shells which may be uncoiled (limpets) or planispirally coiled (Figures 1M, 2C, 2L, and 2M). The limpet-shaped shells were independently developed within all main gastropod groups from the asymmetrically coiled shells of their ancestors. In contrast, planispirally coiled shells are known only in several groups, such as the Palaeozoic Porcellioidea and Bellerophontoidea (Figures 2L and 2M) or the Holocene Planorbioidea. Some gastropods may change the coiling

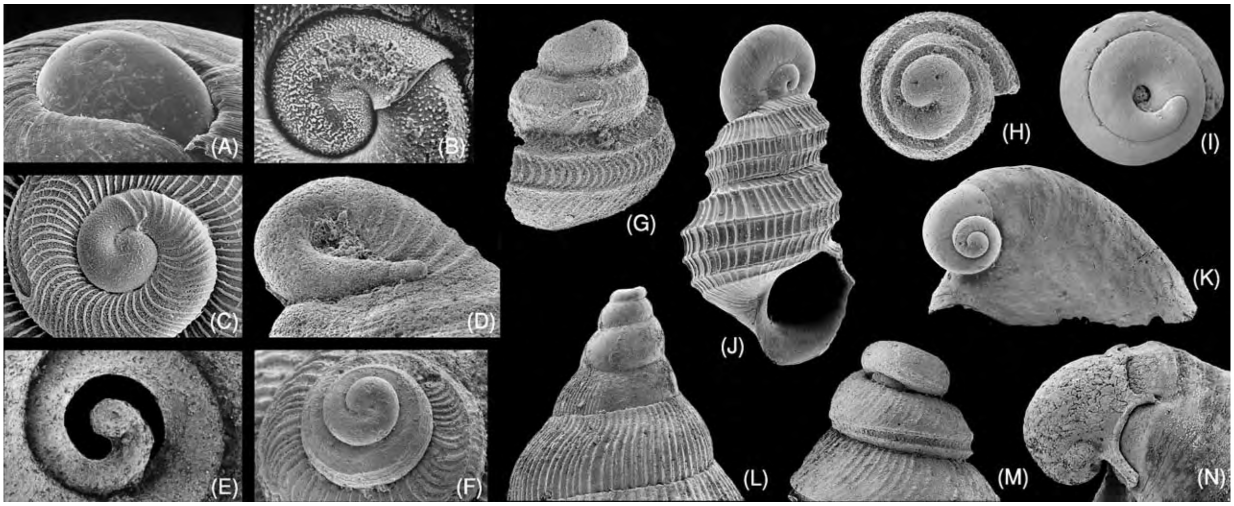


Figure 3 Variety of protoconch shape. Strongly convolute larval shell (protoconch II), Neritimorpha: (A) Holocene *Smaragdia* sp. (Neritoidea); (N) Triassic *Pseudorthonychia alata* (Pseudorthonychiidae). Embryonic shell (protoconch I) followed by teleoconch, Archaeogastropoda: (B) Triassic *Wortheniella coralliophila* (Vetigastropoda); (C) Holocene *Anatoma proxima* (Vetigastropoda); (F) Devonian *Zlichomphalina* sp. (Eotomarioidea); (G, H) Devonian *Diplozone innocens* (Murchisonioidea). (D) Openly coiled larval shell (protoconch II), Cyrtoneritimorpha, Carboniferous *Orthonychia parva* (Orthonychiidae). (E) Openly coiled early shell, Permian *Euomphalus* sp. (Euomphaloidea). (I) Openly coiled larval shell (protoconch II) of the Silurian Peruneloidea. (J) Heterostrophic larval shell (protoconch II), Jurassic *Mathilda* sp. (Architectonicoidea, Heterobranchia). Larval shell (protoconch II), Caenogastropoda: (K) Holocene *Hipponix* sp. (Vanikoroidea); (L) Devonian *Balbiniconcha cerinka* (Subulitoidea). (M) Heterostrophic early shell, Devonian *Alaskiella medfraensis* (Porcellioidea, Archaeogastropoda).

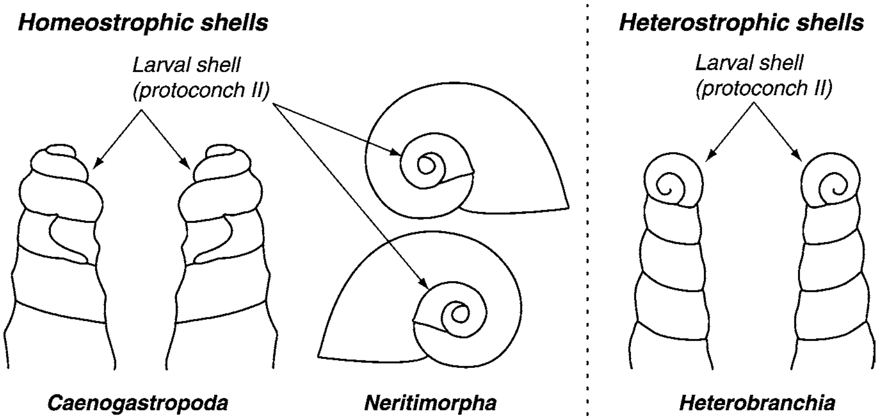


Figure 4 Schematic diagram showing the relationship between the coiling of larval (protoconch II) and post larval (teleoconch) shells in planktotrophic Caenogastropoda, Neritimorpha, and Heterobranchia. Coiling of both shells in the same direction (Caenogastropoda and Neritimorpha) is termed homeostrophic. If the handedness of the shells is opposite (Heterobranchia), the coiling is termed heterostrophic.

of their shells (Figure 4) from sinistral to dextral (dextral heterostrophy), or vice versa (sinistral heterostrophy), during ontogeny. Such a change may occur at a developmental stage, when gastropods undergo a metamorphosis from larval to post-larval stages (e.g., Heterobranchia; Figure 3J), or later (e.g., Porcellioidea; Figure 3M).

Dextrality or sinistrality of the shell is independent of the coiling of the soft body, and the asymmetrical soft body of gastropods may be dextral or sinistral. Anatomical dextrality or sinistrality may be easily recognized, even in fossil gastropods, if they developed a spiral operculum. The spiral operculum of anatomically dextral gastropods is coiled counterclockwise (viewed externally), and vice versa in sinistral gastropods. Thus, there are four possible relationships between shell coiling and body asymmetry in the shell-bearing gastropods (Figure 7). If anatomically dextral (or sinistral) animals occupy dextrally (or sinistrally) coiled shells, such a condition is called dextral (or sinistral) orthostrophy. If the handedness of the shell and soft body is different, the term hyperstrophy is used. All four kinds of coiling (Figure 7) have occurred in gastropods, but their frequencies are very different. The great majority of living gastropods are dextrally orthostrophic, and sinistral orthostrophy is uncommon. Dextral or sinistral hyperstrophy is very rare (e.g., Ordovician Macluritoidea or some Holocene Planorbioidea).

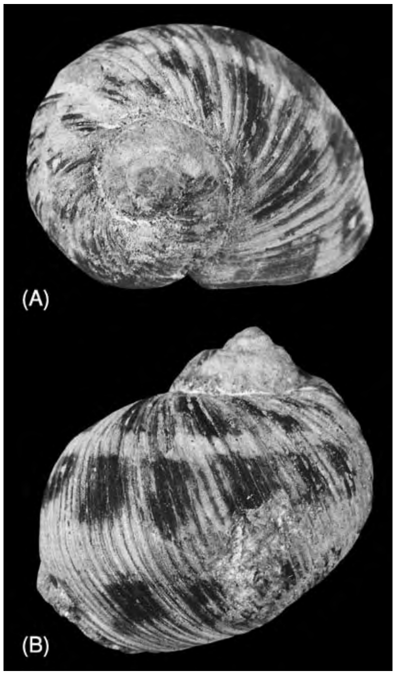


Figure 5 Apical (A) and lateral (B) views of the Middle Devonian (about 400 Ma) neritimorph, *Paffrathopsis subcostata*, showing the colour pattern.

Muscle Scars

Gastropod shells are attached to the soft body by muscles, which may leave distinct scars on the inner shell surface. The geometry of muscle scars has frequently been used as a diagnostic feature for distinction between torted (i.e., gastropods) and untorted states in the Palaeozoic molluscs (Monoplacophora, Helcionelloida, Cyrtonellida, etc.). However, new anatomical studies of living gastropods have shown that the larval muscles taking part in torsion and the post-larval muscles are developed quite independently. Thus, the muscle scar pattern sometimes observable in the fossil molluscan shells may be a good ecological indicator, but has no systematic significance.

Classification of the Gastropoda

Gastropods as an independent group of molluscs were recognized and named by the French naturalist, Georges Cuvier, more than 200 years ago (*see* **Famous Geologists:** Cuvier). Since then, scientists have tried to classify them by using different features of their bodies. However, the classification of such a numerous group with extraordinary morphological and anatomical variability of their bodies and shells has encountered many problems. During the nineteenth century, several different classifications of the Gastropoda were published, based on the shape of the shells, position of the mantle cavity, or on the arrangement of various organs (e.g., gills or head). Generally, these classification schemes used only a limited number of distinguishing characters. At the beginning of the twentieth century, the German zoologist, Johannes

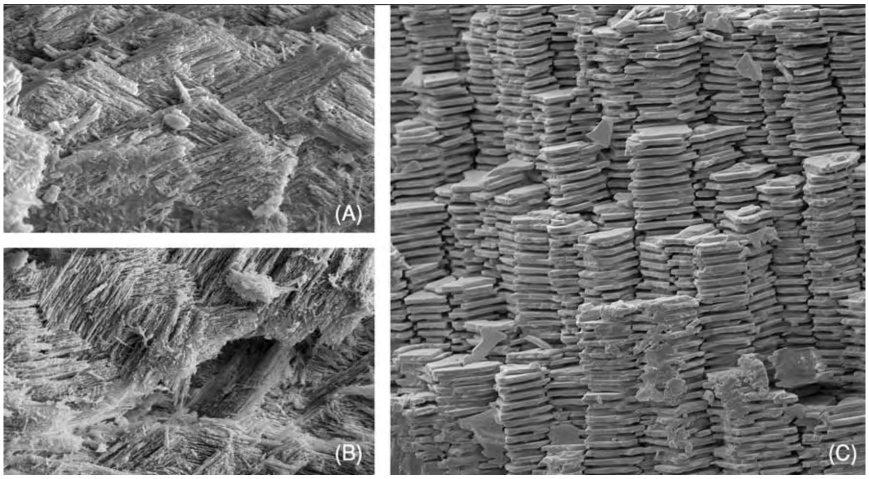


Figure 6 Examples of shell structure in fossil gastropods. Aragonitic crossed lamellar structure in the Carboniferous (about 300 Ma) *Amphiscapha catilloides* (Euomphaloidea): views perpendicular to (A) and parallel to (B) the shell surface. (C) Nacreous structure (columnar nacre) in the Late Cretaceous (about 80 Ma) *Sensuitrochus ferreri* (Porcellioidea, Archaeogastropoda).

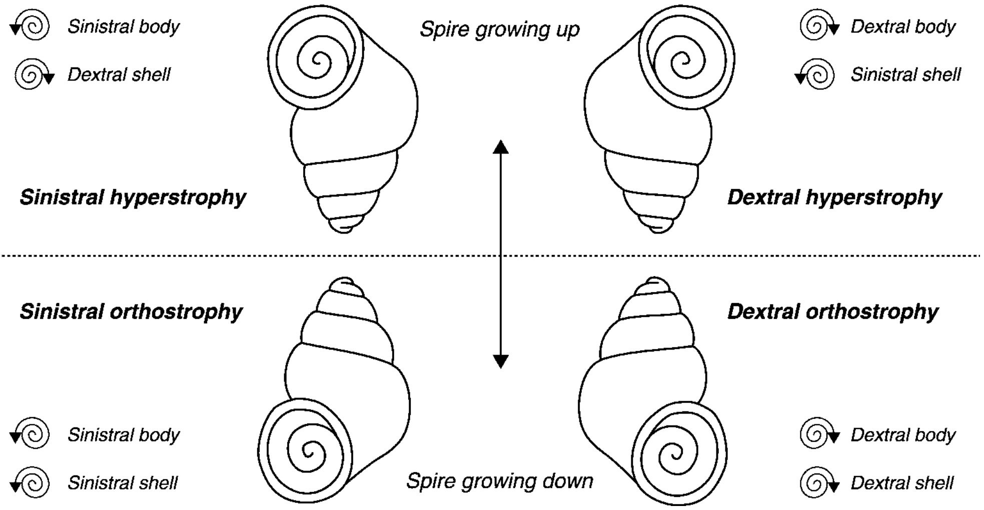


Figure 7 Schematic diagram showing the four possible relationships between shell coiling and body asymmetry. Orthostrophy means that anatomically dextral (or sinistral) animals occupy dextrally (or sinistrally) coiled shells. The term hyperstrophy is used when the handedness of the shell and soft body is different (see text for explanation).

Thiele, integrated earlier classifications and divided the gastropods into three subclasses: Prosobranchia, Opisthobranchia, and Pulmonata. In addition, the Prosobranchia were divided into three orders: Archaeogastropoda, Mesogastropoda, and Neogastropoda. Thiele's system was used by zoologists and palaeontologists for most of the twentieth century.

However, during recent decades, numerous new data on the anatomy of various gastropod groups have been accumulated, mainly by the application of new methods (e.g., transmission electron microscopy). At the same time, studies of the deep-sea faunas associated with hydrothermal vents have brought the discovery of new gastropod groups with unusual anatomical features. The evaluation of this newly gathered data in the light of the existing classification revealed a need for its revision. Recent analyses of numerous morphological and developmental characters of living gastropods have resulted in a new classification scheme (Figure 8), which has been independently supported by results from molecular studies. The placement of fossil gastropods into this classification of living gastropods has been difficult because of the lack of necessary anatomical characters.

Recent studies have revealed that Patellogastropoda (= Docoglossa, Cyclobranchia) represents the sister group to all other living gastropods. Living patellogastropods with limpet-shaped shells are exclusively marine and occur mostly on rocky shores in all continents. The Patellogastropoda and their coiled ancestors have been united into the subclass Eogastropoda. All other living gastropods and their ancestors have been placed in the subclass Orthogastropoda, comprising four main groups of living gastropods: Neritimorpha, Archaeogastropoda, Caenogastropoda, and Heterobranchia (Figure 8).

The Neritimorpha (= Neritopsina) is an ancient gastropod group with a long fossil record (Figures 2, 5, and 9), which colonized many different marine (shallow- and deep-water), freshwater, and terrestrial environments. The Palaeozoic Cyrtoneritimorpha, with openly coiled early shells (Figures 2H and 3D), may represent a closely related group. The living Archaeogastropoda unites the Vetigastropoda (Figures 3B and 3C) and several smaller groups, such as the Neomphaloidea, which occur in faunas associated with deep-sea hydrothermal vents. The Archaeogastropoda have colonized almost all marine and estuarine environments. There are also a number of extinct, mainly Palaeozoic groups (Figure 2) with uncertain relationships to living archaeogastropods. The Palaeozoic Euomphaloidea (= Euomphalomorpha; Figures 2P and 3E), known mainly from shallow-water, marine environments, may be a sister or basal group of the Archaeogastropoda.

The Caenogastropoda and Heterobranchia are sister groups which are united in the taxon Apogastropoda. Both groups are highly diverse and have colonized almost all marine, freshwater, and terrestrial environments. The Palaeozoic Subulitoidea and Peruneloidea (Perunelomorpha) (Figures 2I and 3L) may be ancestral or basal groups of the Caenogastropoda or of all Apogastropoda. The extant Caenogastropoda unites the two orders Architaenioglossa and Sorbeoconcha. Terrestrial Cyclophoroidea and freshwater Ampullarioidea form the Architaenioglossa. On the other hand, the mostly marine Sorbeoconcha represents a highly diverse group uniting

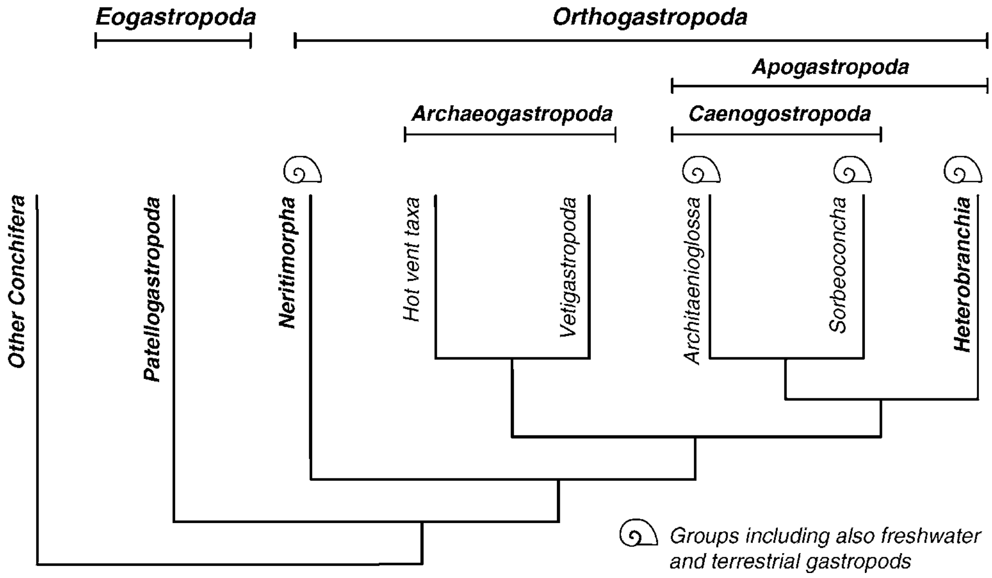


Figure 8 Recent classification scheme of living gastropods, illustrating their phylogenetic relationships and the distribution of freshwater and terrestrial groups. Based on Ponder WF and Lindberg DR (1997) Towards a phylogeny of gastropod molluscs: an analysis using morphological characters. *Zoological Journal of the Linnean Society* 119: 83 256.

more than 25 superfamilies of living gastropods (Figure 1).

The Heterobranchia encompasses the gastropod groups placed by Thiele's classification into the 'Opisthobranchia' and 'Pulmonata', as well as some 'prosobranch' groups, such as the Valvatoidea and Architectonicoidea. The Valvatoidea is an ancient group of freshwater gastropods, but the highly diverse Architectonicoidea represents a marine group (Figure 1H). The majority of the lower Heterobranchia (Opisthobranchia or sea slugs) are also marine gastropods, typically with their shells reduced or absent. They are extraordinarily variable and are divided into about 30 superfamilies of nine orders. The higher Heterobranchia (Pulmonata) form a dominant group of terrestrial gastropods, but also occur in freshwater environments. There are several classifications of the Pulmonata, which may be divided into three orders: Systellommatophora, Basommatophora, and Eupulmonata. The ancient marine Basommatophora have been separated into the Archaeopulmonata and the freshwater Basommatophora into the Brachiopulmonata. The Stylommatophora is a dominant group of terrestrial gastropods and is the most numerous group in the Eupulmonata.

The higher classification of extinct gastropods is less stable than that for living groups. The Palaeozoic, exclusively marine Pelagiellida, Bellerophontida (Figures 2L and 2M), Macluritoidea, and Mimospirina are amongst the most discussed extinct groups, and the gastropod nature of the Bellerophontida and Pelagiellida is still a frequently discussed problem. The phylogenetic relationships of the Macluritoidea and Mimospirina (Figure 2G), with sinistrally coiled shells, are uncertain and both groups may be sister groups to more advanced gastropods.

Evolution of the Gastropoda

The more than 500 million years of evolution of the Gastropoda is still poorly known. The main difficulties are that the phylogenetic positions and relationships of extinct gastropods can be inferred only from the limited number of characters observable in their fossilized hard body parts (i.e., shell and operculum). However, the number of extinct gastropod species and genera is much higher than those living. In addition, some belong to extinct higher taxa of family or order levels with unknown anatomy. Another complication is the development of similar shells in unrelated groups (homoplastic similarity) which has been documented in many living gastropods.

Origin and Early History of the Gastropoda

Since 1970, many new mollusc-like fossils from the Cambrian have been discovered (e.g., *Halkieria*, Merismoconcha, etc.). Their interpretation has given rise to different models of evolutionary relationships within the Mollusca. Even though these models are controversial, the Gastropoda has been generally accepted to be the sister group of the classes Cephalopoda (*see* **Fossil Invertebrates:** Cephalopods (Other Than Ammonites)) or Tryblidiida ('Monoplacophora'). The latter have been combined with the

Scaphopoda and Bivalvia (*see* **Fossil Invertebrates:** Bivalves) within the group Conchifera, which unites the higher Mollusca. Whether or not the Conchifera is monophyletic is uncertain.

Torsion of the soft body has been considered to be one of the main diagnostic characters of the Gastropoda. For this reason, the majority of the models of gastropod origin have been based on different interpretations of this anatomical feature in the extinct gastropod-like molluscs. The Early Palaeozoic Helcionelloida, Bellerophontida (Figures 2L and 2M), and Tryblidiida, with bilaterally symmetrical shells, as well as the Pelagiellida and Macluritida, with asymmetrically coiled shells, are the most frequently discussed groups, and have been variously interpreted as untorted or torted molluscs. However, there is no reliable method of recognizing torsion in extinct fossil molluscs. Thus, the unknown nature of the bodies in the Early Palaeozoic gastropod-like fossils has enabled controversial speculations to be made about the origin of the Gastropoda. Generally, it is accepted that the first undoubted gastropods appeared in the Late Cambrian.

Palaeozoic Era

During the Early Ordovician radiation, the diversity of gastropod groups which had appeared in the Late Cambrian (Archaeogastropoda, Euomphaloidea, Macluritoidea, Mimospirina, Peruneloidea) rapidly increased (Figure 9). The Macluritoidea with large shells, together with different groups of Archaeogastropoda, Euomphaloidea, Bellerophontida, and Mimospirina, were typical elements of gastropod faunas of the tropical regions. In contrast, higher latitude faunas were composed mainly of the Bellerophontida and Archaeogastropoda. This arrangement survived until the early Middle Ordovician, when the diversity of some groups (Macluritoidea and Euomphaloidea) decreased and some new groups appeared (the slit-lacking Archaeogastropoda, Subulitoidea, Platyceratoidea, Loxonematoidea, etc.). During the Middle Ordovician, gastropod diversity rapidly increased and, in the Late Ordovician, reached its maximum. Middle and Late Ordovician faunas consisted of members of all the main groups of Palaeozoic gastropods, except the Heterobranchia (Figure 9). The end of the Ordovician saw a dramatic decrease in gastropod diversity, as well as the extinction of the Macluritoidea.

The Silurian was a period of increasing diversity of many gastropod groups (e.g., Archaeogastropoda, Bellerophontida, and Platyceratoidea), when some gastropods in all marine communities continually increased, together with an increase in the morphological variability of their shells. This suggests an increase in their ecological adaptation to specific

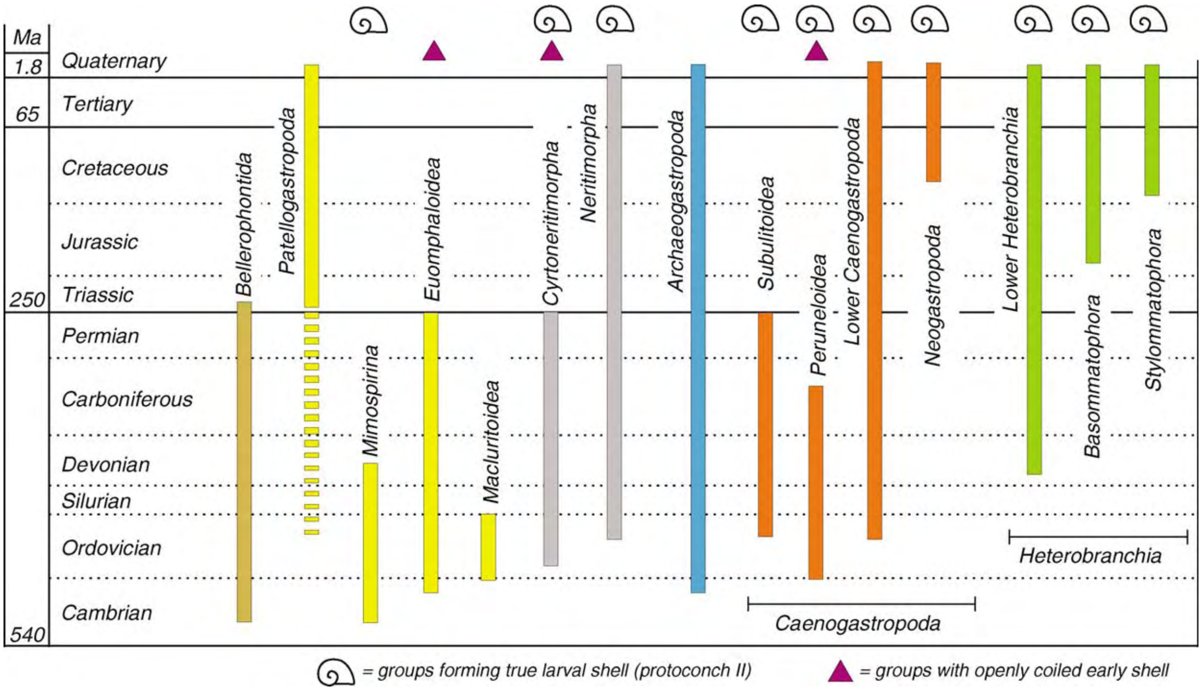


Figure 9 Diagram illustrating the evolution of the Gastropoda and the types of early shell ontogeny. The bars show the stratigraphical ranges of each gastropod group. Based on Bandel K (1997) Higher classification and pattern of evolution of the Gastropoda. *Courier Forschungsinstitut Senckenberg* 201: 57 81, and Frýda J and Rohr DM (2004) Gastropoda, 184 195 In: Webby BD, Droser ML, Paris F, and Percival IG (eds.) *The Great Ordovician Biodiversification Event*, pp. 408. New York: Columbia University Press.

environments. In comparison with the Ordovician, during the Silurian some gastropods with high-spired shells (mainly Loxonematoidea, Murchisonoidea and Subulitoidea; Figure 2) increased considerably.

The Devonian was a time of distinct changes in marine gastropod communities. Some Ordovician–Silurian groups became extinct (Mimospirina; Figure 2G), new groups appeared (Heterobranchia), and many groups underwent rapid radiation and specialization (Caenogastropoda and Neritimorpha). Thus, the Devonian faunas contained representatives of all extant gastropod orders (Archaeogastropoda, Neritimorpha, Caenogastropoda, and Heterobranchia), as well as many Palaeozoic groups (Figure 9). The Devonian was also the time when the protoconch morphology of several gastropod groups underwent considerable change. Gastropods with openly coiled protoconchs (Perunelomorpha, Cyrtoneritimorpha, and Euomphalomorpha; Figures 2H, 2P, 3D, and 3E) formed a considerable, sometimes even dominant, part of the Ordovician and Silurian gastropod communities. During the Early Devonian, their numbers rapidly decreased and none survived the Permian/Triassic extinction.

Carboniferous and Permian faunas had a similar composition of marine gastropod communities. A characteristic feature of Late Palaeozoic gastropod faunas was the fast radiation of different groups of Apogastropoda. The dominance of diverse groups of Caenogastropoda with high-spired shells (mainly Ctenoglossa and Cerithiomorpha) and Heterobranchia (Allogastropoda) was typical of shallow-water, muddy bottom communities.

Mesozoic and Cenozoic Eras

The Permian/Triassic crisis affected gastropods as well as all other marine animals. The Euomphalomorpha and Cyrtoneritimorpha (Figures 2 and 3), as well as many groups of Archaeogastropoda, Neritimorpha, and Caenogastropoda, became extinct. During the Triassic, the last members of the Bellerophontida disappeared. The Late Triassic was a time of fast radiation of neritimorphs (Neritopsoidea and Neritoidea), caenogastropods (Ctenoglossa, Cerithimorpha, Architaenioglossa, and Littorinimorpha), and heterobrachs (Allogastropoda and Archaeopulmonata). From Triassic strata, the oldest limpets of the subclass Patellogastropoda are documented. The Patellogastropoda is considered to represent the most ancient gastropod group, but their ancestors (probably bearing coiled shells) have not yet been recognized amongst Palaeozoic gastropods (Figure 9).

The composition of the Jurassic and Early Cretaceous marine gastropod faunas was roughly the same as in the Late Triassic. The characteristic feature of Mesozoic and Cenozoic gastropods was the development of more ornamented shells in most groups, as well as the lesser occurrence of openly coiled shells, by comparison with Palaeozoic gastropods (Figures 1 and 2). Both macro-evolutionary trends have been interpreted as adaptation to increasing predation activities by other animals. During the Cretaceous, more advanced caenogastropod groups (higher Mesogastropoda and Neogastropoda) appeared (Figure 9), which underwent fast radiation and diversification after the Cretaceous/Tertiary faunal crises. Both groups developed the possibility of extending their planktotrophic larval stages and, from the beginning of the Tertiary, they formed one of the dominant groups of marine gastropods. During the Cretaceous, some gastropods (lower Heterobranchia) started to reduce their shells, enabling their adaptation to holoplanktic life (e.g., pteropods). The Early Cenozoic marine gastropod faunas are very similar to extant gastropods in higher taxonomic composition.

Evolution of Freshwater and Terrestrial Gastropods

In contrast with marine gastropods, the fossil record for freshwater and terrestrial forms is less complete, limiting our knowledge of their evolution. Successful invasion to freshwater and land habitats has been closely linked with the mode of gastropod reproduction. External fertilization, which occurs in the ancient Patellogastropoda and Archaeogastropoda, limited them to marine environments. The freshwater and terrestrial environments were colonized by gastropods with egg capsules and internal fertilization (Neritimorpha, Caenogastropoda, and Heterobranchia). Even though members of these groups are known from the Early (Neritimorpha and Caenogastropoda) or Middle (Heterobranchia) Palaeozoic (Figure 9), the first freshwater and terrestrial gastropods are recorded from Late Palaeozoic strata (Archaeopulmonata). The first freshwater Basommatophora appeared during Jurassic time and, in the Cretaceous, the Stylommatophora started their invasion of the land and soon became the most diversified group of terrestrial gastropods.

Glossary

Archaeogastropoda Group of extant gastropods.

Bellerophontida Extinct group of Palaeozoic molluscs with bilaterally symmetrical shells.

Caenogastropoda Group of extant gastropods.

embryonic shell (protoconch I) Gastropod shell formed during embryonic development.

Heterobranchia Group of extant gastropods.

heterostrophic Condition of the protoconch when its whorls coil in the opposite direction to those of the teleoconch.

homeostrophic Protoconch and teleoconch whorls coil in the same direction.

hyperstrophy Condition in which anatomically dextral animals occupy sinistrally coiled shells, and vice versa.

larval shell (protoconch II) Gastropod shell formed during larval development in members of the Neritimorpha, Caenogastropoda, and Heterobranchia.

lecithotrophic Form of development in which larvae use yolk in egg for their nutrition.

Mimospirina Extinct group of Early and Middle Palaeozoic gastropods with sinistrally coiled, homeostrophic shells.

Neritimorpha Group of extant gastropods.

operculum Lid-like structure used for closing of the aperture in gastropod shells.

Opisthobranchia Gastropod subclass of Thiele's classification.

orthostrophy Condition in which anatomically dextral (or sinistral) animals occupy dextrally (or sinistrally) coiled shells.

Patellogastropoda Group of extant gastropods with limpet-shaped shells.

periostracum Outer organic layer of gastropod shells.

planktotrophic Form of development in which free-swimming larvae use planktic organisms for their nutrition.

Prosobranchia Gastropod subclass of Thiele's classification.

protoconch Gastropod shell formed during larval and/or embryonic development.

Pulmonata Gastropod subclass of Thiele's classification.

teleoconch Post-larval gastropod shell.

trochophore larva Gastropod larva formed during early larval development which may be free swimming or occurs in egg capsules.

veliger Gastropod larva formed during later larval development before metamorphosis to post-larval stages.

See Also

Biological Radiations and Speciation. **Evolution**. **Famous Geologists:** Cuvier. **Fossil Invertebrates:** Molluscs Overview; Bivalves; Cephalopods (Other Than Ammonites). **Palaeoecology**. **Palaeozoic:** End Permian Extinctions.

Further Reading

Bandel K (1997) Higher classification and pattern of evolution of the Gastropoda. *Courier Forschungsinstitut* 201: 57 81.

Beesley PL, Ross GJB, and Wells A (eds.) (1998) *Mollusca: The Southern Synthesis, Part B, Fauna of Australia,* vol. 5, pp. 565 1234. Melbourne: CSIRO Publishing.

Bieler R (1992) Gastropod phylogeny and systematics. *Annual Review of Ecology and Systematics* 23: 311 338.

Fretter V and Graham A (1994) *British Prosobranch Molluscs. Their Functional Anatomy and Ecology.* London: Ray Society.

Frýda J and Rohr DM (2004) Gastropoda, 184 195. In: Webby BD, Droser ML, Paris F, and Percival IG (eds.) *The Great Ordovician Biodiversification Event,* p. 408. New York: Columbia University Press.

Knight JB, Cox LR, Keen AM, *et al.* (1960) Systematic descriptions. In: Moore RC (ed.) *Treatise on Invertebrate Paleontology, Part I, Mollusca 1,* pp. I169 I324. Lawrence, KS: Geological Society of America and University of Kansas Press.

Lindberg DR and Ponder WF (2001) The influence of classification on the evolutionary interpretation of structure a re evaluation of the evolution of the pallial cavity of gastropod molluscs. *Organisms, Diversity and Evolution* 1: 273 299.

Peel JS (1991) The classes Tergomya and Helcionelloida, and early molluscan evolution. *Groenlands Geologiske Undersoegelse, Bulletin* 161: 11 65.

Ponder WF and Lindberg DR (1997) Towards a phylogeny of gastropod molluscs: an analysis using morphological characters. *Zoological Journal of the Linnean Society* 119: 83 256.

von Salvini Plawen L (1990) Origin, phylogeny and classification of the phylum Mollusca. *Iberus* 9: 1 33.

von Salvini Plawen L and Haszprunar G (1987) The Vetigastropoda and the systematics of streptonerous Gastropoda (Mollusca). *Journal of Zoology* 11: 747 770.

Taylor JD (ed.) (1996) *Origin and Evolutionary Radiation of the Mollusca.* Oxford, New York, Tokyo: Oxford University Press.

Waren A and Bouchet P (1993) New records, species, genera, and a new family of gastropods from hydrothermal vents and hydrocarbon seeps. *Zoologica Scripta* 22(1): 1 90.

Wenz W (1938 1944) Gastropoda. In: Schindewolf OH (ed.) *Handbuch der Paläozoologie,* p. 1639. Berlin: Borntraeger.

Zilch A (1959 1960) Gastropoda; Teil 2, Euthyneura. In: Schindewolf OH (ed.) *Handbuch der Paläozoologie,* p. 834. Berlin: Zehlendorf.

Cephalopods (Other Than Ammonites)

P Doyle, University College London, London, UK

Introduction

Molluscs are some of the most successful organisms on Earth, and after arthropods they are the most diverse group of invertebrates. In common with all molluscs, cephalopods have developed from a simple molluscan archetype of a conical shell, a muscular foot, a mantle cavity housing gills, and a basic feeding mechanism or radula. In evolving from this structural blueprint, cephalopods have developed a buoyant shell and a differentiated head, with well-developed brain and heightened sensory apparatus, and have modified their foot to create a set of arms, often equipped with hooks or suckers, that capture prey and draw it to a set of beak-like jaws situated at the centre of arm crown.

Cephalopods have a long geological history, which extends back into the Late Cambrian (Figure 1). The search for the earliest cephalopod has created a list of contenders, many of them remarkably like the posited molluscan archetype, but with one significant difference – the first indications of a chambered shell, creating one of the most important characteristics of the cephalopods, the ability to be buoyant or free-swimming. The first tentative formation of chambering presumably gave the nascent cephalopods an adaptive advantage, allowing the otherwise unremarkable mollusc to rise above the seabed to escape predators. This humble beginning marks the start of the radiation of all cephalopods, which were to become important predators of the Palaeozoic and Mesozoic seas, being replaced by the rise of the bony fishes only in the Late Mesozoic.

Cephalopod Morphology

Within the broad envelope of the cephalopods there is a surprising range of morphologies and in some cases body plans. For example, whereas the vast majority of living cephalopods have either internal (endocochleate) shells or no shell at all, most fossil cephalopods are known from the remains of their external (ectocochleate) shells.

Ectocochleate cephalopods have a complex shell built to withstand water pressure and act as a buoyancy aid (Figure 2). Effectively the shell is a cone, and this cone may be coiled, although numerous bizarre shapes are known. The cone-like shell is primarily aragonitic, with an external organic coating or epithelium. Some authorities have suggested that this external shell had a range of pigments in life, and this is certainly true of the surviving species of *Nautilus*, which are characterized by an ornament of reddish brown and white 'stripes'. Some fossil examples, including orthocones from the Early Palaeozoic, display complex zig-zag patterns. In other forms, a range of knobs and ribs ornament the shell.

Internally, the shell is chambered (Figure 2A). There are two basic types of chamber: buoyancy chambers or camerae, and a body chamber housing the mantle cavity, head, and arms when retracted. That part of the shell containing the buoyancy chambers is known as the phragmocone, and within it the camerae are regularly spaced and separated by chamber walls known as septa. The septa are constructed from multiple layers of aragonite, like the remainder of the shell, and at their simplest they have the strong domed shape of a watch glass. In some cases, particularly at the margins, this shape may be modified, and the junction of each septum with the inner wall of the phragmocone – the suture line – may take a variety of shapes, from a simple sinuous curve through to an extremely convoluted shape. The latter form

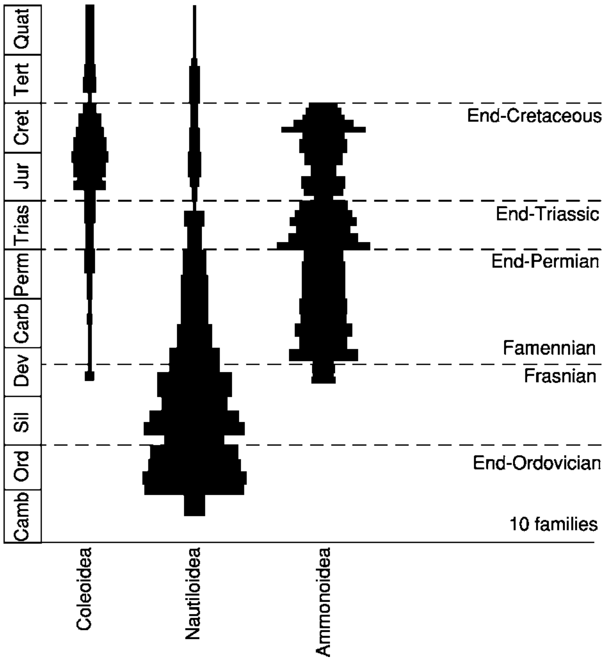


Figure 1 Ranges of the main cephalopod groups. Dotted lines indicate the main Phanerozoic extinction events. (Reproduced from Doyle P (1996) *Understanding Fossils*. Chichester: John Wiley.)

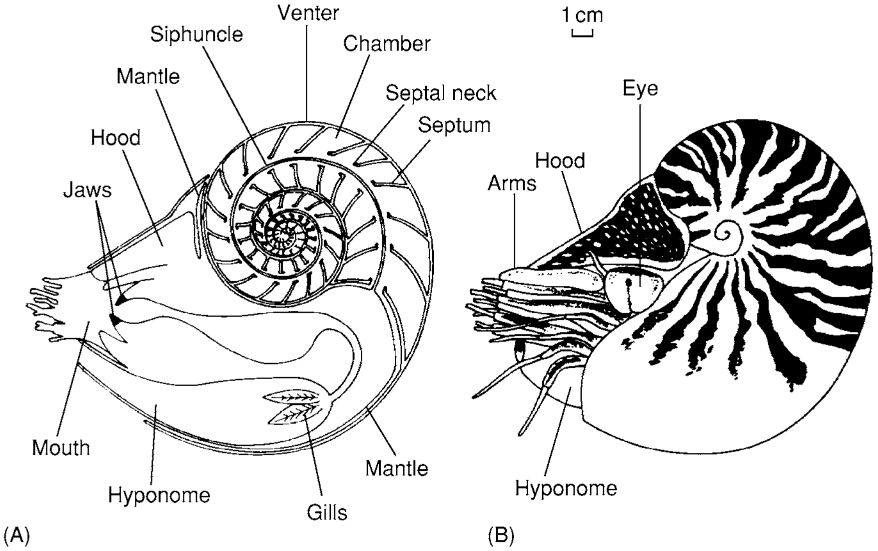


Figure 2 Form and structure of the living *Nautilus*. (Reproduced from Doyle P (1996) *Understanding Fossils*. Chichester: John Wiley.)

is more typical of the Ammonoidea (*see* **Fossil Invertebrates:** Ammonites); simpler sutures are found in the Nautiloidea and, where present, in the Coleoidea. Experiments have shown that the watch-glass shape imparts significant strength to the shell, and it is suggested that the water depth at which implosion will occur may be in excess of 500 m, based on experiments using living *Nautilus*.

The camerae are joined by a continuous porous pipe known as the siphuncle, which passes through each septum by means of a 'bottle-neck' arrangement, the septal neck (Figure 2A). This neck has many forms, most of which point forwards (prochoanitic) or backwards (retrochoanitic), with grades in between, depending on the taxon. Ultimately, the siphuncle joins all the chambers through the septal necks to the body chamber. These septal connectors are often modified with complex structures – difficult to interpret in a functional sense – which are common features of fossil nautiloids. In fact, apart from living *Nautilus*, which has a simple siphuncle, many of the higher nautiloid taxonomic groups are distinguished by such structures. Characteristically, the siphuncle is placed centrally in the living *Nautilus* and several nautiloid groups, but is marginal in other nautiloids and in ammonoids and shell-bearing coleoids.

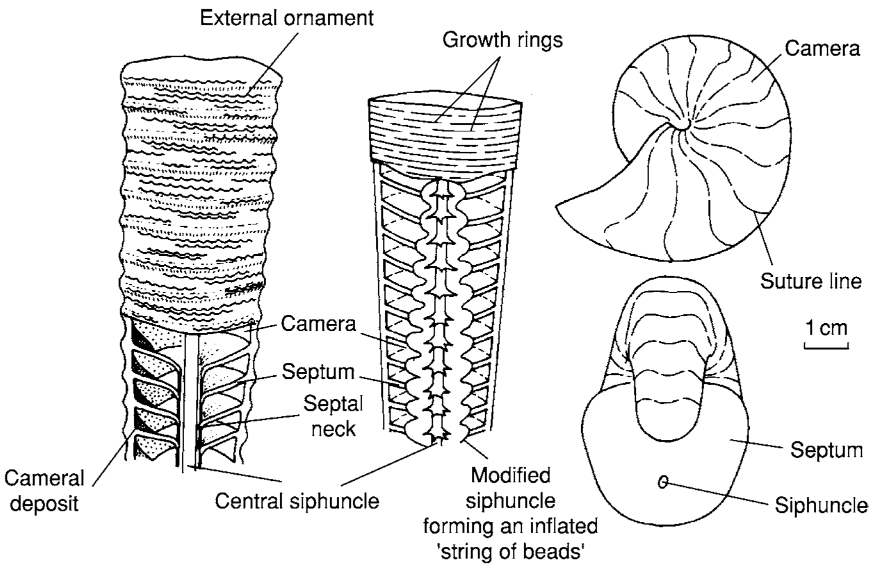
The prime function of the camerae is to provide buoyancy, and it is the role of the siphuncle to regulate the mixture of gas and liquid within each camera in order to create a buoyant state. Regulation of buoyancy is such that during a diurnal cycle, based on recent studies, ectocochleate cephalopods are able to move through the water column. In some cases, the camerae can be modified to improve the buoyancy, with additional shell aragonite being laid down within each chamber – cameral deposits – to aid the balance of the shell. Similar deposits are found within the siphuncles of some fossil nautiloids – endosiphuncular deposits – although, again, they are absent from the relatively uncomplex *Nautilus*.

Cephalopod Classification

The cephalopods are generally treated as a class within the Phylum Mollusca (*see* **Fossil Invertebrates:** Molluscs Overview), but it is here that consensus ends. The class Cephalopoda is further subdivided into subclasses, but there is considerable debate about the relative ranks of the constituent cephalopod groups. For example, following the *Treatise on Invertebrate Paleontology* some authors propose a division of the Nautiloidea into four separate subclasses in their own right: Endoceratoidea, Actinoceratoidea, Orthoceratoidea, and Nautiloidea (in a restricted, yet still very broad, sense). A simple classification, involving subdivision into three more-or-less equal groups, is employed here.

Class Cephalopoda Cuvier, 1797 (Cambrian–Holocene). Fully marine free-swimming self-buoyant molluscs with a differentiated head and well-developed sensory apparatus.

Subclass Nautiloidea Agassiz, 1847 (Cambrian–Holocene). Cephalopods with coiled or straight (orthocone) ectocochleate shells with simple chamber walls; the siphuncle is often complex, marginal to central; living examples have multiple arms; tetrabranchiate.



(A) *Dawsonoceras* (B) *Antinoceras* (C) *Cenoceras*

Figure 3 Fossil nautiloids: (A) the Silurian orthocerid *Dawsonoceras*, (B) the Ordovician actinoceratid *Actinoceras*, (C) and the Jurassic nautilid *Cenoceras*. (Reproduced from Doyle P (1996) *Understanding Fossils*. Chichester: John Wiley.)

Subclass Coleoidea Bather, 1888 (Devonian–Holocene). Cephalopods with endocochleate shells, which are chambered in some and reduced or absent in others; ten or eight arms grouped around the mouth; dibranchiate.

Subclass Ammonoidea Zittel, 1884 (Devonian–Cretaceous). Cephalopods with mostly coiled ectocochleate shells with complex suture walls; siphuncle marginal; number of arms uncertain; number of gills uncertain, presumed tetrabranchiate. This group is dealt with in detail elsewhere in this volume (*see* **Fossil Invertebrates:** Ammonites) and includes the earliest ancestors, the Bactritida.

Subclass Nautiloidea

The Nautiloidea is an important fossil group, represented today by the single living genus *Nautilus*, with its six known species or subspecies, compared with an estimated 11 000 fossil examples. This 'living fossil' is in effect the lone survivor of a group that achieved its maximum diversity in the Palaeozoic, declining through the Mesozoic to the present day (Figure 1). The Nautiloidea are representative of the first true cephalopods, and exhibit a rapid increase in diversity following widespread extinctions at the close of the Cambrian. Following the Ordovician–Silurian maximum, the nautiloids were in steady decline, such that, after the extinctions at the Triassic–Jurassic boundary, only two lineages remained extant; they survived until further extinctions at the Miocene–Pilocene boundary left the few species we know today.

The living genus *Nautilus* is one of the most powerful tools for interpreting the morphology and life history of the nautiloids, although this approach is fraught with difficulties given that *Nautilus* is the sole survivor of a rich diversity of forms from the geological past. This is particularly relevant given the range of form in such aspects as the siphuncle, particularly its size, form, and structure (Figure 3); modern-day *Nautilus*, in common with other members of its restricted group, has only simple structures, which are inadequate to explain the full diversity of the structures found in fossil representatives. The nautiloid shell is ectocochleate and is known in all forms from a straight cone (orthocone), through a curved cone (cyrtocone), to the typical tight planispiral coil of the living *Nautilus*.

Significant structural features of the nautiloid shell are the forms of the siphuncle, its septal necks, and connecting rings, with the connectors being highly variable, including everything from simple tubes (as in *Nautilus*) to highly complex examples with cone structures and endosiphuncular deposits (as in the oncocerids). The typical position of the nautiloid siphuncle is central or subcentral to the camerae, but again this is highly variable, with marginal siphuncles apparent in a number of groups (Figure 3). Its relative worth in taxonomic studies is still under active debate. Cameral deposits and endosiphuncular deposits also vary within the group.

The higher classification of the Nautiloidea is a strongly debated area, hinging mostly on the taxonomic position of the Endoceratidea, the

Actinoceratidea, the Orthoceratidea, and the Nautiloidea proper, which have been treated variously as subclasses, superorders, and orders. Here they are treated as natural but informal groups of sub-subclass rank.

Endoceratids This group comprises mostly medium-sized orthoconic shells, with some rarer cyrtoconic examples. Siphuncles are large and, unusually, marginal rather than central. Septal necks are of variable lengths, generally getting longer with time. The endoceratids typically have deposits within the siphuncle itself, often in the form of conical sheaths. The group is known from the Lower Ordovician to the Silurian, although it is commonest in Lower–Middle Ordovician deposits.

Actinoceratids Actinoceratids are generally orthoconic with short septal necks and a broad siphuncle. The siphuncle segments between the septal necks display many specialized structures, as illustrated by *Actinoceras* itself (Figure 3). Most actinoceratids have cameral deposits. They are known from the Middle Ordovician to the Upper Carboniferous, but are commonest in the Middle and Upper Ordovician.

Orthoceratids Orthoceratids are characterized by their orthoconic or weakly crytoconic slender shells. The genus *Orthoceras* is typical, but this group has been used by many authors to house many similar forms. The orthoceratids have narrow siphuncles, and they typically have both cameral and endosiphuncular deposits, as in *Dawsonoceras* (Figure 3). Orthoceratids are known from the Lower Ordovician to the Triassic, but are commonest in the Middle Ordovician–Silurian.

Nautiloids Nautiloids are diverse, with a great variety of shell forms, from orthoconic to tightly coiled. The siphuncle is variable in position, form, and size. Siphuncular and cameral deposits are similarly variable, present in some and absent in others. The earliest representatives include ellesmerocertids, which are cyrtoconic with marginal siphuncles and found in Upper Cambrian to Upper Ordovician rocks; oncocerids, which have generally short (breviconic) shells in a variety of shapes, usually with thin septa, recumbent septal necks, and expanded siphuncles, usually with endosiphuncular deposits, and are found in Middle Ordovician to Lower Carboniferous rocks; discordids, which have short stout cyrtoconic shells and thick siphuncles with endosiphuncular deposits, and are found in Middle Ordovician to Devonian rocks; tarphycerids, which consist of coiled shells with broad siphuncles and are found in Lower Ordovician to Upper Silurian rocks; and, finally, the nautilids themselves, which have a wide variety of coiled shells in which the siphuncle is narrow and cylindrical, and which range from the Early Devonian to the Holocene, some with a close similarity to *Nautilus*, such as the Jurassic genus *Cenoceras* (Figure 3; compare with Figure 2).

Subclass Coleoidea

The Coleoidea is a diverse group of cephalopods which are united by the possession of an internal or endocochleate shell. The vast majority of cephalopod species alive today are coleoids, and most of these have a greatly reduced internal shell, a ‘torpedo’ body shape, a well-developed head with chitinous beaks, and two gills in the mantle cavity. The geological history of the Coleoidea has been a subject of debate: the earliest definite coleoid is of Carboniferous age, although there are some doubtful records in the Devonian (Figure 1).

The diversification of the Coleoidea took place in the Mesozoic. Three superorders, based on aspects of the shell and soft parts, are recognized, although there is still considerable debate about the higher subdivisions of the group as a whole. The three superorders are Belemnoidea (characterized by ten subequal arms and a multi-layered conotheca), Octobrachia (characterized by ten arms in which the second pair is modified or lost), and Decabrachia (characterized by ten arms in which the fourth pair are modified as tentacles). Representatives of each of these groups have substantial internal skeletons.

The groups considered here are Superorder Belemnoidea (aulacocerids and belemnites), Superorder Decabrachia (spirulids, sepiids, and teuthids), and Superorder Octobrachia (vampyromorphs and octopods).

Superorder Belemnoidea This group comprises ten-armed cephalopods with an internal shell consisting of a multiple-layered phragmocone contained within a concentrically layered structure known as the rostrum (sometimes referred to as the telum in aulacocerids) (Figure 4). This group first appeared in the Devonian and became extinct at the end of the Cretaceous. It contains two major orders: the Aulacocerida, characterized by an entire body chamber and an aragonitic or organic rostrum; and the Belemnitida, characterized by a ventrally open body chamber – the pro-ostracum – and a mostly calcitic rostrum (Figures 4 and 5). Belemnites with soft-parts preserved have been discovered in Jurassic rocks, particularly the Lower Jurassic of southern Germany (*Acrocoelites*, *Passaloteuthis*) and the Middle Jurassic of England (*Belemnotheutis*). These specimens allow the

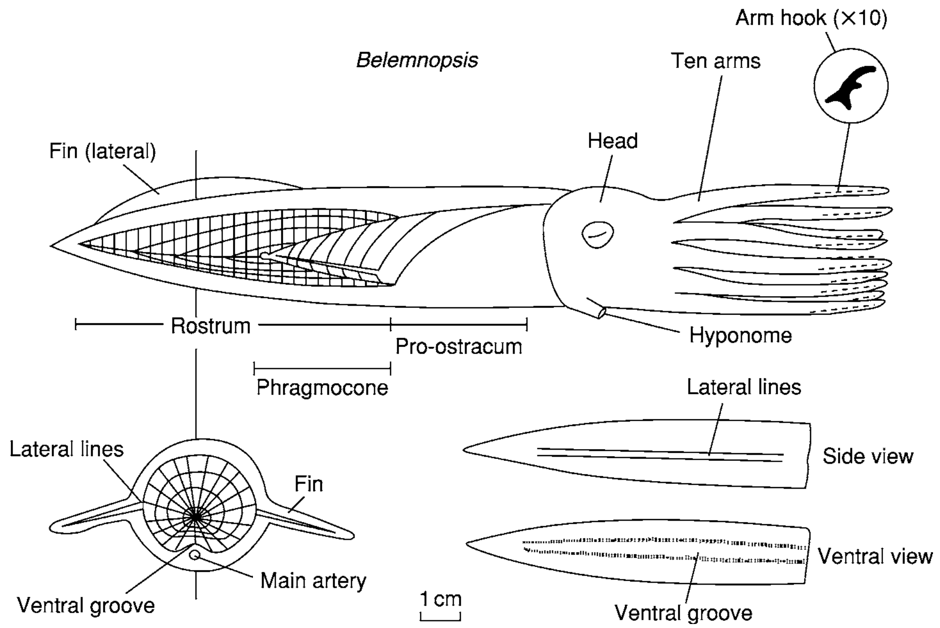


Figure 4 The belemnite shell, based on the Jurassic Cretaceous genus *Belemnopsis*. (Reproduced from Doyle P (1996) *Understanding Fossils*. Chichester: John Wiley.)

interpretation of the form and function of the belemnoid animal and in particular the nature of the proportional relationships between the shell and the soft body parts.

The belemnoid rostrum has a diverse structure and shell chemistry, but it is characterized by concentric incremental growth, the rostrum developing posteriorly from the first-formed primordial rostrum, associated with the protoconch. Continued growth leads to the gradual envelopment of the phragmocone in a conical cavity, the alveolus (Figure 4). The overall similarity in form, ontogeny, and function of all belemnoid rostra highlights the probable evolutionary relationship between the aulacocerid ancestors and belemnitid descendants.

The Aulacocerida is the most primitive group of the Belemnoidea. It first appeared in the Devonian or Carboniferous, and persisted until the Jurassic. Aulacocerids are often distinguished in their host rocks by the presence of large phragmocones, often with a low apical angle (belemnitids usually have a greater apical angle). Phragmocones complete with their outer wall (conotheca) display growth lines that indicate the presence of a complete body chamber, which distinguishes them from the Belemnitida, which have a ventrally open vestige of a body chamber, the pro-ostracum (Figures 4 and 5). Aulacocerid rostra are usually aragonitic and may be extensively furrowed or smooth, leading to some confusion with true belemnites.

The Belemnitida probably evolved from the Aulacocerida in the Early Jurassic, and belemnites are common fossils in Mesozoic sedimentary sequences. However, there are belemnite records from earlier Mesozoic (Triassic), Palaeozoic (Carboniferous–Permian), and Palaeogene rocks, although these are probably representative of other coleoid and even non-cephalopod groups. The belemnite rostrum typically consists of a concentric arrangement of growth rings, which alternate between organic-rich and organic-poor calcareous layers. The original mineralogy of the concentric layers has been a subject of some debate, with both low-magnesium calcite and aragonite being favoured, while some belemnites have been described with a primarily organic rostrum, but these are in need of further study. It is now generally held that the belemnite rostrum is composed of original low-magnesium calcite. However, the first-formed (primordial) rostrum of at least one taxon (*Hibolithes*) is originally aragonitic, and the sheath-like rostrum of the Jurassic belemnite *Belemnotheutis* is also known to be composed of primary aragonite.

The belemnite rostrum displays a variety of shapes and sizes, but most are based on three simple shapes: conical (e.g. *Acrocoelites*), cylindrical (e.g. *Cylindroteuthis*), and club/spear shaped (variously described as hastate, clavate, or lanceolate; e.g. *Hibolithes*), and these have influenced classification. In some rare cases belemnites appear blade-like (e.g. *Duvalia*). The rostrum displays three basic patterns of grooves and furrows (Figure 4): indistinct lateral 'lines' (thought to be the locations of fins), ventral and/or dorsal grooves (thought to be the locations of

Figure 5 Well preserved Jurassic belemnite *Cylindroteuthis*, with rostrum, phragmocone and pro ostracum. (Reproduced from Doyle P (1996) *Understanding Fossils*. Chichester: John Wiley.)

major blood vessels), and vascular marks (found only in belemnotellids and indicative of fine blood vessels). The primary function of the rostrum has traditionally been considered to be a counterweight, a support for musculature and/or fins, a protection for the phragmocone, or a streamlining for the backwards motion of the animal. The majority of authors favour the counterbalancing hypothesis for both belemnites and aulacocerids, and the existence of club-like rostra in particular is relevant to this hypothesis, as the club shape shifts mass posteriorly, thereby increasing the counterbalancing potential.

Superorder Decabrachia The Superorder Decabrachia comprises ten-armed coleoids in which the fourth pair of arms are modified as tentacles for the efficient capture of prey. The group is currently divided into three main orders: the Spirulida, Sepiida, and Teuthida. These encompass the living shelled cephalopods *Spirula* and *Sepia*, in which the shell is plesiomorphic, and the squids, most of which have the apomorphy of reduced shells. The geological history of the group stretches back to the Jurassic, the oldest known representative being the squid *Plesioteuthis* from the Late Jurassic, while the genera *Groenlandibelus* and *Naefia* are the oldest known representatives of the shelled spirulids.

The living genus *Spirula* is characterized by a loosely coiled chambered phragmocone, which is supported by the internal tissues of the body and which provides primary buoyancy. The life attitude of this genus approximates to vertical, with the head and arms hanging from the buoyant posterior, which contains the phragmocone. There are few traces of a rostrum or sheath, or cameral deposits, both of which would act as a counterbalance for the buoyant phragmocone. The geological record of the Spirulida extends back to the Cretaceous. Fossil representatives of the group demonstrate that the coiling of the spirulid phragmocone and its lack of a rostrum are apomorphic: the majority display only weak endogastric curvature and an aragonitic sheath-like rostrum.

The Recent genus *Sepia* has attracted much interest with regard to its mechanism for buoyancy, but fossil sepiids are less well known. The sepiid shell is entirely aragonitic and comprises a dorsal shield, which encompasses the ventrally open phragmocone; the multiple septa are closely spaced, buttressed by a series of pillars that demonstrate annular incremental growth, and the siphuncle is open to the internal tissue of the cuttlefish. The sepiid shell is fragile, and the septa are rarely preserved except in exceptional circumstances, and this has led to a number of debates about the relative antiquity of the group. Current estimates suggest that true sepiids extend as far back as the Cretaceous (with the Actinosepiidae), although claims that the Late Jurassic genus *Trachyteuthis* (Figure 6B) is a sepiid have recently been refuted. The evolutionary development of the sepiids can be traced to spirulid ancestors, with progressive loss of a ventral margin to the shell.

Living teuthids, encompassing many genera of squid, are characterized by an organic gladius, composed of β-crystalline chitin, which has no vestige of a phragmocone structure. The development of the gladius led to the loss of primary buoyancy: squids rely entirely upon the motion of the body for support in the water column. The geological history of the Teuthida extends back to the family Plesioteuthididae, known from the Jurassic. This fossil teuthid has a calcitic gladius, although some geochemical investigations have demonstrated a composition of primary francolite. Although some authors have suggested that this family is in fact a representative of the Vampyromorpha, other studies have indicated the homology of the fossil gladii of this family with living teuthids.

Superorder Octobrachia The Superorder Octobrachia comprises ten-armed coleoids in which the

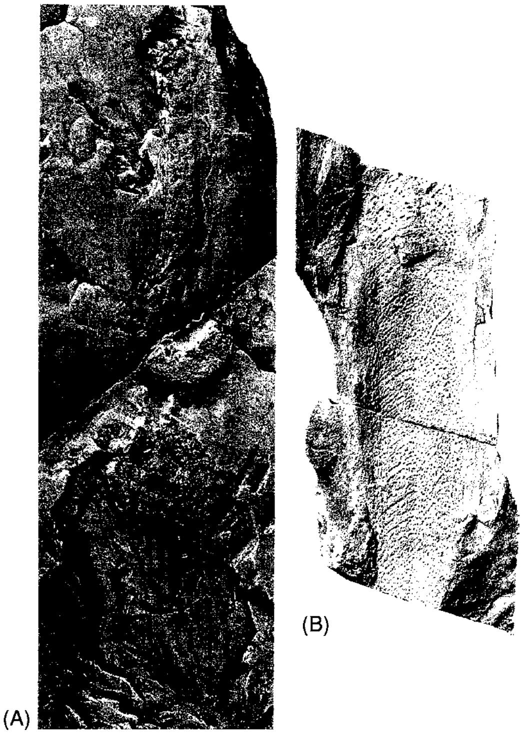


Figure 6 Jurassic vampyromorphs: (A) the Lower Jurassic loligosepiid *Jeletzkyteuthis*, complete with ink sac; and (B) the Upper Jurassic *Trachyteuthis*. (Reproduced from Doyle P (1996) *Understanding Fossils*. Chichester: John Wiley.)

second pair of arms is either modified or lost and there is no trace of a phragmocone forming a primary buoyancy aid. The Octobrachia includes the octopods, with little or no shell, the archaic order Vampyromorpha, and fossil representatives of the Loligosepiina and related families (Teudopsidae and Trachyteuthididae) (Figure 6A). These fossil groups are characterized by a calcareous shell, although some authors have argued for primary francolite and a phosphatic shell.

The geological history of the Octobrachia extends back to at least the Triassic with the first Loligosepinids. However, *Pohlsepia mazonensis* from the Carboniferous Mazon Creek Lagerstätte may, despite its ten arms, be the earliest known octopod. This view is based on the facts that it does not possess a well-defined head, it has a modified arm pair, and it has a sac-like body. *Palaeoctopus* from the Late Cretaceous of Lebanon is an undoubted octopus and has much in common with *Pohlsepia*; it has no identifiable shell, eight subequal arms (with suckers), a sac-like body, a poorly defined head, and free posterior fins.

The living genus *Vampyroteuthis* is an imperfectly known deep-sea coleoid that combines the octobrachian character of eight arms (the second pair of the normal ten-armed configuration being lost) with the essentially teuthid character of an internal gladius, which is not seen in other octobrachians. Like the teuthid gladius, the vampyromorph gladius is composed of β-crystalline chitin, and it corresponds in form to the ladle-like gladius, providing support for the visceral mass of the animal. Fossil representatives of the Teudopsidae and Trachyteuthididae have similar gladii, presumably with the same function. The Loligosepiina, also recognized as possessing vampyromorph affinities, have an arrow-head gladius, which may have had a role in positioning the head during its piston movements on swimming.

See Also

Fossil Invertebrates: Molluscs Overview; Bivalves; Gastropods; Ammonites. **Mesozoic:** Jurassic; Cretaceous.

Further Reading

Bandel K and Leich H (1986) Jurassic Vampyromorpha (dibanchiate cephalopods). *Neues Jahrbuch für Geologie und Paläontologie Monatshefte* 1986: 129 148.

Bandel K, Engeser T, and Reitner J (1984) Die Embryonalentwicklung von Hibolithes (Belemnitida, Cephalopoda). *Neues Jahrbuch für Geologie und Paläontologie Abhandlungen* 167: 275 303.

Donovan DT (1977) Evolution of the dibranchiate Cephalopoda. *Symposis of the Zoological Society of London* 38: 15 48.

Doyle P (1993) Mollusca, Cephalopoda (Coleoidea). In: Benton MJ (ed.) *The Fossil Record 2,* pp. 229 236. London: Chapman and Hall.

Doyle P (1999) The functional morphology of the coleoid shell. In: Savazzi E (ed.) *The Functional Morphology of the Invertebrate Skeleton,* pp. 327 347. Chichester: John Wiley.

Doyle P, Donovan DT, and Nixon M (1994) Phylogeny and systematics of the Coleoidea. *University of Kansas Paleontological Contributions, New Series* 5: 1 15.

Dzik J (1984) Phylogeny of the Nautiloidea. *Palaeontologica Polonica* 43: 1 219.

Engeser T (1988) Fossil 'octopods' a critical review. In: Clarke MR and Trueman ER (eds.) *The Mollusca, Volume 12. Paleontology and Neontology of Cephalopods,* pp. 81 87. San Diego: Academic Press.

Engeser T (1990) Phylogeny of the fossil coleoid Cephalopoda (Mollusca). *Berliner Geowissenschaftlichen Abhandlungen A* 124: 123 191.

Engeser T (1998) Fossil Coleoidea Page. http://userpage.fu berlin.de/~palaeont/fossilcoleoidea/welcome.html.

Engeser T (1988) Fossil Nautiloidea Page. http://userpage.fu berlin.de/~palaeont/fossilnautiloidea/fossilnautpage.html.

Engeser T and Bandel K (1988) Phylogenetic classification of coleoid cephalopods. In: Wiedmann J and Kullmann J (eds.) *Cephalopods present and past,* pp. 105 115. Stuttgart: E. Schweizerbart'sche verlagsbuchhandlung.

Fischer AG and Teichert C (1969) Cameral deposits in cephalopod shells. *University of Kansas Paleontological Contributions* 37: 1 30.

Holland CH (1987) The nautiloid cephalopods, a strange success. *Journal of the Geological Society, London* 144: 1 15.

Jeletzky JA (1966) Comparative morphology, phylogeny and classification of fossil Coleoidea. *University of Kansas Paleontological Contributions, Mollusca* 7: 1 162.

King A (1993) Mollusca, Cephalopoda (Nautiloidea). In: Benton MJ (ed.) *The Fossil Record 2*. London: Chapman and Hall.

Kluessendorf J and Doyle P (2000) *Pohlsepia mazonensis*, an early 'octopus' from the Carboniferous of Illinois, USA. *Palaeontology* 43: 919 926.

Macleod N (ed.) (2003) *Paleobase Macrofossils Part 2.0, Ammonoids, Bivalves, Coleoids, Gastropods and other Mollusca.* Oxford: Blackwell.

Riegraf W and Hauf R (1983) Belemnitenfunde mit Weichkörper, Fangarmen und Gladius aus dem Untertoarcium (Posidonienschiefer) und Unteraalenium (Opalinuston) Südwestdeutschlands. *Neues Jahrbuch für Geologie und Paläontologie Abhandlungen* 165: 466 483.

Sælen G (1989) Diagenesis and construction of the belemnite rostrum. *Palaeontology* 32: 765 798.

Teichert C (1988) Main features of cephalopod evolution. In: Clarke MR and Trueman ER (eds.) *The Mollusca, Volume 12. Paleontology and Neontology of Cephalopods.* London: Academic Press.

Yochelson EL, Flower RH, and Webers GF (1973) The bearing of the new later Cambrian monoplacophoran genus *Knightoconus* on the origin of the Cephalopoda. *Lethaia* 6: 275 310.

Ammonites

G E G Westermann, McMaster University, Hamilton, ON, Canada

Introduction

The term 'ammonite' (from genus *Ammonites*, Latin = lover/team of Ammon) was derived from *cornu Ammonis* (Latin = Horn of Ammon) because of its resemblance to the curved horns of the ram-headed sphinxes at the Roman temple of Jupiter Ammon in the Ammon oasis of Libya, a latinized form of the much older Egyptian temple Amun-Re at Thebes.

'Ammonites' is the popular name for the shells of the extinct Subclass Ammonoidea (Class Cephalopoda), which lived from Devonian to Cretaceous times. They are the most important fossils for Mesozoic biostratigraphy and biogeography. The animal that secreted the protective, external shell presumably had ten arms, two of which may have been modified into tentacles, but there is no direct evidence for the structure of the soft parts. The body chamber ('living chamber') housed the soft parts, whereas the chambered, essentially empty phragmocone provided the uplift of the neutrally buoyant organism. Some ammonites resemble the shell of the extant *Nautilus*, but all differ from the Subclass Nautiloidea by their internal structure, i.e., folded septa and a marginal, thin siphuncle. *Nautilus* shell is structurally very similar to ammonite shell.

This article treats, in sequence, ammonite shape and architecture, phylogeny, growth and sexual dimorphism, buoyancy and poise, functional morphology, and ecology.

Shape and Architecture

Shell size is usually from 5 to 30 cm, but ranges from 1 to 150 cm. The most common shape is ammonitocone, a closed planar spiral with involute to evolute coiling and varying whorl section (Figures 1A–C). Later ammonoids included a great variety of heteromorphs (= other shapes), including orthocones (straight), ancylocone (hooked), gyrocones and cyrtocones (open-curved), and helicones (snail-like). All may be smooth or ornamented with ribs (costae/plicae), tubercles, or spines, often changing with growth. Prominence of ornament tended to be higher in inflated than in compressed forms and increased during ammonoid evolution. Ribs may be single, branching, intercalating, or fasciculating (bundled). Nodes are rounded and spines prominent. Constrictions are periodic, deep folds.

The shell consists of high-strength nacre (mother-of-pearl), a bicomponent material of aragonite platelets separated by organic sheets. Fully grown, adult shells are recognized by the shape of body chamber and peristome. The former is often modified by partial uncoiling and inflation, both lowering the centre of mass; the peristome usually thickens with age, and may carry lateral lappets or a ventral rostrum. Other ammonoids grew until death.

Internally, the shell is divided into the phragmocone providing buoyancy and the body chamber protecting the animal. The phragmocone has transverse

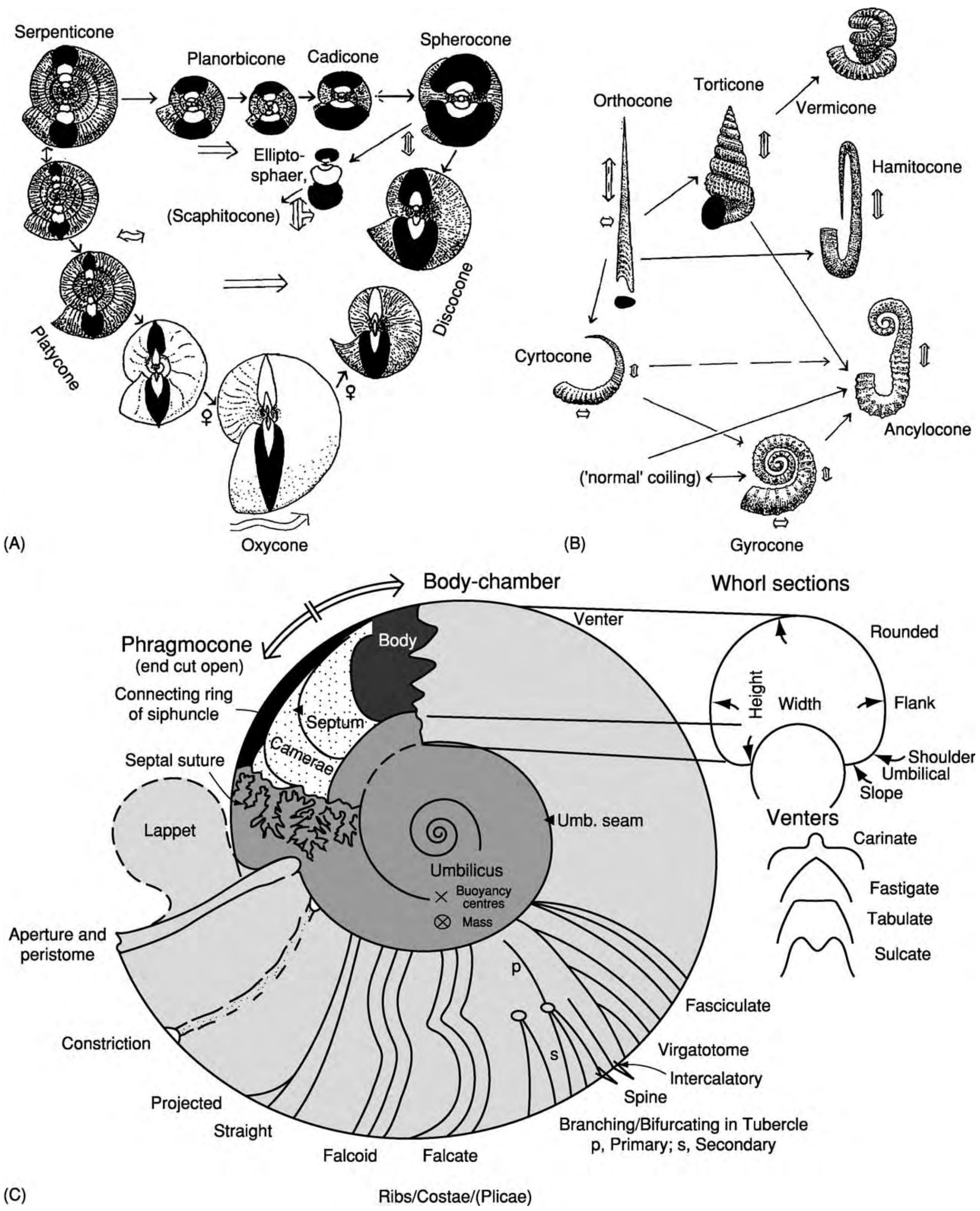


Figure 1 Ammonite morphology. (A) Planispiral shell shapes, illustrating the morphologic continuum (adapted from Westermann (1996)). (B) Major types of heteromorphs (adapted from Westermann (1996)). (C) Terminology of the ammonite shell; cut open at upper left to show interior of phragmocone.

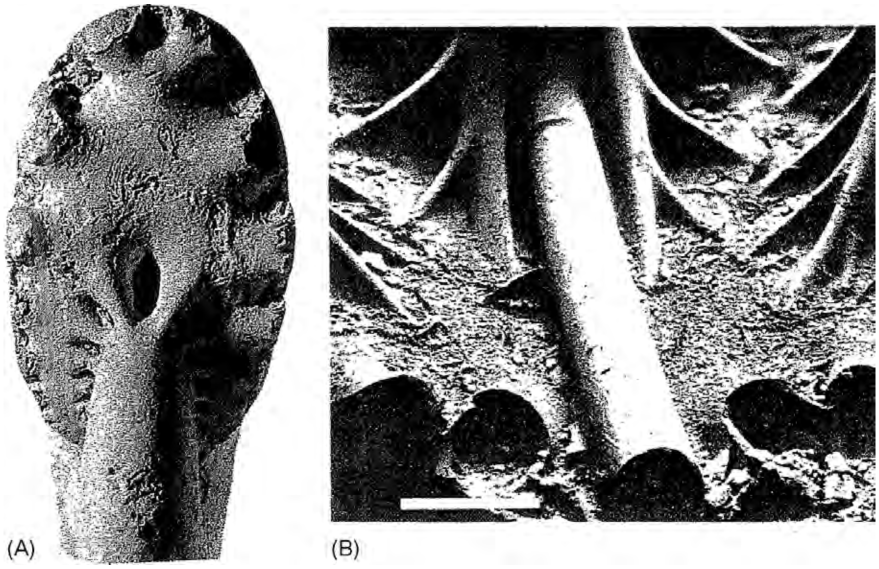


Figure 2 Ammonite septum and siphuncle. (A) Anterior view of the last septum of a Jurassic phylloceratid at natural size; from the central, flat region the flutes increase in prominence towards the margin but decrease on the narrow whorl overlaps; the positive flutes are the saddles, the negative ones the lobes. (B) Frontal view into an empty chamber of another Jurassic phylloceratid, scale bar 1 mm; shown are the siphuncle and 3rd and 4th order 'frills' of two septa that attach to the wall, forming the medium part of the external lobe. (Adapted from Westermann and Tjujita (1999).)

septa that separate the camerae. As in all living cephalopods, these contained gas at less than atmospheric pressure, together with small amounts of water (cameral liquid). The ventro-marginal siphuncle (but dorso-marginal in Clymeniida), with phosphatised organic connecting rings, connected all camerae to the body and once contained the siphon.

Ammonoid septa are folded into a series of anticlastic, concavo-convex arches or 'flutes' (Figure 2), which abut the wall in the sutures ('suture lines') consisting of a series of adorally concave lobes separated by convex saddles (Figure 3). Sutures with undivided lobes and saddles are called goniatitic and usually are of either Z type (for 'zigzag') or M type (for 'meander'). Both reduced the free (unsupported) wall space, the Z type by large, subtriangular lobes and saddles that telescope with neighbouring sutures; the M type by more numerous lobes and saddles with parallel sides, without overlap. Ammonitic sutures developed when the original (primary) flutes formed secondary flutes, called lobules and folioles, during evolution or ontogeny; further (3rd- and 4th-order) subdivision of lobules and folioles are typical of later ammonoids. This resulted in 'complex' sutures with fine 'frills'. Significantly, size and spacing of the sutural elements are highly organised, providing improved wall support compared with goniatitic sutures. Sutures have been used in classification for two centuries, and their function is much debated (see below).

The only commonly preserved structures not attached to the shell are the aptychi (Figure 4), which are strongly calcified lower jaws. Aptychi are found mainly in Late Mesozoic ammonites and consist of paired plates hinged together by the organic 'horny' layer forming the ancestral jaw. They more or less fitted the whorl cross-section near the aperture and acted as opercula or protective lids. During active periods they could be folded and withdrawn for possible duplicate use as lower jaws.

Phylogeny

The evolution of the Subclass Ammonoidea (Figure 5) began with slender orthocones and cyrtocones, the long-ranging (Silurian-Triassic) Bactritoidea, which evolved into Palaeozoic nautiloids and lived as plankton with a vertical poise. From them evolved the Devonian Anarcestida, which retained orthoconic innermost whorls before coiling and had sinuous to goniatitic-Z sutures. The Goniatitida were mostly smooth spherocones and discocones with goniatitic-Z sutures that became goniatitic-M and ammonitic near their Permian extinction. The short-lived, Late Devonian Clymeniida, with goniatitic-Z sutures and diverse shapes including subtriangular coiling, were unique in their dorso-marginal siphuncle. The Prolecanitida were commonly smooth, platyconic, and discoconic brevidomes, suggesting nektic habitats, with goniatitic-M or ceratitic sutures. They were ancestral

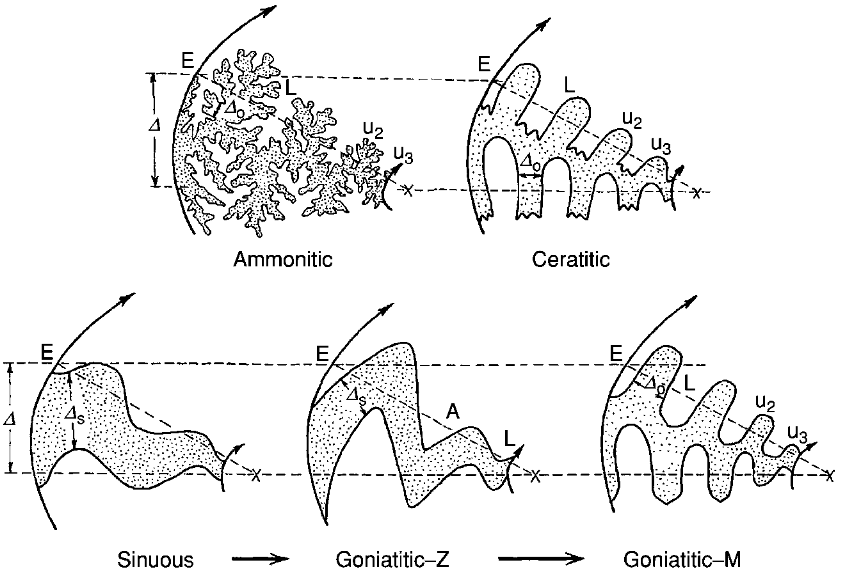


Figure 3 Schematic illustration of the major sutural types according to their function as wall supports, i.e., minimizing unsupported shell wall (between sutures Δ_s; between parts of same suture Δ_o), with septal spacing (Δ (Δ)) constant. Arrows indicate evolutionary trends. The lobes from venter to umbilical seam are: E, external; L, lateral; U, umbilical; A, adventive. Saddles are named after the lobes forming them, e.g., E/L saddle. (Adapted from Westermann (1975).)

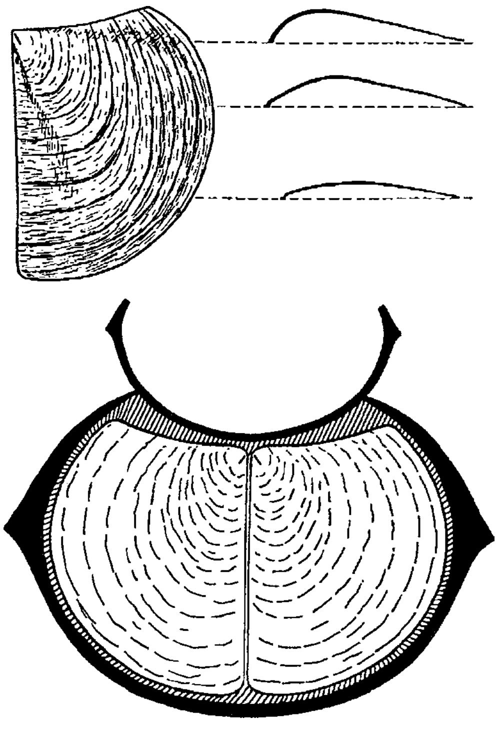


Figure 4 Jurassic aptychus (*Praestriaptychus*) and its reconstruction as a hinged pair in the aperture of the associated ammonite (*Stephanoceras*, Ammonitida); note excellent fit. Space at dorsal overlap presumably enabled water intake (3X).

to all Mesozoic ammonoids, beginning with the mainly Triassic Ceratitida, which had ceratitic to ammonitic sutures, closely followed by the long-ranging Phylloceratida, with broadly rounded saddles or folioles. In the earliest Jurassic, the Prolecanitida gave rise to the Lytoceratida and Ammonitida; the former distinguished by bipartite (not tripartite) saddles and lobes; the latter being the most common, coiled and often ornamented ammonites with complex sutures. Near the end of the Jurassic arose the last ammonoid order, the Ancyloceratida, probably also from the Lytoceratida. They included a host of heteromorphs and, remarkably, began with uncoiled forms. All four orders became extinct at the great end-Cretaceous mass-extinction.

Growth, Longevity, and Sexual Dimorphism

Growth of the shell began with the ammonitella in the minute egg (1–2 mm). After hatching, the shell grew by terminal secretion at the persistome, so that all growth stages are preserved in the fully grown shell. During periodic growth, the posterior mantle secreted cameral liquid in front of the last septum, followed firstly by a membrane and secondly by the nacreous septum. On completion, the (incompressible) liquid, possibly a gel as in the cuttlefish *Sepia*, was withdrawn through the siphuncle into the body. Salinity differences in the epithelium of the vascular siphon generated sufficient osmotic pressure to transport the fluid through the porous connecting rings into the body cavity, against the hydrostatic pressure of the ambient seawater. The new camera was at first

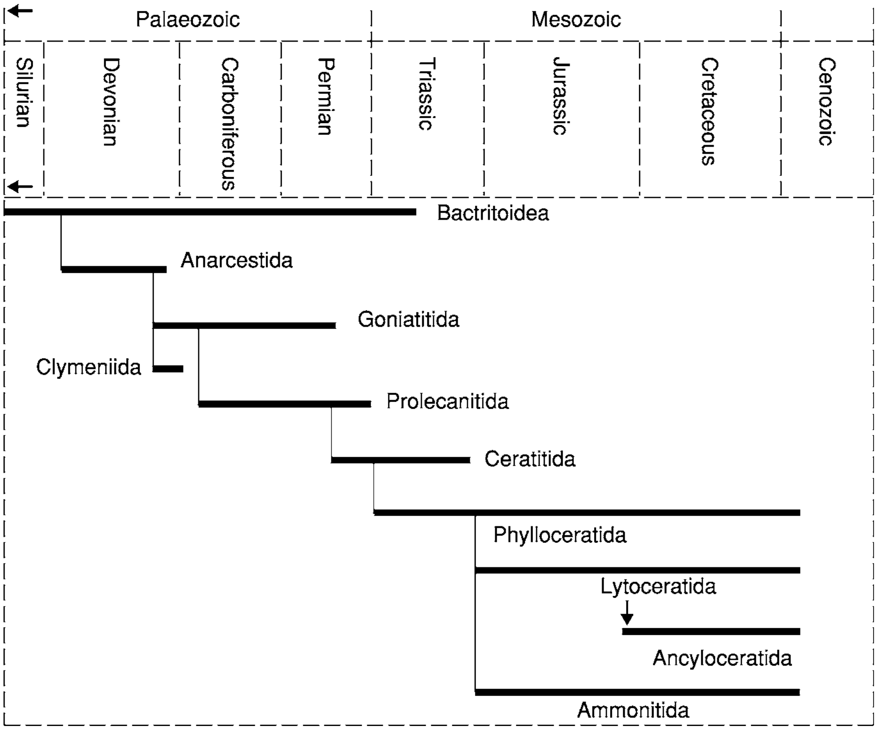


Figure 5 Phylogeny of the Ammonoidea orders.

under vacuum which was later partially replaced by gas at up to 1 atmosphere. Thus, the entire ambient pressure acted (directly) on the outer phragmocone wall and (indirectly through the soft body) on the latest/adoral septum risking its implosion, as well as on the organic connecting rings risking their explosion.

Growth of the pre-adult shell was mainly exponential (secretion time of each whorl increasing with the expansion rate W), followed by a slowing down, as indicated by the more closely spaced (approximated) last several septa. The total duration of growth depended on: (i) size; (ii) shell thickness, especially of the septa; (iii) diameter and wall thickness of the connecting rings: liquid transport depends on surface area and permeability; (iv) habitat depth: ambient pressure slowed cameral emptying; and (v) temperature: shell secretion is more rapid in warm than in cold water. The oceanic *Nautilus*, living at 200–400 m depth, is useful for comparison: at 20–25 cm adult diameter, the thickness of wall and septa are about 1 mm, and the connecting rings are narrow and thick-walled (i.e., strong); growth takes 10–15 years; adult/gerontic life is several more years. Most ammonoids were smaller or similar in size, had thinner septa and shell, broader and thinner-walled (i.e., weaker) connecting rings, and lived in mostly warm epeiric seas or as oceanic epiplankton at 30–150 m; growth took from 2 to 7 years. At the other extreme, some oceanic ammonoids of similar size lived at 250–500 m depth (e.g., mesopelagic lytoceratids, phylloceratids, desmoceratids), resembled *Nautilus* in shell and siphuncle properties, and presumably grew at similar slow rates. Some ammonoids, however, were much larger and became proportionately older, perhaps 50 years or more. Several more years of adult/gerontic life are indicated for ammonites with strongly modified peristomes. But in some ammonoids, especially among oceanic forms, growth was indeterminate up to death and their size became gigantic (‘megaconchs’).

For a century and a half, consistently associated pairs of distinctly modified shells have been observed in many Jurassic and Cretaceous Ammonitida and interpreted as sexual dismorphism. In extreme cases, the two forms, called antidimorphs, differed so strongly that they have been placed in different families. The most obvious distinction is in adult size, with the larger shell or macroconch usually two to three times larger than the smaller shell or microconch (Figure 6). Because in living animals with size ratios between the sexes of at least 2, the larger one is always the female; the macroconch is considered the female shell and the microconch the male shell. Furthermore, egg capsules have been found in macroconchs only. In the classic sexual dimorphism of Ammonitida, the peristomes differ in that only the

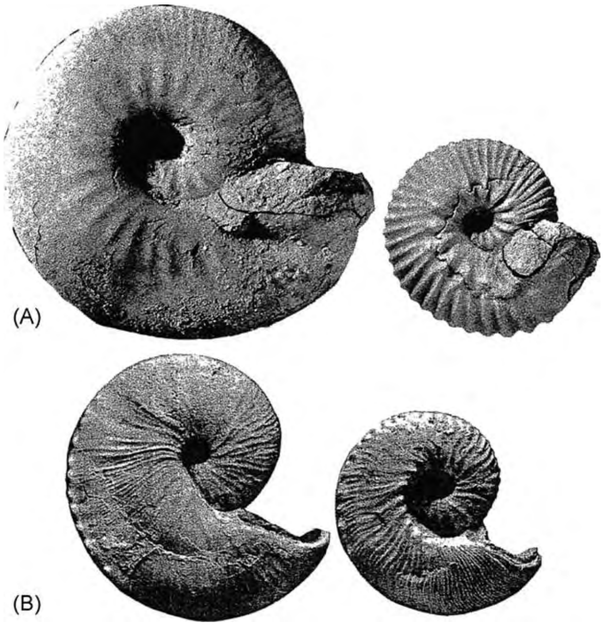


Figure 6 Sexual dimorphism. (A) Difference in size, body chamber and aperture in Jurassic Ammonitida (*Emileia giebeli* ♂ and ♀, peristomes traced). (B) Difference in size and body chamber in Cretaceous Ancyloceratida (*Hoploscaphites nicolletii*). (Adapted from Davis *et al.* (1996).)

microconchs or males have lateral extensions called lappets. In other Mesozoic ammonoids (e.g., Ancyloceratida), dimorphism is developed only in size and major to minor body-chamber modification, whereas dimorphism was rare in Palaeozoic ammonites and less clearly developed. Antidimorphs are recognized as conspecific because the inner whorls are identical and the animals have the same stratigraphic and geographic distributions. However, their numbers are often grossly mismatched; in many cases, the males are 10 (and up to 100) times more frequent, in others the females. The most probable causes for the mismatches are sexual segregation except during mating periods, and differences in postmortem drift (see below). Although the existence of sexual dimorphism is now universally acknowledged, at least for the Ammonitida, disagreement remains in classification; some ammonitologists still place all antidimorphs in different genera, even naming new ones; most place them in the same genus when specific pairing is impossible, and in the same species when pairing is highly probable. In the latter case, the sex symbols or 'macroconch (M)' and 'microconch (m)' are appended. The choice of classification is obviously of great consequence to phylogenetic reconstruction, biogeography, etc.

Disputed Functions of Ammonitic Septa and Complex Sutures

The main function of complex ammonitic sutures remains the most controversial topic among ammonitologist. The oldest ('orthodox') and still prevalent hypothesis resembles that of the Oxford professor William Buckland, who in 1827 concluded that this structure strengthened the shell (phragmocone) against implosion by ambient water pressure. *Nautilus*, which lives deeper than most ammonites did, has only simple, sub-hemispherical septa, but these are placed within well rounded, ovoid whorls with thick walls. Many typical ammonites, on the other hand, had flattened thin-shelled flanks with little (membrane) strength of their own, for example, platycones, oxycones, and compressed discocones. Their sutures were a functional improvement over goniatitic sutures by providing more closely spaced and somewhat elastic support for the weak, flat parts of the phragmocone wall (Figure 3), as well as some protection against predators. This prevented the shell from bending and fracturing (under tension) over stiff, distant sutures and from fracturing from tooth bites. Whorl section (e.g., laterally compressed for 'streamlining') was, therefore, the dominating factor, rather than habitat depth. Among related taxa with similar whorl shape, however, epeiric species often had simpler sutures (and weaker septa) than their oceanic relatives. Computer modelling and analysis (Finite-Elements) have shown that, with increasing complexity, the septa became stronger against the hydrostatic pressure transmitted through the body (as earlier theory had predicted). Recent claims to the contrary were based on faulty modelling.

Other hypotheses attempting to explain sutural complexity, include: (i) improved muscle attachment in the lobules and 'frills' for (a) better attachment of the mantle and/or muscle fibres, or (b) for a pre-septal gas bladder that could change overall buoyancy by expansion and contraction; (ii) improved liquid transport out of and/or into the chambers; and (iii) the retention of cameral liquid in the microcavities created by the lobules, folioles, and 'frills'. Hypotheses (i) and (ii) are not feasible because: (ia) The principal muscles were attached elsewhere; (ib) a preseptal gas bladder was inoperable because of basic gas laws and limitation of muscular forces; (ii) improved liquid transport would be ill served by marginal fluting because of difficult transport to the siphuncle, creating a 'bottleneck'; (iii) liquid storage was probably a *secondary* function of marginal fluting, keeping ballast water safe from swishing around.

Hydrostatics and Hydrodynamics: Reconstructing the Living Organism

Buoyancy, poise, stability, and bathymetry can be calculated for the shell and approximated for the living ammonoid, although the soft parts remain essentially unknown. The arms, which were presumably 10 in number, could not have been very muscular as in extinct coleoids (squid, belemnites, sepiids) which are well known from bituminous shales, and they must have been small to fit into the bodychamber (e.g., behind the aptychi). The head, arms, and hyponomes of ammonoids can therefore only be conjectured. They are here shown adapted to diverse feeding strategies (Figure 7).

The neutral buoyancy necessary for a free-floating organism, and which is present in all living cephalopods, requires strict correlations between shell thickness and the volumes of phragmocone and body chamber. The shell may be envisaged as a coiled cone: assuming similar shell thickness, the relative volumes and, hence, the relative lengths of phragmocone and body chamber must remain constant, whether the cone is slender or thick. When coiled into a logarithmic spiral, the slender cone (longicone) is a multiwhorled 'serpenticone' with long, thin body chamber (longidome; low expansion rate W *ca.* 1.5); the thick cone (brevicone) expands more rapidly (W *ca.* 2.5), and phragmocone and body-chamber become shorter (brevidome). In the longidomes, the body chamber was 1 to 2 whorls long; about 3/4 whorls in the abundant mesodomes (W *ca.* 2, i.e., shell diameter doubles with each whorl); and only about 1/2 of a whorl in the brevidomes, which resemble *Nautilus* (W 3–3.5).

Poise, i.e., the orientation of the aperture, and stability were obviously of great importance to the animal. Both were controlled by the positions of the centres of buoyancy and mass. The centre of buoyancy is the three-dimensional centre of the seawater displaced by the entire organism; the centre of mass is that of *all masses* within the same volume, with the principal variable being body-chamber length (measured in whorls). The buoyancy centre lies above the mass centre, and the distance between them determines the degree of stability, i.e., against the torque produced by the hyponome that threatens to rotate the organism. Stability limits the force of jet propulsion, depending on jet direction (vector force) relative to the centre of rotation (Figure 7). The densities of the phragmocone (*ca.* 0.2 kg/l) and body chamber with body (*ca.* 1.2 kg/l) differed greatly, so that brevidomes, with their body chamber mainly below the phragmocone, were more stable than mesodomes, and longidomes, with the body chamber completely surrounding the phragmocones, were highly unstable. Stability sufficient for jet propulsion existed only in brevidomes and mesodomes. But torque could be prevented only in mesodomes, when the jet force would pass through the centre of rotation – but the animal was limited to swimming backward, as well as up- and downward. Forward swimming explains the apparently wasteful 'rocking' of *Nautilus*. Its extremely brevidomic shell places the hyponome so low that it can curve backward below the shell, although this creates a torque; *Nautilus* is able to swim forward as well as backward and upward simply by curving the hyponome. 'Rocking' results from jet pulsation: the power phase rotates the animal backward and the inherently high static stability makes its return to rest position.

Backward swimming, however, required good steerage for maneouvrability; spherocones could, at most, have used long, trailing tentacles as rudders, whereas in oxycones and keeled platycones the shell allowed good steerage.

Hydrodynamic potential varied greatly among ammonoids. Involute, compressed shells (oxycones, platycones, discocones), more or less smooth, are obviously 'streamlined', i.e., they produce relatively little drag and friction during locomotion; brevidomes, sometimes with their body chambers slightly uncoiled and inflated to lower the centre of mass, were among the best swimmers among ammonoids. Surface roughness provided by fine riblets (the golf-ball effect) may have further reduced drag in some cases. Like *Nautilus*, brevidomes could swim forward and backward. Velocity increased with size, but swimming was sluggish in all externally shelled (ectocochliate) cephalopods because the body-chamber limited contraction of the mantle cavity for pumping. Only some oxyconic brevidomes among the Ammonitida (e.g., *Aconeceras*, *Quenstedtoceras*) show large muscle scars that indicate the presence of effective head retractors as in *Nautilus*. These were either nektic hunters or demersal ('bentho-pelagic'), feeding from the seafloor. At the other extreme were the serpenticones and longidomic sphaerocones, ribbed or smooth, which were unable to swim because the slightest jet force would have rotated the unstable, high-drag shells. They belonged to the mega-plankton and, according to their shell strength, were either limited to drifting near the surface or able to dive into deep water, presumably diurnally. The many intermediate forms, i.e., the abundant planorbicones and cadicones, usually mesodomic and with prominent ornamention, also tended to increase stability at maturity. They were probably very sluggish

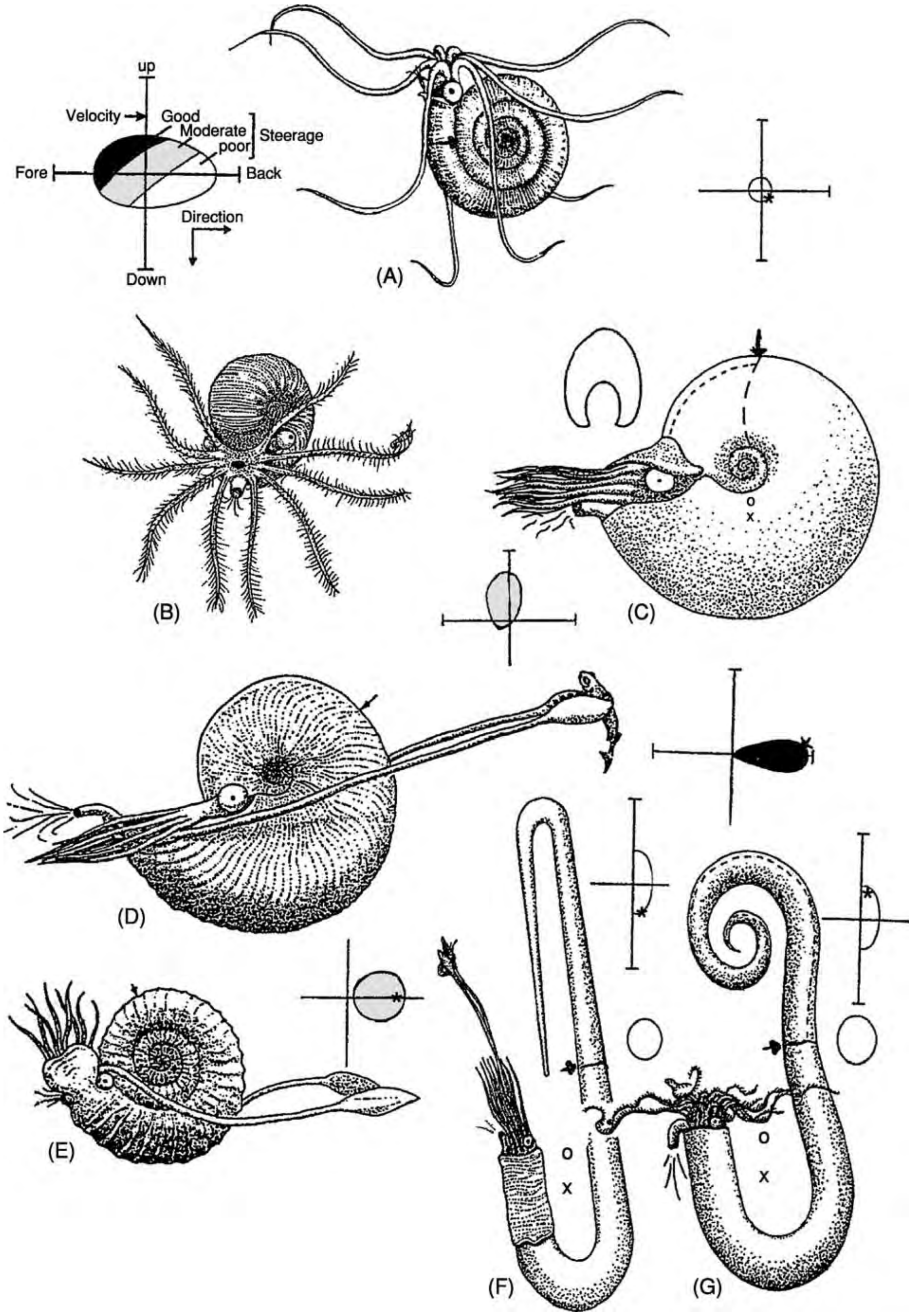


Figure 7 Reconstructions of poise and swimming potential according to shell form, body chamber length, and stability based on estimated centres of mass (x) and buoyancy (o); head, arms, and hyponome conjectured for various feeding strategies. (A) serpenticone; (B) longidomic spherocone; (C) mesodomic discocone; (D) brevidomic oxycone; (E) mesodomic platycone, microconch; (F) hamitocone, with mantle overlap; (G) ancylocone. Small arrows indicate end of body chamber. (Adapted from Westermann (1996) and Westermann and Tsujita (1999).)

backward swimmers and/or vertical migrants, mostly in shallow water, where their coarse ribs and/or spines would have provided some protection from predators and, perhaps, camouflage. Ellipto-spherocones possessed an unstable immature stage followed by a stable adult stage. Extension of this trend led to the heteromorph scaphitocones. The result was greatly improved stability, with the aperture turned upward, i.e., an adaptation for vertical migration (Figure 7). Whereas the open coils of gyrocones lie in one plane, the wider coils of cyrtocones are irregularly helicoidal. The former grew large and may have been demersal; the latter were small and perhaps lived as pseudoplankton coiled around floating seaweed (e.g., *Sargasso* Community). Hamitocones, with a series of parallel segments, must have rotated during growth. The adults had stable near-vertical poise with upward directed apertures, implying diving habits as megaplankton. Orthocones also had stable near-vertical poise, but with the aperture downward, implying that they were sluggish benthos feeders with rapid escape potential. The torticones were also (diurnal) vertical migrants, whereas the rare vermicones, with a complex growth programme, were planktic. Buoyancy calculations have shown that all heteromorphs were neutrally buoyant.

Habitat depth limits can be calculated from the strength of septa, phragmocone wall, and connecting rings, because cameral gas pressure was maximally that of the atmosphere. Physical properties of ammonoid shell (test) are calibrated on *Nautilus*. Ammonoid connecting rings resemble the chitinous inner tubes of *Nautilus*, but ammonoid septa differ greatly in their architecture, being folded and not basically hemispheres. Stresses in ammonite septa are therefore calculated using curvature and thickness of lobes or lobules. Wall strength between sutures has also been calculated (Figure 8). The resulting depth limits range from 50 to over 1000 m. Actual habitat depth was at about one-third to two-thirds of the limit.

Ecology: Food, Predators, and Habitat

Feeding habits are known only for a few ammonoids, but were presumably as diverse as their shells. The general presence of weak jaws and radulae and the apparent absence of muscular arms suggest a number of possibilities, such as: (i) filter-feeding in mid-water on passing microplankton and detritus; (ii) predation on macroplankton and/or sluggish nekton, including smaller or young ammonoids, by pursuit or ambush; and (iii) preying on benthos while hovering. The radula aided in swallowing soft organisms whole. The known rare records of crop or stomach contents are of pelagic ostracods, microcrinoids, larger crustacean appendices, and juvenile aptychi indicating cannibalism.

Many ammonite shells suffered broken body chambers and some of these injuries were healed by the mantle before embedding. The predators are usually not identified, but large crustaceans, fish, and marine reptiles are strongly suspect. The best evidence comes from large, shallow-water oxycones (*Placenticeras*) that lived in the murky surface waters of the Cretaceous Western Interior Seaway of North America. These shells are perforated by numerous round holes that sometimes have the exact spacing and jaw angle of mosasaurs, large marine reptiles with conical teeth that are known from the same formation. Some authors believe that the holes are the diagenetically collapsed home scars (resting places) of limpets grazing on the empty shell, but this is improbable because of the shape and orientation of the holes.

The minute eggs and early juvenile (neanic) growth stages are rare in bituminous shales, which were deposited in lethal, oxygen-starved environments. This suggests that eggs (in gel?) and hatchlings floated in mid-water, where they were protected from active predators by the slight oxygen deficiency; they died and sank to the floor when dysoxic bottom waters rose periodically (Figure 9A). As in living cephalopods, some ammonoids may have spawned on oxygenated seafloors, but their minute aragonitic shells would have dissolved. Juvenile and immature ammonites are much scarcer than expected and the adults tend to range through a variety of lithofacies of epeiric seas upto 150–300 m deep. Most ammonoids did not depend on the seafloor and were pelagic: only a few species were demersal bottom feeders. A single case of brackish habitat has been documented by stable isotopes for the Late Cretaceous *Placenticeras*. Deep-water, oceanic sediments did not preserve ammonites, although epipelagic and some mesopelagic ammonoids lived in most oceans (Figure 9B). Most Jurassic-Cretaceous Phylloceratida and Lytoceratida have long been recognized as mostly deep-water ocean dwellers; planktic Ammonitida drifted and heteromorphic Ancyloceratida dived in the upper waters, and were dispersed far and wide by surface and subsurface currents.

Life Versus Death Assemblages: Migration and Post-Mortem Drift

There are two main reasons for numerical mismatches between sexual partners and for the common scarcity or even absence of immature shells from most fossiliferous deposits; spawning migrations and post-mortem transport. Most living cephalopods feed in open, deeper waters, but breed and die in shallow, warm waters. This often involves segregation of the sexes before and/or after mating, causing frequent mismatches (see Dimorphism above). The second reason for uneven sex ratios, as well as for the scarcity of immature shells, is selective surface drift of the empty shells. A strong natural bias for large shells of deep-water ammonoids over small, shallow-water species was caused by size and ambient pressure. Reflooding of the phragmocone causing sinking was much slower in large shells than in small ones,

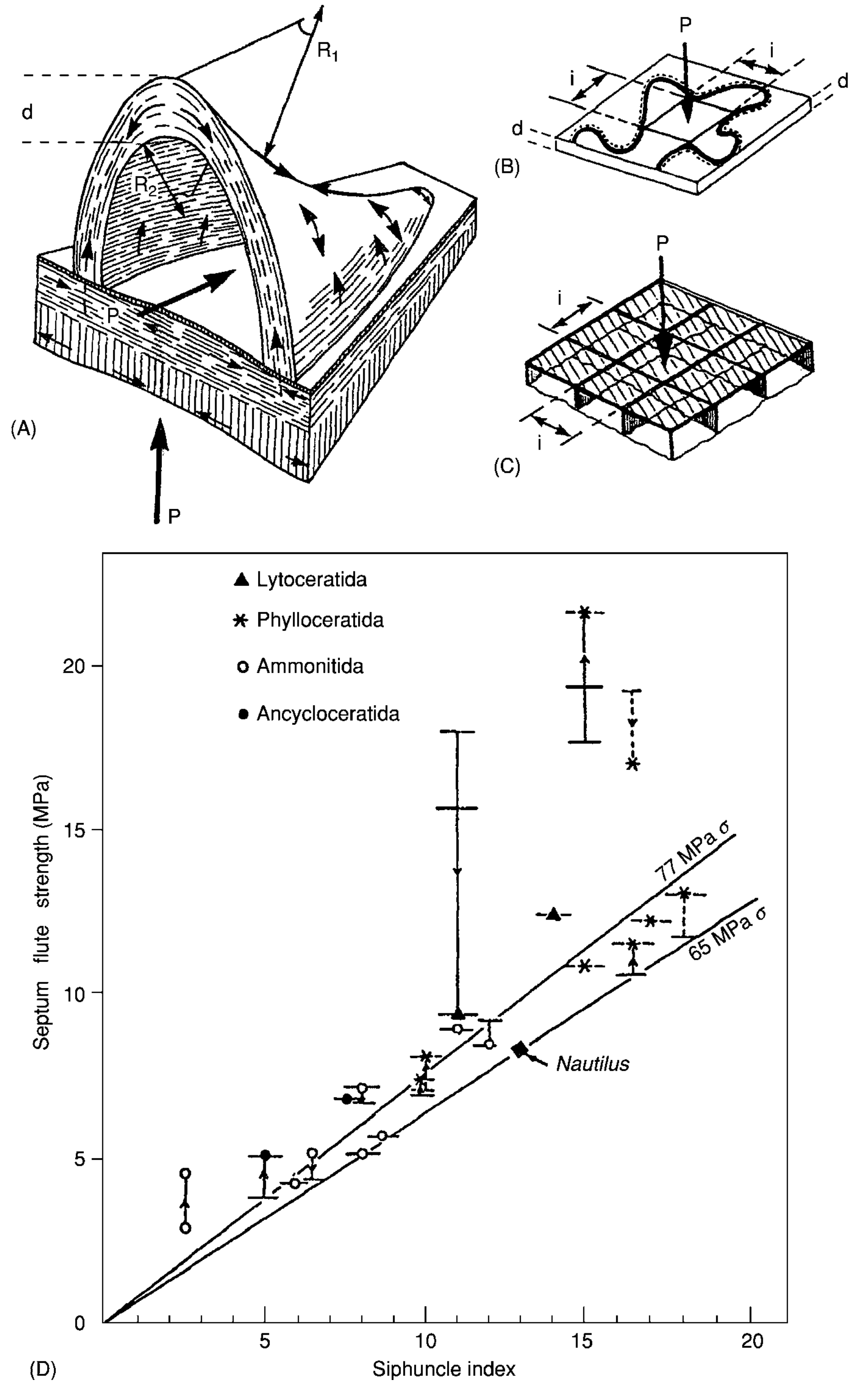


Figure 8 The shell parameters used for bathymetry, based on the ammonitic septum, phragmocone wall, and siphuncle. (A) Lobule of last septum with anticlastic curvature used to calculate Septum Flute Strength; R_1 and R_2, orthogonal radii; P, hydrostatic pressure; d, septum thickness; arrows, compressive and tensile stresses. (B) Wall with lobule to calculate wall strength, the implosion depth based on the unsupported wall distance; i, length of largest unsupported square; d, wall thickness. (C) Connecting ring of siphuncle, cross section (Siphuncle Strength Index 100X d_s/r_i). (D) Habitat limits based on septum and siphuncle of adult ammonoids and *Nautilus* (ontogenies indicated): correlation is good except for some compressed, deep water Phylloceratida, which have thicker septa that functioned significantly as lateral braces; depth limit calibrated on *Nautilus*. (Adapted from Hewitt (1996).)

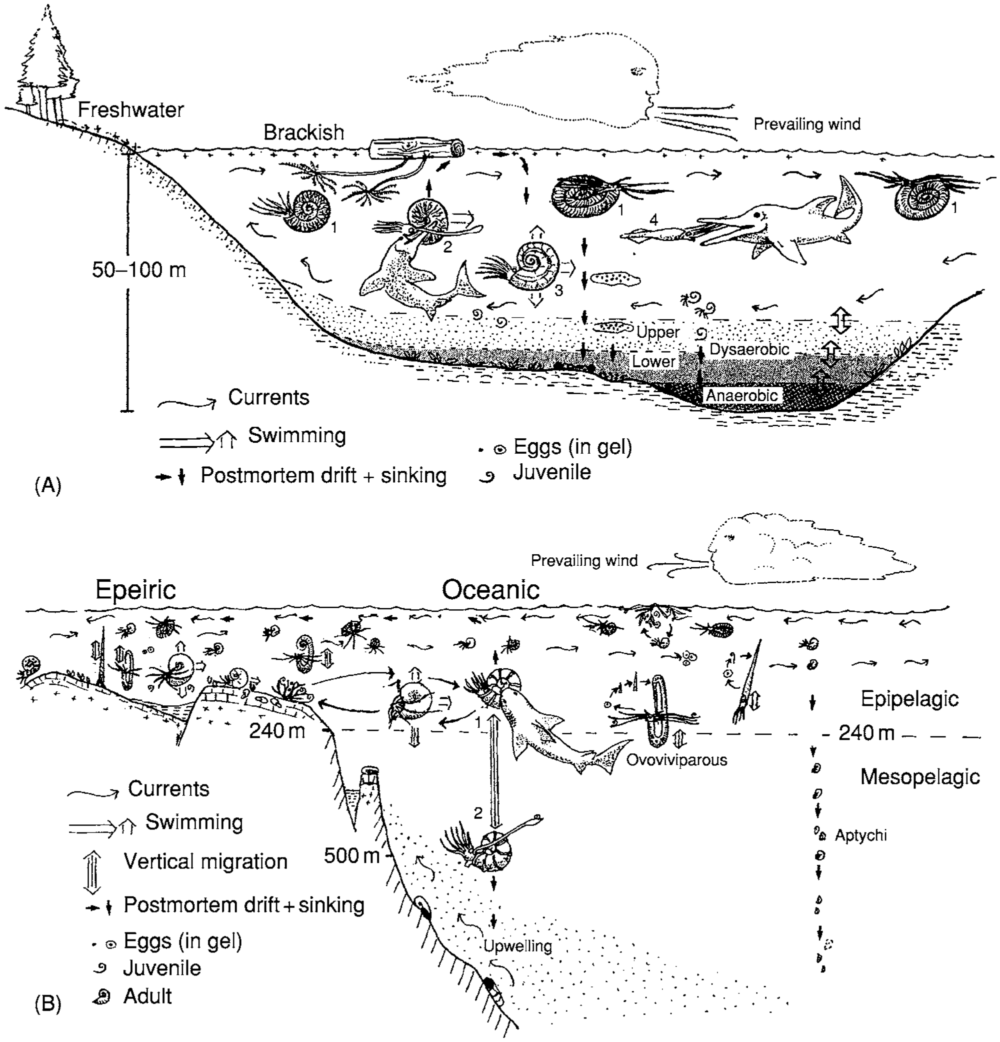


Figure 9 Possible scenarios for Jurassic Cretaceous ammonoid habitats, with life cycles and post mortem drift. (A) Epeiric basin, black shale facies, e.g., Early Jurassic *Posidonia* Shale or Jet Rock; 1, serpenticonic, epiplanktic drifters (*Dactylioceras*); 2 and 3, discoconic predators (*Harpoceras, Hildoceras*), prey of ichthyosaurs; 4, belemnite animal. (B) Ocean, slope, shelf, and epeiric sea; serpenticonic Ammonitida drift and heteromorphic Ancyloceratida dive in both biomes; most Ammonitida are restricted to epeiric seas; typical Phylloceratida (1) were deep water swimmers and typical Lytoceratida (2) among the deepest vertical migrants; only calcitic aptychi reach the deep sea floor. (Adapted from Westermann (1996).)

especially those with thick-walled, narrow connecting rings (siphuncle surface grows with the square of diameter, chamber volume with the cube) – just as cameral emptying rates in growing shells. Similarly, the reflooding rate depended on the depth at death: it was rapid only under high ambient pressure, so that the shell sank immediately; during slower refilling in shallower water the shell would rise to the surface (Figure 9) before sinking after weeks or months of drifting. Drifting distance has been much debated, but it was probably rarely more than a few tens of kilometres in epicontinental seas. An exceptional case, however, was observed in the Andes at the edge of the palaeo-Pacific: several gigantic lytoceratids of almost certainly meso-pelagic habitat were found in a nearshore assemblage. The sinking aragonitic shells of dead oceanic ammonoids were dissolved below compensation depth, but their calcitic aptychi have been found in red deepsea clay.

Summary

Most ammonoids were pelagic and a few demersal between 30 and 150 m depth in epeiric seas; others were epi- to mesopelagic in the oceans; and some lived in epeiric seas as well as oceans, for example, cosmopolitan longidomes. Their principal food was micro- and mesoplankton, including young ammonoids, presumably caught with weak arms or tentacles, and sinking organic detritus netted with an umbrella-like arm crown, as well as benthos. Their predators were fish, reptiles, and larger ammonoids. The most diverse ammonoid assemblages are found in sediments of warm tropical and subtropical seas and typically include mainly adults. Gray and black shales that originated in temperate or basinal waters, on the other hand, tend to have low-diversity faunas that may include ammonitellas and juveniles; single, highly variable species dominate, sometimes ranging from ribbed spherocones to smooth oxycones or platycones. Their contrasting hydrodynamic properties did not apparently function; these ammonoids were mega-plankton without significant locomotion.

Ocean currents, rather than swimming, were the main means of wide dispersal, because planktic species tend to be more cosmopolitan than nektic species. Prime examples are longidomic spherocones (e.g., arcestid Ceratitida) and serpenticones (e.g., dactylioceratid and psiloceratid Ammonitida) as well as many heteromorphs (Ancyloceratida), which lived in the shallow waters of all seas. The mainly oceanic Phylloceratida and Lytoceratida evolved much more slowly, which may account for their wide species distributions. The importance of most ammonites in biostratigraphy (*see* **Biozones**) and biogeography resulted from rapid evolution combined with high potential for planktic dispersal.

See Also

Biozones. **Fossil Invertebrates:** Molluscs Overview; Cephalopods (Other Than Ammonites).

Further Reading

Becker RT and Kullmann J (1996) Paleozoic ammonoids in space and time. In: Landman, *et al.* (eds.) *Ammonoid Paleobiology,* pp. 711 754. New York: Plenum Press.

Bucher H, Landman NH, Klofak SM, and Guex J (1996) Mode and growth in ammonoids. In: Landman, *et al.* (eds.) *Ammonoid Paleobiology,* pp. 408 462. New York: Plenum Press.

Davis RA, Landman NH, Dommergues J L, Marchand D, and Bucher H (1996) Mature modifications and dimorphism in ammonoid cephalopods. In: Landman, *et al.* (eds.) *Ammonoid Paleobiology,* pp. 463 539. New York: Plenum Press.

Doguzhaeva LH and Mutvey H (1996) Attachment of the body to the shell in ammonoids. In: Landmann, *et al.* (eds.) *Ammonoid Paleobiology,* pp. 44 63. New York: Plenum Press.

Hewitt RA (1996) Architecture and Strength of the Ammonoid Shell. In: Landman, *et al.* (eds.) *Ammonoid Paleobiology,* pp. 297 339. New York: Plenum Press.

House MR and Senior JR (eds.) (1981) The Ammonoidea. *The Systematics Association*, Spec. Vol. 18. London: Academic Press.

Jacobs DK and Chamberlain JA, Jr. (1996) Buoyancy and hydrodynamics in ammonoids. In: Landman, *et al.* (eds.) *Ammonoid Paleobiology,* pp. 169 224. New York: Plenum Press.

Landman NH, Tanabe K, and Davis RA (eds.) (1996) *Ammonoid Paleobiology.* New York: Plenum Press.

Page KN (1996) Mesozoic ammonoids in space and time. In: Landman, *et al.* (eds.) *Ammonoid Paleobiology,* pp. 755 794. New York: Plenum Press.

Tanabe K and Fucuda Y (1996) Morphology and function of cephalopod buccal mass. In: Savazzi E (ed.) *Functional Morphology of the Invertebrate Skeleton,* pp. 245 262. Chichester: Wiley.

Westermann GEG (1971) Form, structure and function of shell and siphuncle in coiled Mesozoic ammonoids. *Royal Ontario Museum, Life Sciences Contributions* 78: 1 39.

Westermann GEG (1975) Model for origin, function and fabrication of fluted cephalopod septa. *Palaeontologische Zeitschrift* 49: 235 253.

Westermann GEG (1996) Ammonid life and habitat. In: Landman, *et al.* (eds.) *Ammonoid Paleobiology,* pp. 607 707. New York: Plenum Press.

Westermann GEG and Tsujita CJ (1999) Life habits of ammonoids. In: Savazzi E (ed.) *Functional Morphology of the Invertebrate Skeleton,* pp. 299 325. Chichester: Wiley.

Porifera

J K Rigby, Brigham Young University, Provo, UT, USA

Introduction

The phylum Porifera, or sponges, includes simple or primitive multicellular sedentary organisms, and is among the most diverse and successful of extant phyla. Living sponges are filter feeders and are characterized by body plans of three structural grades (Figure 1), which have been recognized also in fossil forms. Water is pumped in through numerous smaller inhalant openings, or ostia, by the uncoordinated beating of flagella of choanocytes, or collar cells. These cells are located in interior chambers within the sponge, and cause the water to circulate via numerous canals throughout the sponge body, and to exit through larger exhalant openings, or oscula. Food particles are filtered from the inhalant water virtually on a cell-by-cell basis, whilst waste products are emitted into the outflowing exhalant currents. Soft parts or tissues of sponges are supported by internal skeletons made of organic fibres, mineralized needle-like or multirayed spicules, or a combination of fibres and spicules. It is primarily the mineralized skeletons that make up most of the geological record of the group. The nature, shapes, and interrelationships of spicules and other structures within their skeletons are used to classify fossil sponges.

Skeleton elements of living sponges are secreted by specialized cells, termed spongocytes, which secrete the spongin fibres, or sclerocytes, which secrete the mineralized spicules. These spicules are composed of opaline silica or crystalline to microgranular calcium carbonate. Their compositions and shapes are used to differentiate taxa at several levels. Calcium carbonate may also occur as layered, granular to crystalline aragonite or calcite in the Calcarea, Archaeocyatha, and Sclerospongiae. Intermixed siliceous and carbonate skeletal elements are known in the Sclerospongiae and in hypercalcified sponges in two subclasses of the Demospongea.

The nomenclature of spicules is based upon their size and the numbers of rays or axes that they have. One sponge may have several kinds of spicules, and the same kind of spicule may occur in several sponges. Megascleres (Figure 2) are the large principal skeletal elements and range from monaxons, with a single axis, to triaxons or tetraxons, with three or four axes of ray growth. Monaxial spicules may be termed monactine or diactine, depending upon whether there are one or two directions of growth. Similarly, triaxons may have three directions of growth and be triactines, which are generally of calcareous composition, or they may be hexactines with six directions of ray growth and of siliceous composition. Tetractines have four rays and directions of growth. Octactines are specialized calcareous tetractines with eight rays, and they characterize the Heteractinida, a class of Palaeozoic fossils. Some sponges have spicules with many rays and axes of growth. Such spicules are termed polyactines or polyaxons, or sphaeractines.

Microscleres are small spicules used in the classification of modern sponges (Figure 3), but they are only occasionally preserved in fossil sponges, and are consequently of minor use in palaeontology.

Living sponges are aqueous organisms and are dominantly marine, although they do occur in a few freshwater lakes and streams. Marine forms occur from polar to tropical environments, and from shallow tidal environments to abyssal depths. They are characteristically bottom dwellers. Fossil sponges occupied the same general environments, based on their occurrences in various kinds of rock in the geological record.

Sponges are relatively minor fossils when the total palaeontological spectrum is examined, but they played major roles locally, as in the construction of the famed Permian reefs of western Texas and New Mexico in the USA, or the Jurassic reefs of Europe, for example. Following death, some forms with unfused skeletons broke up and their dissociated mineralized spicule elements locally accumulated to form sedimentary deposits, known as spiculites or spiculitic chert.

Classification

Fossil sponges are classified on the basis of the composition and forms of their spicules, their canal systems, and their structural grades. Traditionally, five classes have been recognized in the phylum, including the Demospongea, Calcarea, Hexactinellida, Sclerospongia, and the extinct Heteractinida. To these have recently been added the Archaeocyatha, which had been considered as a separate phylum for many years, and the Stromatoporoidea and possibly related groups, which have been grouped with the Coelenterata for many years.

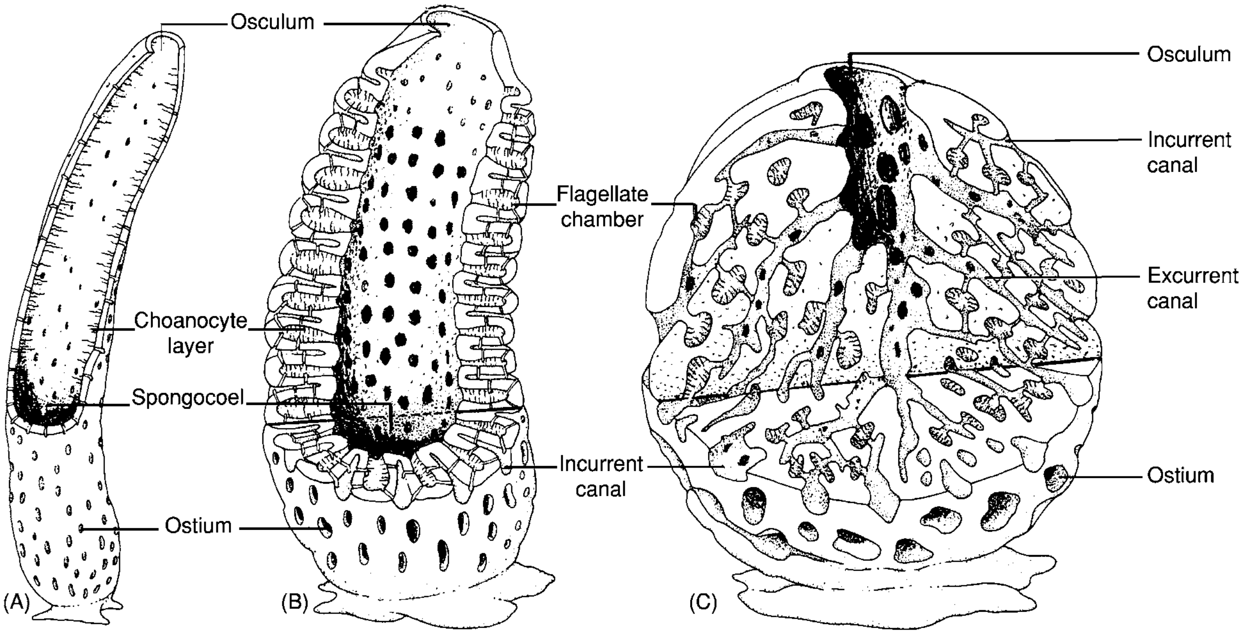


Figure 1 Structural grades of living and fossil sponges. (A) Ascon grade sponges have thin walls and a large spongocoel, or central cavity, lined by flagellate choanocytes. (B) Sycon grade sponges have somewhat thicker walls that contain distinct chambers of flagellate choanocytes, chambers which open directly into the spongocoel. (C) Leucon grade sponges are the most complex and have isolated flagellate chambers that are interconnected with inhalant and exhalant canals by numerous small canals. Not to scale. Adapted with permission from Rigby JK (1987) Phylum Porifera. In: Boardman RS, Cheetham AH, and Rowell AJ (eds.) *Fossil Invertebrates*. Palo Alto, Oxford, London: Blackwell Scientific Publications.

Demospongea

Sponges of the Class Demospongea are characterized by skeletons composed of organic spongin fibres, siliceous spicules, or mixed spongin and siliceous spicules. Those with spongin skeletons have left a poor record and essentially all recognizable fossil demosponges are characterized by their shapes, canal patterns, and spicule shapes and relationships (Figure 4). Spicules of the class range from loose monaxons to tetractines whose rays do not join at right angles, or to irregular root-like forms (Figure 2). Many Palaeozoic and Mesozoic demosponges belong to the Order Lithistida, whose skeletons are formed of fused and cross-braced spicules that have articulated ray tips or rough exteriors with nodes or spines that united to form rigid, resistant skeletons that, once buried, were commonly preserved.

Demosponges first appear in the geological record in the Late Precambrian, and diversified moderately quickly so that, by the Middle Cambrian, several genera are known that represent the initiation of important distinct major Palaeozoic lineages. During the Lower and Middle Ordovician, they underwent great diversification and, for example, became major reef builders and also occupied diverse environments. Their Silurian record shows that they remained common and diversified, so that they are relatively abundant fossils in both carbonate and clastic sequences that accumulated in both shallow and deeper marine environments. Although locally abundant, they became less dominant in later Palaeozoic sequences. Demosponges again played major roles in the Jurassic and Cretaceous, when they reached a second peak of diversity and abundance in a variety of environments.

Demosponges are the dominant class of living sponges, where several thousand species have been described. Although mainly shallow marine organisms, the class also includes the only known living and fossil freshwater sponges. To date, nearly 500 genera of fossil demosponges have been described.

Class Demospongea Sollas, 1875, Precambrian–Holocene
  Subclass Clavaxinellida Lévi, 1956, Precambrian–Holocene
  Subclass Choristida Sollas, 1880, Upper Ordovician–Holocene
  Subclass Tetractinomorpha, Lévi, 1953, Middle Ordovician–Holocene

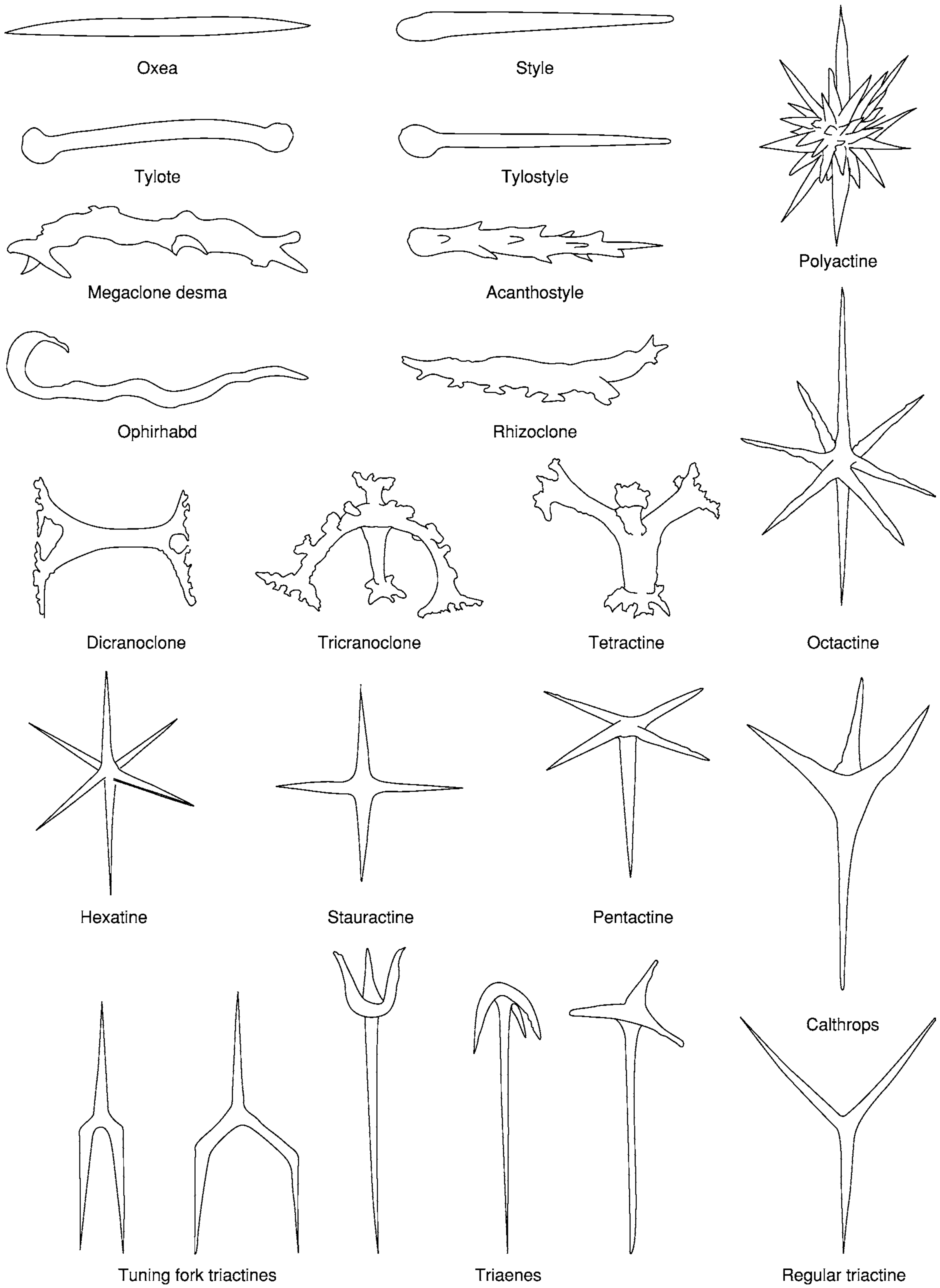


Figure 2 Nomenclature of common megascleres or main body spicules, which may range from a fraction of a millimetre to several millimetres in length. Shapes and compositions of megascleres are important elements in classification. Triactines, for example, are commonly calcareous and representative of the Calcarea, and hexactines and derivative spicules, such as stauractines and pentactines, are siliceous and characterize the Hexactinellida.

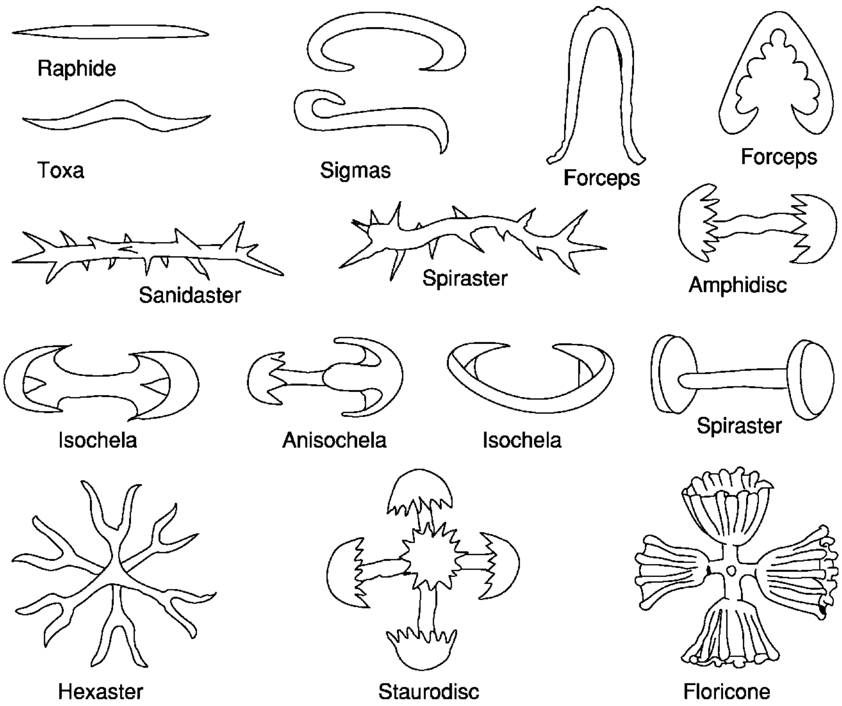


Figure 3 Representative microscleres of the demosponges and hexactinellid sponges. They are considerable smaller than associated megascleres and are usually only a fraction of a millimetre in length.

- Subclass Ceractinomorpha Lévi, 1953, Middle Cambrian, Middle–Upper Ordovician, ? Pennsylvanian, Holocene
- Subclass Lithistida Schmidt, 1870, Cambrian–Holocene

Hexactinellida

The Class Hexactinellida is characterized by skeletons composed of siliceous hexactine-based spicules, and has a geological range that extends from the Late Precambrian to the Holocene. It was, and is, an exclusively marine class of sponges. Two major Palaeozoic groups are documented in the geological record (Figure 5): the asconoid protosponges of the Cambrian and Ordovician, which have thin walls of unfused hexactines that are regularly arranged, and the syconoid lyssacinosid dictyosponges, which have thicker walls of quadrangular arranged to irregularly arranged hexactines with unfused to weakly fused overlapping rays (Figure 6A). The latter group is particularly well known in Devonian to Carboniferous rocks from New York and Pennsylvania to Indiana (Figure 5B). By the latest Permian, and certainly by the Triassic, the Order Hexactinosa developed with rectangularly based, rigid, fused, hexactine skeletons. These reticular skeletons (Figure 6B) were produced where overlapping rays of spicules were fused by cementation, by enclosure of overlapping rays in a siliceous coating, by interconnecting small nodes, or, more rarely, by tip-to-tip fusion. The Hexactinosa reached their peak of development in the Cretaceous. More or less concurrently, sponges of the Order Lychniscosa developed lantern-like, cross-braced, nodal octahedra at spicule centres (Figure 6C) in their more or less rectangular fused hexactinellid skeletons. That order reached its maximum development in the Jurassic and Cretaceous of Europe (Figure 5D).

- Class Hexactinellida Schmidt, 1870, Precambrian–Holocene
  - Subclass Amphidiscophora Schulze, 1887, Precambrian–Holocene
    - Order Amphidiscosa Schrammen, 1924, Ordovician–Holocene
    - Order Reticulosa Reid, 1958, Precambrian–Upper Permian
    - Order Hemidiscosa Schrammen, 1924, Upper Pennsylvanian–Cretaceous
  - Subclass Hexasterophora Schulze, 1887, Ordovician–Holocene
    - Order Lyssacinosa Zittel, 1877, Ordovician–Holocene
    - Order Hexactinosa Schrammen, 1903, Permian–Holocene
    - Order Lychniscosa Schrammen, 1903, Upper Triassic–Holocene

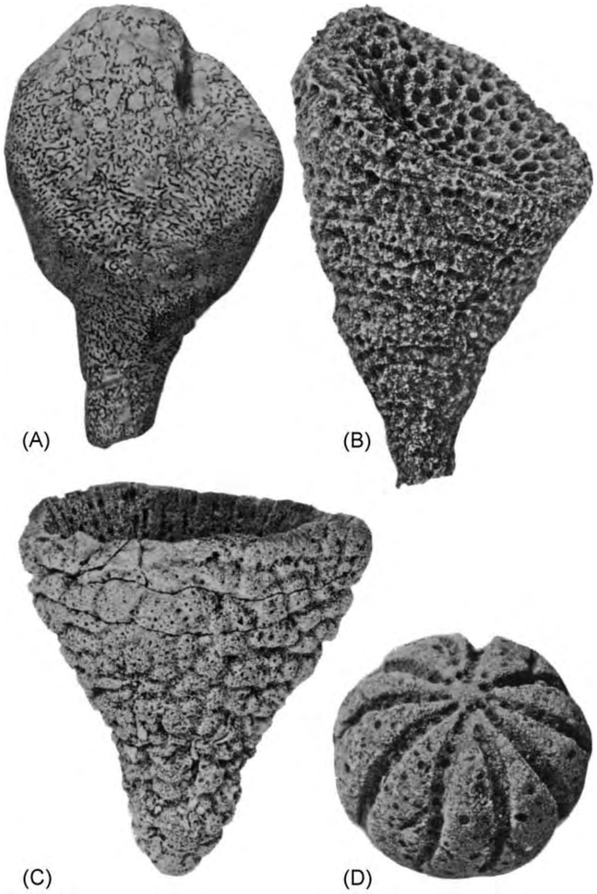


Figure 4 Representative fossil demosponges. (A) *Phymatella*, Cretaceous, Germany. (B) *Aulocopoides*, Devonian, Western Australia. (C) *Camellaspongia*, Ordovician, Minnesota, USA. (D) *Caryospongia*, Silurian, Tennessee, USA. All approximately normal size. Adapted with permission from Rigby JK (1987) Phylum Porifera. In: Boardman RS, Cheetham AH, and Rowell AJ (eds.) *Fossil Invertebrates*. Palo Alto, Oxford, London: Blackwell Scientific Publications.

Hexactinellid sponges are exclusively marine forms and, in modern seas, they are most common on seafloors from 200 to 2000 m deep, although many species have been reported from lower bathyal depths. They also occur, but are less abundant, in hadal depths over 6000 m deep, where they may form dense reef-like clusters. On the other hand, they are known to range up into water as shallow as 25–30 m off the south-western Canadian Pacific Coast, where they also form distinct reef-like structures.

Calcarea

The Class Calcarea includes sponges with calcareous skeletal elements that range from those with distinct three-rayed spicules of calcite or aragonite (Figure 2), to those with rigid skeletons of fused polygonal elements or imbricate calcitic plates. Living calcareous sponges are exclusively marine and are most common in shallow tropical environments. Those included in the Orders Murrayonida and Lithonida are the traditional ‘pharetronids’ and the other orders are the traditional ‘inozoids.’ The class has a geological record ranging questionably from the Late Precambrian, but certainly from the Lower Cambrian to the Recent or Holocene.

- Class Calcarea Bowerbank 1864, Cambrian–Holocene
  - Subclass Calcinea Bidder, 1898,?Precambrian, Cambrian–Holocene
    - Order Clathrinida Hartman, 1958, Holocene
    - Order Murrayonida Vacelet, 1981,?Precambrian, Cambrian–Holocene
  - Subclass Calcaronea Bidder, 1898, ?Cambrian,?Triassic, Jurassic–Holocene
    - Order Leucosoleniida Hartman, 1958, Holocene
    - Order Sycettida Bidder, 1898, Holocene
    - Order Sphaerocoeliida Vacelet, 1977, Cretaceous
    - Order Lithonida Doederlein, 1892, Jurassic–Holocene

Many genera of chambered fossil sponges with calcareous skeletons (Figure 7) had been included in the Calcarea, in the Order Sphinctozoa, until recently, when the polyphyletic origins of the sphinctozoans and their development principally as a structural grade were recognized. They are now treated as a group of ‘hypercalcified’ sponges drawn from other classes. Sphinctozoan grade sponges appeared in the Middle Cambrian, but played a minor role until the Carboniferous and Permian, when they helped to produce massive reefs, such as those in the Guadalupe Mountains in western Texas and south-eastern New Mexico, and in Tunisia and China, and during the Triassic, when they were similarly involved in the construction of reefs in what is now the Alpine region of southern Europe. They were thought to have become extinct by the end of the Cretaceous, but a living form was discovered in 1977.

Their chambered construction, in which ‘living space’ was limited by an outer rigid skeleton, required that new chambers be added for growth of the sponge. Such chambered skeletons developed at various times in several groups, so that a few calcareous and hexactinellid genera and many demosponge genera all had ‘sphinctozoan’ skeletal patterns. Chambered hexactinellid ‘sphinctozoans’ have been recognized, for example, because their chambered skeletons contain reticulate hexactines. Chambered demosponge sphinctozoans with rigid skeletons of aragonite or high-magnesium calcite are similarly included here. Skeletons of these latter

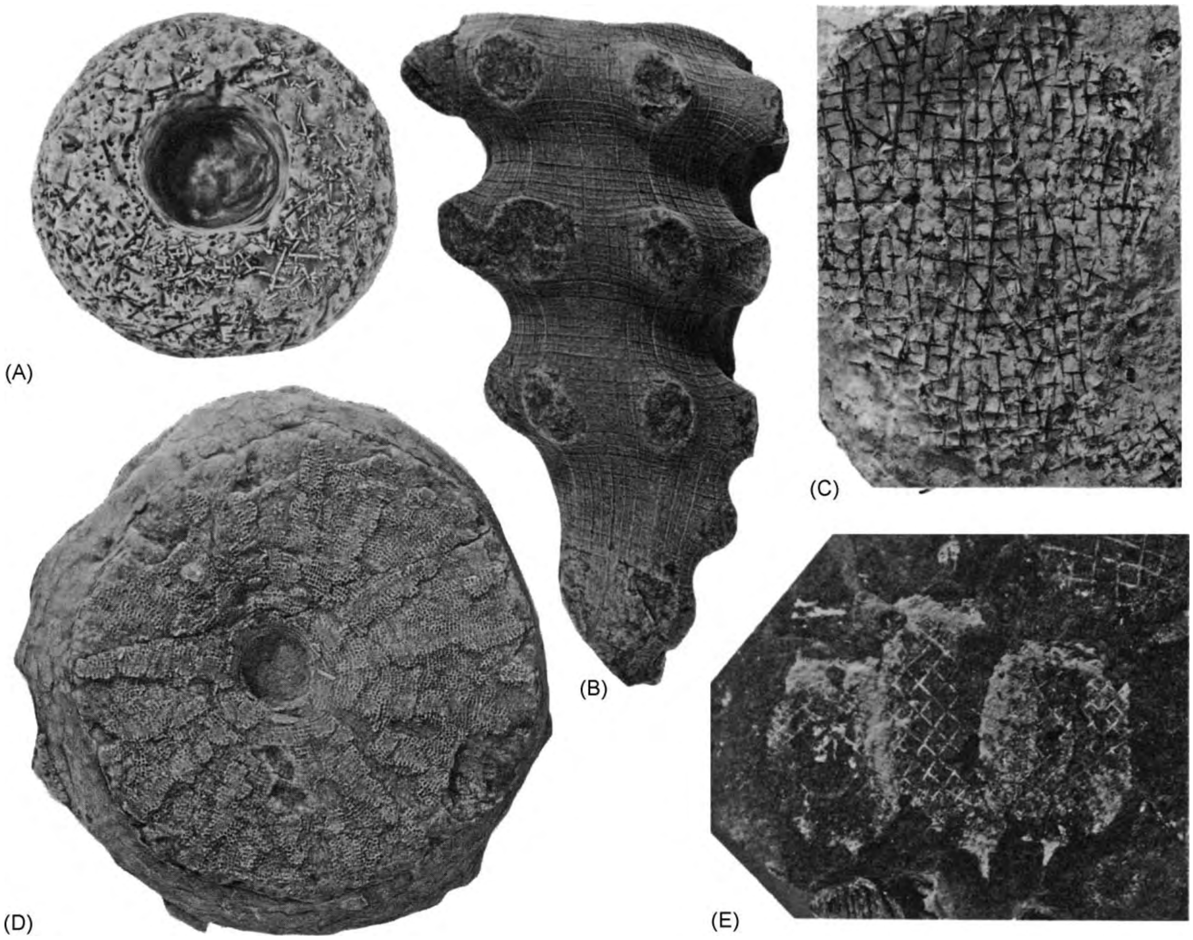


Figure 5 Representative hexactinellid sponges. (A) *Twenhofelella*, Silurian, Quebec, Canada. (B) *Hydnoceras*, Devonian, New York, USA. (C) *Mattaspongia*, Devonian, Alberta, Canada. (D) *Coeloptychium*, Cretaceous, France. (E) *Diagoniella*, Cambrian, Utah, USA. All are natural size. Adapted with permission from Rigby JK (1987) Phylum Porifera. In: Boardman RS, Cheetham AH, and Rowell AJ (eds.) *Fossil Invertebrates*. Palo Alto, Oxford, London: Blackwell Scientific Publications.

forms may or may not contain primary siliceous monaxon spicules, or pseudomorphs of such spicules, within their calcareous rigid elements.

Taxa within the polyphyletic group are differentiated by their shapes and chamber arrangements, their spicule content, as well as by the mineralogical compositions of their skeletons, internal segmentation, canal systems, and filling structures. They range from simple beaded moniliform sponges to polyglomerate cylindrical forms, or to mound-like or stratiform multichambered sponges (Figure 7). Filling structures within chambers may include septae, vesiculae, spore-like elements, or fine reticular structures, amongst others, or the chambers may have been hollow and lacked such skeletal structures.

Heteractinida

The Class Heteractinida, considered by some palaeontologists to be an order of the Calcarea, is characterized by skeletons made of large calcareous octactine spicules. The small class has a geological record that extends from the Lower Cambrian to the Lower Permian, and is one of only two classes of sponges to have become extinct. The Heteractinida includes such forms as the broad saucer-shaped octactinellid *Astraeospongium* (Figure 8), which is a common genus from the Silurian. The class is particularly common in Silurian and Carboniferous rocks of Europe and North America.

Class Heteractinida de Laubenfels, 1955, Lower Cambrian–Lower Permian
  Order Octactinellida Hinde, 1887, Lower Cambrian–Lower Permian
  ?Order Hetairacyathida Bedford and Bedford, 1937, Lower Cambrian

Sclerospongiae

The Class Sclerospongiae was proposed to include a few living sponges, and is now thought by some

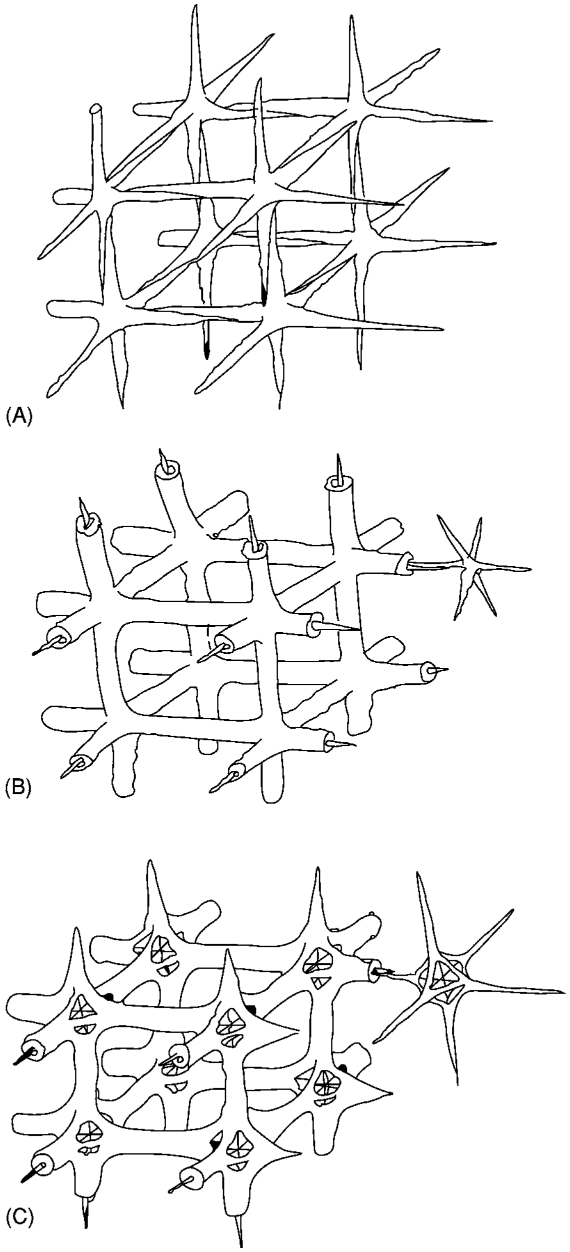


Figure 6 Diagrams of enlarged hexactinellid skeletal structures. (A) Lyssacid grade reticulate skeletal structure with unfused or weakly fused overlapping hexactine spicule rays. (B) Dictyid or hexactinosan grade with overlapping spicule rays fused into a rigid, three dimensional, quadrangular skeletal meshwork. (C) Lychniscoid grade with overlapping rays fused into a rigid skeletal meshwork and with diagnostic octahedral 'lanterns' at the centre of each spicule.

workers to also include the taxonomically limited fossil chaetetids and the diverse stromatoporoids. Stromatoporoids had long been considered as coelenterates related to hydrozoans, but recently they have been moved to the Porifera. These sponges have a laminated calcareous skeleton that is perforated by

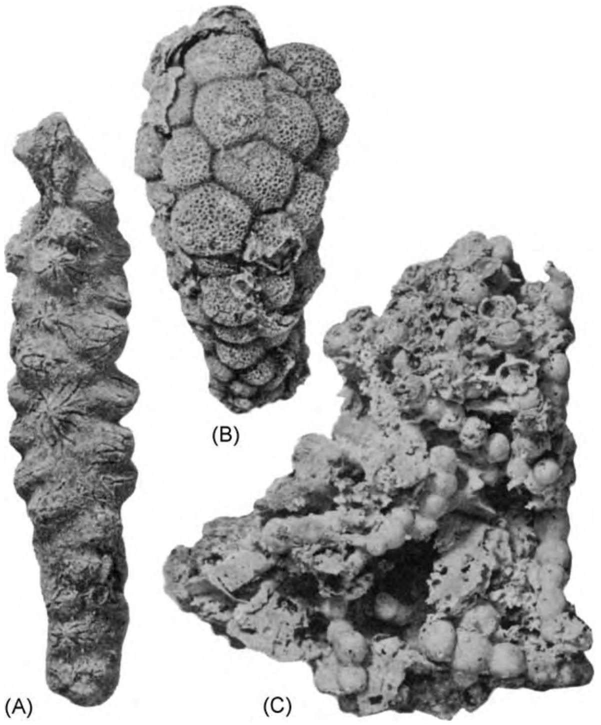


Figure 7 Representative calcareous and 'sphinctozoan' sponges. (A) *Stellispongia*, Permian, Tunisia, an inozoan calcareous sponge. (B) *Cystothalamia*, 'sphinctozoan' hypercalcified demosponge, Permian, Texas, USA. (C) *Girtyocoelia*, 'sphintozoan' hypercalcified demosponge, Permian, Texas, USA. All are natural size. Adapted with permission from Rigby JK (1987) Phylum Porifera. In: Boardman RS, Cheetham AH, and Rowell AJ (eds.) *Fossil Invertebrates*. Palo Alto, Oxford, London: Blackwell Scientific Publications.

astrorhizal canals, and the living animal is thought to have occupied essentially only the upper surface. If the Stromatoporoidea are included, the class Sclerospongiae has a stratigraphical range from the Cambrian to the Holocene, with a major period of development in the Silurian and Devonian, when these organisms were important reef formers. The class became a minor marine element in the Late Palaeozoic, however, and remains so in modern marine faunas.

Stromatoporoids have domal, tabular, branching, or bulbous growth forms. They are commonly laminated to vertically tubular or pillared, and have generally aspiculate, calcareous skeletons in early forms and more cystose and vertically tubular skeletons in later forms. Laminate forms commonly show cyclicity, or growth interruptions, with thicker latilaminae separated by several thinner sheet-like porous laminae (Figure 9). These laminae are interconnected or supported by vertical pillars that may extend only between adjacent laminae or may be superimposed through several laminae. They range from long and

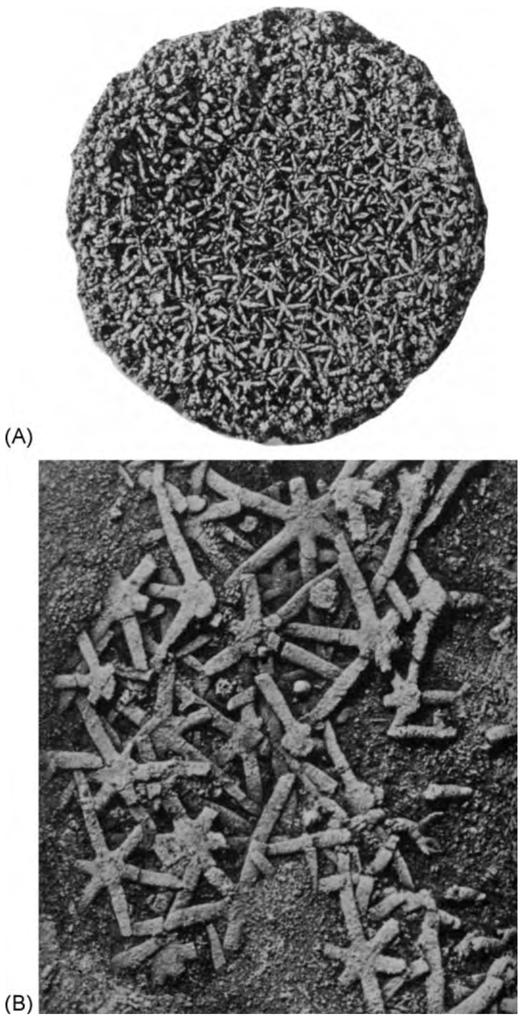


Figure 8 Representative heteractinid sponges. (A) *Astraeos pongium*, Silurian, Tennessee, USA, coarse spiculed, broad, saucer shaped sponge as seen from above, with large octactine spicules, natural size. (B) *Ensiferites*, Devonian, New York, USA, large spicules with six horizontal tangential rays, (×5). Adapted with permission from Rigby JK (1987) Phylum Porifera. In: Board man RS, Cheetham AH, and Rowell AJ (eds.) *Fossil Invertebrates*. Palo Alto, Oxford, London: Blackwell Scientific Publications.

continuous to more spool-shaped and irregular, or they may be rod-like in an interconnected network. Laminate skeletons are commonly perforated by vertical astrorhizal canals with transverse or horizontal stellate tributary canals which show best as interruptions in laminae. Several different types of microstructure are present in the calcareous skeletal structures, apparently a result of variations in the secretion of calcium carbonate skeletons, and in their subsequent diagenesis.

Stromatoporoids first appeared in the geological record in the Ordovician, and by the Middle Devonian they occurred worldwide in sediments of shallow marine environments, where they were significant contributors to mid-Palaeozoic reefs. They left a discontinuous Carboniferous record, but became more common in the Permian and thrived during the mid-Mesozoic, when they again formed reefs. They became extinct at the end of the Cretaceous.

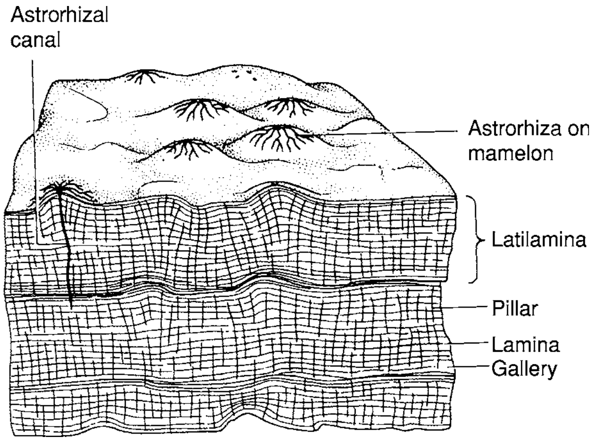


Figure 9 Block diagram of the stromatoporoid, *Actinostroma*, showing internal horizontal thin laminae and coarser latilaminae, with vertical pillars that combine with laminae to define openings (galleries) in the skeleton. Mound like mamelons on the surface are marked by convergent canals at the upper ends of astrorhizal canals (×5). Adapted with permission from Rigby JK (1987) Phylum Porifera. In: Boardman RS, Cheetham AH, and Rowell AJ (eds.) *Fossil Invertebrates*. Palo Alto, Oxford, London: Blackwell Scientific Publications.

The possibly related chaetetids are a moderately small group of colonial organisms that were included with tabulate corals for many years, and only recently have been considered to be sclerosponges, once spicule pseudomorphs and astrorhizal canals were observed in Palaeozoic representatives of the group. The chaetetids were shallow-water marine organisms. They range stratigraphically from the Ordovician to the Miocene, and are locally abundant organisms in Pennsylvanian formations. For example, they form small mound-like reefs in Middle Pennsylvanian rocks of Kansas.

They have compound skeletons that range from plate-like or domed to columnar (Figure 10), and are composed of clustered narrow tubes, or calicles, that are less than 1 mm in diameter and commonly polygonal to irregular in cross-section. These calicles lack true septae, but have common horizontal tabulae, and their shared or joined walls are imperforate. The colonies appear to have enlarged by budding or longitudinal fission of calicles. The growth form, skeletal structure, and the rare occurrence of spicules indicate a strong relationship to other sclerosponges.

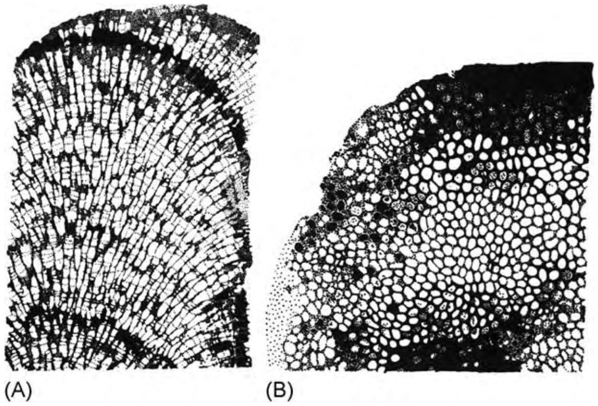


Figure 10 Longitudinal (A) and transverse (B) sections of *Chaetetes ?septosus* Fleming, Lower Carboniferous, Europe, showing narrow calicles, or tubes, with polygonal transverse sections, that become rounded where growth was inhibited, and perforate horizontal tabulae (×4). Adapted from Hill and Stumm, (1956), Treatise on Invertebrate Paleontology, Part F, with permission from, The Geological Society of America and The University of Kansas Press.

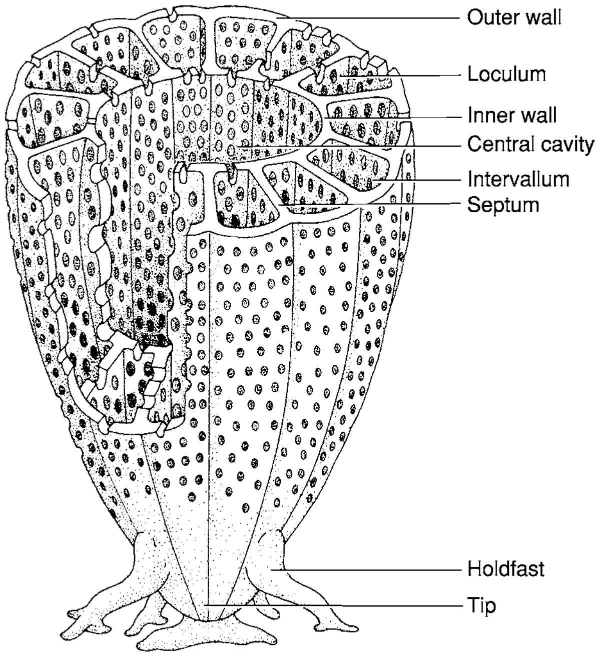


Figure 11 Generalized archaeocyathan showing characteristic skeletal elements of a double walled, steeply obconical, form (×2). Adapted with permission from Rigby JK (1987) Phylum Porifera. In: Boardman RS, Cheetham AH, and Rowell AJ (eds.) *Fossil Invertebrates*. Palo Alto, Oxford, London: Blackwell Scientific Publications.

Archaeocyatha

Fossils of the Class Archaeocyatha are known essentially from the Lower Cambrian, where they became relatively abundant as the first metazoan reef formers, and from which over 300 genera have been described. They appear in the geological record near the base of the Cambrian, but were extinct by Middle Cambrian time, except in Russia and Antarctica where two genera persisted into the Early–Middle Cambrian.

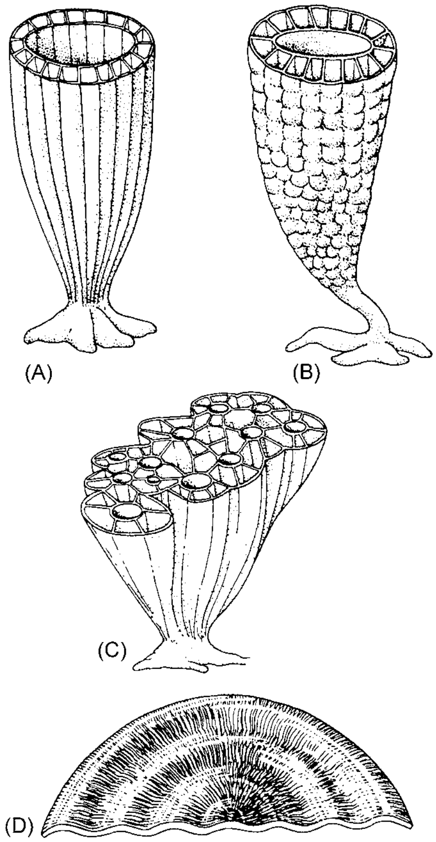


Figure 12 Range in growth forms of representative solitary and colonial archaeocyathans. (A) *Ajacicyathus*, a simple obconical erect form. (B) *Kotuyicyathus*, sculptured obconical form. (C) *Ajacicyathus*, massive colonial form. (D) *Okulitchicyathus*, wavy, discoidal flabellate form. All approximately natural size and all of Lower Cambrian age. Adapted with permission from Rigby JK (1987) Phylum Porifera. In: Boardman RS, Cheetham AH, and Rowell AJ (eds.) *Fossil Invertebrates*. Palo Alto, Oxford, London: Blackwell Scientific Publications.

Their skeletons are typically solitary conical to branching structures and, in early simple forms, are structures in which a porous inner wall surrounds the central cavity and is separated from a porous outer wall by the intervallum, which may be divided into longitudinal openings by radial septa that bridge between the two walls (Figure 11). Basal tips may have holdfasts, which are root-like attachment structures. Simple single-walled forms and more complex thalamid chambered forms are also known, as are less common fan-shaped and bowl-shaped genera and branched colonial forms with moderately complex walls (Figure 12).

Class Archaeocyatha Bornemann, 1884
  Order Monocyathida Okulitch, 1935, Lower Cambrian
  Order Ajacicyathida Bedford & Bedford, 1939, Lower Cambrian
  Order Archaeocyathida Okulitch, 1935, Lower–Middle Cambrian
  Order Capsulocyathida Zhuravleva, 1964, Lower Cambrian
  Order Kazachstanicyathida Konyushkov, 1967, Lower Cambrian
  Order Tabulacyathida Vologdin, 1956, Lower Cambrian

The archaeocyathans were an exclusively marine group of organisms, and are most common and diverse in argillaceous and carbonate rocks that appear to have accumulated in normal marine waters up to 20–30 m deep.

See Also

Mesozoic: Jurassic. **Palaeoclimates**. **Palaeozoic:** Cambrian; Permian. **Sedimentary Environments:** Reefs ('Build-Ups').

Further Reading

Bergquist PR (1978) *Sponges*. Berkeley and Los Angeles: University of California Press.

Broadhead TW (ed.) (1983) *Sponges and Spongiomorphs*, Notes for a short course, organized by J. K. Rigby and C. W. Stearn, *University of Tennessee Studies in Geology* 7. Knoxville: University of Tennessee. (Contains several chapters concerning various groups now included within the Porifera, each written by an authority on the group.)

De Laubenfels MW (1955) Porifera. In: Moore RC (ed.) *Archaeocyatha and Porifera,* Part E, *Treatise on Invertebrate Paleontology,* pp. E21 E112. Boulder, Colorado and Lawrence: Geological Society of America and University of Kansas Press.

Finks RM (1960) Late Paleozoic sponge faunas of the Texas region: the siliceous sponges. *American Museum of Natural History Bulletin* 120.

Hall J and Clarke JM (1899) A memoir of the Paleozoic reticulate sponges constituting the family Dictyospongidae. *New York State Museum Memoir* 2.

Hill D (1972) Archaeocyatha. In: Teichert C (ed.) *Treatise on Invertebrate Paleontology,* Part E (revised), 2nd edn, pp. E1 E158. Boulder, Colorado and Lawrence: Geological Society of America and The University of Kansas.

Hill D and Stumm EC (1956) Tabulata. In: Moore RC (ed.) *Coelenterata,* Part F, *Treatise on Invertebrate Paleontology,* pp. F444 F475. Boulder, Colorado and Lawrence: Geological Society of America and University of Kansas Press.

Hooper JNA and Van Soest RWM (eds.) (2002) *Systema Porifera, A Guide to the Classification of Sponges*. New York: Kluwer Academic/Plenum Publishers (This major work has many authoritative chapters, written by a variety of authors, on the spectrum of groups currently considered within the Porifera, but with an emphasis on living forms).

Pisera AA (1997) Upper Jurassic siliceous sponges from the Swabian Alb: taxonomy and paleoecology. *Palaeontologie Polonica* 57: 3 219.

Reid REH (1958 1964) A monograph on the Upper Cretaceous Hexactinellida of Great Britain and Northern Ireland. *Palaeontographical Society (London) Monograph* 1958, Part 1, pp. i xlvi; 1959, Part 2, pp. xlvii xlviii, 1 26; 1961, Part 3, pp. 27 48; 1964, Part 4, pp. xlix cliv.

Rigby JK (1986) Sponges of the Burgess Shale (Middle Cambrian) British Columbia. *Palaeontographica Canadiana* 2.

Rigby JK (1987) Phylum Porifera. In: Boardman RS, Cheetham AH, and Rowell AJ (eds.) *Fossil Invertebrates,* pp. 116 139. Palo Alto, Oxford, London: Blackwell Scientific Publications.

Rigby JK (1991) Evolution of Paleozoic heteractinid calcareous sponges and demosponges patterns and records. In: Reitner J and Keupp H (eds.) *Fossil and Recent sponges,* pp. 83 101. Berlin and Heidelberg: Springer Verlag.

Rigby JK and Senowbari Daryan B (1996) Upper Permian inozoid, demospongid, and hexactinellid sponges from Djebel Tebaga, Tunisia. *The University of Kansas Paleontological Contributions, New Series, 7.*

Rigby JK, Senowbari Daryan B, and Liu H (1998) Sponges of the Permian Upper Capitan Limestone, Guadalupe Mountains, New Mexico and Texas. *Brigham Young University Geology Studies* 43: 19 117.

FOSSIL PLANTS

Contents

Angiosperms
Calcareous Algae
Fungi and Lichens
Gymnosperms

Angiosperms

P Kenrick, The Natural History Museum, London, UK

Introduction

The angiosperms, or flowering plants, have been the dominant vegetation on land for over 60 million years. The group encompasses an enormous diversity of forms, which reflects the variety of habitats that they occupy as well as their interactions with other organisms – in particular, animals. Conservative estimates indicate that there are in excess of 220 000 known living species of flowering plant, which is much more than all other land plants combined. Despite the ecological supremacy of this group and the huge diversity of living forms, in geological terms the angiosperms are comparative newcomers. Fossil evidence shows that angiosperms first appeared and began to diversify towards the end of the Mesozoic Era, about 130 million years ago. Whereas the early diversification of angiosperms is increasingly well documented, the origin of the group is a subject that has been shrouded in mystery and much debated. For many years, investigation has been hampered by a seemingly uninformative early fossil record, uncertainty about relationships among living species, and apparently insuperable gaps between certain aspects of flowering plant morphology and that of the closest relatives, the gymnosperms. Recently, there have been significant advances in our understanding of the relationships of flowering plants. Of particular importance has been the application of cladistic methods to the time-honoured approach of comparative morphology and the increasing use of molecular methods. From a geological perspective, the use of maceration and sieving techniques on Cretaceous clays has brought to light a wealth of minute but exquisitely preserved fossil flowers, documenting new early evidence on floral diversity. Together with analyses of the fossil record of pollen, these new data are providing a much clearer picture of the early evolution of flowering plants and their rapid rise to dominance.

Characteristics

Angiosperms are, for the most part, an easily recognizable group of plants: they are characterized by the production of flowers. The flower consists of four basic elements arranged in series around a central axis. These include (1) the sepals, which are an outer series of modified bracts, (2) the petals, an inner series of modified bracts that are often coloured, (3) the stamens, which contain the pollen, and (4) the carpels, which contain the ovules (seeds when fertilized). Not all of these units are present in all flowers. For example, within the magnoliids, which are a basal assemblage within angiosperms, there are both large multipart flowers (e.g., *Magnolia*) as well as minute, simple flowers comprising a single carpel and two stamens (e.g., *Piper*) (Figure 1). The enormous complexity of form, colour, and scent in flowers is linked to the evolution of animal pollinators, especially insects. A further major characteristic of angiosperms is that the ovules are enclosed within an ovary. Fertilization takes place when pollen lands on a special surface (the stigma) and germinates, producing a pollen tube that grows into the ovary and delivers the male nuclear material to the ovules. Angiosperms come in both woody and herbaceous forms. They possess a type of pollen with a distinctive cell wall. Broad leaves with reticulate venation are typical of many members of the group. The wood is also typically diagnostic. Seeds, wood, leaves, and pollen are frequently preserved as fossils (Figures 2 and 3), and, together, serve as principal sources of evidence on the geological history of the group. Flowers are more rarely preserved (Figure 4), because they are more likely to fragment into their constituent pieces.

Classification

Classifications of angiosperms based on comparative morphology are still widely used, but these long-standing and familiar systems are now under major revision due to the impact of molecular systematics.

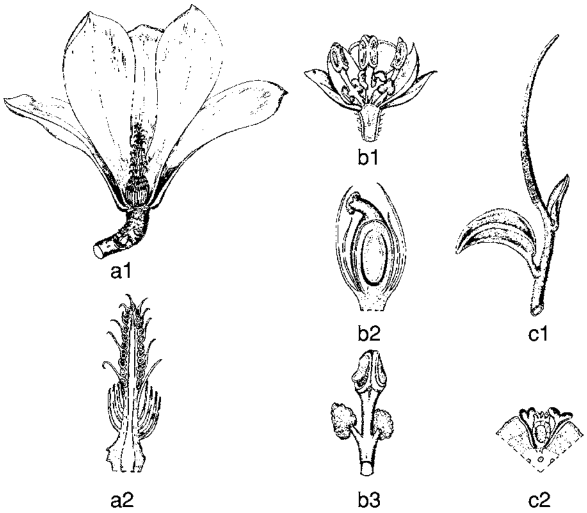


Figure 1 Floral diversity in magnoliid angiosperms. Left: *Magnolia* (Magnoliaceae). (a1) Section through flower showing petals and central receptacle bearing numerous stamens and ovules. (a2) Section through receptacle showing details of ovules in upper part and stamens in lower part. Middle: *Laurus* (Lauraceae). (b1) Male flower, showing petals and stamens. (b2) Female flower, showing ovary with one ovule. (b3) Single stamen detached from flower, showing flaplike valve at apex and two glands near the base. Right: *Peperomia* (Piperaceae). (c1) Elongate terminal inflorescence composed of numerous closely set simple flowers. (c2) Transverse section through inflorescence, showing single flower composed of central ovule and two stamens. Reproduced with permission from Hickey M and King C (1981) *100 Families of Flowering Plants*, p. 567. Cambridge: Cambridge University Press.

The family unit is the focal point of higher level classifications, and from 320 to 590 or more families are recognized, depending on taxonomist and system. Orders and higher level groupings have until recently been of little importance. With a few exceptions, these are difficult to recognize on morphological criteria, and the monophyletic status of such groupings is frequently questioned.

It is no exaggeration to say that the application of gene sequencing technology is revolutionizing the understanding of relationships among angiosperms at all levels. The use of molecular methods is more advanced in plants than in any other major group of organisms, and new classifications are based on a molecular systematic framework. A recent classification published by the Angiosperm Phylogeny Research Group recognized 462 families. These have been classified into 40 monophyletic orders, which are further categorized into a number of informally named higher groups (Figure 5). The outline classification illustrated in Figure 5 represents a major advance in knowledge of the relationships among angiosperms, and much of this has been worked out since 1990. Conservative estimates indicate that at least 10% of families are not monophyletic groups. Furthermore, although many families can be placed within higher groups, their relationships within these are still unclear. Many families are not yet included within orders. Much work remains to be done in fleshing out the details within and between families and in some cases orders, but the currently broad picture is unlikely to change in a radical way.

The root of the angiosperm phylogenetic tree lies within a basal grade of organization that has been dubbed the magnoliid dicots. Included in the magnoliids are families such as Amborellaceae, Austrobaileyaceae, Canellaceae, Chloranthaceae, Hydnoraceaceae,

Figure 2 *Acer trilobatum* (Miocene; Oeningen, Baden, Germany). The leaves of angiosperms are common in the fossil record beginning in the mid Cretaceous.

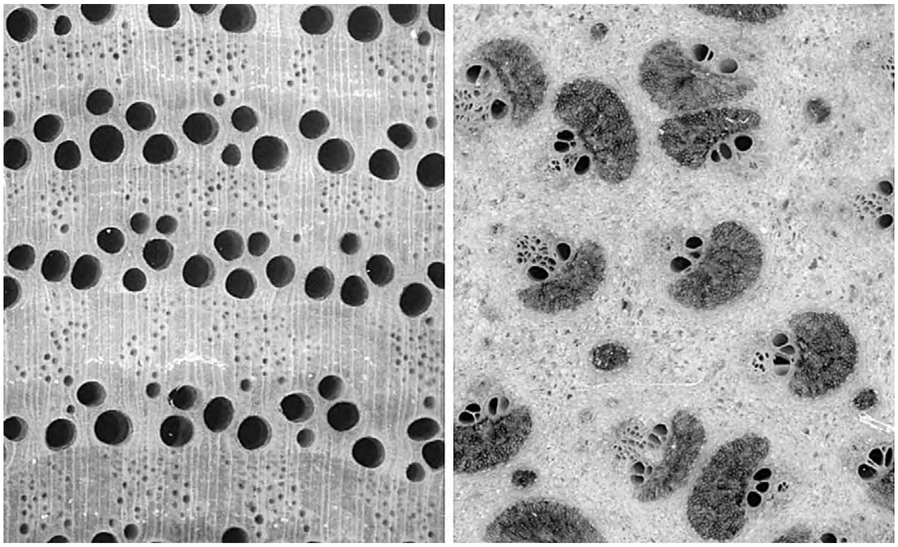


Figure 3 Angiosperm wood with its distinctive cell structure is first recorded in sediments of the mid Cretaceous (Albian stage). The images show two transverse sections through fossilized angiosperm wood. Left: *Quercus* (oak, Miocene; Saddle Mountain, Washington, USA). Right: *Palmoxylon* (palm, Oligocene; Antigua, Eastern Caribbean).

Figure 4 Fossil flower. *Porana oeningensis* (Miocene; Oeningen, Baden, Germany).

Illiciaceae, Nymphaeaceae [+ Cabombaceae], Rafflesiaceae, Schisandraceae, Trimeniaceae, and Winteraceae. The group also encompasses the orders Ceratophyllales, Laurales, Magnoliales, and Piperales. Flowers are basically simple in the magnoliids. The large, complex flowers of magnolias and water lilies are exceptions (Figure 1). The monotypic Amborellaceae is probably the most basal named taxon and sister group, compared to all other angiosperms. The interrelations among the basal orders of angiosperms remain elusive. These groups include the magnolias, peppers, and laurels. The informally named 'monocots' (monocotyledons) is one of the few higher groups that has been widely recognized on morphological and molecular criteria (Figure 5). The monocots are clearly monophyletic, and this large clade comprises ~22% of living species arrayed in 10 orders, including familiar groups such as the lilies, palms, and grasses. All other flowering plants are in the informally named 'eudicots' clade (Figure 5). Within eudicots there is increasing support for a large subgroup with predominantly five-part flowers, termed 'core eudicots'. This group contains two large subclades, the rosids and asterids. Over 75% of living species diversity is contained within the eudicots.

Angiosperm Origins

The origin of flowering plants is one of the most widely discussed and enduring mysteries in evolutionary botany. This whole area has remained controversial, and it is plagued by competing and contradictory ideas. There are several reasons for this. First, until recently there was no clear consensus about which of a diverse group of primitive living families is actually basal within angiosperms. In other words, the root of the angiosperm family tree was unknown, and therefore what constitutes the primitive condition in key

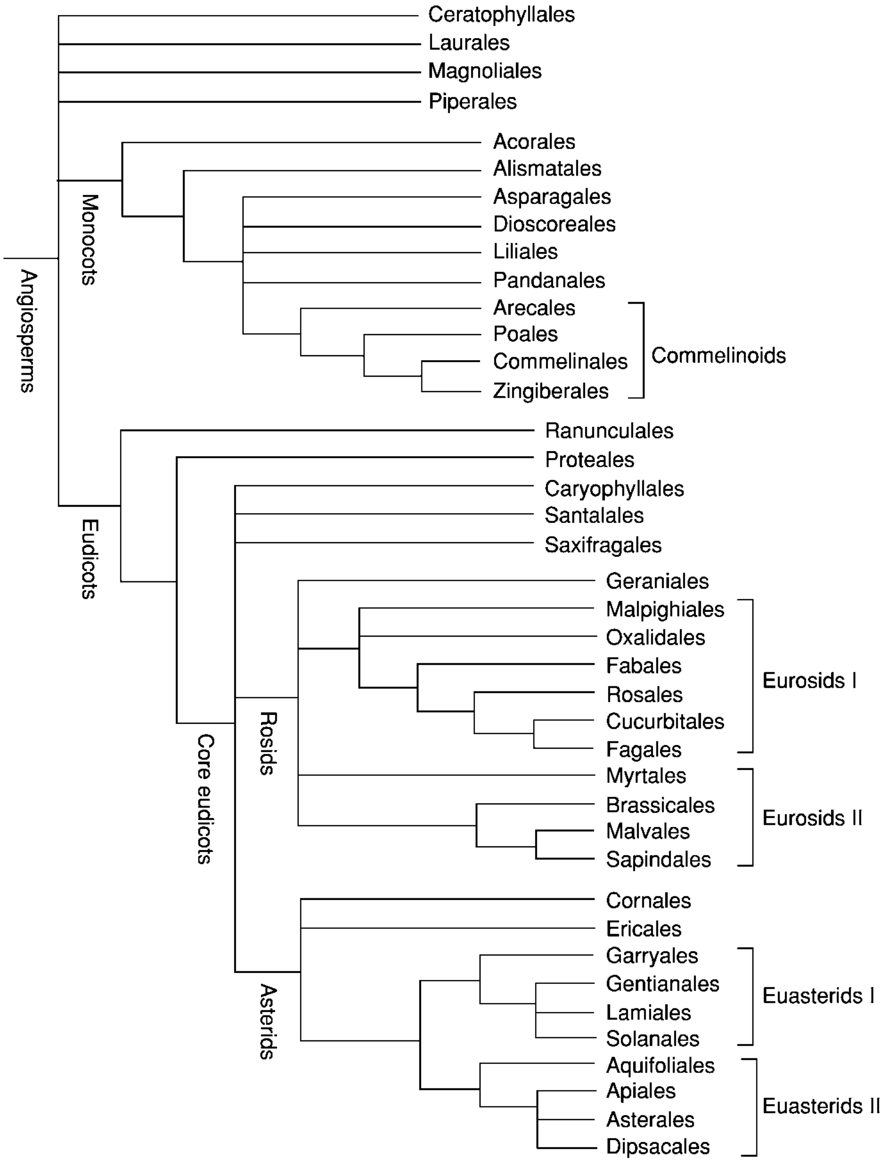


Figure 5 The 1998 Angiosperm Phylogeny Group classification of living angiosperms at the ordinal level. The recognized groupings are based on both molecular and morphological characteristics. Some families (not shown) are not placed within orders.

features, such as the flower, was difficult to assess. Second, although the fossil record is replete with information on pollen, leaves, fruit, and wood, the record of flowers is much sparser. First-hand information on the nature of early Cretaceous flowers has been hard to find. This means that one of the crucial features of flowering plants has remained poorly characterized in early members of the group. Also, because most fossils represent organs or plant fragments rather than whole individuals, their implications for angiosperm origins are not always easily interpreted. Third, although it is clear that angiosperms arose from within the gymnosperms, there are conflicting ideas on which living or extinct groups of gymnosperms constitute the closest relatives. Furthermore, living angiosperms have diverged substantially from related plants in a number of ways. This makes it difficult to draw comparisons between, for example, aspects of the angiosperm flower and putative equivalent structures in gymnosperms. In other words, angiosperms are divergent, and this obscures their relationships to other groups of plants. Fourth, 'angiosperm origins' is really a discussion about the origins of two groups, not one, and this can lead to confusion and argument (Figure 6). There is the issue of the origin of the crown group, which is the group that contains all living species and their most recent common ancestor. Members of this group typically

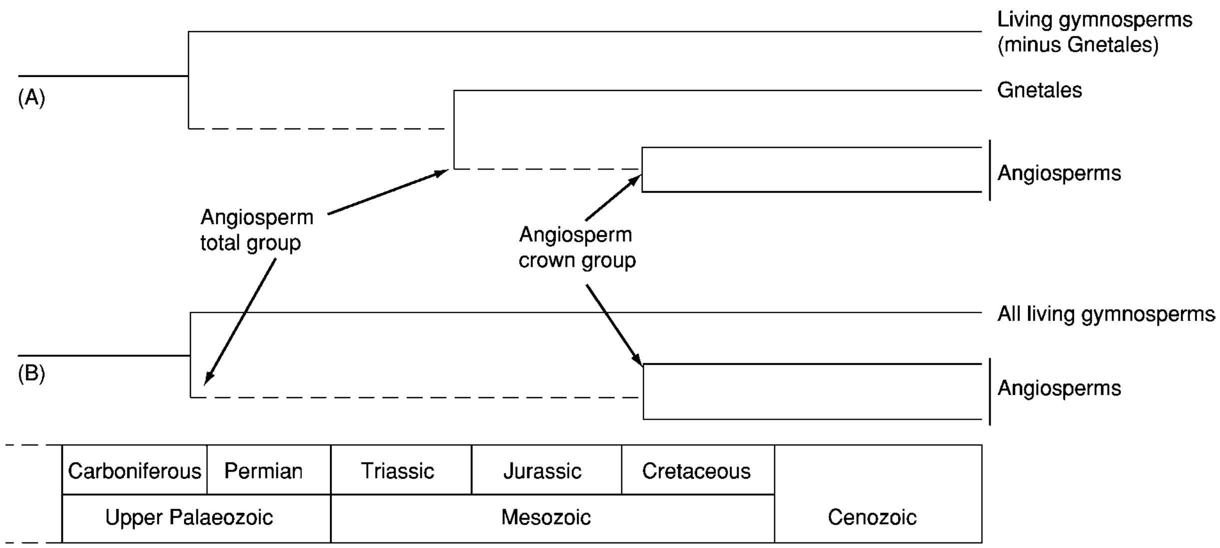


Figure 6 Two hypotheses (A and B) of angiosperm origins. In both hypotheses, the angiosperm crown group (the group that contains the most recent common ancestor of all living species) originated and diversified during the Early to mid Cretaceous. The two alternatives have implications for the origin of the angiosperm total group. (A) Anthophyte hypothesis. Gymnosperms in the Gnetales are the living sister group to angiosperms. Gnetales are documented in the fossil record long before unequivocal angiosperms. This hypothesis implies an angiosperm ghost lineage extending into the Late Triassic. (B) Gne pine hypothesis. Living gymnosperms are a monophyletic group the Gnetales (not shown) are most closely related to conifers, not angiosperms. This whole gymnosperm clade is sister group to angiosperms. This hypothesis implies an angiosperm ghost lineage extending back as far as the Carboniferous Period. Note that direct fossil evidence of angiosperm ghost lineage members is scant. However, there are a number of candidates among Mesozoic gymnosperms.

possess all of the familiar features that we associate with modern angiosperms. There is also the question of the origin of the total group, which contains the crown group plus closely related extinct plants that possess some, but not all, of the attributes of modern angiosperms. This is the group that formed when the angiosperm lineage split from the lineage leading to its sister group in living gymnosperms. It is important to understand that these are two different but related issues, and that although both group concepts encapsulate aspects of the origin of angiosperms, they do not mean quite the same thing. Recent advances in phylogenetic research and new discoveries in the fossil record have gone a long way towards addressing some of these problems.

One of the unequivocal findings of recent phylogenetic research is that angiosperms are a monophyletic group. This means that it is correct to dispense with earlier theories of multiple independent origins of flowering plants. There is no longer any doubt that the group had a single origin: all angiosperms share a common ancestor. Molecular phylogenetic analyses have identified the water lilies (Nymphaeales: Cabombaceae and Nymphaeaceae), together with four other small groups of flowering plants (Amborellaceae, Illiciales, Trimeniaceae, and Austrobaileyaceae), as the first diverging lineages from the main branch of the angiosperm phylogenetic tree. This means that researchers are now very close to understanding exactly where the root of angiosperms lies within the vast number of living species. From the perspective of the fossil record, the most interesting data have come through the application of a new method of looking for early fossil evidence for flowers. This involves the disaggregation and sieving of Cretaceous clays to separate mineral from organic constituents. This approach has yielded remarkable new information on microscopic fossilized flowers preserved either as charcoal or in a mummified state. These data provide clear evidence of the Chloranthaceae and the Nymphaeales in Early Cretaceous sediments. Both groups are thought to be basal based on phylogenetic research. Together with evidence from the fossil record of dispersed pollen, a conservative minimum age estimate for the origin of the angiosperm crown group would be Early Cretaceous (Valanginian). Furthermore, the absence of distinctive triaperturate pollen (diagnostic of a large subclade within angiosperms: eudicots) in well-documented Triassic and Jurassic palynofloras argues against a long period of cryptic evolution of the crown group implied by estimates based on molecular clocks.

The origin of angiosperms also encompasses the dating of the age at which the lineage leading to flowering plants split from living gymnosperms (i.e., the total group) as well as the discovery and description of fossils of an intermediate nature (Figure 6). Dating the origin of the total group provides only minimum age estimates. Counterintuitively, this does not depend entirely on finding early angiosperm-like fossils. Age estimates of the angiosperm total group can also be made by direct inference from the stratigraphic range of the angiosperm sister group (Figure 6). To do this, it is essential to know how angiosperms are related to gymnosperms (*see* **Fossil Plants:** Gymnosperms), but this is one of the most controversial areas in systematic botany. There are two main ideas on relationship and these have radically different implications for the origin of angiosperms and for the relevance of extinct Mesozoic gymnosperms (Figure 6). Comparative morphology indicates that angiosperms are most closely related to living gymnosperms in the Gnetales and to extinct gymnosperms in the Pentoxylales and the Bennettitales. This has been termed the 'anthophyte hypothesis'. Molecular data tell a very different story: they place Gnetales within Coniferales as sister group to Pinaceae. Angiosperms emerge as sister group to a monophyletic group comprising all living gymnosperms. This has been termed the 'gne-pine hypothesis'. Neither hypothesis, though, is particularly strongly supported by the current data. Furthermore, because molecular data cannot deal with the fossil groups, it is, of course, unclear where in molecular phylogenetic trees the fossils would in fact fall out.

The anthophyte and the gne-pine hypotheses have very different implications for the origin of angiosperms and for the relevance to the question of extinct Mesozoic gymnosperms (Figure 6). The anthophyte hypothesis indicates that the lineage leading to angiosperms originated in the mid-Triassic, because this is when fossils attributable to living gymnosperms in the Gnetales first appear. If this hypothesis is correct, the search for angiosperm relatives should be confined to extinct gymnosperms of Late Triassic and Jurassic ages. The gne-pine hypothesis, however, removes Gnetales as sister group to angiosperms and replaces it with the entire gymnosperm crown group. Living gymnosperms are much more remote in a phylogenetic sense from angiosperms. This would indicate a much more ancient origin of the angiosperm lineage, which could reach back as far as the Late Carboniferous. Under this hypothesis, the search for angiosperm relatives should cast a wider net that encompasses gymnosperms of Early Mesozoic as well as Late Palaeozoic ages. Even though it is likely that the lineage leading to angiosperms diverged from other plants during the Triassic or earlier, current knowledge of Mesozoic floras makes it unlikely that angiosperms were diverse or abundant before the Cretaceous.

The search for early angiosperm-like fossils has to confront the problem of recognition. In most respects, early fossils are likely to be more similar to gymnosperms of one sort or another than what might currently be considered as a typical member of the angiosperm crown group. This is undoubtedly one reason why such fossils have proven to be so elusive. There are several plausible candidates among the Mesozoic gymnosperms, but questions remain in all cases due to missing data on key organs. The most recent fossil to emerge as potential sister group to the angiosperm crown group is *Archaefructus* (Figure 7). This plant comes from the famous Early Cretaceous Jehol Biota of north-eastern China. *Archaefructus* is thought to have been a herbaceous aquatic. It had thin stems and finely dissected compound leaves. The 'flower' presents a unique combination of features. It bore clusters of structures interpreted as carpels at the shoot apex. These were subtended by a zone that bore stamens. The pollen was monosulcate, resembling common forms found dispersed in Cretaceous sediments. Unlike most modern angiosperms, though, the 'flower' of *Archaefructus* did

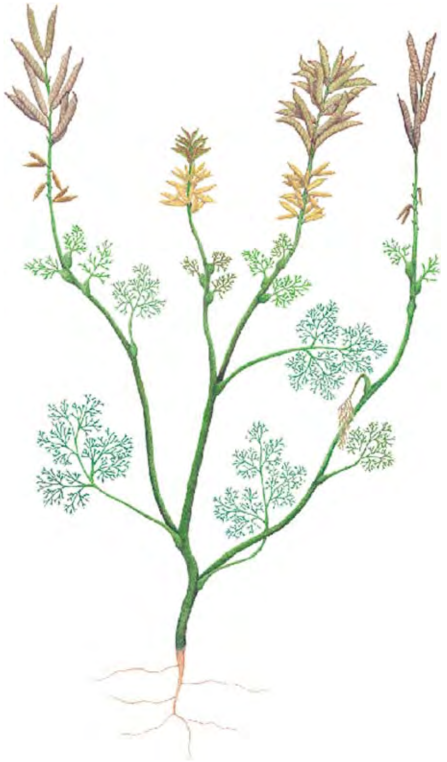


Figure 7 *Archaefructus liaoningensis.* An Early Cretaceous plant thought to be an early angiosperm or close relative of the angio sperms. Reproduced with permission from Sun G, Ji Q, Dilcher DL, *et al.* (2002) Archaefructacea, a new basal angiosperm family. *Science* 296(5569): 899 904.

not have petals and sepals. On the face of it, *Archaefructus* would seem to fall within the angiosperm total group but outside the crown group. As such, it could provide the first clear evidence of a fossil intermediate between living gymnosperms and living angiosperms. Some aspects of the interpretation of *Archaefructus* are, however, controversial, and the precise nature of its relationship with angiosperms is still in question.

An important source of evidence on the origin of angiosperms is the fossil record of pollen (Figure 8). Because pollen is produced in vast quantities and is dispersed over wide areas, the pollen 'footprint' of plants is relatively large. Pollen could therefore provide the most precise constraint on age as well as providing a window onto floral changes through time. The disadvantage of pollen is that its taxonomic resolution is generally low, due to the small suite of available characters. Whereas some pollen seems to be diagnostic of higher level groups, other pollen is more difficult to place within families or orders. Furthermore, it is not always easy and may not be possible in some cases to distinguish clearly the angiosperm crown group or total group from closely related extinct gymnosperms based on pollen alone. Angiosperm-like pollen has been reported from sediments of Triassic and Jurassic ages. The so-called Crinopolles group (Late Triassic) has a very angiosperm-like monosulcate aperture combined with a reticulate–columellar exine. Analysis of exine structure using transmission electron microscopy indicates some differences from angiosperms. A relationship between Crinopolles pollen grains and angiosperms is possible, but it is unlikely to be indicative of the crown group.

Angiosperm Diversification

The early diversification of angiosperms was dominated by plants related to a phylogenetically basal grade (the magnoliid dicots; Figure 1). These are represented in Early Cretaceous sediments by leaves, pollen, and, increasingly, fossilized floral organs, fruits, and seeds. The maceration of clays from Early Cretaceous localities in Portugal and eastern North America, followed by sieving of the organic component, has been a key methodological breakthrough. This technique, combined with scanning electron microscopy of the organics, is revealing a huge diversity of plant organs attributable to charcoal-preserved or mummified angiosperms. One striking feature of the plant assemblages recovered in this way is that all of the early flowers are small – usually less than 2 mm long (Figure 9). It seems unlikely that this is due to depositional biases such as current sorting, because other larger plant parts belonging to conifers are usually found in association. Early fossil flowers are simple, being composed of relatively few parts and an undifferentiated perianth. Stamens are small and the pollen sacs typically disperse their pollen through a hinged flap or valve. The

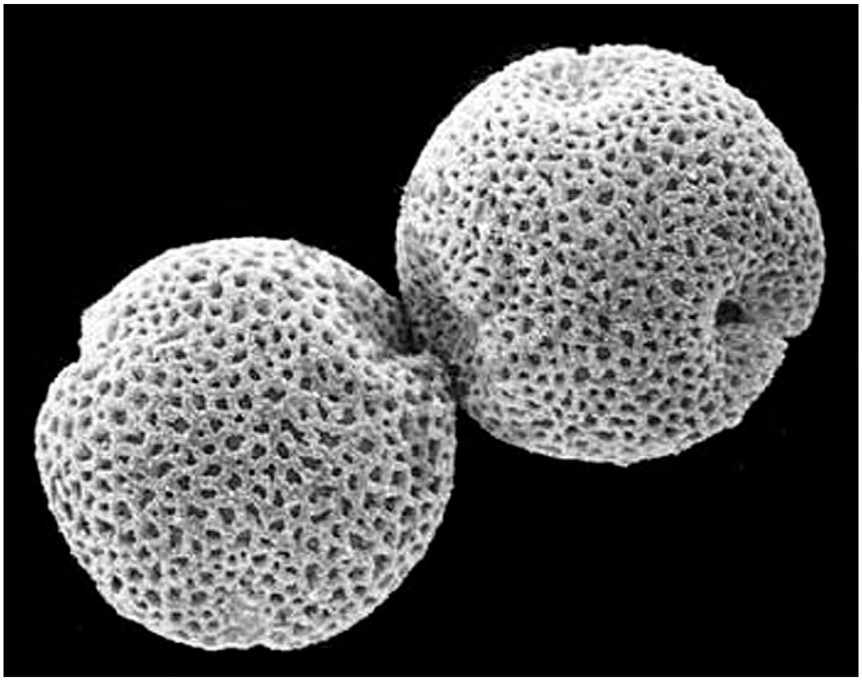


Figure 8 Pollen of living ash (*Fraxinus*).

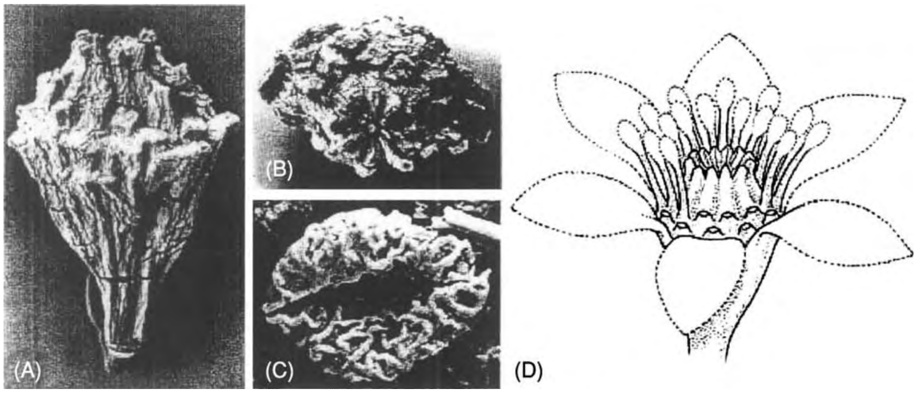


Figure 9 Fossil water lily from the Early Cretaceous (Barremian or Aptian) of Portugal. Scanning electron micrographs of flower (A, B) and pollen (C), and reconstruction of flower (D). Reproduced with permission from Friis EM, Pedersen KR, and Crane PR (2001) Fossil evidence of water lilies (Nyphaeales) in the Early Cretaceous. *Nature* 410: 357 360.

Figure 10 Water lilies, one of the most basal lineages of flowering plants.

stigmatic surface of the carpels is poorly differentiated. Early Cretaceous flowers and fruits are rarely assignable to living families or orders. They show a blend of features of the general magnoliid type.

The earliest fossils assignable to modern families have been recovered from the Barremian–Aptian stages of the Early Cretaceous. These floral organs, fruits, and seeds are attributable to water lilies (Nymphaeaceae) (Figures 9 and 10) and to Chloranthaceae. In marked contrast to the magnoliid dicots, the Cretaceous fossil record of monocots is depauperate. This is surprising in view of the basal phylogenetic position of the monocots, and it may reflect a representation bias and perhaps also recognition bias. Monocots are clearly present by the Late Cretaceous. The earliest fossils include fruits of the ginger family (Zingiberales) and leaves and stems of palms (Arecales). Many modern families of monocots are recognizable by the Early Cenozoic. The earliest fossil evidence for the eudicots comes from grains of triaperturate pollen dispersed in sediments of the Barremian–Aptian stages. The origins of eudicots therefore marginally predate the major diversification and ecological radiation of angiosperms.

Among the first modern families to appear during the Albian stage are the plane and buttonwood trees (Platanaceae) and probably also the box family (Buxaceae). Many more living families are recognizable by the Turonian stage. These include the witch hazels and sweet gums (Hamamelidaceae) and the saxifrages (Saxifragales). The rosids also first appear in the Turonian stage. There is evidence for the capers (Capparales), and somewhat later in the Santonian stage are records of the Myrtales; walnuts, hickories, and pecan nuts group (Juglandales); Myricales; and the birches, alders, beeches, and oaks group (Fagales). The main asterid lineage is also present by the Turonian stage. These are represented by flowers of ericalean affinity, and members of the Hydrangeaceae have been recorded in the Coniacian–Santonian stages. Nearly all of the main eudicot clades are represented by at least one member of their lineage in the Upper Cretaceous. Some notable groups do not, however, appear until the Maastrichtian stage or the Early Cenozoic. These include the pea or bean family (Fabaceae), the sunflower family (Asteraceae), the Lamiales, and the Gentianales, which together comprise 45% of living species diversity in eudicots. The three subfamilies of pea are well documented by flowers and fruits in the Eocene, but pollen grains of Caesalpinioideae have been recorded from the Maastrichtian stage of the Late Cretaceous. The earliest unequivocal records of the Asteraceae, Lamiales, and Gentianales come from the Palaeogene. The fact that most species-rich groups are known from relatively young fossils indicates that a significant proportion of eudicot diversity is the product of relatively recent radiations that occurred during the second half of angiosperm evolution. The evolutionary basis for the rapid diversification of specific eudicot clades remains unknown.

The rapid rise to prominence of angiosperms during the mid-Cretaceous marks the transition from Mesozoic ecosystems dominated by ferns and gymnosperms to those of the Late Cretaceous and Cenozoic, when flowering plants predominate. The record of fossil pollen in the northern hemisphere shows that the initial increase in angiosperm diversity occurred at low palaeolatitudes (Figure 11). During the Early Cretaceous, low-latitude areas experienced semiarid or seasonally arid climates. These conditions may have promoted a weedy life history with precocious reproduction favouring herbaceous annuals rather than long-lived woody perennials. There is some fossil evidence to support this. With the exception of one disputed record, angiosperm wood has not been recorded from rocks that predate the Albian. Angiosperm leaves are also extremely rare in these older sediments. Furthermore, it seems likely that some of the angiosperms from other Lower Cretaceous assemblages (e.g., Crato Formation, Brazil, and Portugal) were aquatic. These data indicate that the earliest angiosperms were predominantly herbaceous and that there may have been considerable diversity in aquatic habitats.

By the end of the Cretaceous period, flowering plants represented about 50–80% of land plant species diversity, but pockets of gymnosperm- and fern-dominated vegetation persisted. This notable change in floral composition predates the end of the Mesozoic Era by some 30 million years and was more radical than that which occurred at the Cretaceous–Tertiary boundary. Some groups of flowering plants that are ecologically prominent today were completely absent from the Cretaceous Period. Grasses, for example, did not become abundant and widespread until the Late Oligocene and Miocene. Ferns may well have played a more prominent role as herbaceous ground cover in prairie vegetation prior to the mid-Cenozoic.

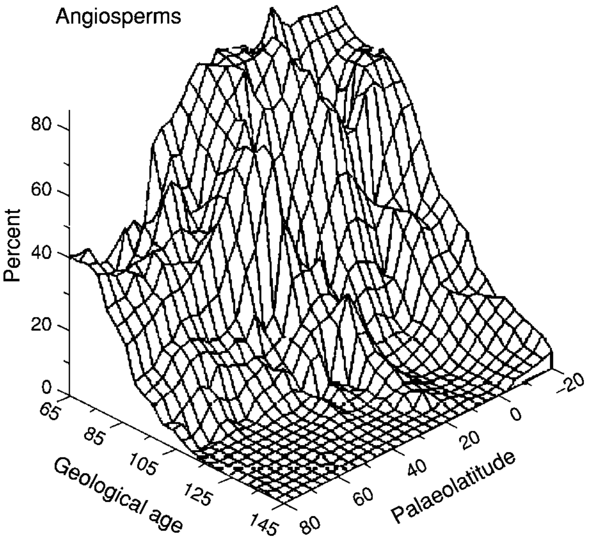


Figure 11 Analyses of dispersed pollen grains indicate that the diversification of angiosperms began at low palaeolatitude. Reproduced with permission from Crane PR and Lidgard S (1989) Angiosperm diversification and palaeolatitudinal gradients in Cretaceous floristic diversity. *Science* 246: 675 678.

The diversification of angiosperms has often been linked to the evolution of various animal groups. But proving cause and effect in the fossil record is often very difficult. For example, the rapid diversification of angiosperms during the mid-Cretaceous coincides broadly with a transition from sauropod-dominated to ornithopod-dominated dinosaur faunas (*see* **Fossil Vertebrates:** Dinosaurs). It is difficult to show whether these changes are in some way linked or just coincidental. A stronger case can be made for a link between insects (*see* **Fossil Invertebrates:** Insects) and flowering plants. The diversification of pollen- and nectar-collecting insects has been linked to flowering plants on the supposition that insect pollination provides new possibilities for reproductive isolation and therefore elevation of speciation rates. Compared to wind pollination, which is widespread in gymnosperms, insect pollination may permit more effective out-crossing at lower population densities and in a greater range of environments, thereby reducing extinction rates. Pollination biology may provide a plausible explanation for spectacular diversity in some angiosperms families (e.g., orchids), but the extent to which this provides a general explanation for the mid-Cretaceous rise of angiosperms is uncertain. There is evidence that insect pollination was an aspect of the biology of some extinct gymnosperms (e.g., Bennettitales) and some living gymnosperms (e.g., Gnetales). Early angiosperms may therefore have had the opportunity to co-opt pollinators from previously established relationships with other groups of seed plants.

The fossil record of flowers from the Early Cretaceous clearly indicates that early angiosperms were insect pollinated. Many of the features that have been documented in early fossil flowers are comparable to those found in modern insect-pollinated groups. The anthers show that pollen output was low, and the pollen grains are often smaller than the most effective size for wind dispersal. Release of pollen from anthers occurred via hinged flaps, and the individual grains are often covered by a pollenkitt-like substance. The stigmatic surfaces of the carpels are generally unelaborated. These features are consistent with

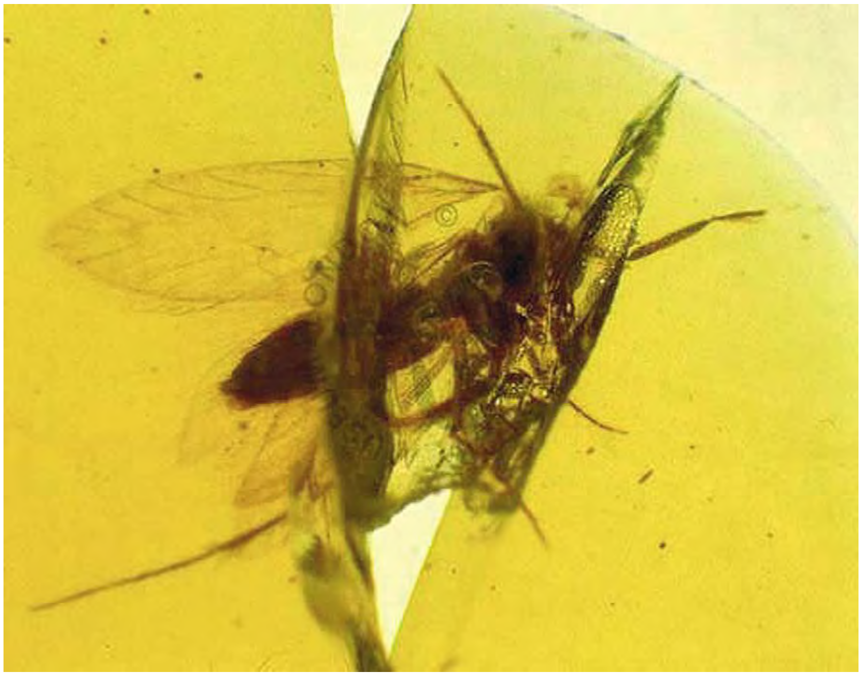


Figure 12 Moth belonging to the living pollen feeding family Micropterygidae, in Burmese amber of Late Cretaceous age.

pollination by pollen-collecting or pollen-eating insects (Figure 12). Flowers pollinated by nectar-collecting Hymenoptera and Lepidoptera occur in more derived groups of angiosperms and appear later in the fossil record. Even though early angiosperms may have been very similar to their living relatives in pollination syndrome, modes of fruit and seed dispersal were probably very different. Cretaceous fruits and seeds are generally very small compared to their modern relatives, and there is no evidence of specialized mammal or bird dispersal of fruits and seeds. Among basal groups of angiosperms, the evolution of fleshy fruits, arillate seeds, and other adaptations for dispersal by animals seems to be correlated with the evolution of fruit- and seed-eating birds and mammals, perhaps during the latest Cretaceous but most strikingly during the Early Tertiary.

See Also

Fossil Invertebrates: Insects. **Fossil Plants:** Gymnosperms. **Fossil Vertebrates:** Dinosaurs.

Further Reading

Angiosperm Phylogeny Research Group (APG) (1998) An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden* 85: 531 553.

Chase MW, Fay MF, and Savolainen V (2000) Higher level classification in the angiosperms: new insights from the perspective of DNA sequence data. *Taxon* 49(4): 685 704.

Crane PR, Friis EM, and Pedersen KR (1995) The origin and early diversification of angiosperms. *Nature* 374: 27 33.

Crane PR and Lidgard S (1989) Angiosperm diversification and palaeolatitudinal gradients in Cretaceous floristic diversity. *Science* 246: 675 678.

Friis EM, Chaloner WG, and Crane PR (eds.) (1987) *The Origin of Angiosperms and Their Biological Consequences*. Cambridge: Cambridge University Press.

Friis EM, Pedersen KR, and Crane PR (1999) Angiosperm origin and radiation. In: Briggs DEG and Crowther PR (eds.) *Palaeobiology II*, pp. 97 102. Oxford: Blackwell Science.

Herendeen PS and Crane PR (1995) The fossil history of the monocotyledons. In: Rudall PJ, Cribb PJ, Cuttler DF, and Humphries CJ (eds.) *Monocotyledons: Systematics and Evolution*, pp. 1 21. Kew: Royal Botanical Gardens.

Magallón S, Crane PR, and Herendeen PS (1999) Phylogenetic pattern, diversity, and diversification of eudicots. *Annals of the Missouri Botanical Garden* 86: 297 372.

Sun G, Ji Q, Dilcher DL, *et al.* (2002) Archaefructaceae, a new basal angiosperm family. *Science* 296(5569): 899 904.

Taylor DW and Hickey LJ (eds.) (1995) *Flowering Plant Origin, Evolution Phylogeny.* New York: Chapman & Hall.

Wikstrom N, Savolainen V, and Chase MW (2001) Evolution of the angiosperms: calibrating the family tree. *Proceedings of the Royal Society of London, Series B, Biological Sciences* 268(1482): 2211 2220.

Calcareous Algae

J C Braga, University of Granada, Granada, Spain
R Riding, Cardiff University, Cardiff, UK

Introduction

Algae are anatomically simple photosynthetic plants, lacking vascular tissues, roots, and leaves. They constitute a heterogeneous group comprising many divisions (*see* **Palaeontology**) (botanically equivalent to phyla). In addition, it is convenient to consider some photosynthetic bacteria (*see* **Biosediments and Biofilms**), such as Cyanobacteria (blue-green algae) along with eukaryotic algae. The algal body (thallus) can be unicellular, or multicellular with a certain degree of complexity due to cell differentiation. Although most algal groups are aquatic, some inhabit soils and subaerial environments.

Calcareous algae precipitate $CaCO_3$ during life at particular sites on or within their thalli. Calcification is a cross-systematic feature, occurring in diverse divisions, in freshwater and marine environments, and in planktic and benthic forms. However, fewer than 10% of extant benthic marine algae calcify, and they principally belong to red and green algae. Only one extant calcified genus of brown algae (*Padina*) is known, and cyanobacteria rarely calcify in marine environments at the present day although they did so at times in the past. In freshwater, charaleans and a few other chlorophytes calcify, and cyanobacterial calcification is common in calcareous streams and lakes. Photosynthesis raises pH and increases saturation state with respect to $CaCO_3$ minerals. However, the degree of biological control over calcification depends mainly on the location of sites for precipitation within the organism. In general, cyanobacterial calcification is environmentally controlled, and calcification by chlorophytes is also environmentally dependent. In contrast, coralline red algae and coccolithophores calcify intracellularly and relatively closely control their calcification. This is reflected in their wider latitudinal distribution in marine environments.

Fossils that may be calcified algae have been reported from the Neoproterozoic but the substantial record of calcified cyanobacteria and algae is mainly Phanerozoic (Figure 1). Their long geological history and wide environmental distribution result in calcified bacteria and algae being significant producers of calcium carbonate sediment in marine and freshwater deposits of many ages. Sedimentary products range from in-place reefal masses (*see* **Sedimentary Environments:** Reefs ('Build-Ups')), through nodules to coarse and fine bioclastic fragments. Postmortem disintegration of a wide variety of calcified algae and cyanobacteria has probably extensively produced mud-sand-gravel sediment, although often only coarser components are readily recognizable.

Rhodophyta (Red Algae)

Corallinales

Coralline algae are the major extant group of calcified marine red algae. Their thalli consist of branched cell filaments with a coherent unified, pseudoparenchymatous, organization. Growth is achieved by addition of new cells through cell division at the tip of each filament. Most present-day corallinaleans are heavily calcified by high magnesium calcite precipitated in the cell wall (Figure 2). This consists of a layer of crystals parallel to the wall from which radial crystals grow towards the cell interior. In non-geniculate (non-articulated) corallines all vegetative cells,

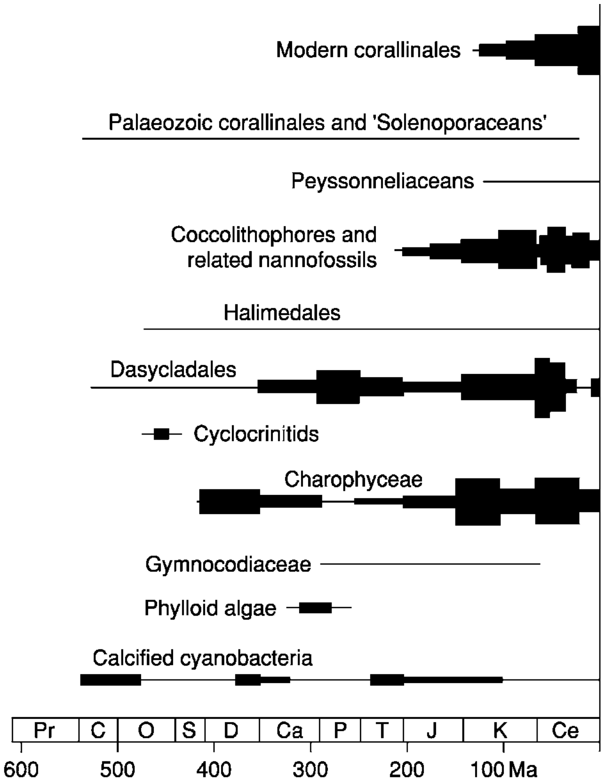


Figure 1 Stratigraphic range of the major groups of calcareous algae. Width of bars indicates the relative diversity of the group in each period. Pr: Proterozoic, C: Cambrian, O: Ordovician, S: Silurian, D: Devonian, Ca: Carboniferous, P: Permian, T: Triassic, J: Jurassic, K: Cretaceous, Ce: Cenozoic.

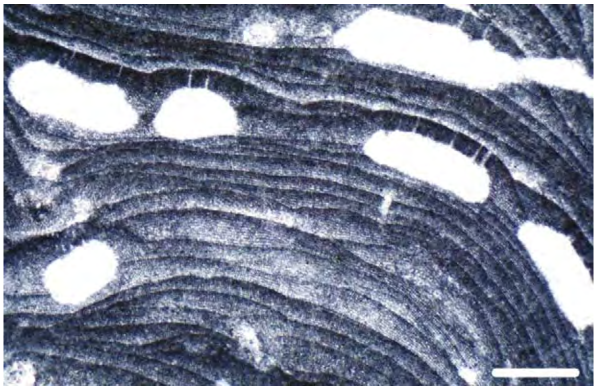


Figure 2 Section of the coralline alga *Lithothamnion*. The plant body consists of closely packed cell filaments with calcitic walls. Reproductive structures (conceptacles, multiporate in this example) remain empty after release of the spores. Dark and lighter areas represent seasonal growth zones. Upper Miocene, south eastern Spain. Scale bar 500 microns.

except those at the surface (epithallial cells), are calcified. Consequently, the entire plant has a rigid structure with high preservation potential. Non-geniculate corallines can grow on one another, and on firm substrates, including other plants, or live unattached on mobile sediments. They can form thick crusts and develop long protuberances and branches. Individual plants can reach several centimetres in size. In geniculate (articulated) species, calcified segments of the thallus are separated from one another by non-calcified joints (genicula). These articulated corallines form erect branching thalli usually attached to the substrate by an encrusting holdfast. They have a bushy appearance and grow up to 10 cm in height. After death, articulated corallines disintegrate and calcified segments are shed as sand and gravel particles.

Sexual and asexual reproduction both occur in the life cycle of most extant species. Gametes (male and female) and spores are produced in separate plants that otherwise exhibit similar vegetative morphology. Gametes and spores are produced in conceptacles within the thallus. The shape and size of conceptacles producing male gametes (and sometimes those producing female gametes too) differ from those in which spores develop. These differences can be recognised in fossil plants (Figure 2). The nature and number of openings (pores) of spore-producing conceptacles are key characters in corallinalean taxonomy.

Coralline algae constitute a cosmopolitan marine group with tropic to polar distribution and from the intertidal zone to depths of 270 m. In addition to this wide environmental range, they can tolerate very variable levels of nutrient supply and turbulence, as well as hypersaline to brackish coastal settings. Corallines are significant components of tropical shallow-water reefs (*see* **Sedimentary Environments:** Reefs ('Build-Ups')) but are also common on seafloors with low terrigenous supply at depths to 100–120 m in all climatic belts. In both shallow and deeper settings, they build rigid frameworks or lie unattached on loose substrates where they form nodules called rhodoliths.

Despite the wide geographic and environmental distribution of corallines as a whole, individual families and sub-families exhibit substantially differing depth-temperature distributions. Sporolithaceans are almost entirely restricted to low latitudes where they mainly occupy deep-water or cryptic habitats. Melobesioids dominate deep-water coralline assemblages in low–mid latitudes and occupy shallow water in high latitudes. Lithophylloids and mastophoroids live mainly in shallow water in low–mid latitudes, mastophoroids predominate in the tropics, and lithophylloids are more common in sub-tropical and warm temperate conditions. These habitat preferences of different taxa, together with variations in growth morphology according to levels of turbulence and light intensity, make fossil corallines valuable palaeoenvironmental indicators.

Coralline-like algae are known in the Ordovician (e.g., *Arenigiphyllum*) and Silurian (e.g., *Graticula*). The evolutionary relationships of these fossils, and of Late Palaeozoic fossils that have been called ancestral corallines (e.g., *Archaeolithophyllum*, see Phylloid Algae), to younger examples are still being elucidated. The continuous well-documented subsequent history of corallines commenced in the Early Cretaceous, and the group diversified throughout the Late Cretaceous and Cenozoic (Figure 1). About 30 extant genera are recognized.

'Solenoporaceae'

Solenoporaceans are traditionally regarded as an extinct family of red algae, morphologically similar to but simpler than corallines, ranging from the Cambrian to the Palaeogene. The superficial similarities that unite the 'Solenoporaceae' are simple, nodular form and an internal structure composed of narrow, juxtaposed branching tubes with diameters up to 100 microns (Figure 3). Details that indicate systematic differences include the cross-sectional shape of the tubes and the presence or absence of cross-partitions. However, the group is heterogeneous and contains some sponges as well as red algae, together with fossils resembling cyanobacteria. The type species of *Solenopora* is a chaetetid sponge. Some Ordovician and Silurian fossils formerly attributed to *Solenopora*, such as *Graticula*, closely resemble the extant coralline *Sporolithon* (see Corallinales above).

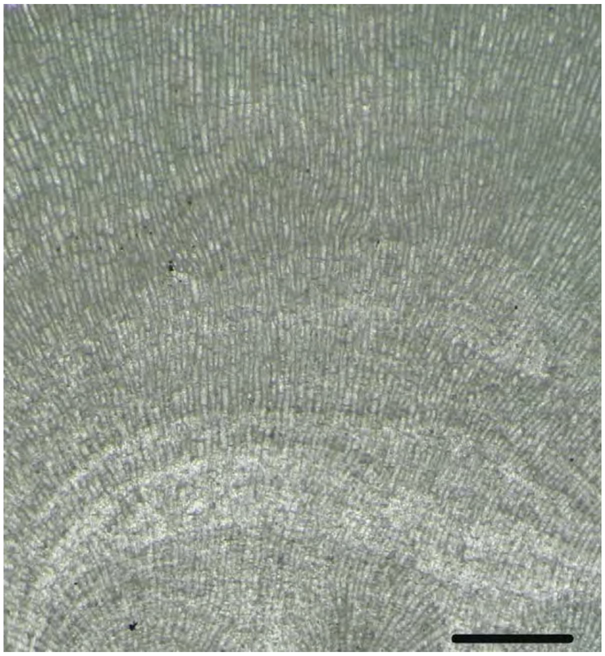


Figure 3 Section of *Solenoporella*, fossil attributed to 'solenoporaceans'. The densely packed cell filaments with transverse cell partitions suggest an affinity with red algae similar to Corallinales. Upper Jurassic, southern England. Scale bar 1 mm.

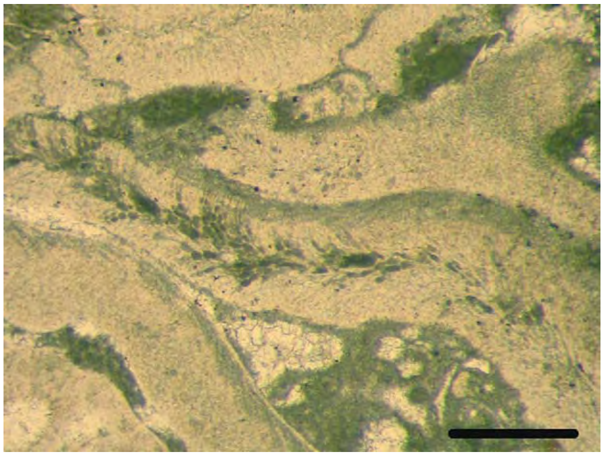


Figure 4 Section of *Polystrata*, a peysonneliacean alga. The thalli consist of cell filaments. The aragonitic nature of the original cell walls usually results in poor preservation of the internal structure. Paleocene, western Pyrenees, Spain. Scale bar 500 microns.

Peyssonneliaceae

Calcification in peyssonneliaceans (formerly squamaricaeans) is more variable than in corallines and is aragonitic, but otherwise peyssonneliaceans and corallines are broadly similar in consisting of branched cell filaments, forming a coherent thallus structure, and having intracellular calcification (Figure 4). Peyssonneliaceans occur as encrusting and leaf-like thalli up to 2 mm thick. Reproductive structures are rarely preserved in fossils. Aragonite may be densely or weakly deposited in the cell wall. Consequently, geological preservation of peyssonneliaceans ranges from good, with faithful preservation of the multicellular thallus, to very poor. The fossil record of peyssonneliaceans is consequently less well known than that of corallines, but they are common in the Palaeogene to present-day carbonate deposits (*see* **Sedimentary Environments:** Carbonate Shorelines and Shelves), and also build rhodolith nodules. External basal cement-like crusts commonly form, similar to those seen on crustose corallines. The confirmed range of peyssonneliaceans is Early Cretaceous to present-day, but they show similarities with some Carboniferous phylloids (see Phylloid Algae below) such as *Archaeolithophyllum* (see Corallinales above). Extant genera include *Polystrata* (= *Ethelia*) and *Peyssonnelia*. The depth and temperature ranges of present-day peyssonneliaceans are broadly similar to those of coralline algae.

Haptophyta

Coccolithophorales

Coccolithophores are a major group of nannoplankton (size 60 microns or less) and are the most important extant primary producers precipitating $CaCO_3$. They are mostly unicellular coccoid (non-motile) algae, although some species possess life stages of motile flagellate cells, non-motile colonies, or filaments. Most coccolithophores have one or more layers of organic (polysaccharide) scales surrounding the plasma membrane. Calcified scales, called coccoliths, form a cover termed the coccosphere (Figure 5) that is external to any additional organic scales that are present. The coccoliths are low-magnesium calcite in an organic matrix and range 1–25 microns in size. In different species, coccoliths differ widely in shape and arrangement, generating distinctive coccospheres composed of 10–100 coccoliths. Vegetative (fission) and sexual reproduction both occur. Life cycles are diverse among coccolithophoraleans but include a non-motile stage alternating with one or more motile phases. The non-motile stage can be benthic and colonial. Both motile and non-motile phases can bear coccoliths.

There are two basic types of coccolith: heterococcoliths composed of crystal elements of diverse size and shape (Figure 5), and holococcoliths made up of small identical crystals. Individual coccoliths range from simple disks, with various sculpture including

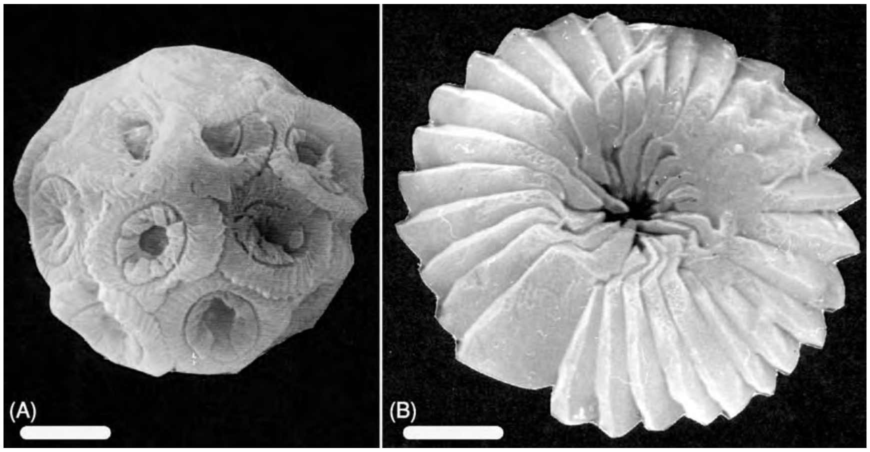


Figure 5 Coccolithophoralean coccosphere (A) and individual coccolith (B). Calcitic coccoliths are a major source of pelagic calcium carbonate sediment in modern oceans. Eocene, southern Spain. Scale bars 2 microns in (A) and 1 micron in (B). (Courtesy of B. El Mamoune.)

relatively long projections, to two plates joined by a central column. Many coccolithophores have different types of coccoliths on different parts of the cell surface. Others possess different coccoliths in different life stages.

Present-day coccolithophores are important phytoplankton, both in open oceans and in low-salinity seas such as the Black Sea. They extend from tropical to subpolar regions, although species assemblages differ according to latitude. Assemblages also show seasonal changes in relative abundance and species composition. Diversity is highest in warm oligotrophic waters and tends to decrease in coastal and restricted seas. Coccolithophores mostly live in the photic zone. This is deeper in low latitudes, particularly in the subtropics. Consequently, coccolithophore depth assemblages are well-developed in subtropical regions where shallow assemblages dominated by species bearing holococcoliths can be distinguished from deeper assemblages down to 220 m. These assemblages have differing temperature and nutrient, as well as light requirements. The high surface-to-volume ratio of individual coccospheres and coccoliths results in slow settling through the water column that allows dissolution of the calcite crystals. Faster settling is accomplished in zooplankton faecal pellets and other aggregates. Selective dissolution modifies the composition of sedimented coccolith assemblages and their diversity tends to reduce with increasing water depth.

Coccolithophores are first known from the Late Triassic. They gradually increased to a diversification maximum in the Late Cretaceous. Following major extinction at the Cretaceous/Tertiary boundary, the group regained high diversity during the Eocene but this has since decreased (Figure 1) (*see* **Palaeontology**). Several other types of nannofossils of uncertain affinities co-occur with coccoliths in pelagic oozes. Spherical tests made up of stacked calcite crystals, e.g., *Prinsiosphera*, *Schizosphaerella* and *Thoracosphaera*, are among the first nannofossils recorded in the Late Triassic. Nannoconids occur as tubes of calcite plates in the Late Jurassic to Cretaceous. Discoasterids have star or rosette shape and range from Paleocene to Late Pliocene.

Coccolithophores and related nannofossils have important applications in biostratigraphy and palaeo-oceanography. Nannofossil-based biostratigraphic zonations exist for the Jurassic to the present day. Most of these zonations consist of interval zones with boundaries characterized by first appearance or extinction datums of nannofossil species that are considered to be synchronous. Relative to other biostratigraphic markers in pelagic deposits, some nannofossils possess a cosmopolitan distribution and are regarded as relatively independent of water mass characters (i.e., temperature, nutrient content, salinity). Their first appearances and extinctions are therefore used for long-distance correlation. Nannofossil zones have good time resolution, in particular from the Late Cretaceous onwards, ranging from a few hundred thousand to a few million years. Correlation of calcareous nannoplankton zones to standard chronostratigraphic units is well established, facilitating widespread use of these fossils in biostratigraphy. Coccolith assemblages preserved in seafloor sediments in present-day oceans reflect the communities living in overlying near-surface waters. Since calcareous nannofossil assemblages vary with temperature over latitudinal gradients modified by currents and other oceanographic factors, they can be used as proxies for palaeo-oceanographic conditions in ancient pelagic sediments.

Chlorophyta (Green Algae)

Halimedales

This order (formerly included in the Codiaceae or Udoteaceae) comprises marine algae of siphonous organisation, i.e., the plant is a single multinucleate cell forming a branching tube or siphon (the parts of the tube resulting from branching are also called siphons). In *Halimeda* the thallus is a bundle of interwoven, subparallel siphons that branch into swollen vesicles (utricles) to form the outer surface. The thallus is composed of segments encrusted with aragonite needles that form in the spaces between siphons and vesicles (Figure 6). Thalli, typically a few decimetres in size, can reach 1 m or more in length. They are articulated by having uncalcified connections between calcified segments. *Halimeda* attaches to hard substrates or uses a large bulbous holdfast (aggregate of rhizoidal siphons) to anchor itself in mobile sand-gravel substrates. Plants can be erect or, especially when larger, sprawling with several points of attachment. Male and female gametes are produced in separate plants in branched stalks at the upper margins of segments. In sexual reproduction the cell contents of the entire plant are transformed into gametes (mass spawning), leaving the thallus empty and dead.

Halimeda occurs in tropical to warm temperate seas. In temperate waters such as the Mediterranean Sea the thallus is weakly calcified. In tropical carbonate environments *Halimeda* is common in coral reefs and lagoons from depths of less than 1 m to 150 m. Its fast growth produces copious quantities of gravel-size segments that are shed both during life and after death. *Halimeda* sediment production is particularly important in relatively deep water, at depths of 20–50 m. *Halimeda* mounds at this depth cover areas of hundreds of square kilometres on the outer Queensland shelf of the Great Barrier Reef. In the Florida Keys, *Halimeda* is the single most important component of carbonate sand and gravel, rivalling coral in abundance.

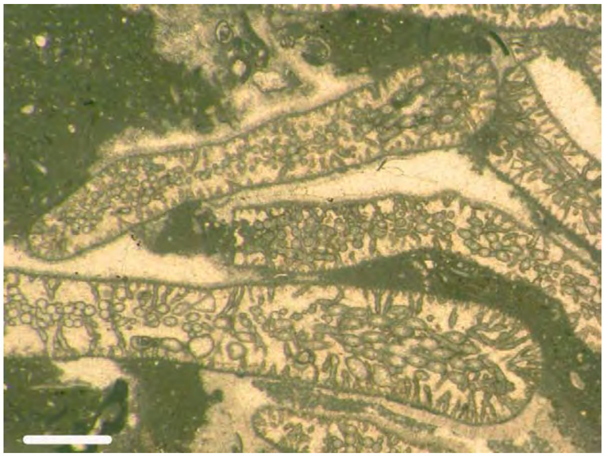


Figure 6 *Halimeda* segments. Aragonitic crystals (now calcite) form a cast of the complex, interwoven cell tubes of the alga, which remain as voids later filled with cement. Upper Miocene, south eastern Spain. Scale bar 1 mm.

Halimeda-like fossils are present in the Palaeozoic, e.g., the Ordovician *Dimorphosiphon* and the Devonian *Litanaia*. The group diversified in the Mesozoic with *Boueina* (Late Triassic–Late Cretaceous) and *Arabicodium* (Mid-Jurassic–Early Tertiary). *Halimeda* ranges from Early Cretaceous to present. Other extant halimedaleans, such as *Udotea*, *Penicillus*, *Rhipocephalus*, and *Tydemania*, are also important carbonate producers in tropical shallow water environments. *Udotea* has a fan-shaped thallus with a simple stalk. *Penicillus* and *Rhipocephalus* have brush-like thalli anchored by rhizoidal siphons. *Tydemania* is globular. Greater post-mortem skeletal disaggregation than in *Halimeda* limits the preservation potential of these genera as recognizable fossils.

Dasycladales

As in halimedaleans, the thallus is siphonous, consisting of a single multinucleate cell with a large central vacuole surrounded by cytoplasm. Plants range from a few millimetres to 20 cm in height, and are anchored by a rhizoidal holdfast. They have a single erect axis (dichotomously divided in a few species) with lateral branches. Gametes (zooids, no sexual differentiation occurs) are produced in fertile laterals. Dasycladalean taxonomy is complex. Six major families can be recognized, four of which (Seletonellaceae, Beresellaceae, Diploporaceae, Triploporellaceae) are extinct. Dasycladaceae and Polyphysaceae (formerly Acetabulariaceae) are extant.

Present-day Dasycladaceae (six genera) occupy shallow marine environments often less than 5 m in depth, in sheltered areas in tropical and subtropical waters. Spaces between the main axis and lateral branches are filled to varying degrees by aragonite. The resulting calcareous skeleton forms a cast of the plant that preserves the shapes of the main axis and laterals as voids (Figure 7). Degree and site of calcification vary with the species. This strongly influences its potential to preserve details of thallus anatomy. Calcification generally decreases with illumination and temperature. A few poorly calcified species grow in warm-temperate seas. *Batophora*, which does not calcify, can also be found in brackish and fresh water.

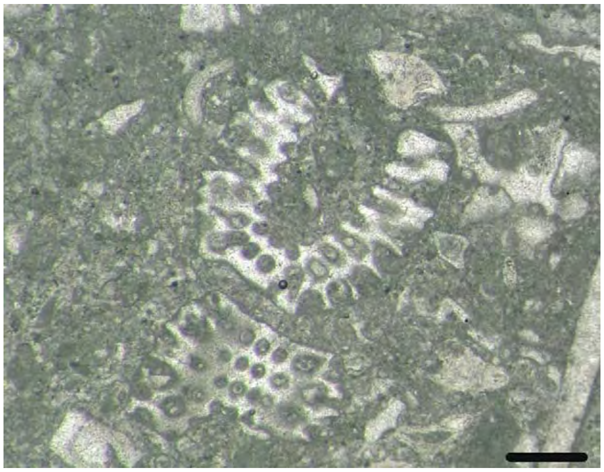


Figure 7 Dasycladalean sections. The aragonitic skeletons (now calcite) form a cast of the cell axis and lateral branches, which remain as matrix filled voids. The oblique section (upper) partly shows the size and shape of axis and laterals whilst the tangential section (lower) indicates the arrangement of the laterals. Paleocene, western Pyrenees, Spain. Scale bar 500 microns.

Present-day Polyphysaceae (five genera) also occupy shallow tropical to warm temperate seas. Some species tolerate high salinity fluctuations in shallow-water lagoons. Members of this family are characterised by the presence of a cap of gamete-producing laterals at the top of the plant as in *Acetabularia*, or alternating with sterile laterals in *Halycorne*. Calcification mostly takes place in the cap.

Possible dasycladaleans (e.g., *Yakutina*) are known from the Cambrian, but the earliest firm records are Ordovician. Dasycladaleans are common in platform carbonate deposits since the Carboniferous (*see* **Sedimentary Environments:** Carbonate Shorelines and Shelves). They reached high diversity in the Late Jurassic to Early Cretaceous and Eocene. The species richness of the group decreased after the Eocene, and extant species can almost be considered 'living fossils' (Figure 1). Dasycladaleans are used in biostratigraphy of shallow-water carbonates from the Late Palaeozoic to the Palaeogene. Zonations based upon dasycladalean genera and species have relatively poor time-resolution and their relation to standard chronostratigraphic scales is imprecise. However, together with benthic foraminifers, dasycladaleans are often the only fossils of biostratigraphic value found in extensive shallow platform carbonates.

Cyclocriniteae

Cyclocrinitids, or Cyclocriniteae, are a group of probable dasycladaleans (see Dasycladales above) mainly of Mid-Ordovician to Early Silurian age. They are characterized by an apically-inflated main axis giving rise to a cluster of radiating lateral branches. The tips of the laterals expand into hexagonal facets that unite to form an unbroken outer globular surface. The resulting swollen thallus is usually several centimetres in diameter. Common genera, including *Apidium*, *Coelosphaeridium*, and *Cyclocrinites*, are locally abundant. The affinities of Cyclocriniteae have been complicated because they have also been compared with, and sometimes placed in, receptaculitids – a problematic Palaeozoic group that shows some similarities with sponges. However, although cyclocrinitids were originally regarded as animals, their morphology is similar to that of the modern weakly calcified dasycladalean *Bornetella*.

Charophyceae

These chlorophytes are the closest algal relatives of embryophytes (mosses and vascular plants). Living charophyceans are grouped in the order Charales. Two additional extinct orders occur in the Palaeozoic. The group is characterized by complex bushy thalli, up to many decimetres in height, consisting of alternating nodes and internodes. The internodes are mostly made up of elongate cells. The nodes exhibit whorls of short branches ('leaves', and spine cells in some species) and branches of unlimited growth (Figure 8). Most charophyceans are anchored in soft sediment by rhizoids and have the appearance of subaqueous shrubs. The stem and branches can be encrusted by low-magnesium calcite and, more rarely, aragonite. Reproductive structures occur at nodes. Male gametes (spermatozoids) are produced in spherical structures called antheridia. Female gametes form in an ovoid oogonium that consists of an egg cell surrounded by sterile cells (Figure 8). In some species, following fertilization of the egg the surrounding cells become encrusted by calcite, while the remainder of the oogonium decays. The resulting calcified structures, gyrogonites, are the most common fossil remains of charophyceans (Figure 9).

Gyrogonites are made up of vertically elongate cells in primitive members of the Sycidiales, ranging from Late Silurian to Early Carboniferous; of dextral spiral cells in Trochiliscales, ranging latest Silurian-Permian, and of sinistral spiral cells in the extant order Charales, first recorded in the Devonian. The number of cells and the type of opening in gyrogonites are family-level taxonomic characters. Members of the Mesozoic family, Clavatoraceae possess gyrogonites with an external calcified cover (utricle) of various shapes. The Charales achieved greatest diversity in the Cretaceous to Palaeogene. It is a small, almost relict, group at the present day (Figure 1).

Charaleans grow in freshwater streams, ponds, and lakes and are especially abundant in calcareous

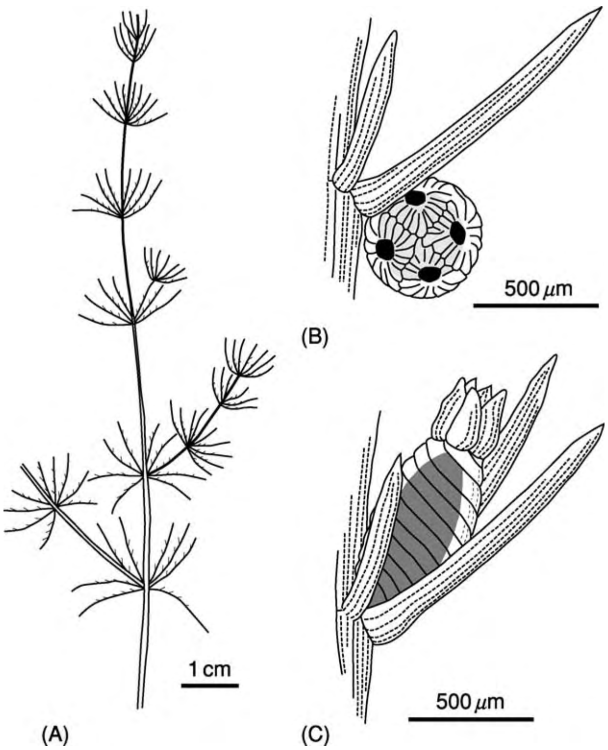


Figure 8 Charalean plant (A), showing the male reproductive structure (antheridium) at a node (B), and a female reproductive structure (oogonium) (C). The spiral cells in the oogonium are encrusted by calcite after fertilization. The resulting calcified structures, gyrogonites, are the most common fossil remains of Charales. (Adapted from Engler and Prantl (1898 1900).)

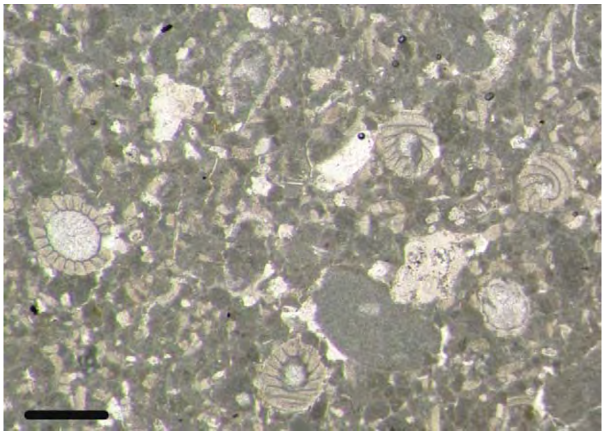


Figure 9 Sections of charalean gyrogonites. Lower Cretaceous, Transylvania, Romania. Scale bar 500 microns.

oligotrophic waters. Some species live in brackish lagoons. Fossil gyrogonites are used in biostratigraphy of non-marine deposits, especially from the Early Cretaceous to the Palaeogene. The biostratigraphic resolution of charalean zones is relatively low and correlation with marine chronostratigraphy is imprecise. Nonetheless, charalean fossils permit stage, or even sub-stage, level dating of deposits that often lack other biostratigraphic fossils.

Groups of Uncertain Affinity

Gymnocodiaceae

In vegetative construction, gymnocodiaceans are similar to halimedaleans (see Halimedales above), but rounded cavities within the cortex of some specimens resemble reproductive structures and suggest a red algal affinity. Gymnocodiaceans have been compared with the modern weakly calcified chaetangiacean *Galaxaura*. This enigmatic group, including *Gymnocodium, Permocalculus, Nanjinophycus*, and possibly *Succodium*, is most widespread and diverse in the Late Permian. *Permocalculus* also occurs in the Cretaceous and has been reported from the Paleocene and Miocene. Gymnocodiaceans appear to be shallow tropical marine in distribution, and locally are important bioclastic sediment producers.

Phylloid Algae

The term ‘phylloid algae’ describes fossils with flattened leaf-like shape that are especially common in Upper Carboniferous and Lower Permian reefs. Their recognition as a group is practical because the internal structure of these fossils is commonly poorly preserved, but it overlooks important morphological and systematic variations, and phylloids are a highly heterogeneous group that probably includes both green and red algae. Their internal structure is gradually being elucidated by discovery of well-preserved specimens and by use of cathodo-luminescence. Two morphological subgroups can be distinguished: (i) erect blades with medullary and cortical structure resembling that of halimedalean green algae, e.g., *Anchicodium, Eugonophyllum*, *Ivanovia* and, doubtfully, *Calcifolium* (see Halimedales); (ii) prostrate crusts with internal cells and possible conceptacles, e.g., *Archaeolithophyllum* that has differentiated multicellular filaments together with surficial conceptacle-like structures (Figure 10). This latter subgroup is compared with red algae, particularly peyssonneliaceans (see Peyssonneliaceae above). Phylloid algae formed skeletal frame reefs in the Late Carboniferous to Early Permian.

Calcified Cyanobacteria

Originally named blue-green algae, the prokaryotic cell organization of this group clearly allies it with bacteria. In natural environments, cyanobacteria extrude polymeric substances to form protective

Figure 10 Sections *Archaeolithophyllum*, a phylloid alga. Despite recrystallization, the multicellular structure of the crusts can be discerned. Upper Carboniferous, Cantabrian Mountains, northern Spain. Scale bar 500 microns.

sheaths. These can create diffusion limited sites in which photosynthetic uptake of HCO_3 generates pH gradients that can result in $CaCO_3$ precipitation. However, calcification only occurs if environmental conditions favour precipitation. Since cyanobacterial calcification is localised in the external mucilaginous sheath, the resulting microfossils are simple in form and include bush-like forms (e.g., *Angusticellularia*), tube-like filaments (e.g., *Girvanella*, *Hedstroemia*, *Cayeuxia*) (Figure 11), and chambered clusters (e.g., *Renalcis*).

Whereas some fossils are readily recognized as calcified cyanobacteria, others are more problematic. These include sedimentologically important Cambrian and Late Devonian fossils such as *Epiphyton* and *Renalcis* that have some resemblances to cyanobacteria but are not identical to modern examples, and are generally referred to as calcimicrobes. *Renalcis* and *Epiphyton*, together with the cyanobacterium *Angulocellularia*, were major reef builders in the Cambrian.

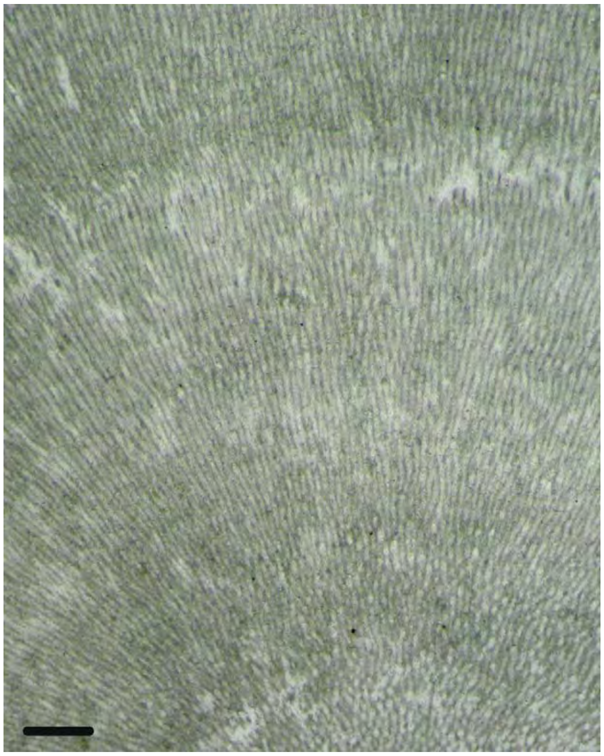


Figure 11 Section of *Cayeuxia*, a calcified cyanobacterium. The branching tubes are the result of calcification of the external mucilaginous sheath surrounding cell filaments. Middle Triassic, Betic Cordillera, southern Spain. Scale bar 500 microns.

Cyanobacteria have a geological record from the Archaean, but are not conspicuous as calcified fossils in marine environments until the Phanerozoic and even then their secular distribution is episodic. This could reflect variation in the saturation state of seawater over geological time. Calcified cyanobacteria assumed major importance in the Cambrian and Early Ordovician. They reappeared in abundance in the Late Devonian, and were common until the mid-Cretaceous, but are scarce or absent in marine environments during the Cenozoic (Figure 1). During episodes of abundance, calcified cyanobacteria were major reef components and in the Palaeozoic they often rivalled calcareous sponges such as archaeocyaths and stromatoporoids in importance. In present-day lakes and rivers, calcified cyanobacteria can form thick tufa bioherms and dams, together with oncoids (spherical stromatolites).

Small unicellular picoplanktic cyanobacteria, such as *Synechococcus*, form seasonal blooms in lakes along with diatoms and other planktic algae. Their photosynthesis can stimulate water column precipitation of small calcite crystals (whitings). Marine whitings in tropical seas may have a similar origin and potentially account for abundant lime mud production on ancient carbonate platforms.

See Also

Biosediments and Biofilms. **Palaeontology**. **Sedimentary Environments:** Carbonate Shorelines and Shelves; Reefs ('Build-Ups'). **Sedimentary Rocks:** Limestones.

Further Reading

Adey WH (1986) Coralline algae as indicators of sea level. In: Van de Plassche O (ed.) *Sea level research: a manual for the collection and evaluation of data,* pp. 229 279. Amsterdam, The Netherlands: Free University of Amsterdam.

Bassoullet JP, Bernier P, Deloffre R, Génot P, Poncet J, and Roux A (1983) Les algues udoteacées du Paléozoique au Cénozoïque. *Bulletin des Centres de Recherche, Exploration Productionn, Elf Aquitaine* 7: 449 621.

Berger S and Kaever MJ (1992) *Dasycladales. An Illustrated Monograph of a Fascinating Algal Order.* Stuttgart, Germany: Georg Thieme.

Braga JC, Bosence DWJ, and Steneck RS (1993) New anatomical characters in fossil coralline algae and their taxonomic implications. *Palaeontology* 36: 535 547.

Flügel E (ed.) (1977) *Fossil Algae. Recent Results and Developments*. Berlin, Germany: Springer.

Graham LE and Wilcox LW (2000) *Algae*. Upper Saddle River, NJ: Prentice Hall.

Jordan RW (2002) Environmental applications of calcareous nannofossils. In: Haslett SK (ed.) *Quaternary Environmental Micropalaeontology,* pp. 185 206. London, UK: Arnold.

Martín Closas C and Schudack M (1991) Phylogenetic analysis and systematization of post Paleozoic charophytes. *Revue de la Societé Botanique de France* 138: 53 71.

Riding R (ed.) (1991) *Calcareous Algae and Stromatolites*. Berlin, Germany: Springer.

Tappan H (1980) *The Paleobiology of Plant Protists*. San Francisco: Freeman. W.H.

Toomey DF and Nitecki MH (eds.) (1985) *Paleoalgology. Contemporary Research and Applications*. Berlin, Germany: Springer.

Van den Hoek C, Mann DG, and Jahns HM (1997) *Algae. An Introduction to Phycology.* Cambridge: Cambridge University Press.

Winter A and Siesser WG (eds.) (1994) *Coccolithophores*. Cambridge University Press: Cambridge.

Woelkerling WJ (1988) *The Coralline Red Algae: An Analysis of the Genera and Subfamilies of Nongeniculate Corallinaceae*. Oxford: Oxford University Press.

Fungi and Lichens

T N Taylor, University of Kansas, Lawrence, KS, USA
M Krings, Bayerische Staatssammlung für Paläontologie und Geologie, Geo-Bio Center, Munich, Germany

Introduction

Our understanding of the ecology of ancient ecosystems has progressed rapidly in recent years, so it is now possible to examine highly complex interactions involving previously understudied organisms, such as fungi and lichens. Based on the biological and ecological significance of fungi and lichens in many ecosystems today, a detailed knowledge of their fossil record, evolution, and the roles they played in biological and ecological processes in the past is important in understanding the evolutionary history of ancient and modern ecosystems. However, the fossil record of fungi and lichens remains incomplete. This article surveys the information gathered on fossil fungi and lichens. In some instances, preservation of the specimens is so extraordinary that it is possible to make direct comparisons with modern organisms and to examine interactions between fungi and other saprophytic, parasitic, and mutualistic organisms. Examples of enigmatic fossils, in which the fungus or lichen identity is still being debated, are also presented.

Fungi

Fungi are a unique group of eukaryotic organisms that are typically constructed of delicate filaments (hyphae), which may be loosely arranged or tightly packed to form complex structures. Fungi are heterotrophic, which means that they lack the ability to manufacture their own food and must rely on external sources to obtain nourishment. Many fungi cause serious diseases in plants, animals, and humans, whereas others are useful in their ability to produce antibiotics and various organic substances and to metabolize sugars into alcohol. Moreover, fungi can serve as potential agents for clearing toxic substances from the environment. Today, fungi and bacteria are the primary decomposers in the biosphere: they break down organic materials, such as the lignin in wood, and make the carbon, nitrogen, and phosphorous available to other organisms. Despite the fact that an estimated 1.5 million different types of fungi exist on Earth today, less than 100 000 species have been described and named to date.

Information about the geological history of fungi has been slow to accumulate because it has generally been assumed that the delicate filamentous bodies of most fungi (the mycelium) would not be adequately preserved in the fossil record. In addition, there is a gulf between palaeobotanists, who collect fossil fungi, and mycologists, who possess the experience to describe their structure and organization adequately. Finally, there is an inherent bias against collecting fossil fungi because the symptoms that they typically cause (e.g. rot, decay, disease) make the host specimens unattractive and often, therefore, not collected. Despite these obstacles, there is an ever-increasing awareness that fungi are adequately preserved in certain rocks and thus are a critical component in understanding the evolution of other organisms and the ecosystems in which they lived.

The extant fungi are classed into four major groups: chytridiomycetes, which are microscopic forms that occur in both terrestrial and aquatic habitats and produce motile cells; zygomycetes, which form thick-walled spores as a result of sexual reproduction and are constructed of coenocytic hyphae (included in this group are the mycorrhizae, which have formed intricate symbiotic associations in the roots of most terrestrial plants); ascomycetes, which have specialized cells (asci) that produce internally formed spores (ascospores); and basidiomycetes, which produce spores externally on a specialized cell (basidium). Although other structural, biochemical, and physiological characters are used to subdivide these major groups further, the recognition of fossil fungi is based on certain morphological features and, where sufficiently preserved, the symptoms they cause.

Fossil Fungi

Although evidence of life extends well back into the Proterozoic, the first unequivocal fungal remains come from the Palaeozoic. These consist of branched tubular hyphae with specialized cross-walls and multicelled spores recovered from Lower and Middle Silurian rocks. Based on similarities to structures seen in modern fungi, the Silurian remains have been suggested to have affinities with the ascomycetes. These early reports are based on fossils that were discovered after macerating rock fragments, and thus nothing is known about the role that the fungi played in the ecosystem in which they lived. Fungal remains in the form of spores have also been reported from the Ordovician; however, these spores look so modern that there is some question as to whether they may be contaminants of living fungi.

The Early Devonian Rhynie Chert ecosystem in Scotland has provided the most complete view of Palaeozoic fungi to date. Because this freshwater hot-springs ecosystem is nearly completely preserved *in situ* in a siliceous matrix, it contains a large number of different types of fungi, which were present at the time of fossilization; many of these fungi demonstrate a wide variety of interactions with other organisms. The most diverse group of fungi in the Rhynie Chert are the chytridiomycetes, which were probably the principal decomposers of organic matter in the Early Devonian. Like their modern equivalents, the chytridiomycetes from the Rhynie Chert include forms that are found on the surfaces of other organisms as well as inside cells. Because of their small size and great abundance, it has been possible to detail not only the morphology but also the life history of several forms. Other Rhynie Chert fungi (members of the Zygomycetes) formed symbiotic associations with a number of land plants as vesicular–arbuscular mycorrhizae (Figures 1E and 1F). In these associations, the fungus (*Glomites* sp.) has the ability to penetrate certain cells of the plant to form highly branched absorbing structures termed arbuscules (Figure 1F), which act as physiological exchange sites where the fungus obtains carbon. The host plant benefits from the greater ability of the fungus to obtain nutrients and water. Many believe that this fungus–land-plant relationship allowed plants to become established in the terrestrial realm, since it is estimated that today more than 90% of all land plants share a beneficial relationship with certain fungi.

A large number of extant fungi are parasites and pathogens of other organisms, and some of these associations are represented in the Rhynie Chert. Plant-parasitizing fungi in the Rhynie Chert ecosystem typically invaded the softer tissues of the plants, and it is possible to observe areas where, as a result of fungal infection, cells are broken down, sometimes resulting in easily observable necroses. Other parasitic fungi (predominantly chytridiomycetes) in the Rhynie Chert ecosystem colonized the spores of land plants (Figure 2B); as a result, the spores lost their capability to germinate and produce a gametophyte. Still other fungi were mycoparasites, i.e. they obtained nourishment from other fungi and/or fungal spores (Figures 1A and 1B). In some instances, when a plant-parasitizing fungus attempted to enter a cell, the plant responded by increasing the size of the cell, increasing the number of cells in the immediate region, and/or producing chemical deterrents in order to encapsulate or ward off the infection. Responses of this type are common in plants today and indicate not only that the Rhynie Chert plants were alive when the fungi attacked but also that the signalling mechanisms between these two organisms were well-established more than 400 Ma ago.

To date, three of the four major groups of fungi have been found to be represented in the Rhynie Chert; only the basidiomycetes are not present or have not yet been discovered. Especially noteworthy is the occurrence of an advanced saprophytic or parasitic perithecial ascomycete. Beneath the epidermis of upright stems and rhizomes of the land plant *Asteroxylon* (an early representative of the lycophytes) occur spherical to flask-shaped opaque structures (Figure 1C), which represent the reproductive structures (perithecia) of this fungus. The perithecia are characterized by short ostiolate necks that typically protrude from the epidermis of the host plant through stomatal openings. Lining the interior of the perithecium are elongate thin-walled hairs interspersed with

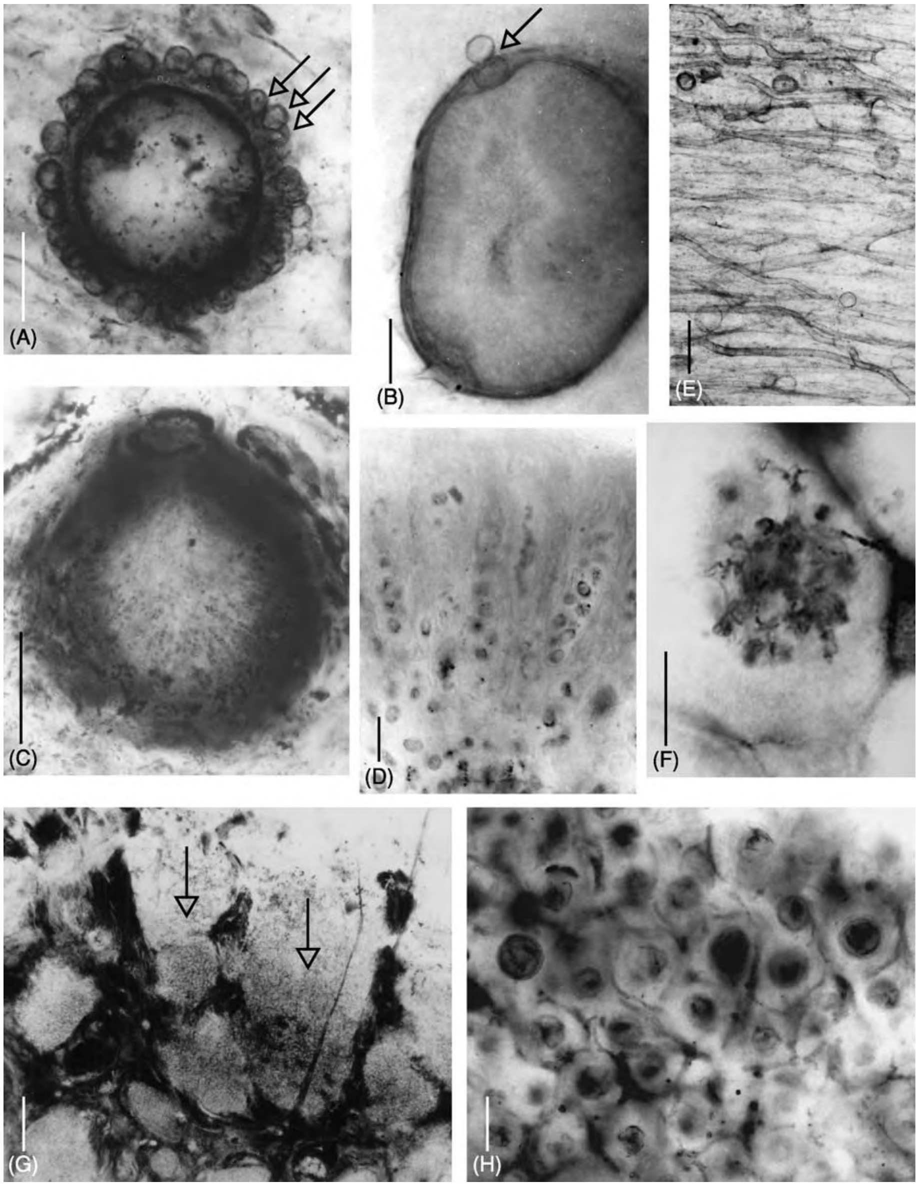


Figure 1 (A) Fungal spore from the Early Devonian Rhynie Chert, with its outer surface densely covered by parasitic chytridiomycetes (arrows); scale bar: 50 μm (From Taylor TN, Remy W, and Hass H (1992) Fungi from the Lower Devonian Rhynie Chert: Chytridiomycetes. *American Journal of Botany* 79: 1233 1241.) (B) Parasitic chytridiomycete (arrow) that has developed between the wall layers of a fungal spore from the Rhynie Chert; scale bar: 10 μm. (C) Longitudinal section through a perithecium of a Rhynie Chert ascomycete; scale bar: 100 μm. (D) Detail of Figure 1C, showing part of the perithecium with several asci containing ascospores; scale bar: 15 μm. (E, F) Vesicular arbuscular mycorrhiza in the Rhynie Chert land plant *Aglaophyton major*; (E and F adapted from Taylor TN, Remy W, Hass H, and Kerp H (1995) Fossil arbuscular mycorrhizae from the Early Devonian. *American Journal of Botany* 87: 560 573.) (E) hyphae with vesicles of the mycorrhiza fungus *Glomites* sp.; scale bar: 100 μm; and (F) *Glomites* arbuscule within a cortical cell from an aerial axis of *A. major*; scale bar: 10 μm. (G) Section of the thallus of the Rhynie Chert lichen *Winfrenatia*, showing depressions that contain cyanobacteria (arrows) surrounded by more a opaque zone that represents the mycobiont; scale bar: 50 μm. (H) Detail of Figure 1G, showing part of a single depression containing unicellular cyanobacteria and fungal hyphae that form a net like structure; scale bar: 25 μm.

sacs (asci) that contain unicellular to up to five times septate sexual spores (ascospores) (Figure 1D).

Plants with abundant woody tissues first appear in the Middle to Late Devonian, and it is at this time that the first wood-decomposing fungi – basidiomycetes – appear. The earliest fossil evidence for the possible existence of basidiomycetes occurs in the form of branching hyphae of varying diameters

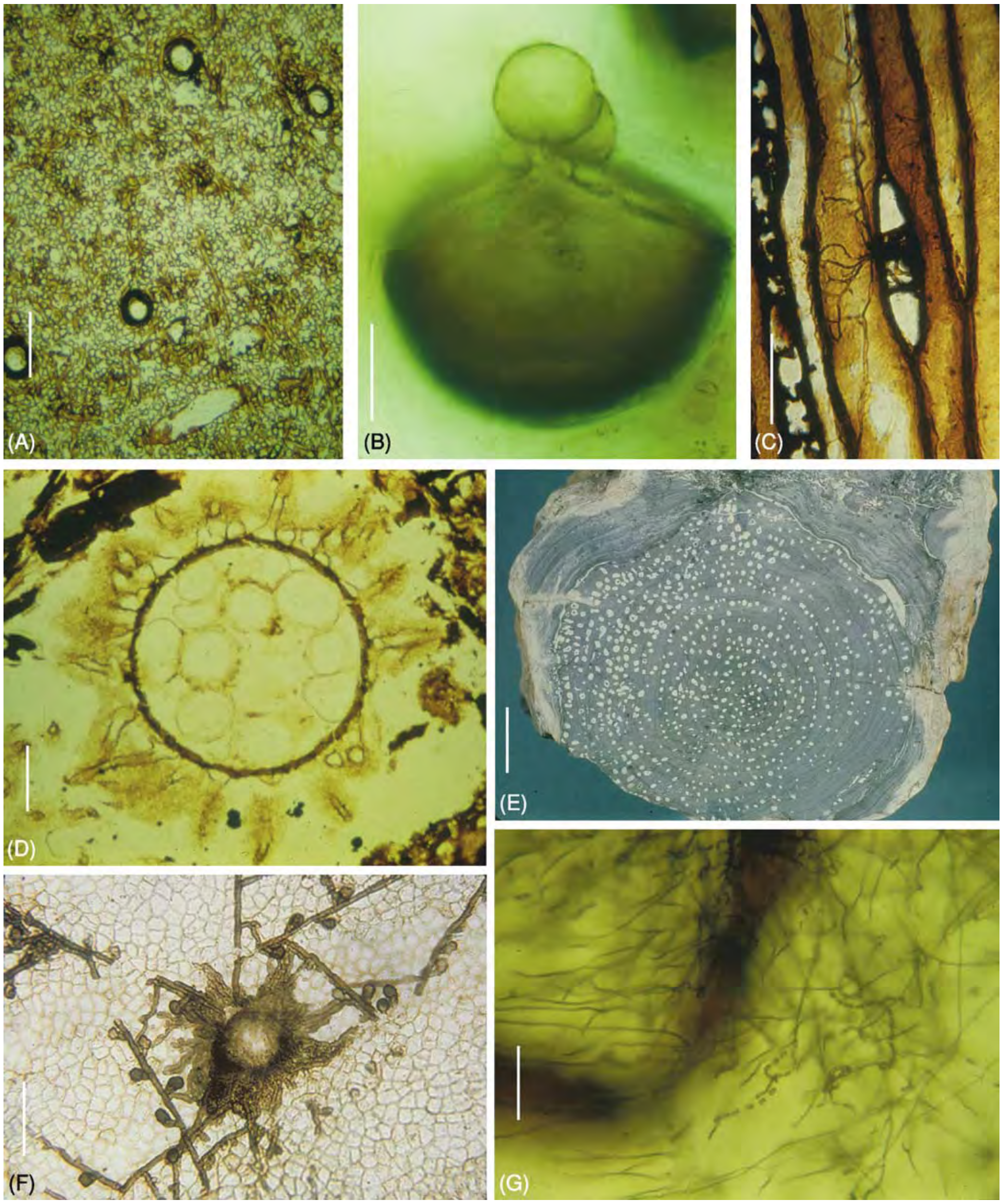


Figure 2 (A) *Prototaxites*; transverse section showing the internal anatomy, which is entirely constructed of interwoven septate tubules (?hyphae); scale bar: 100 μm. (B) Parasitic Chytridiomycetes on a spore of the Rhynie Chert plant *Aglaophyton major*; scale bar: 10 μm. (From Taylor TN, Remy W, and Hass H (1992) Fungi from the Lower Devonian Rhynie Chert: Chytridiomycetes. *American Journal of Botany* 79: 1233 1241.) (C) Radial section of *Callixylon newberryi* (progymnospermous wood) from the Upper Devonian of Indiana, showing branching fungal hyphae in one of the tracheids; scale bar: 35 μm. (D) *Traquairia williamsonii*, an ornamented sporocarp from the Upper Carboniferous of Kentucky, containing several thin walled spores; scale bar: 80 μm. (Adapted from Stubblefield S and Taylor TN (1983) Studies of Paleozoic fungi.1. The structure and organization of *Traquairia* (Astomycota). *American Journal of Botany:* 70: 387 399.) (E) Cross section through an *Araucarioxylon* (gymnospermous wood) stem from the Triassic of Antarctica, showing numerous pockets of decay (lighter areas); scale bar: 1.2 cm. (From Stubblefield S and Taylor TN (1986) Wood decay in silicified gymnosperms from Antarctica. *Botanical Gazette* 147: 116 125.) (F) Leaf colonizing fungus on an angiosperm leaf from the Eocene (Tertiary) of the USA; scale bar: 60 μm. (G) *Geotrichites glaesarius*, a saprophytic fungus from the surface of a spider that is preserved in Tertiary amber from the Dominican Republic; scale bar: 60 μm.

(e.g. Figure 2C), some possessing rounded knobs (so-called clamp-connections), and spores; indirect evidence of the presence of these fungi includes solution troughs along the inner surfaces of water-conducting cell walls. Because modern basidiomycetes are the primary degrading agents of lignin and cellulose, dissolution patterns in fossil wood that resemble those seen in modern wood indicate the existence of wood-degrading fungi, even though there is no direct evidence of the fungus. In Mesozoic fossils from Antarctica and other southern-hemisphere continents, there is another type of pathology caused by fungi in plants. When certain fossil logs are examined in cross-section, the wood contains numerous circular areas that are devoid of cells (Figure 2E). These areas, usually 1–2 mm in diameter, represent spindle-shaped cavities that were formed by the enzymatic activities of a fungus. When examined in detail, the cells at the margins of these areas show evidence of selective dissolution of the cell-wall components. Patterns of this type, termed white rot and white pocket rot, are common in woody plants today, where they are caused by several basidiomycetous fungi. What is interesting is that the trees attacked during the Mesozoic are all now extinct, but the fungus has survived and produces exactly the same symptoms today in different hosts. Another example of fungal attack in woody plants comes from the Triassic Petrified Forest in Arizona. Sections of wood show regions of highly disrupted cells. Although evidence of the fungus responsible for this degradation is generally lacking, the large number of trees with the same symptom has led to the suggestion that perhaps much of this forest was attacked by a fungus, which caused a level of devastation comparable to that seen in modern trees infected with Dutch elm disease. Pathogenic fungi are also known from several types of seed, in which the fungus has greatly altered the internal tissues. For example, the disruption of nutritive cells in the seed and the formation of fungal reproductive structures in their place means that the seed was no longer capable of producing a new plant.

Fungi can also be found on the leaf surfaces of many different fossil plants (e.g. Figure 2F). Excellent examples are common on Eocene flowering plants, where the fungus produces specialized structures that penetrate the leaf surface to obtain nutrients. Based on the types of plants and the fungi on the leaves, these fossil fungus–plant interactions provide information about fungal host preferences and can perhaps also be used as a proxy record for some climate variables.

Fungi, like all organisms, must be capable of producing large numbers of spores and other propagules that can be disseminated and serve as reinfecting agents. One interesting fossil example comes from the Eocene Princeton Chert of British Columbia, in which a fungus invaded the normal pollen-producing structures of a flowering plant. This fungus, which is most like a modern smut fungus, replaced the pollen with its own spores and thus was able to colonize new host plants effectively because the spores were distributed by pollinating insects that visited the flowers in search of nectar.

Today several major groups of fungi have successfully exploited animals, including humans, as hosts. Modern fungi are found in the hind guts of various arthropods, and at least one record suggests that this association extends back to the Triassic. In this association, the fungus is able to exploit new niches through the mobility of the host animal; it is not fully understood, however, whether there was a benefit for the host. Some animals use fungi as a food source (fungivory). Aggregations of fungal hyphae in the form of coprolites date from the Middle Silurian onwards and indicate that fungi were a principal component of the diet of some early animals. Coprolites are a potential new source of information about fungus–animal associations in the fossil record. Another excellent preservation matrix for fossil fungi is plant resin (amber), which dates back to the Carboniferous. Fungus–animal associations are especially well documented from Tertiary amber and include fungi on a partially decomposed spider from the Upper Oligocene–Lower Miocene (Figure 2G) and pathogenic fungi on a termite, an ant, and within the body cavity of a nematode.

Fungi with Uncertain Affinities

Throughout the Carboniferous a variety of conceptacle-like structures are found that have been interpreted as sporocarps. They occur singly or in clusters and are common in the permineralized peat that forms Carboniferous coal balls. The walls of these circular (up to 1 mm in diameter) structures are composed of interlaced hyphae, and the outer surface may be smooth or ornamented by processes (Figure 2D). Some contain thin-walled spores, which in turn contain additional spores. These sporocarps have been assigned to the Ascomycetes by some, whereas an alternative hypothesis places them within the Zygomycetes. Another interesting fungus from the Carboniferous with problematic affinities combines features of the Ascomycetes with those of perhaps a basidiomycete. While this interesting fungus may be a form that shares features of two groups, another idea is that one fungus has parasitized the other. Perhaps the most unusual life form considered by many scholars to be a fungus is the Nematophyte *Prototaxites*, which existed until the Late Devonian. This enigmatic trunk-like organism grew up to 1 m in diameter and its internal anatomy was entirely constructed of interwoven septate tubules (?hyphae) of three different sizes (Figure 2A); the arrangement of tissues suggests a periodicity in growth. The reproductive parts of *Prototaxites* remain equivocal; however, there is some suggestion that its affinities lie with the Basidiomycetes. Whether *Prototaxites* was a saprophytic fungus, as has been hypothesized, or some other type of life form with an unusual combination of features, perhaps not even remotely related to any modern organism, it must have presented an imposing structure, extending more than 8 m above the very small plants that made up the Early Devonian landscape.

Lichens

Lichens may be interpreted as 'dual organisms' since they do not represent a single organism, in the way that most other life forms on Earth do, but rather a combination of two or three entirely different organisms: a fungus (mycobiont) that lives in an obligate mutualistic association with a green alga or cyanobacterium or both (photobiont). The lichen symbiosis is unique since it represents a new 'hybrid' life form with a distinct body plan (lichen thallus) and structural and physiological properties that are different from those of either of the partners. An estimated 13 000–17 000 species of lichen exist today, extending from the tropics to the polar regions and growing on a highly diverse array of substrates, including soil, bark, bare rock surfaces, leaves of vascular plants, barnacle shells, and other lichens. The vast majority of lichen-forming (lichenized) fungi are members of the Ascomycetes, but some 20 species of basidiomycete and one zygomycete are also known to enter into lichen symbioses. In contrast to the many thousands of lichenized fungi, there are only about 100 different photobionts, most of which are unicellular green algae (Chlorophyta).

It is still not fully understood how this unique assemblage of organisms evolved. The failure to resolve more fully the evolutionary history of lichens is due primarily to a meagre fossil record. Moreover, in order to establish the existence of a fossil lichen, it is necessary to demonstrate not only the individual partners but also that there is some degree of interdependence between the separate organisms. Thus, even if an entire association is well preserved, it may be difficult to determine whether it is a lichen or simply an asymptomatic aggregate of algae or cyanobacteria and fungi that are closely associated but do not interact. Alternatively, the aggregation may represent a fungus that has conveniently parasitized an alga or cyanobacterium. Molecular and genetic studies of extant lichens suggest that the lichen symbiosis has arisen several times over the course of geological time.

Fossil Lichens

It has been hypothesized that one of the first steps in the colonization of land by eukaryotes may have been the formation of a lichen symbiosis. However, compelling fossil evidence supporting this hypothesis remains elusive. The earliest fossil (from the Early Proterozoic Witwatersrand Group, South Africa) that has been considered to be a lichen is a thallus-like structure comprising erect columns made up of branched and apparently septate filaments. The lichen nature of this fossil is, however, questionable since remains of the photobiont are not observed. Moreover, the filaments are much smaller than most fungal hyphae and are in the size range of certain filamentous bacteria. Some believe that the filaments are not fossils because similar objects can be produced artificially in the laboratory. Other Precambrian life forms, which have been suggested to be lichens, are the so-called 'Ediacara fossils' or 'Vendobionta' (*see* **Precambrian:** Vendian and Ediacaran). These unusual 600 Ma old organisms, which presumably lived on soft mud or sands at the bottom of shallow coastal seas, are usually considered to be primitive animals, but one hypothesis considers them to be lichens or lichen-like based on the fact that they were relatively resistant to being compressed after burial. This idea has not received much support from either the palaeontological community or the lichenological community.

Several groups of Palaeozoic fossils have variously been interpreted as lichens or lichen-like associations. One group is the Nematophytes, which occurred from the Late Ordovician to the Late Devonian. The most impressive Nematophyte, *Prototaxites* (Figure 2A), is usually considered to be a giant fungus (see above), but some workers have speculated that it may have been a lichen-like association of two or more different kinds of organism. A second member of the Nematophytes, *Nematothallus*, was a small foliose organism that consisted of intermixed large and small tubules, covered by a thick cuticle-like layer. *Nematothallus* possesses a number of striking features that are also found in certain extant lichens. The cuticle-like surface layer, for example, displays impressions of a distinct cell pattern, which is virtually indistinguishable from that seen on the thallus surface in certain extant cyanolichens. Unfortunately, as is the case with other early lichen-like fossils, the affinities of *Nematothallus* cannot be established with certainty because a photobiont partner has not yet been discovered. Another example of a Palaeozoic life form, the identity of which is still being debated, is *Spongiophyton* (Middle Devonian). This thalloid organism is typically preserved as a flattened cuticle-like sheath with circular pores on one surface. Some have suggested that *Spongiophyton* was an early bryophyte (perhaps a primitive liverwort), whereas others speculate that it may have been a lichen.

The most convincing fossil example of a Palaeozoic lichen comes from the Early Devonian Rhynie Chert.

Winfrenatia (Figures 1G and 1H) consists of a mycelial mat, constructed of interwoven aseptate hyphae. Along the upper surface of the mat are numerous shallow relatively uniform depressions (Figure 1G). Within many of these depressions are coccoid unicells that are morphologically similar to certain extant cyanobacteria; also present are clusters of the same cells within mucilaginous investments that are interpreted to be stages in the life history of the cyanobacterium. Hyphae of the fungus extend into the depressions and become intertwined with the cyanobacteria (Figure 1H). The range of features observed in *Winfrenatia*, including the size of the thallus and the number of depressions on the surface, have allowed speculations as to the life-history strategy, which included the production of new cells of the photobiont to maintain the symbiosis and support the fungus and, at the same time, an increase in the size of the fungal mycelial mat. It has been suggested that *Winfrenatia* is not a true lichen but rather an unstable association in which a fungus parasitized a cyanobacterial colony. While the physiological stability of a symbiotic association in the fossil record may never be fully deciphered, the definition of a lichen as a controlled parasitism is perhaps more accurate, as it underscores the varying degrees of lichenization that appear in modern ecosystems. In that context, *Winfrenatia* may quite accurately be regarded as an Early Devonian lichen. The systematic position of the fungal partner of *Winfrenatia* has not been conclusively identified. The presence of aseptate hyphae and certain thick-walled sculptured spores associated with the mycelial mats suggest that the affinities of the fungus may lie close to the Zygomycetes. This is especially noteworthy since there is only one extant lichen with a zygomycetous mycobiont and a cyanobacterial photobiont, *Geosiphon pyriforme*. In this unique endosymbiotic association, cyanobacteria of the genus *Nostoc* become encapsulated in pear-shaped bladders formed by the fungus.

A few fossil lichens have been reported from the Cenozoic, most of which are preserved in Tertiary amber. That these fossils represent lichens is relatively easy to demonstrate, since their morphology closely resembles that of extant lichens; some forms can even be assigned to modern families and genera. In addition to these lichens, there are a few reports of adpressed lichen thalli, including a fossil *Lobaria* from Miocene rocks at Redding Creek (California) and the foliicolous (leaf-colonizing) lichen *Pelicothallos*, which was discovered on an angiosperm leaf from the Eocene of Tennessee.

It remains puzzling why lichens are so rare in the fossil record, or have been so infrequently described. One reason may be that ancient lichens were quite different in appearance from those seen today, and, as a result, it is difficult to recognize these organisms in the fossil record. Moreover, some workers have pointed out that many lichen-dominated habitats, such as tundra and mountains, have a relatively low potential for yielding fossils. However, lichens today also thrive in many environments that could lead to preservation, including swamp forests. Still, even in the extensive Carboniferous coal swamp forests, which have been studied extensively for more than 150 years, not a single compelling lichen fossil has ever been recorded.

Concluding Remarks

As is the case in many areas of palaeontology, the discovery and recognition of ‘unique’ fossils requires an intersection between the vagaries of preservation and the necessary expertise to interpret the fossil. Historically, relatively little attention has been directed at fossil bacteria, algae, fungi, and other miscellaneous organisms in which taxonomic placement has been difficult. As palaeontologists strive to understand the complexities of the biological and physical world of the past, it is becoming increasingly clear that organisms such as fungi and lichens were not only present but played pivotal roles in shaping the ecosystems of the past.

Glossary

Clamp-connections Outgrowths of fungal hyphae that, at cell division, make a connection between the resulting two cells by fusion with the lower

Coal balls Preservation type in which mineral matter has filled the cell lumens and intercellular spaces but has not replaced the cell walls.

Coprolites Fossilized faecal material.

Hypha (plural hyphae) Tubular filament that is the structural component of the majority of fungi and may be coenocytic or septate.

Mycelium (plural mycelia) The total mass of hyphae that constitutes the vegetative portion or thallus of a fungus.

Mycoparasite A fungus that is parasitic on another fungus.

Mycorrhiza Mutualistic association of a fungus with (the roots of) a higher plant.

Necrosis Localized death of living cells due to infection or injury.

Parasite An organism that obtains nourishment from another living organism.

Pathogen Any organism capable of causing a disease.

Saprophyte An organism that obtains nutrients from dead organic material.

Sporocarp A closed body or conceptacle containing one or more masses of spores or sporangia.

Symbiosis A form of interaction where one organism lives in an intimate association with another.

Thallus The vegetative body of a plant or lichen that is not differentiated into organs such as leaves and stems.

See Also

Biosediments and Biofilms. **Carbon Cycle**. **Fossil Plants:** Calcareous Algae. **Palaeozoic:** Devonian. **Precambrian:** Vendian and Ediacaran. **Sedimentary Rocks:** Chert. **Tectonics:** Hydrothermal Activity. **Tertiary To Present:** Eocene.

Further Reading

Hallbauer DK and van Warmelo KT (1974) Fossilized plants in thucholite from Precambrian rocks of Witwatersrand, South Africa. *Precambrian Research* 1: 199 212.

Hass H, Taylor TN, and Remy W (1994) Fungi from the Lower Devonian Rhynie Chert: mycoparasitism. *American Journal of Botany* 81: 29 37.

Heckman DS, Geiser DM, Eidell BR, *et al.* (2001) Molecular evidence for the early colonization of land by fungi and plants. *Science* 293: 1129 1133.

Hueber FM (2001) Rotted wood alga fungus: the history and life of *Prototaxites* Dawson 1859. *Review of Palaeobotany and Palynology* 116: 123 158.

Poinar GO (1992) *Life in Amber*. Stanford, CA: Stanford University Press.

Purvis W (2000) *Lichens*. Washington, DC and London: Smithsonian Institution Press and The Natural History Museum.

Stubblefield S and Taylor TN (1986) Wood decay in silicified gymnosperms from Antarctica. *Botanical Gazette* 147: 116 125.

Stubblefield S and Taylor TN (1988) Recent advances in palaeomycology. *New Phytologist* 108: 3 25.

Taylor TN (1993) Fungi. In: Benton MJ (ed.) *The Fossil Record,* vol. 2, pp. 9 13. London: Chapman & Hall.

Taylor TN and Taylor EL (1993) *The Biology and Evolution of Fossil Plants*. Englewood Cliffs, NJ: Prentice Hall.

Taylor TN and Taylor EL (2000) The Rhynie Chert ecosystem: a model for understanding fungal interactions. In: Bacon CW and White JF (eds.) *Microbial Endophytes,* pp. 31 47. New York: Marcel Dekker, Inc.

Taylor TN, Remy W, and Hass H (1992) Fungi from the Lower Devonian Rhynie Chert: Chytridiomycetes. *American Journal of Botany* 79: 1233 1241.

Taylor TN, Hass H, and Kerp H (1997) A cyanolichen from the Lower Devonian Rhynie Chert. *American Journal of Botany* 84: 992 1004.

Tiffney BH and Barghoorn ES (1974) The fossil record of the fungi. *Occasional Papers of the Farlow Herbarium of Cryptogamic Botany* 7: 1 42.

Waggoner BM (1995) Ediacaran lichens: a critique. *Paleobiology* 21: 393 397.

Gymnosperms

P Kenrick, The Natural History Museum, London, UK

Introduction

Gymnosperms are an ancient group of plants that include some familiar living forms such as conifers, cycads, and ginkgos as well as a great variety of far less well-known and long-extinct kinds. The modest number of living species (approx. 750) belies the geological importance of the group. Gymnosperms dominated the Mesozoic era. It was during this time that many of the major subgroups originated, diversified, and finally succumbed to their ultimate fate, extinction. The evolution of gymnosperms and in particular their relationships with the angiosperms (flowering plants) is also the subject of much debate. Phylogenetic studies clearly show that these two groups are close relatives, but the exact nature of their kinship is still unclear. Knowing how these two groups are related is essential to developing a family tree of gymnosperms and to understanding the origin of the flowering plants. Because much of the diversity of gymnosperms is extinct, the fossil record continues to play a crucial role in elucidating evolutionary patterns. Recently, molecular systematics has introduced an enormous amount of new comparative data. This has brought unprecedented precision in our ability to reconstruct family trees, shedding new light on the evolution of gymnosperms and their lengthy and chequered fossil history.

Characteristics

Gymnosperms are a varied assemblage of plants whose principal botanical characteristic is reproduction by means of ovules, or seeds as they are called in their fertilized state. This is an aspect of their biology that they share with the angiosperms. One of the main differences between gymnosperms and angiosperms relates to the appendages associated with the ovules. In the angiosperms, ovules are part of the flower. One or more ovules is wrapped within a layer of tissue, the whole structure being called the carpel. Flowers usually comprise carpels and additional organs such as pollen-bearing sacs, petals, and sepals. Gymnosperms do not have flowers. The ovules are borne in cones or singly on more or less modified branches. Furthermore, the ovules are not enclosed within a carpel, they are 'naked', hence the origin of the term gymnosperm (Greek 'gumnospermos': 'gumnos', naked + 'sperma', seed). Modern gymnosperms are generally also woody shrubs and trees, but some have a vine-like growth form. A more diverse array of growth forms, including groups with fern-like leaves, is known from the fossil record. Ovules and seeds, wood, leaves, and pollen are frequently preserved as fossils, and together these provide our principal sources of evidence on the geological history of the group.

Classification

Recent classifications of gymnosperms are based upon phylogenetic studies. Because many aspects of gymnosperm phylogeny still remain unclear, modern classifications at the family or ordinal level must be regarded as very tentative. There are problems in placing some extinct groups because they are still very poorly characterized. In other instances, there are conflicts between the phylogenetic trees generated from comparative morphology and those produced from molecular systematic studies. Despite these problems there is reason to be optimistic in the longer term. The history of gymnosperm classification shows that better characterization of problematic fossils tends to lead to more stable ideas on relationships, and in the molecular world there is a seemingly inexhaustible supply of new data that should lead ultimately to a firm backbone phylogeny for the living groups. Both approaches should converge on a mutually consistent and stable family tree.

Table 1 Classification of seed plants (Spermatophyta)

Spermatophyta				
	'Calamopityales' *incertae sedis*[a, b]			
	Hyrdaspermales[a, b]			
	'Lyginopteridales'[a, b]			
	Medullosales[a, b]			
	Euspermatales (taxa *sedis mutabilis*)			
		Cycadales		
		Callistophytales[a, b]		
		Coniferophytales		
			Cordaitiales[b]	
			Coniferales (Pinales)	'Gymnosperms'
		Glossopteridales[a, b]		
		Czekanowskiales[b]		
		Ginkgoales		
		'Peltaspermales'[b]		
		'Corystospermales' ('Umkomastales')[b]		
		Caytoniales[b]		
		Pentoxylales[b]		
		Bennettitales[b]		
		Gnetales		
		Archaefructaceae[b]		? Angiosperm
		Angiospermopsida		Angiosperms

Grouping is indicated through the use of indentation and the sequence a name occupies in the list. Commonly applied names for individual groups are used without regard to Linnean rank order convention. Questionably monophyletic groups are marked with quotes. *Sedis mutabilis* indicates that groups which follow at the next level of the hierarchy are of uncertain relationship (i.e., form a polytomy or conflict strongly in different phylogenetic analyses). *Incertae sedis* indicates that the taxon bearing this epithet is of uncertain relationship to those at the same level. Gymnosperms is a paraphyletic group. It is equivalent to seed plants minus flowering plants (i.e., Spermatophyta minus Angiospermopsida). Modified after Crane and Kenrick (1997) Problems in cladistic classification: Higher level relationships in land plants. *Aliso* 15: 87 104.

[a] Groups typically included as Pteridosperms, or seed ferns, which are a heterogenous assemblage of gymnosperms with fern like foliage.

[b] Extinct group.

One of the findings of recent phylogenetic studies is that some groups of gymnosperms are more closely related to angiosperms than they are to other gymnosperms (Table 1). In technical terms, gymnosperms are paraphyletic to angiosperms. The precise nature of this relationship, however, remains unclear, and this is one major source of conflict between family trees based on comparative morphology and those derived from molecular data. Comparative morphology indicates that angiosperms are most closely related to living gymnosperms in the Gnetales and extinct gymnosperms in the Pentoxylales and the Bennettitales. This has been dubbed the 'anthophyte hypothesis' (Figure 1). Molecular data tell a very different story. They place Gnetales within Coniferales. Angiosperms emerge as sister group to a monophyletic group comprising all living gymnosperms. This has been dubbed the 'gne-pine' hypothesis (Figure 1). Neither hypothesis though is particularly strongly supported by the current data. Furthermore, because molecular data cannot deal with the fossil groups, it is of course unclear where in molecular

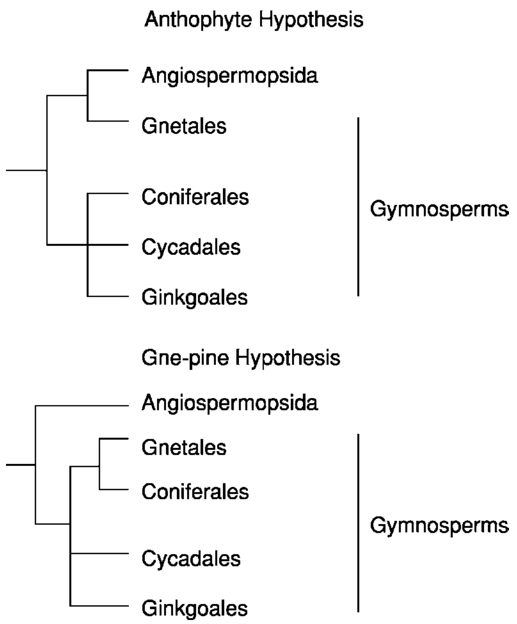


Figure 1 Two hypotheses of how living gymnosperms are related to angiosperms (flowering plants). Morphological data favour the anthophyte hypothesis, whereas molecular data favour the gne-pine hypothesis.

Figure 2 Frond of the progymnosperm *Archaeopteris hibernica* (Upper Devonian, Kiltorkan, Kilkenny, Ireland).

phylogenetic trees the fossils would in fact fall out. However, it seems likely that whichever hypothesis ultimately wins out, Early Palaeozoic gymnosperms such as Calamopityales, Hydraspermales, Lyginopteridales, and possibly Medullosales (Table 1) will remain basal in a phylogenetic sense to both angiosperms and to the more derived living gymnosperms. Also, it is plausible that extinct Mesozoic groups such as Pentoxylales, Bennettitales, and possible Caytoniales are more closely related to angiosperms than they are to other gymnosperms. From a phylogenetic perspective, angiosperms are in reality little more than a morphologically divergent group of gymnosperms.

Gymnosperm Origins

Gymnosperms originated from a grade of Late Devonian plants called the progymnosperms. The best-known exemplar is *Archaeopteris*. This plant had large fern-like fronds, some of which are known to have been well over 1 m in length (Figure 2). Like ferns, *Archaeopteris* produced spores rather than seeds. However, unlike true ferns, the fronds were borne on woody branches. In this respect, *Archaeopteris* resembled gymnosperms. Some species of *Archaeopteris* were undoubtedly large trees, as trunks of the distinctive wood (*Callixylon*) are known to exceed several metres in diameter. This extraordinary extinct plant therefore possessed a unique combination of characteristics that are intermediate between gymnosperms and early relatives of the ferns. It shows that the woody architecture that characterizes gymnosperms evolved before other defining features, such as the ovule.

The earliest gymnosperms are found in Upper Devonian sediments of Europe and North America. These are known from plant fragments such as dispersed seeds and seed-bearing branches (Figure 3) as well as a handful of more completely known plants. *Elkinsia polymorpha* (Hydraspermales) from the Upper Devonian of West Virginia, USA, is one of the more completely characterized forms. Petrified ovules in which cellular anatomy is preserved show gymnosperm features alongside other features that differ from modern forms. Specifically, many early ovules lacked a completely formed integument (enveloping tissue layer), so they are sometimes called ‘preovules’. Some also had additional specialization to aid fertilization (pollen capture). In many, ovules were borne on the ends of branches in ‘cupules’. The earliest gymnosperms were small woody shrubs that would have had a fern-like appearance (Figure 4).

Figure 3 Ovule (seed) bearing cupules of the early gymnosperm *Xenotheca devonica* (Upper Devonian, Devon, England).

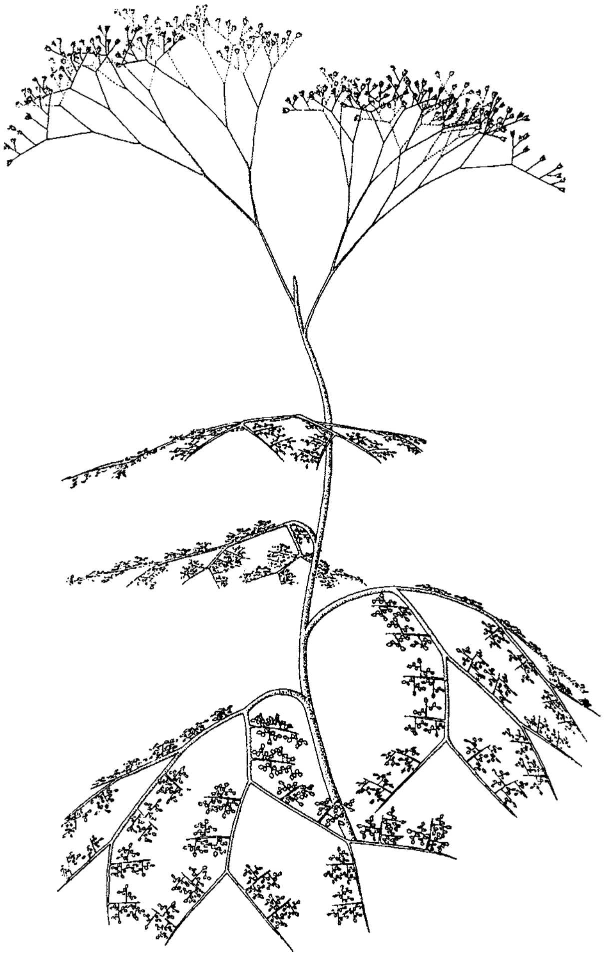


Figure 4 Reconstruction of the early gymnosperm *Elkinsia polymorpha* (Upper Devonian, USA). After Serbet and Rothwell (1992). *International Journal of Plant Sciences* 153: 602–621.

This fern-like theme in gymnosperm evolution continued into the Carboniferous period. Much of the fern-like foliage of coal swamp forests is actually attributable to extinct gymnosperms. These are the so-called ‘seed ferns’ or Pteridosperms (Figure 5). As the name implies, these were seed-bearing plants that possessed fern-like leaves, which is a combination of characteristics seen only in extinct gymnosperms. Groups common but not exclusive to the Carboniferous period include Calamopityales, Hydraspermales, Medullosales, Lyginopteridales, and Callistophytales. For the most part, these were understory shrubs and small trees. Some are also known to have been vine-like climbers. The earliest gymnosperms of more modern aspect were an extinct group of conifers called the Voltziales (Figure 6). This group first appeared in the Late Carboniferous and the plants are thought to have resembled in general habit and leaf morphology living conifers such as *Araucaria heterophylla* (Norfolk Island Pine).

During the Late Palaeozoic, gymnosperms were an important component of terrestrial floras, but there was a significant increase in species diversity during the Early Mesozoic. Numerous petrifactions of stem and trunk show that most Mesozoic gymnosperms, like their modern relatives, were woody shrubs and trees. Evidence from fossil forests deposits reveal landscapes dominated by gymnosperms, which formed the main canopy element (Figure 7). In particular, the conifers underwent a major radiation during the Triassic period. Studies of the pollen record show that the living Gnetales diversified alongside angiosperms during the mid-Cretaceous, but then underwent a decline towards the end of this period.

Molecular phylogenetic studies provide an alternative perspective on the fossil history of some groups of gymnosperms. These are beginning to yield an unprecedented level of phylogenetic resolution in the most species-rich living groups, and they are providing additional insights into long-standing phylogenetic conundrums. Molecular data support the hypothesis that conifers had a single origin, clearing up a long-standing controversy over the phylogenetic position of the Taxaceae (Yews) (Figure 8). Also, within conifers, the pines (Pinaceae) come out as sister group to all other living species. As long suspected, the monkey puzzle family Araucariaceae is closely related to the Yellow-Wood family, Podocarpaceae. In addition to

Figure 5 Fern like frond of the extinct gymnosperm *Neuropteris heterophylla* (Carboniferous, Clay Cross, Derbyshire, England).

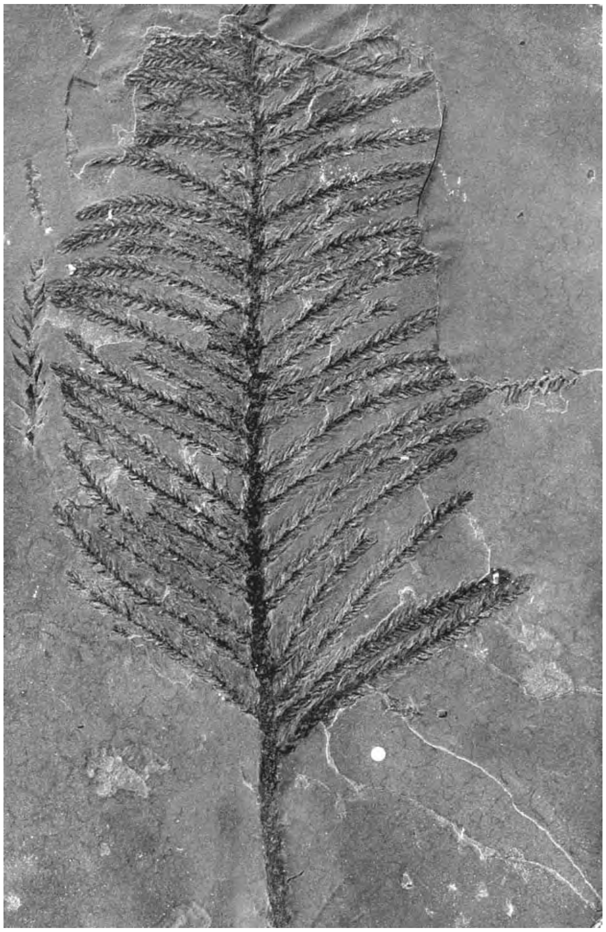


Figure 6 Leafy branch of *Walchia piniformis* (Permian, Montpellier, France).

providing phylogenetic information, one can estimate the divergence times of these groups based on a molecular clock approach. Despite some shortcomings, this approach provides an additional perspective on the evolutionary history of gymnosperms. The results of these gene-based calibrations of various family trees for living gymnosperms can and should be compared against direct evidence from the fossil record. Within cycads, family trees calibrated using the chloroplast gene *rbc*L indicate significantly later origins of genera and other major groupings than fossil evidence would support. This has been used to suggest that the disjunct distributions of modern genera are the result of Miocene or later dispersals. The fossil record indicates that this late dating of groups is unlikely because of the wider geographic distributions of modern genera and their longer histories through the Tertiary.

Major Gymnosperm Groups

Hydraspermales (Upper Devonian to Lower Carboniferous)

This extinct group contains the earliest known gymnosperms. The Hydraspermales are characterized by a very particular suite of morphological characteristics associated with their seed-like reproduction and pollination biology. The ovules of Hydraspermales are technically called preovules because they are not completely enclosed within an integument—the outer coat of later true ovules. Detailed information on the morphology and anatomy of these preovules is providing insights into the early evolution of the seed. The preovules of Hydraspermales had a unique structure termed the lagenostome, which takes the form of a funnel, ring, or cup situated at the apex. This structure was intimately involved with the capture of pollen. Preovules were borne in clusters in lobed cupulate structures. Many Hydraspermales are known only as plant parts, such as isolated seeds or cupules (Figure 3). One of the most completely characterized plants is *Elkinsia polymorpha* from the Upper Devonian of West Virginia, USA. This was a small, slender shrub that bore delicate fronds of the *Sphenopteridium* type (Figure 4).

Calamopityales (Upper Devonian to Lower Carboniferous)

This is an extinct group of rather poorly characterized small- to medium-size shrubs. The concept of Calamopityales is based largely on fragments of petrified stem, and because of this our knowledge of the whole

Figure 7 210 My old silicified tree trunks from the Petrified Forest National Park, Arizona, USA. Most of the wood is attributable to the conifer *Araucarioxylon* (Late Triassic, Chinle Formation, Arizona, USA). © Paul Davis.

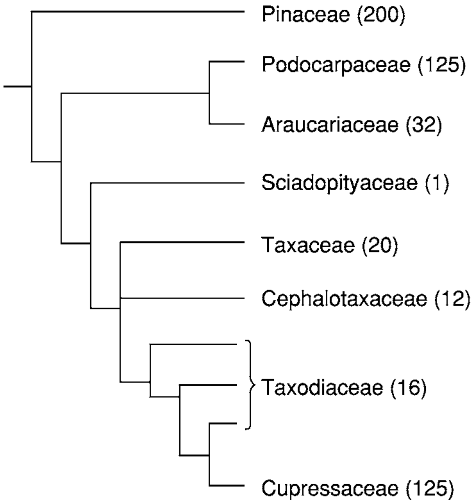


Figure 8 Phylogeny of the conifers inferred from partial 28S rRNA gene sequences. Approximate numbers of living species are given in parentheses. After Stefanovic *et al.* (1998) Phylogenetic relationships of conifers inferred from partial 28S rRNA gene sequences. *American Journal of Botany* 85: 688 697.

plants is poorly developed. The seeds, pollen organs, and foliage of Calamopityales are for the most part unknown. On the basis of frequent association, it seems likely that stems called *Stenomyelon* bore seeds known as *Lyrasperma*, which are found dispersed in sediments. Also, it is probable that Calamopityales foliage was of the general fern-like frond variety, as exemplified by form genera such as *Triphyllopteris*, *Adiantites*, and *Sphenopteridium*.

Lyginopteridales (Mid- to Late Carboniferous)

This extinct group contained small slender shrubs and possibly also trees of substantial size. The group is known mainly from petrifactions documented in coal balls, and many of the defining characters come from the often exquisitely preserved anatomy of the stem. *Lyginopteris* was probably a small shrub or a vine-like plant. It bore large fronds of the *Sphenopteris* or *Pecopteris* type. Ovules were borne in lobed cupules, and pollen developed within cylindrical sacs that were borne fused in clusters. Other possible member of the group is *Pitys*, which was a large tree with trunks in excess of 1 m diameter and 5 m in height.

Medullosales (Lower Carboniferous to Permian)

The Medullosales were among the largest of the seed ferns. This extinct group comprised small- to medium-size trees, some of which exceeded 3.5 m in height. In overall appearance, *Medullosa* bore a superficial resemblance to a modern tree fern. The trunk bore a crown of large fern-like leaves (Figure 5). Large seeds—up to 7 cm long—were borne on the underside or at the tips of fronds (Figure 9). The Medullosales produced large characteristic pollen grains in morphologically complex spore-bearing organs. Some species may have occupied mangrove-like habitats.

Cycadales (? Carboniferous to Recent)

This group contains about 140 living species classified into 11 genera and up to 4 families (Cycadaceae, Zamiaceae, Boweniaceae, Stangeriaceae) (Figure 10).

Figure 9 Large nutlike seeds (about 2.5 cm long) of *Trigonocarpus parkinsoni* (Upper Carboniferous, England).

Figure 10 Living cycad in the Fairchild Botanical Garden, Florida, USA.

The group has a predominantly tropical to subtropical distribution. Some species reach 18 m in height, whereas others are small with tuberous stems. Cycads bear one or more crowns of large compound leaves. Seeds and pollen are borne in cone-like structures on separate plants. Large species of cycads bear a superficial resemblance to palm trees with stout stems. They are also closely similar in overall appearance to extinct Mesozoic gymnosperms in the Bennettitales. Significant differences in the structure of the seed and pollen-bearing organs indicate that the two are only distant relatives. Cycads probably originated in the Late Palaeozoic. Plants such as *Archaeocycas* are known from the Lower Permian, and the ancestors of modern cycads are thought to extend back into the Carboniferous. The group reached its maximum diversity during the Mesozoic era.

Callistophytales (Late Carboniferous)

This small extinct group was first characterized in detail in 1975. The best known species is *Callistophyton poroxyloides*. The stem anatomy of the group resembles that of Lyginopteridales, whereas the reproductive structures are closer to those of Cordaitales. *Callistophyton* is portrayed as a small shrubby plant with stems up to 3 cm diameter, bipinnate fronds, and a scrambling habit. The leaves are similar to fossil foliage of *Medullopteris* type. Ovules and pollen organs were borne on the underside of unmodified leaves. Pollen was monosulcate with a bladder, resembling the pollen of conifers.

Cordaitales (Late Carboniferous to Early Permian)

This distinctive extinct group had a cosmopolitan distribution. Some species were large trees that attained heights of as much as 30 m, whereas at least one species is known to have been a prostrate scrambler some 2 m in height. The leaves of many species were large and strap-like, bearing a striking resemblance to the leaves of flowering plant monocotyledons such as *Hippeastrum* (*Amaryllis*) or living conifers in the genus *Agathis*. Pollen- and ovule-forming structures were borne on shoots scattered among the leaves. The wood was similar in structure to that of *Araucarioxylon*. The centre of the stems contained a large pith cavity with transverse septa. Fossilized infillings of this pith (*Artisia*) are commonly found. The reproductive structures are compound fructifications comprising cones subtended by bracts borne along a central axis.

Voltziales (Late Carboniferous to Early Jurassic)

This extinct group of conifer-like plants comprises three families: Utrechtiaceae (Lebachiaceae), Emporiaceae, and Majoniaceae (Figure 6). On the whole, these are thought to resemble in general habit and leaf morphology living conifers such as *Araucaria heterophylla* (Norfolk Island Pine). Some species had bisaccate pollen like that of living Pinaceae. The structure of the wood also resembles that of living conifers in the Araucariaceae. The fructifications of Voltziales resemble cones, and they were borne at the tips of leafy branches. The major difference between the cones of true conifers lies in the nature of the structure that bore the ovules. The ovule-bearing scales are attached to a short shoot rather than directly to the bract as in conifers. This condition is interpreted as intermediate between the fructifications of Cordaitales and the cones of Coniferales, to which Voltziales are clearly closely related.

Coniferales (Triassic to Recent)

This major group of gymnosperms contains some 537 living species classified into 69 genera and 9 families (Pinaceae, Podocarpaceae, Araucariaceae, Phyllocladaceae, Sciadopityaceae, Cephalotaxaceae, Cupressaceae, Taxaceae, Taxodiaceae). Most living members are medium to large evergreen trees, but the group contains shrubs and one rare parasitic species (*Parasitaxus*, Podocarpaceae). Several extinct family level units are recognized, of which the best known and most diverse is the Mesozoic Cheirolepidiaceae. The leaves of conifers are on the whole simple, needle-shaped or flattened and strap-shaped and more rarely ovate-lanceolate-elliptical. Pollen and ovules are produced in cones, except for Taxales in which single ovules terminate branches. Araucariaceae (~32 living species) have a tropical–subtropical Old World distribution, but they are absent from Africa. This is an ancient family that probably originated during the Late Permian. Araucariaceae had a much wider geographic distribution during the Mesozoic in both southern and northern hemispheres (Figure 11). Cephalotaxaceae (~12 living species) are plants of temperate to subtropical south-east Asia. They have a fossil record that extends possibly as far back as the Triassic. Cupressaceae (~125 living species) is a cosmopolitan group. Leaf shoots assignable to this family have been described from the Early Jurassic. The extinct Cheirolepidiaceae range from Triassic to Mid-Cretaceous. These were probably mostly trees with scale-like foliage of the Cupressaceae type (Figure 12). They produced a distinctive type of pollen called *Classopollis*. Pinaceae (~200 living species) are almost exclusively northern hemisphere. This is the largest living family of Coniferales. The fossil record of Pinaceae might extend back into the Triassic, but it seems likely that the crown group began to diversify comparatively late during the Cretaceous period. Podocarpaceae (~125 living species) are tropical–subtropical mostly montane conifers of the southern hemisphere. This ancient family is first documented in the Late Triassic. The monotypic Sciadopityaceae is confined to temperate south-east Asia. A Late Cretaceous or Palaeocene origin of this group seems likely. Taxaceae (~20 living species) occur mainly in the northern hemisphere. The earliest fossils are Early Jurassic. Taxodiaceae (~16 living species) are mainly northern temperate—subtropical. This family is known from the Jurassic (Figure 13). Fossils leaves in the form genus *Elatides* are most like living *Cunninghamia*. Much fossil wood of Mesozoic age is assignable to the Coniferales.

Glossopteridales (Permian to Triassic)

This is an extinct group of Gondwanan gymnosperms. Over 50 species have been described based on leaf morphology. Many of these are probably variants of

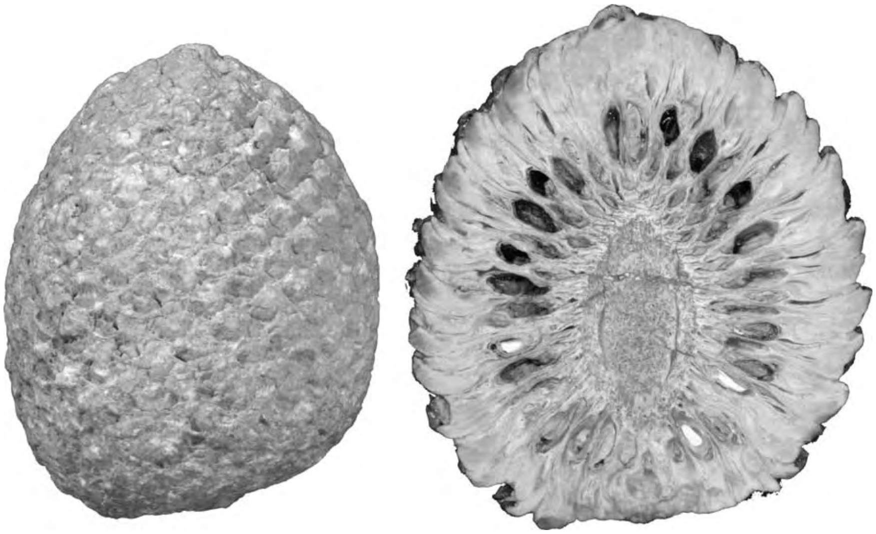


Figure 11 Silicified cones of *Araucaria mirabilis* (Upper Jurassic, Cerro Alto, Santa Cruz, Argentina). Specimen on right cut and polished to reveal inner structure.

Figure 12 Thick, robust leaves of the extinct conifer *Pagiophyllum peregrynum* (Upper Jurassic, Lyme Regis, Dorset, England). Foliage of this type could belong to several families of conifers including Araucariaceae and Cheirolepidiaceae.

Figure 13 Foliage of the Dawn Redwood *Metasequoia* was documented in the fossil record long before living specimens of these large conifers were discovered in the Chinese province of Szechuan in 1944 (Eocene, Driftwood Canyon, Smithers, British Columbia, Canada).

Figure 14 Tongue shaped leaves of *Glossopteris* are characteristic of the Permian and Triassic rocks of Gondwana: *G. browniana* (Permian, Nagpur, India).

the same species or individuals. Some are known to have been large trees. The leaves of Glossopteridales are highly characteristic. They are large and tongue-shaped with a conspicuous midrib and a network of veins (Figure 14). Pollen organs and ovules were borne on special branches on the upper surfaces of some leaves. Some fructifications are foliar, whereas others are thought to resemble cones.

Czekanowskiales (Jurassic to Cretaceous)

This is an extinct and poorly understood group. The leaves of Czekanowskiales are distinctive. They are highly dissected, in some species almost like a bundle of fibres. Ovules are thought to have been borne in cones composed of capsulate appendages. The habit of the plant remains poorly understood.

Ginkgoales (? Permian to Recent)

This group contains one living species *Ginkgo biloba*, which is a large tree that bears distinctive fan-shaped leaves. Pollen organs are loose catkin-like cones, whereas ovules are in pairs on stalks. They are borne on short leafy shoots on separate plants. The natural distribution of modern *Ginkgo* is restricted to a small area of China, but leaves of *Ginkgo* type are widespread in the fossil record of the Mesozoic and Tertiary eras. The leaf morphology ranges from highly dissected resembling leaves of Czekanowskiales to forms very similar to those of the modern species (Figure 15). Even though ovules attributable to the Ginkgoales were common in their dispersed form during the Mesozoic very few intact reproductive structures are known. Ovulate structures and catkin-like pollen cones have been reported from the Jurassic of Yorkshire, England.

Peltaspermales (Permian to Late Triassic)

The concept of Peltaspermales is based on leaves, pollen organs, and ovulate organs. Knowledge of whole plants is scant, but organ associations and characteristic cuticular features have been used to piece together some aspects of whole plant morphology. Clusters of pollen sacs and ovule-bearing structures were borne in a pinnate arrangement on branches. Plants bore large fern-like leaves, which were probably deciduous. This is consistent with a woody habit.

Corystospermales (Mid-Triassic to Mid-Jurassic)

The concept of this group is based on leaves, pollen organs, and ovulate organs. Knowledge of whole plants is scant, but organ associations, characteristic cuticular features, and the association of pollen with ovules have been used to piece together some aspects of whole plant morphology. Woody stems bore fern-like foliage. Ovules were borne in cupules, whereas pollen was borne in elongated sacs on the underside of oval laminae. Woody stems bore fern-like foliage.

Caytoniales (Upper Triassic to Lower Cretaceous)

This extinct group is widely recognized on the basis of distinctive leaves. The leaves (*Sagenopteris*) are palmate, composed of three to six lanceolate blades attached to a petiole (Figure 16). Each blade has an anastomosing network of veins. Knowledge of whole plants is scant, but the concept of the group has been built from well-established organ associations. Ovules were borne in cupules along a branch (*Caytonia*). The pollen was produced in elongated anther-like structures (*Caytonanthus*). *Caytonia* is though to have been a small tree with a woody trunk.

Pentoxylales (? Jurassic to Cretaceous)

This extinct group is known from India, New Zealand, and Australia. Knowledge of whole plants is scant, but the stems (*Pentoxylon*) are known to have borne a distinctive type of wood, resembling that of conifers. The leaves were large (up to 20 cm long), strap-shaped, and borne on short lateral shoots.

Figure 15 Leaf of the Maidenhair tree, *Ginkgo gardneri* (Palaeo cene, Isle of Mull, Scotland).

Figure 16 The extinct gymnosperm *Caytonia* bore palmate leaves: *Sagenopteris phillipsi* (Jurassic, Cayton Bay, Yorkshire, England).

Figure 17 Stout, barrel shaped trunk of the extinct Bennettitalean gymnosperm *Cycadeoidea microphylla* (Jurassic, Isle of Portland, Dorset, England).

Ovules were borne in clusters of compact cones. Pollen organs were borne in clusters on a receptacle. These plants were probably small trees or shrubs.

Bennettitales (Triassic to Upper Cretaceous)

This extinct group had a cosmopolitan distribution. In habit, Bennettitales were shrubs or small trees with stout, sparsely branched trunks or squat, unbranched, and barrel-shaped trunks (Figure 17). The cones of Bennettitales were highly distinctive flower-like structures. Cones in the genus *Williamsoniella* comprised a central ovule bearing receptacle subtended by a whorl of rather broad, leaf-like pollen-bearing segments. Whorls of leaf-like bracts subtend the pollen and ovule-bearing parts. Cones were borne either along branches (Williamsoniaceae) or along the squat trunks embedded and partly protruding from a thick layer of old leaf bases (Cycadeoidaceae). In some, pollen- and ovule-bearing parts were in separate cones (e.g., *Williamsonia*), whereas in others they were borne within the same cone (e.g., *Williamsoniella*). Leaves were large and either pinnate or flat bladed (Figure 18). The striking superficial similarity to modern cycads masks fundamental differences in anatomy and especially the structure of the cones.

Gnetales (Late Triassic to Recent)

This group of gymnosperms contains approximately 70 living species classified into three genera and families (Ephedraceae, Gnetaceae, Welwitschiaceae). Living members are predominantly shrubs or vines, but the group contains some small trees. Of the living groups of gymnosperms, the fossil record of gnetales is most poorly understood. Macrofossil remains are sparse, but there is a lengthy pollen record that begins in the Late Triassic. Fossil pollen resembles that of living Ephedraceae and Welwitschiaceae. The record of dispersed pollen indicates that living Gnetales are relicts of a group that was once more widespread and much more diverse. Maximum diversity was achieved during the mid-Cretaceous. Some putative early Gnetales macrofossils resemble conifers, raising the possibility that the poor fossil record might in part be explained by misidentification (Crane, 1996).

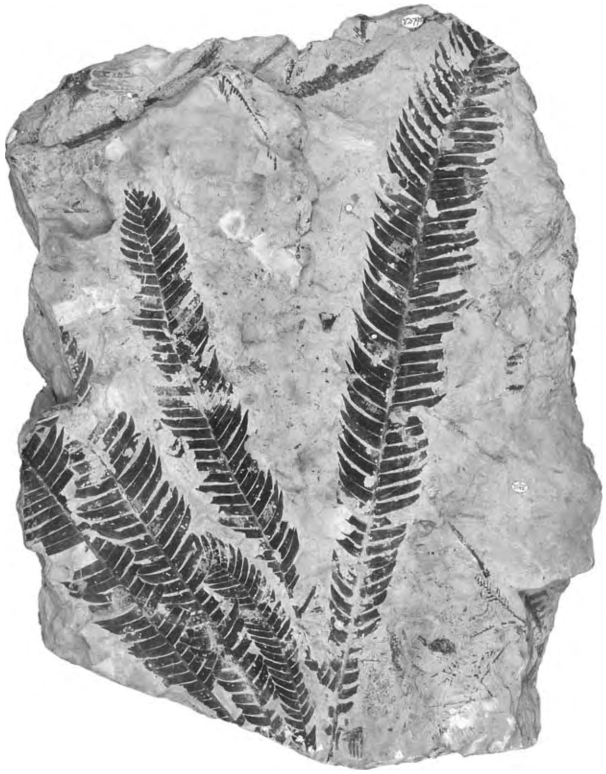


Figure 18 Leaves of the extinct cycad like Bennettitalean gymnosperm *Zamites gigas* (Jurassic, Yorkshire, England).

See Also

Dendrochronology. **Fossil Plants:** Angiosperms. **Mesozoic:** Triassic; Jurassic; Cretaceous. **Microfossils:** Palynology. **Palaeozoic:** Devonian; Carboniferous. **Tertiary To Present:** Oligocene.

Further Reading

Beck CB (ed.) (1988) *Origin and Evolution of Gymnosperms*. New York: Columbia University Press.

Crane PR (1996) The fossil history of the Gnetales. *International Journal of Plant Sciences* 157: S50 S57.

Crane PR and Kenrick P (1997) Problems in cladistic classification: higher level relationships in land plants. *Aliso* 15: 87 104.

Donoghue MJ and Doyle JA (2000) Seed plant phylogeny: demise of the anthophyte hypothesis. *Current Biology* 10: R106 R109.

Kramer KU and Green PS (eds.) (1990) *The Families and Genera of Vascular Plants. I. Pteridophytes and Gymnosperms*. Berlin: Springer.

Meyen SV (1984) Basic features of gymnosperm systematics and phylogeny as evidenced by the fossil record. *Botanical Review* 50: 1 111.

Rydin C, Kallersjo M, and Friis EM (2002) Seed plant relationships and the systematic position of Gnetales based on nuclear and chloroplast DNA: Conflicting data, rooting problems, and the monophyly of conifers. *International Journal of Plant Sciences* 163(2): 197 214.

Stefanovic S, Jager M, Deutsch J, Broutin J, and Masselot M (1998) Phylogenetic relationships of conifers inferred from partial 28S rRNA gene sequences. *American Journal of Botany* 85: 688 697.

Stewart WN and Rothwell GW (1993) *Paleobotany and the Evolution of Plants*. Cambridge: Cambridge University Press.

Treutlein J and Wink M (2002) Molecular phylogeny of cycads inferred from rbcL sequences. *Naturwissenschaften* 89(5): 221 225.

FOSSIL VERTEBRATES

Contents

Jawless Fish-Like Vertebrates
Fish
Palaeozoic Non-Amniote Tetrapods
Reptiles Other Than Dinosaurs
Dinosaurs
Birds
Swimming Reptiles
Flying Reptiles
Mesozoic Amphibians and Other Non-Amniote Tetrapods
Cenozoic Amphibians
Mesozoic Mammals
Placental Mammals
Hominids

Jawless Fish-Like Vertebrates

D K Elliott, Northern Arizona University, Flagstaff, AZ, USA

Introduction

Jawless fishes (agnathans) have a fossil record that spans almost 500 million years, now extending back to the Early Cambrian of China (*see* **Palaeozoic:** Cambrian). They were particularly prominent during the Late Silurian and Devonian (*see* **Palaeozoic:** Devonian; Silurian) when they radiated to form a variety of diverse groups, most of them characterised by the presence of bony armour plating over the head and body. They are valuable in biostratigraphy, helping to determine the placement (*see* **Stratigraphical Principles**) of continents during the Palaeozoic and they chart the development of many features important in vertebrate history such as the appearance of bone, paired fins, lateral line systems, and an inner ear with two semicircular canals. During the last decade there has been a dramatic increase in the amount and quality of information about the origin of vertebrates and the diversification of the earliest forms. This has stemmed partly from new finds that have provided us with examples of early forms and also from new methods of analysis, which have enabled the development of a better understanding of relationships (*see* **Micropalaeontological Techniques**).

Origin and Relationships of Vertebrates

Vertebrates belong to the phylum Chordata, which also includes the marine cephalochordates and tunicates. Vertebrates have the basic chordate characters of a notochord (a tough flexible rod that runs down the back) and V-shaped muscle blocks or myotomes along the length of their bodies but they also have the hard tissue bone and elements of the vertebral column, characters not shared by the other chordates. It is generally considered that the cephalochordates, small eel-like marine filter feeders exemplified by amphioxus, are the closest relatives of the vertebrates while the tunicates, attached marine suspension feeders, are considered to be less closely related. Within the vertebrates the significance of the development of the jaw is recognised by the classification of vertebrates into gnathostomes ('jaw mouths') and agnathans ('without jaws'). The jawless vertebrates are considered to be the most primitive and were originally separated as the Agnatha by Cope in 1889. At that time the Agnatha included the modern lampreys and hagfishes, together with a number of fossil groups often termed ostracoderms ('shell skin'). The modern forms share a similarity in their primitive eel-like appearance and their feeding habits (lampreys are parasites on other fish, sucking their blood and rasping their flesh, while hagfish eat their way through dead or dying fishes). The absence of jaws also assigns the ostracoderms to the Agnatha, however, it is clear that many of them also share characters with the gnathostomes, thus leading to the problem of their relationship to modern vertebrates. Developing an understanding of these relationships has been difficult because of the generally poor preservation of the internal anatomy of ostracoderms and because careful studies of their anatomy were not carried out until recently. However, analysis of some forms using serial grinding techniques in the early 1900s, together with recent discoveries of many new kinds of ostracoderms and the development of new acid preparation techniques have greatly improved our knowledge of these organisms.

It has become clear more recently that the lampreys share a large number of characters with the gnathostomes and they are now considered to have a sister-group relationship with them, the hagfishes being the sister group of lampreys plus vertebrates. The grouping of lampreys plus gnathostomes is now more correctly referred to as the Vertebrata because elements of the vertebral column are present while these are missing in the hagfishes. The more inclusive group including hagfishes is termed the Craniata because they all have a skull but frequently the two terms are used interchangeably and in this article I will use Vertebrata.

Early Chordates

The earliest chordates are known from the Chengjiang and Burgess Shale lagerstätte which are Early and Middle Cambrian in age, respectively (*see* **Palaeozoic:** Cambrian). Of these the best known is *Pikaia gracilens* which is a leaf-shaped organism with clearly defined sigmoidal muscle blocks and a notochord. There is some doubt about its cephalochordate affinities, however, as there is no evidence for any gill-like structures. A similar animal from the Chengjiang fauna, *Cathaymyrus*, is also thought to be a cephalochordate and both have been compared with the modern cephalochordate amphioxus. Thus, there may be evidence of the closest relatives of vertebrates as far down as the Early Cambrian. Very recently however, two additional species have been reported from Chengjiang, *Myllokunmingia* and *Haikouichthys*, both of which are described as vertebrates. Both show zig-zag shaped muscle blocks, relatively complex and presumably cartilaginous skulls, gill arches, heart, and fin supports but the presence of serially arranged gonads and a dorsal fin with rostrally tilted radials casts doubt on their inclusion within vertebrates.

Prior to the discovery of the Chengjiang vertebrates, the evidence for the presence of vertebrates in the Cambrian rested on the preservation of fragments of phosphatic hard tissue. Of these *Anatolepis* has been reported from the Late Cambrian although initially described from the Early Ordovician of Spitsbergen (*see* **Palaeozoic:** Ordovician). It consists of microscopic plates and spines with scale-like ornamentation and a layered internal structure. Although there has been some controversy over the vertebrate attribution of this material, recent studies of the histology of these fragments show that they do contain the characteristic vertebrate hard tissue dentine. Although fragments of purported vertebrate hard tissue have also been recently reported from the Late Cambrian of Australia there is some doubt as to their affinity as thin sections show a resemblance to some arthropod cuticles.

Conodonts

Conodonts are small tooth-like structures made of apatite and have been known since the mid-1800s.

Although extremely abundant in marine rocks and used extensively in biostratigraphy their vertebrate affinities have only recently been understood (*see* **Microfossils:** Conodonts). The hard tissue of the crown consists of an enamel-like material while the underlying basal body includes the mineralised tissue dentine. The discovery of complete conodont animals from the Mississippian Granton Shrimp Beds of Scotland in 1983 demonstrated that they were small and eel-like (22–25 mm long) with a feeding basket composed of the conodont elements that functioned in seizing prey and cutting it up (Figure 1). Additional specimens from the Silurian of Wisconsin and the Upper Ordovician (*see* **Palaeozoic:** Silurian; Ordovician) of South Africa are larger (up to 40 cm long) and have demonstrated the presence of a number of chordate characters including a notochord, a dorsal nerve cord, the presence of myotomes or muscle blocks, the presence of a midline tail fin, and a bilaterally symmetrical body. Further characters that place the conodonts within the vertebrates include a head lying in front of the notochord, a caudal fin with radial supports, extrinsic eye musculature, and an apatitic skeleton. Interestingly, the presence of hard tissues in conodonts indicates that they have more advanced characters than either the hagfishes or the lampreys.

Agnathan Diversity

At least 600 species of ostracoderm are known and most species were covered in a bony armour. This armour may have developed as a storage area for calcium and phosphate or to prevent loss of water from the body tissues, or it may have developed as a protection from predators. In early vertebrates two principal types of bone occur: acellular bone (aspidin) which was laid down by the dermis and is often laminated, and cellular bone in which there are cell spaces for osteocytes (bone-forming cells) throughout the bone. The ostracoderms show a bewildering variety of types of bone and plate and tail shapes. In spite of this variety though most ostracoderms can be classified into one of a small number of groups.

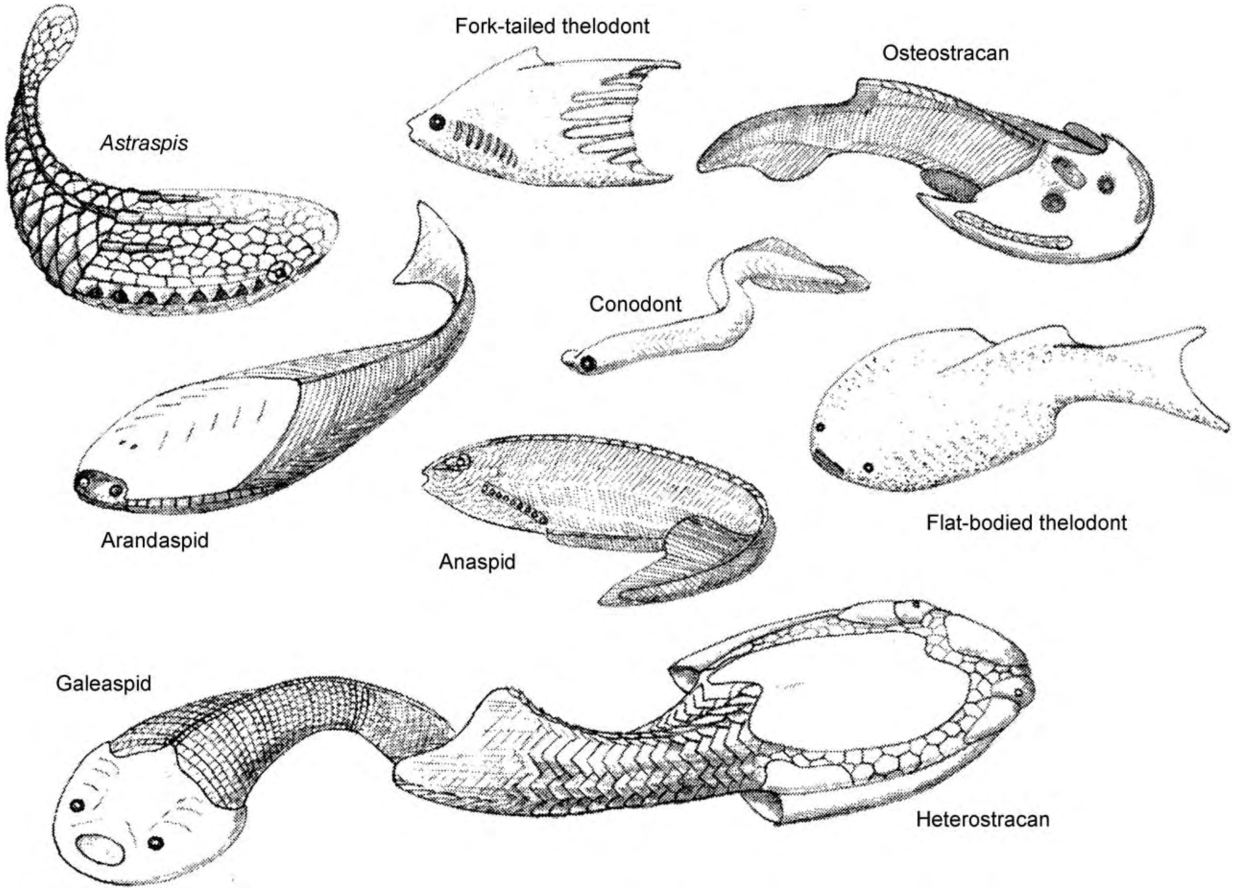


Figure 1 Reconstructions of some of the main groups of jawless vertebrates. (Based partly on Elliott, 1987 and Janvier, 1998).

Ordovician Ostracoderms

The North American fishes *Astraspis* and *Eriptychius* were long known as the earliest fossil fish, having been initially described by Charles Walcott in the late nineteenth century. *Astraspis* (Figures 1–3) is known from a number of localities flanking the Ordovician Transcontinental arch and was covered by tesserae (small bony plates) with a three-layered structure of honeycombed sheets of aspidin covered by tubercles formed from dentine with an enameloid cap. There were eight pairs of gill openings behind the laterally positioned eyes, no paired or dorsal fins, and a tail covered by large overlapping scales. The structure of the anterior of the headshield and of the tail termination are currently unknown. *Eriptychius* is known only from isolated tesserae and a single partially articulated head, which demonstrates the presence of an endoskeleton formed of globular calcified cartilage. There is little information on *Eriptychius* and its affinities have yet to be resolved.

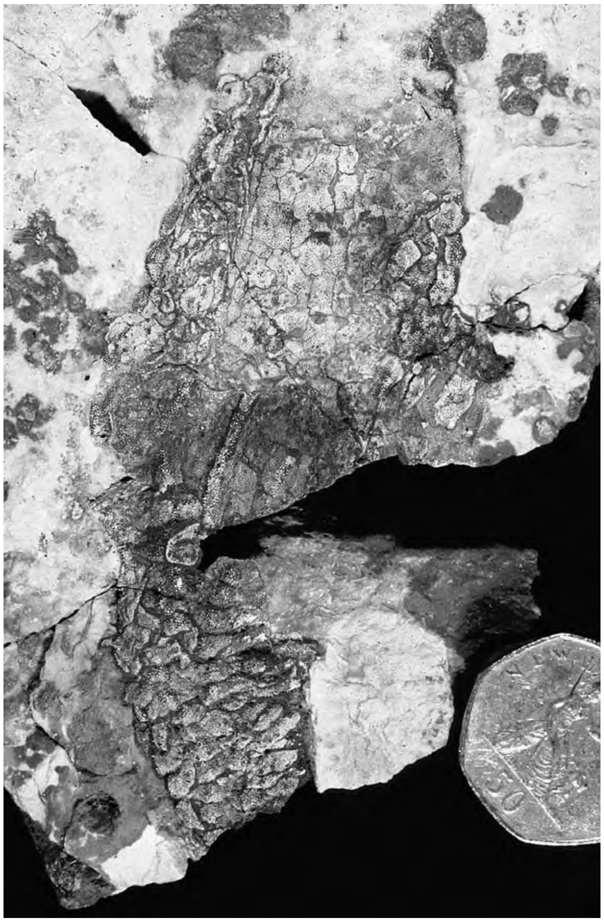


Figure 2 An articulated specimen of *Astraspis desiderata* from the Middle Ordovician Harding Sandstone of Colorado. The animal is exposed on the dorsal surface and the orbits can be seen on either side at the top of the preserved part of the speci men. The anterior part of the head is lost to erosion and the tail is also missing. The specimen is 120 mm long.

The arandaspids are known from the articulated remains of two genera, *Arandaspis* (Australia) and *Sacabambaspis* (Bolivia) (Figure 1). Both animals have headshields formed from roughly oval dorsal and ventral plates ornamented with oak-leaf or teardrop shaped tubercles. The gill openings are numerous and form a sloping row, covered by square branchial plates, between the two main plates. The eyes are anterior and paired pineal and parapineal openings are present, a feature unique among vertebrates. The trunk scales are narrow and elongated. Although several other genera are known from scales or tesserae from the Early Ordovician of central Australia their affinities are currently unknown.

Silurian and Devonian Ostracoderms

The ostracoderms flourished during the Late Silurian and Devonian during which period they were found worldwide. The osteostracans (Figure 4) were a successful group of ostracoderms that lived in North America, Europe, Siberia, and central Asia. A semicircular headshield encased the dorsal surface of the skull while the undersurface was covered with small tesserae that surrounded the gill openings and mouth. The brain was surrounded by perichondral bone (present in many gnathostomes but not lampreys and hagfishes) and its details are thus well known. The eyes were situated on top of the head and were separated by openings for the pineal and nasohypophysial openings. Central and lateral 'sensory fields' on the head are characteristic of this group and were probably a specialised part of the lateral line system that enables fishes to detect vibrations in the water. Paired pectoral fins and a flexible tail made these ostracoderms very maneuverable.

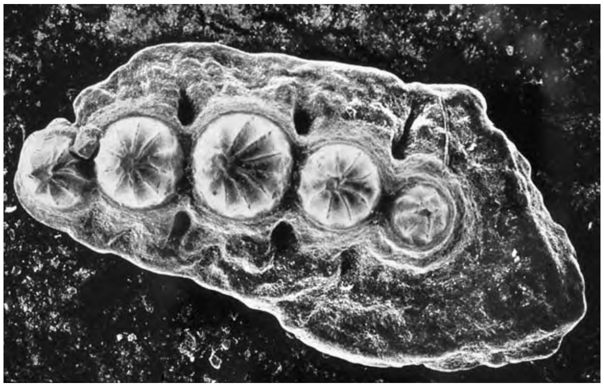


Figure 3 SEM photograph of a single tessera (4 mm long) from the headshield of *Astraspis desiderata* showing the stellate tuber cles that give it its name.

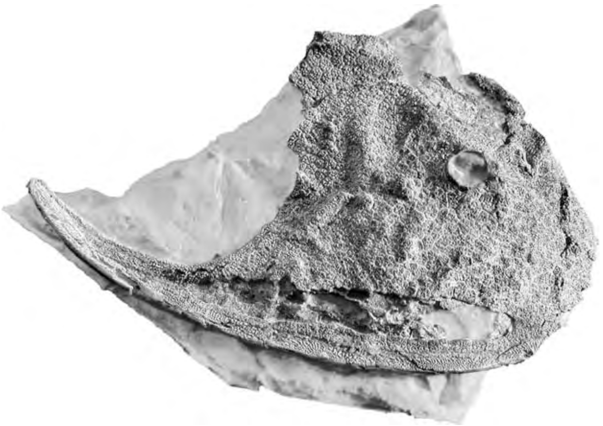


Figure 4 Lateral view of the headshield of the Early Devonian osteostracan *Cephalaspis utahensis* showing the location of the orbits and the specialized sensory areas. The headshield is 130 mm long.

A related group of ostracoderms, the galeaspids (Figure 1), inhabited parts of China and North Vietnam during the Late Silurian and Devonian. Although similar to osteostracans in general appearance they differ in having a central opening in the headshield that was used to take water into the gill chamber. Galeaspids also lacked paired fins.

A small and enigmatic group of agnathans from the Middle Devonian of Australia have been named the pituriaspids ('hallucinatory shield'). These animals had an elongated headshield with small, paired areas similar to the sensory areas in osteostracans. They also appear to have had paired pectoral fins that projected through openings in the headshield. Their relationship to other agnathans is unclear, although they are generally placed close to osteostracans and galeaspids.

Anaspids lived at the same time as the osteostracans and galeaspids in North America and Europe, and had fusiform bodies with downturned tails (Figure 1). They may have had ribbon-like paired fins behind the row of gill openings that allowed them to manoeuvre slowly forward and back. Some anaspids were scaled while others lacked scales and the latter type sometimes had circular cartilage surrounding the mouth together with large eyes and a basket-like branchial skeleton. Although these features have been cited in the past as suggesting a relationship with lampreys they are now generally considered to be convergences.

Figure 5 Articulated specimen of the Early Devonian hetero stracan *Errivaspis waynensis* in ventral view. the oral area is formed of finger like oral plates whose function is currently unknown. The animal is 140 mm long.

Figure 6 A dorsal shield of the Early Devonian heterostracan *Tuberculaspis elyensis* showing the single pair of branchial ducts opening posteriorly and the posterior dorsal spine. The shield is 65 mm long.

Heterostracans were among the first of the armoured agnathans to be discovered although it was not until 1858 that Thomas Huxley recognised them to be vertebrates. They are a group of armoured agnathans that are characterised by the presence of a common external branchial opening on either side of the head (Figures 1, 5 and 6). The armour consists of apidin with a well-developed honeycomb structure and a surface of dentine ridges or tubercles. The earliest forms are Early Silurian in age and the youngest are found in the Late Devonian (Late Frasnian) where they reached 1.5 meters in length making them the largest known agnathans. Heterostracans are usually found as disarticulated plates and scales but some complete animals are known from the Late Silurian and Early Devonian of northern Canada and from the Early Devonian of Germany and the Welsh Borders. None of these specimens shows any trace of the endoskeleton beyond impressions on the internal surface of some of the dermal plates and it is assumed to have been cartilaginous, however, the impressions have enabled some reconstruction of the internal anatomy to be made. Because of this lack of internal information the arrangement and structure of the dermal armour is the main feature used in discussions of classification, phylogeny, or affinities.

The head armour of heterostracans consists of a series of large plates that also contain the canals of the sensory canal system. These plates are almost always arranged to form a dorsal and a ventral shield. The dorsal shield generally consists of a series of median and lateral plates while the ventral shield is generally a single plate. There are variations to this arrangement as some heterostracans have a ventral shield composed of small platelets while other forms have a headshield in which all the plates are fused into a continuous capsule. The body is covered by large scales which become smaller on to the caudal fin where they are arranged in radial rows. They seem to have been animals that lived in brackish and freshwater environments perhaps feeding on algal layers or microorganisms suspended in the water column.

Thelodonts (Figure 1) are an interesting group of agnathans that occur from the Ordovician through the Devonian worldwide. They are rarely found complete but the tiny scales that covered their bodies (Figure 7) can be extremely numerous and are valuable tools in biostratigraphic correlation. Rare complete specimens show that most were flattened fish with slanting gill openings and broad pectoral areas. However, new material from the Mackenzie Mountains in Canada shows that some were laterally compressed with large forked tails and small triangular dorsal fins. These Furcacaudiformes have a recognisable stomach which is an organ previously thought to be absent in jawless fishes as it is lacking in the living forms. The structure of thelodont scales is in some cases very similar to the placoid scales of sharks and this together with the presence of a stomach has given rise to the suggestion that some thelodonts may be closely related to gnathostomes.

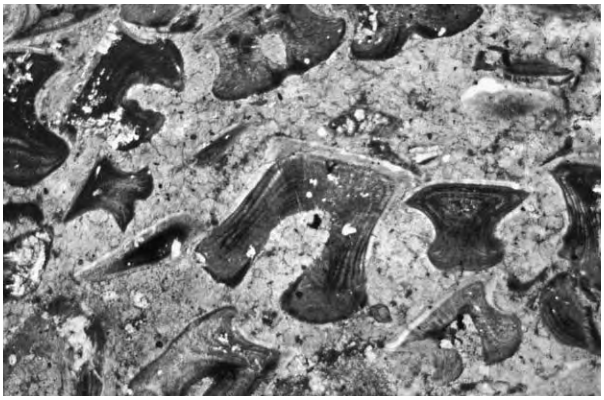


Figure 7 Scales of thelodont *Thelodus macintoshi* in cross section and showing the central pulp cavity surrounded by lam inar aspidin and capped by a clear crown of dentine. The large central scale is about 1 mm across.

The Last Ostracoderms

Although lampreys and hagfishes have a very poor fossil record, they are known from the Pennsylvanian and thus must have continued through the Palaeozoic and Mesozoic to the present. The armoured ostracoderms, however, had almost all become extinct by the start of the Late Devonian. At that point one osteostracan and three anaspids are known from Canada, one galeaspid from northern China, a single Australian thelodont, and some of the last heterostracans from Europe. The decline is attributed mostly to the rise of the gnathostomes as all the major groups of jawed fishes evolved during the Devonian and by the Middle Devonian many had reached a peak of diversity. It is clear that in many cases the niches once inhabited by agnathans were taken over by gnathostomes as they succumbed to increasing predation pressure.

Agnathan Relationships

The enormous range of features in the agnathans makes it difficult to sort out the relationships between them. Because of this a variety of schemes have been proposed, starting with the idea that the lampreys and hagfishes were most closely related to each other (the 'cyclostomes') because both were scale-less and elongated and had pouch-like gills and median nostrils. By the 1970s, however, it became clear that lampreys shared a suite of characters with gnathostomes including the presence of neural arches along the notochord, large eyes with associated muscles, nervous control of the heart, osmotic regulation of body fluids, and a brain with a cerebellum and other advanced features. It was proposed at this time that lampreys were more closely related to the gnathostomes than either was to the hagfishes. Although molecular data has now been brought to bear on this problem, it only supports the close relationship of lampreys and gnathostomes if the cephalochordates are used as the group for comparison. If the tunicates are used then the cyclostome relationship is supported. These results suggest that more molecular analyses are necessary to clarify the historical message. Fossil lampreys and hagfishes do not help with this problem as they are virtually similar to the modern forms.

Analysis of the fossil agnathans has depended very much on the methodology termed 'phylogenetic systematics' or 'cladistics'. (*see* **Palaeontology**) This method was first introduced in the 1960s by the German entomologist Willi Hennig and has been improved most recently by the introduction of computer algorithms. The goal of cladistics is to recognise sister-group relationships as in the statement that taxon A is more closely related to taxon B than to taxon C. These relationships are recognised by the presence of derived (advanced) characters that are unique to the sister-group pair and so are assumed to have arisen in the most recent common ancestor of both. Primitive characters present in taxa other than the sister-group are thus not relevant to the relationship. Analysis of the distribution of characters results in groupings that are most parsimonious, i.e., groups in which the number of changes is kept to a minimum.

An analysis of features within modern and fossil agnathans confirms that among living forms lampreys and gnathostomes share a sister-group relationship (Figure 8). The hagfishes are the most primitive known vertebrates and therefore many hagfish characters must indicate the primitive condition including the lensless eye, simple ear, lack of nervous regulation of the heart, the absence of lateral line system, electroreceptors, and cerebellum, and an inability to regulate internal ionic concentrations. Among the ostracoderms, the osteostracans are most closely related to the gnathostomes while conodonts and anaspids are the most distantly related. That the dermal skeleton

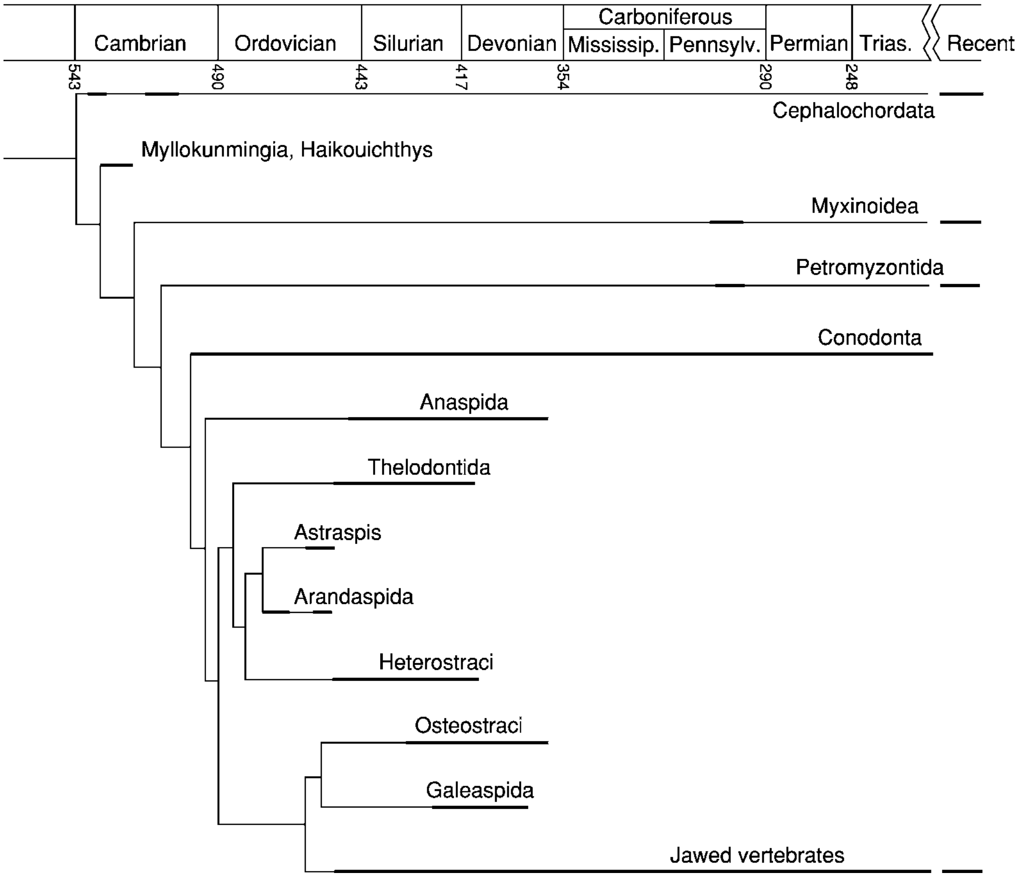


Figure 8 The evolutionary relationships of jawless vertebrates are here superimposed on a stratigraphic scheme. Broad lines indicate the fossil record of each lineage. The relationships of the Conodonta and of the Cambrian genera *Myllokunmingia* and *Haikouichthys* are a matter of some debate at present. Although the earliest accepted gnathostomes are Silurian acanthodians there are a number of scales of chondrichthyan, acanthodian, and placoderm type recognized from the Harding Sandstone. (Compiled from Donoghue *et al.*, 2000; Forey, 1995; Janvier, 2003).

arose later is suggested by the fact that all ostracoderms seem to be more derived than lampreys or hagfishes. Initially it would have consisted of dentine and acellular bone (aspidin) with cellular bone appearing later. A calcified internal skeleton also appeared in the ancestors of gnathostomes after the divergence of lampreys from hagfishes. Paired fins, characteristic of gnathostomes, were presumably present in the ancestors of osteostracans and possibly pituriaspids as they were the only other fossil group that might have possessed them.

Following this analysis of relationships it is possible to reconstruct the evolution of some of the vertebrate characters. An example is the lateral line system, electroreceptors, and the labyrinth of canals in the inner ear used in balance. Hagfishes lack both electroreceptors and neuromasts (lateral line cells) while the lampreys have neuromasts in short lines on the head and electroreceptors scattered over the head and body. In the extinct arandaspids and *Astraspis* from the Ordovician grooves in the surface of the headshields are thought to have held neuromasts while the heterostracans, thelodonts, osteostracans, and galeaspids have a lateral line system that consists of a complex of canals that runs within the plates and scales and opens to the surface through pores. This system is similar to that found in gnathostomes and so development seems to have been from isolated neuromasts to neuromasts in grooves on the head and body to neuromasts embedded in the plates and scales. Similarly the lampreys, osteostracans, and galeaspids have two vertical canals in the inner ear, as do the heterostracans, while the gnathostome ear has two vertical semicircular canals together with a horizontal canal. Unfortunately it is unknown what the brains of arandaspids, thelodonts, and anaspids were like.

Many questions concerning the early development of vertebrate structures still remain unanswered but it is clear that the ostracoderms have provided information unobtainable in any other way. New information will certainly be gained as further searches are made in the Cambrian and Ordovician rocks that are now yielding early vertebrates and perhaps into rocks below the Cambrian that may contain even more ancient forms.

See Also

Fossil Vertebrates: Fish; Palaeozoic Non-Amniote Tetrapods. **Microfossils:** Conodonts. **Micropalaeontological Techniques**. **Palaeoecology**. **Palaeontology**. **Palaeozoic:** Cambrian; Ordovician; Silurian; Devonian. **Stratigraphical Principles**. **Time Scale**.

Further Reading

Blieck A, Elliott DK, and Gagnier P Y (1991) Some questions concerning the phylogenetic relationships of the heterostracans, Ordovician to Devonian jawless vertebrates. In: Chang MM, Liu YH, and Zhang GR (eds.) *Early Vertebrates and Related Problems of Evolutionary Biology*, pp. 1–17. Beijing: Science Press.

Donoghue PCJ, Forey PL, and Aldridge RJ (2000) Conodont affinity and chordate phylogeny. *Biological Review* 75: 191–251.

Elliott DK (1987) A reassessment of *Astraspis desiderata*, the oldest North American vertebrate. *Science* 237: 190–192.

Elliott DK, Blieck A, and Gagnier P Y (1991) Ordovician Vertebrates. In: Barnes CR and Williams SH (eds.) *Advances in Ordovician Geology*, Geological Survey of Canada Papers 90, pp. 93–106. Canada: Ottawa.

Forey PL (1995) Agnathans recent and fossil, and the origin of jawed vertebrates. *Reviews in Fish Biology and Fisheries* 5: 267–303.

Janvier P (1998) *Early Vertebrates*. Oxford: Clarendon Press.

Janvier P (2003) Vertebrate characters and the Cambrian vertebrates. *C. R. Paleovol* 2: 1523–1531.

Long JA (1993) *Palaeozoic Vertebrate Biostratigraphy and Biogeography*. London: Belhaven Press.

Sansom IJ and Elliott DK (2002) A thelodont from the Ordovician of Canada. *Journal of Vertebrate Paleontology* 22: 867–870.

Sansom IJ, Smith MP, Smith MM, and Turner P (1997) *Astraspis* the anatomy and histology of an Ordovician fish. *Palaeontology* 40: 625–643.

Sansom IJ, Smith MM, and Smith PM (2001) The Cambrian origin of vertebrates. In: Ahlberg PE (ed.) *Major Events in Early Vertebrate Evolution*, Systematics Association Special Volume Series 61, pp. 67–84.

Sansom IJ, Smith MM, and Smith PM (2001) The Ordovician radiation of vertebrates. In: Ahlberg PE (ed.) *Major Events in Early Vertebrate Evolution*, Systematics Association Special Volume Series 61, pp. 156–171.

Fish

J A Long, The Western Australian Museum, Perth, WA, Australia

Introduction

Fishes, loosely defined as free-swimming aquatic vertebrates, have a fossil record spanning some 540 million years (Figure 1). The fossil remains of early fishes are useful for elucidating the major anatomical stages of vertebrate evolution, from the origins of bone, jaws, and teeth through to the development of powerful limbs leading to the tetrapod invasion of land. Some Palaeozoic fish remains are biostratigraphically useful, and in some instances can provide valuable palaeoecological information.

The earliest fishes are boneless forms like *Haikouichthys* and *Myllokungmingia* from the Early Cambrian Chengjiang fauna, Yunnan, China. *Haikouichthys* is regarded as the first fish as it has a fusiform body with notochord, gill slits, V-shaped muscle bands, and fin-rays supporting a median fin and tail fin. These are now recognized as the first true fishes even though they lack bone.

The evolutionary development of organically mineralized tissue which supports and protects the soft tissues and allows for more efficient attachment of muscles marks the beginning of a great diversity of the first jawless fishes, the agnathans.

The oldest vertebrates remains of scales composed of bone-like tissues come from the Late Cambrian Georgina Basin of central Queensland, Australia. Early Ordovician fish remains from central Australia include several forms known from scales, such as *Areyongia* and *Apedolepis*. The first well-preserved impressions of complete jawless fish armours occur in the Middle Ordovician of Australia. *Arandaspis* (Figure 2) had well-developed dorsal and ventral bony shields, with distinct rows of squarish branchial plates over the gills. It was a precursor to the more successful heterostracan fishes, such as *Pteraspis*, which had a single branchial plate covering the gills. Such fishes were prevalent in Silurian and Early Devonian seas, and to a lesser extent, freshwater habitats, largely occupying the niches of benthic detrital and filter feeders. The bone of heterostracans lacked bone cells, and has been termed ‘aspidin’.

Agnathan Diversity

Many kinds of jawless fishes evolved throughout the Silurian and Devonian. These include armoured forms with a scoop-shaped solid bony shield such as the osteostracans (e.g., *Cephalaspis*), the naked lamprey-like anaspids (e.g., *Jaymoytius*), and the heavily-scaled thelodonts (e.g., *Thelodus* (Figure 2)). Finds from Canada indicate that some thelodonts, the Furcacaudiformes, had laterally compressed bodies with well-developed stomachs. Thelodonts, in general, had an internal soft anatomy (evidenced by impressions in *Thelodus*) suggesting close affinity to jawed fishes (gnathostomes). All early agnathans had well-developed sensory-line systems, which are seen as lines of pores or grooves in the dermal bones. Osteostracans possessed complex fields of sensory organs along the sides and top of their shields.

Osteostracan fishes appear to be the closest group to the gnathostomes because they share perichondral bone enveloping the braincase, a large median head vein, and true bone with cell sites. The head shields of osteostracan fishes indicates that the first gill arches were well forward of the eyes, immediately above the mouth. Osteostracans presumably had cartilaginous gill arch supports, as shown by impressions of gill structures on the visceral surface of their shield. Primitive agnathans, like heterostracans, lacked pectoral and pelvic fins. These are found in anaspids as extensions of the fin-folds, but only osteostracans had muscular pectoral fins which were internally attached to a cartilaginous ossification, the scapulocoracoid. A recent observation that osteostracans, acanthodians, and possibly chondrichthyans all had similar inner ear physiology with open endolymphatic ducts further reinforces a close phylogenetic relationship between osteostracans and gnathotomes.

The Appearance of Jaws

Primitive jawless fishes had many more paired gill arches than the jawed fishes; some like the anaspid *Legendrelepis*, possessed twenty or more paired gill pouches. Jaws may have first originated in fishes by modification of the front gill arch support bones. The first primitive set of jaws and teeth could have formed when dermal scales, with a tooth-like structure and shape, invaded the pharynx and developed as mucous membrane denticles. Some agnathans had tooth-like scales lining the buccal cavity (e.g., thelodonts) which could have acted as ‘prototeeth’ in the reduction of food. Furthermore, the dermal scales of most jawless fishes are constructed of dentinous tissues underlying the enameloid crown, with a bony base. Thelodonts even have a pulp cavity in the base of the scales, making these scales resemble ‘primitive teeth’ in terms of their histology.

The appearance of jaws should naturally herald the abundant appearance of teeth in the fossil record at the same time. Gnathostomes (jawed vertebrates) include the first fishes with jaws and teeth: chondrichthyans, acanthodians, placoderms, and osteichthyans, and their evolutionary descendants, the tetrapods (amphibians, reptiles, birds, mammals). Sharks, for example, have hundreds of teeth in their mouths which they grow and shed continuously through life. The average modern shark may shed up to 20 000 teeth into the sediment, although Palaeozoic sharks may not necessarily have shed as many teeth as modern forms as they often exhibit fewer tooth rows. Shark skin contains thousands of tiny placoid scales which are also shed into the sediment after death, and readily identify the fossils.

There has been much recent debate over the appearance of true teeth in certain placoderms, mainly as some workers see ‘teeth’ as having arisen from patterns of dermal denticles whereas others think that placoderm teeth arose from dental lamina, and are not related to the patterning of dermal denticles seen on the plates.

Early Sharks and Their Relatives

The oldest shark-like scales come from the Late Ordovician Harding Sandstone of North America and the Early Silurian of Mongolia, yet the oldest shark teeth are of lowest Devonian age, some 30 million years later. As the numbers of teeth and scales are both high per individual shark, the absence of shark teeth in Silurian strata would seem to be a real observation, suggesting that the first sharks had scales of recognizable morphology, akin to those of modern sharks, yet they probably lacked teeth. The appearance of the first teeth most likely correlates with the evolution of jaws in sharks.

The most complete early shark is *Doliodus problematicus*, from the Early Emsian (*ca*. 400 mya) of New Brunswick, Canada. It shows tooth families *in situ*, a typical early shark braincase and the pectoral fins have paired spines preceding them, a feature previously only seen in acanthodians, placoderms, and the basal osteichthyan *Psarolepis*. This implies that paired pectoral spines may be a primitive character of all jawed fishes.

Sharks possibly first arose by the Early Devonian in Laurentia (*Doliodus*), and could have spread via

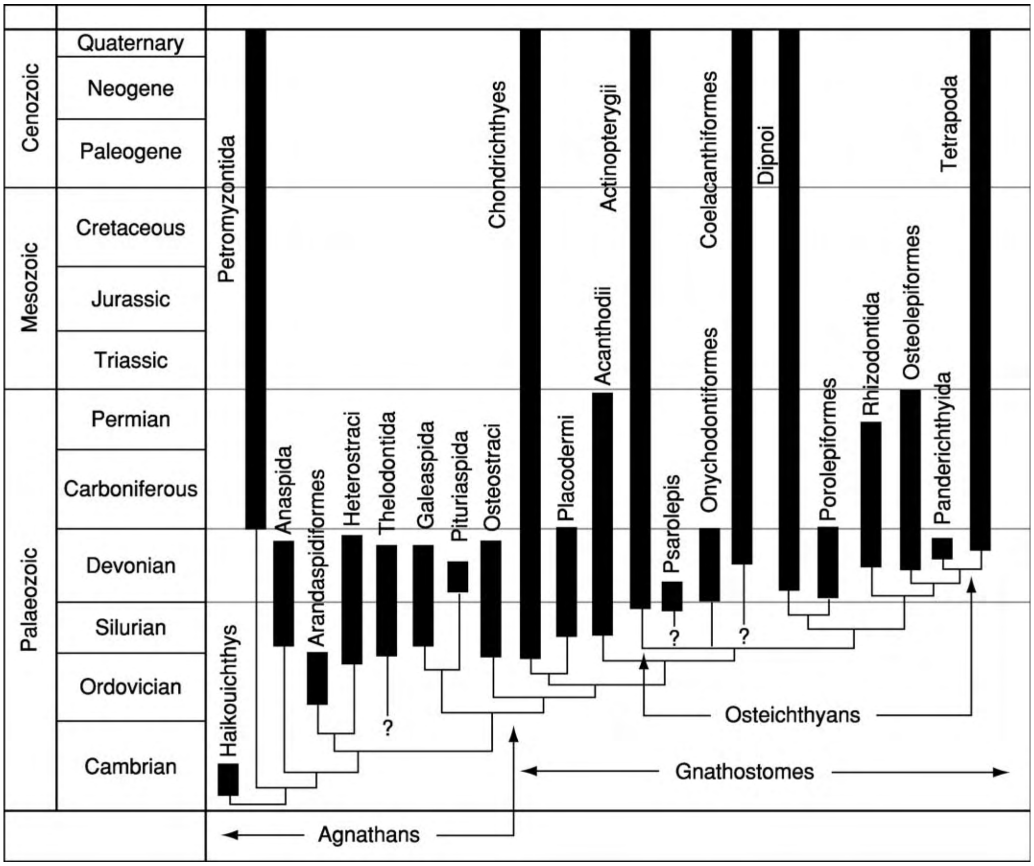


Figure 1 Major radiations of fish groups through time.

shallow seaways into northern Gondwana, as their remains are commonly found in late Early Devonian and Middle Devonian deposits in Spain, Australia, Antarctica, South Africa, and South America. The Middle Devonian Aztec Siltstone of Antarctica has an unusually high diversity of sharks within one stratigraphic unit in both species' numbers and the size of teeth, suggesting that Gondwana may well have been the place where toothed sharks underwent their first major radiation. By the Middle Devonian, sharks teeth are commonly found throughout the world, and by the Late Devonian more than forty species are known. The first of the modern sharks, or neoselachians, may date back as far as the Middle Devonian, represented by teeth of *Mcmurdodus* which have multilayered enameloid. *Hamiltonichthys*, from the Lower Carboniferous, was the oldest of the hybodontid shark lineage, a group prevalent in the Mesozoic Era. In the Lower Carboniferous some very large sharks had evolved, such as *Edestus giganteus*, from North America, which possessed serrated teeth up to 7–8 cm high on a continuous whorl. Such mega-predators would have reached 6 metres or more in length. Another major chondrichthyan group, the holocephalomorphs, represented today by the chimaerids and rabbitfishes, appeared in the Early

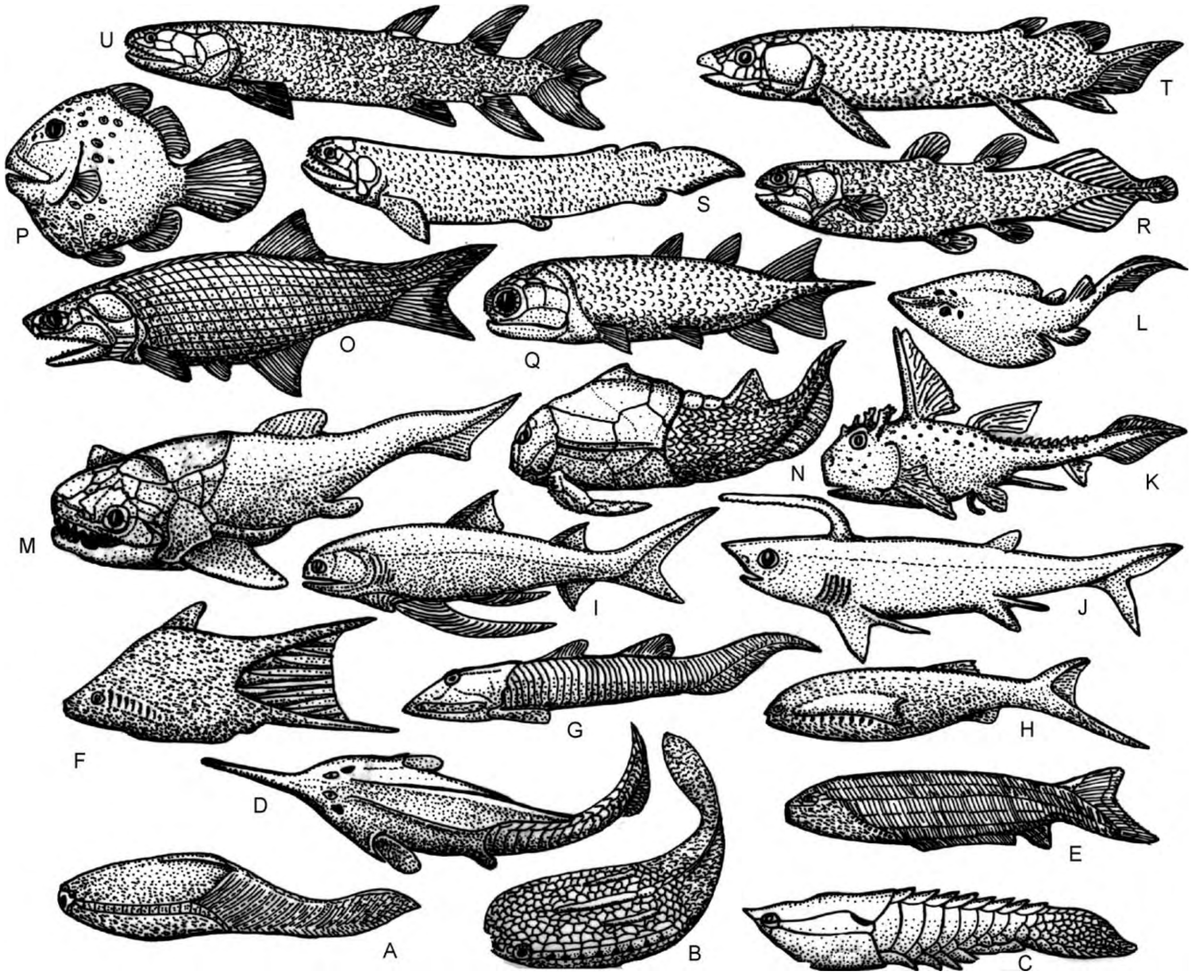


Figure 2 Diversity of major fish groups through time. (A H, agnathans; I U, gnathostomes). A, *Arandaspis* (Arandaspidiformes; Ordovician). B, *Eriptychius*, (Astraspidiformes; Ordovician). C, *Anglaspis* (Heterostraci; Silurian). D, *Pituriaspis* (representing Pituriaspida and Galeaspida; Devonian). E, *Birkenia* (Anaspida; Silurian). F, *Furcacauda* (Furcacaudiformes; Devonian). G, *Ateleaspis* (Osteostraci; Silurian). H, *Thelodus* (Thelodontida; Silurian). I, *Gyracanthides* (Acanthodii, Climatiiformes; Devonian Carboniferous). J, *Falcatus* (Chondrichthyes, Selachii; Carboniferous). K, *Echinochimaera* (Chondrichthyes, Holocephalomorphi; Carboniferous). L, *Rhinobatis* (Chondrichthyes, Selachii, Batoidea; Jurassic Recent). M, *Mcnamaraspis* (Placodermi, Arthrodira; Devonian). N, *Sherbonaspis* (Placodermi, Antiarchi; Devonian). O, *Mimia* (Osteichthyes, basal Actinopterygii; Devonian). P, *Eoplectus* (Osteichthyes, Actinopterygii, Teleostei; Eocene). Q, *Strunius* (Osteichthyes, Onychodontiformes; Devonian). R, *Hadronector* (Osteichthyes, Coelacanthiformes; Carboniferous). S, *Strepsodus* (Osteichthyes, Rhizodontiformes; Carboniferous). T, *Dipterus* (Osteichthyes, Dipnoi; Devonian). U, *Eusthenopteron* (Osteichthyes, Osteolepiformes; Devonian).

Carboniferous (Figure 2). Some sharks of this time, like the petalodonts, had unusual crushing or nipping teeth, possessing few sets in the jaws.

Sharks radiated and diversified throughout the Late Palaeozoic and Mesozoic. The largest Palaeozoic predatory sharks may have been the edestids, such as *Helicoprion*, whose coiled tooth whorls occur in the Middle Permian. If these large whorls sat on the tip of the lower jaws, estimated sizes of 10 metres would not be unlikely for these sharks. Huge lamnid sharks evolved in the mid-Tertiary, such as *Carcharocles megalodon* with teeth 18 cm high, suggesting a total body length of about 15 metres.

The rays originated from benthic sharks back in the Late Jurassic and by the end of the Mesozoic were a widespread and diverse group.

Acanthodians

The oldest fishes with jaws and teeth preserved intact are the acanthodians, which date back to the Early Silurian. They are often represented in microscopic residues from dissolved limestone as isolated teeth, scales, and fin spines dating back to the Early Silurian. Acanthodians have been likened to 'spiny sharks' in past literature although their affinities appear to lie closer to the higher jawed fishes, such as osteichthyans (true bony fishes) because of scale structure and gross morphology of their braincase. The oldest acanthodians include forms with distinct gnathal bones on to which strong teeth are ankylozed. This group, the ischnacanthids, were moderately large predators in the Early Devonian seas; like *Xylacanthus grandis* from Spitsbergen, reached sizes of 2–3 metres. The climatiforms (e.g., *Climatius Gyracanthides* (Figure 2)) had elaborate dermal shoulder girdle armour and some had many additional spines along the ventrolateral ridge of the body. Acanthodiforms, such as *Acanthodes*, were filter-feeding forms which lacked shoulder girdle armour and had extensive gill rakers for sifting food, plus possessed otoliths (statoconia, ear stones) presumably for refining balance and orientation during swimming manoeuvres. The group survived until the end of the Permian Period.

Placoderms

The most successful of all the Middle Palaeozoic jawed fishes were undoubtedly the placoderms, which appeared in the Early Silurian and reached a peak of diversity during the Middle–Late Devonian. Their name means 'plated skin', referring to their characteristic mosaic of overlapping dermal plates that enveloped the head and trunk region (Figure 3). They possessed shark-like heterocercal tails and in their overall anatomy were very much akin to chondrichthyans. Placoderms have alternatively been placed as possible relatives of the bony fishes (osteichthyans) because of their dermal bones and internally ossified braincases. However, the teeth which developed on placoderm jaw bones were not real teeth with roots set into a discrete jaw bone, like those of the osteichthyan fishes, but well-developed pointed cusps that protruded from the jaw bone itself, with semidentine present. Placoderms exhibit a multitude of dentition types, revolving around rows of pointed cusps on gnathal bones, or clusters of small denticles or pointed cusps on sheets of bone. As discussed above, there is ongoing debate as to whether

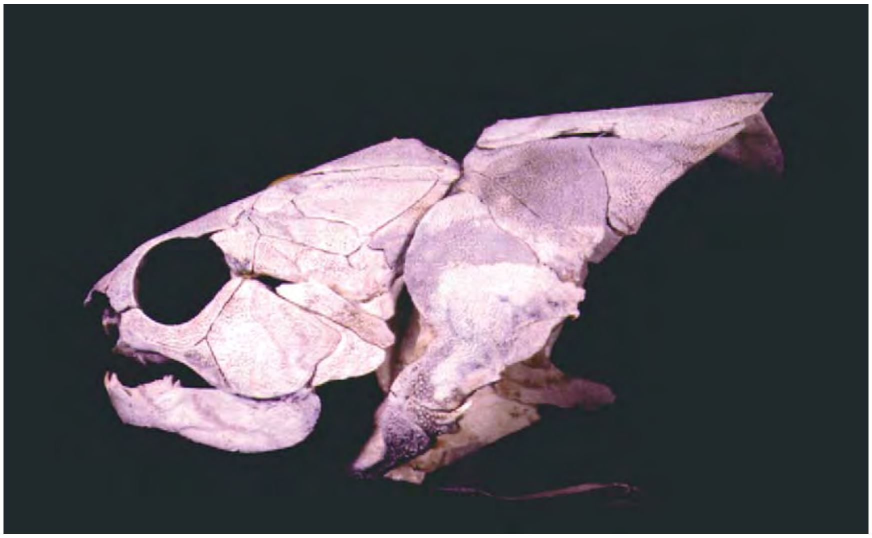


Figure 3 The armour of a Late Devonian placoderm fish *Mcnamaraspis kaprios*, from the Gogo Formation of Western Australia.

placoderm have true teeth or whether they have 'teeth' derived from dermal denticles.

The largest group of placoderms were the arthrodires, which includes the biggest forms, the dinichthyids (e.g., *Gorgonosteus*) reaching 6 metres or more. The small antiarchs had external bone-covered props for pectoral fins (e.g., sherbonaspis, Figure 2). One of these, *Bothriolepis*, was ubiquitous in the Middle–Late Devonian, and is represented by over one hundred species worldwide. Other groups include the flattened ray-like rhenanids and their heavily armoured relatives, the acanthothoracids, the petalichthyids, and the flattened phyllolepids. Placoderms can often be identified from isolated remains because of their unusual plate shapes and dermal ornamentation, making them useful in Devonian biostratigraphic and biogeographic studies. Placoderms inhabited both marine and freshwater environments, with some genera found in both fluvial and marine deposits (e.g., *Bothriolepis, Remigolepis, Groenlandaspis*). The placoderms became extinct at the end of the Devonian Period, possibly being displaced by the steadily increasing radiation of chondrichthyans.

Osteichthyans

The osteichthyans (true bony fishes) appeared by the Late Silurian, represented by isolated scales, teeth, and rare isolated bones of small ray-finned and lobe-finned fishes (Figure 2). The Osteichthyes comprise the ray-fins (Actinopterygii) which are more than 99% of all living fish species and are represented by over 23 000 species (e.g., trout, salmon, goldfish); and the Sarcopterygii, which have muscular, fleshy lobed fins and include the lungfishes (Dipnoi), and what was commonly grouped as the 'crossopterygians' (now recognized as a paraphyletic mixture of groups). In palaeontological terms, this group is better refered to by Order, and includes mostly extinct groups apart from one surviving species, the coelacanth *Latimeria chalumnae* (within the Order Actinistia or Coelacanthiformes). The many extinct groups of non-dipnoan sarcopterygians that diversified during the Middle Palaeozoic include the orders Onychodontiformes, Porolepiformes, Rhizodontiformes and Osteolepiformes.

Actinopterygians

The ray-finned fishes were represented in the Devonian by primitive palaeoniscoids such as *Cheirolepis* and *Moythomasia*. The group radiated rapidly by the beginning of the Carboniferous, and over fifty families are known during the Late Palaeozoic, typified by forms such as *Palaeoniscus, Elonichthys*, and *Gonatodus*. Most of the heavily scaled primitive forms gave way to those having thinner round scales, with less rigid cheek and jaw bone arrangements; this enabled more efficient feeding and respiratory mechanisms, such as the buccal-pump system, to evolve. This line of evolution saw the rise of holostean

Figure 4 *Leptolepis koonwarri*, an Early Cretaceous teleost fish from Australia.

and subholostean fishes in the Late Palaeozoic (today represented by gars and bowfins, e.g., *Amia, Lepisosteus*) and the first of the modern ray-fins, the teleosteans, by the Late Triassic. Teleosteans have advanced tail fin skeletons with many supporting bones, plus mobile cheeks and detached jaw bones. Their first gill arch (hyoid), which supports the jaws, is often vertically or forwardly inclined. Many of the modern families of fishes had appeared by the close of the Cretaceous Period (Figure 4), and the first teleosteans adapted specifically to life on reefs had appeared by the Eocene Period, represented by species from Monte Bolca, Italy (e.g., Eoplectus, Figure 2).

Sarcopterygians

In the Devonian Period, the most diverse group of bony fishes were the sarcopterygians (over one hundred species); the ray-fins were only a minor component of fish faunas (about ten species). The earliest sarcopterygians, such as *Psarolepis* from China had a shiny tissue, cosmine, that covered the dermal bones and scales and was lost in more advanced lineages. Cosmine may have functioned as an electrosensory tissue.

One of the basal groups of sarcopterygians is the dagger-toothed fishes or Onychodontiformes. These are a poorly known group whose best representative is *Onychodus* from the Late Devonian Gogo Formation of Western Australia. *Onychodus* is known from isolated bones and tooth whorls in the Middle–Late Devonian worldwide, and was the largest of all Devonian osteichthyans, *O. sigmoides*, reaching estimated lengths of 4 metres.

The ability to gulp air and transgress environmental boundaries evolved within lungfishes during the Middle Devonian (e.g., *Dipterus*), becoming prominent in the Late Devonian. This gave them an edge on other groups, with later forms developing a further enhanced survival strategy, the ability to aestivate (e.g., *Gnathorhiza*). Lungfishes diversified into forms with crushing tooth plates (e.g., *Chirodipterus*) and others with denticulated palates (e.g., *Griphognathus*). Post-Devonian lungfishes were largely freshwater forms. Today there are three living lungfish genera, the most primitive of which is *Neoceratodus forsteri* from Australia, which dates back to the Cretaceous Period.

Some of the advanced osteolepiform fishes appear to have developed the ability to breathe air independently, as they had evolved a palatal nostril (a choana) by the Middle Devonian (e.g., *Gogonasus*). Osteolepidids were a common component of most Devonian fish faunas, in many cases taking the role of top predator. More advanced osteolepiforms, like *Eusthenopteron* (Figure 2), from the Late Devonian of Canada, lost the cosmine layer and had rounded scales. The largest of these predatory lobe-fins were the rhizodontiforms, which dominated the lakes and rivers of the Carboniferous Period. Forms like *Rhizodus* may have reached 6–7 metres in length.

The panderichthyid fishes are morphologically closest to the elginerpetonid amphibians, known from Scotland and Russia, in having identical skull-roof patterns, pectoral and pelvic girdles, and endoskeletons which approach the morphology of the basic tetrapod limbs more closely than any other fish. A primitive rhizodontid pectoral skeleton (*Sauripterus*), with eight fin radials supporting the pectoral fin, has been described from North America. This matches the eight digits occurring in the limbs of the Devonian amphibian *Acanthostega*. The first tetrapods evolved by the end of the Frasnian stage of the Late Devonian, and at least nine genera had appeared by the close of the Devonian Period.

See Also

Fossil Vertebrates: Jawless Fish-Like Vertebrates.

Further Reading

Campbell KSW and Barwick RE (1987) Palaeozoic lung fishes a review. *Journal of Morphology, Supplement* 1: 93 132.

Forey PL and Janvier P (1993) Agnathans and the origin of jawed vertebrates. *Nature* 361: 129 134.

Jarvik E (1980) *Basic structure and evolution of verte brates*. New York: Academic Press.

Long JA (1993) Cranial ribs and the origin of air breathing in lungfishes. *Memoirs of the Australisian Association of Palaeontologists* 15: 199 210.

Long JA (1995) *The rise of fishes 500 million years of evolution*. University of New South Wales: Sydney, Johns Hopkins University Press: Baltimore.

Long JA, Campbell KSW, and Barwick RE (1997) Oste ology and functional morphology of the osteolepiform fish *Gogonasus andrewsae* Long, 1985, from the Upper Devonian Gogo Formation, Western Australia. *Records of the Western Australian Museum, Supplement* 53: 1 89.

Miller RF, Cloutier R, and Turner S (2003) The oldest articulated chondrichthyan from the Early Devonian Period. *Nature* 425: 501 504.

Sahney S and Wilson MVH (2001) Extrinsic labyrinth infillings imply open endolymphatic ducts in Lower Dev onian osteostracans, acanthodians, and putative chon drichthyans. *Journal of Vertebrate Palaeontology* 21: 660 669.

Shu D G (2003) A palaeontological perspective of vertebrate origin. *Chinese Science Bulletin 2003,* 48: 725 735.

Shu D G, Conway Morris S, Han J, *et al.* (2003) Head and backbone of the Early Cambrian vertebrate *Haikouichthys*. *Nature* 421: 526 529.

Vorobyeva E and Schultze H P (1996) Description and systematics of panderichthyid fishes with comments on their relationship to tetrapods. In: Schultze H P and Trueb L (eds.) *Origins of the higher groups of tetrapods: controversy and consensus,* pp. 68 109. Cornell: Cornell University Press.

Wilson MVH and Caldwell MW (1998) The Furcacaudiformes: a new order of jawless vertebrates with thelodont scales, based on articulated Silurian and Devonian fossils from northern Canada. *Journal of Vertebrate Palaeontology* 18: 10 29.

Young GC (1997) Ordovician microvertebrate remains from the Amadeus Basin, central Australia. *Journal of Vertebrate Paleontology* 17: 1 25.

Young GC, Karatajute Talimaa VN, and Smith MM (1996) A possible Late Cambrian vertebrate from Australia. *Nature* 383: 810 812.

Zhu M, Yu X B, and Janvier P (1999) A primitive fossil fish sheds light on the origin of bony fishes. *Nature* 397: 607 610.

Palaeozoic Non-Amniote Tetrapods

J A Clack, University of Cambridge, Cambridge, UK

Introduction

The Late Palaeozoic saw the emergence of land vertebrates from their aquatic ancestors. Vertebrates with legs and digits are usually referred to as 'tetrapods', meaning 'four-legged'. In fact, the term not only refers to ancient extinct forms, but also embraces all modern forms which either possess legs (or arms, wings, or flippers) or whose ancestors possessed them (such as snakes and others that have lost them). This great 'family tree' of animals has its roots in the Late Palaeozoic, specifically the Late Devonian, when the first legged vertebrates evolved. Because of the nature of evolution, it is, in practice, sometimes difficult to draw the distinction between a 'fish' and a 'tetrapod', especially when there are no limbs or fins preserved in the fossils we have; therefore, to understand the emergence of tetrapods, it is also necessary to look at the 'fish' relatives from whom the 'tetrapods' evolved. This article therefore starts by describing a few 'tetrapod-like fish' and their relationships to the 'fish-like tetrapods' that evolved from them, before going on to see what kinds of animals emerged from the swamps and began their colonization of the land during the Carboniferous and Permian periods.

It was during the Late Palaeozoic that the foundations were laid for fully terrestrial living vertebrates, when they adapted their skeletons and physiologies for a very different environment from that in which they first evolved. Breathing mechanisms, feeding mechanisms and strategies, sensory systems, reproductive modes, and locomotory techniques all had to be modified during this greatest of evolutionary transitions, and many of these changes are reflected in skeletal changes that can be picked up in the fossil record. The differing means by which this was achieved is also reflected today in the legacy of modern tetrapod anatomy and physiology.

The general reader may be aware that modern tetrapods fall into a number of major groups, namely the amphibians, reptiles, birds, and mammals, but in fact these can be clumped into two coherent but biologically divergent groups, with the amphibians (known as 'lissamphibians') forming one lineage and the rest forming the other, known as the amniotes. Amphibians and amniotes each trace their origins to the Late Palaeozoic and, within the amniotes, each of the subgroups can also be traced back to a Carboniferous or Permian origin. By contrast, one of the least understood episodes of tetrapod evolution is the origin of modern lissamphibian groups (of which there are only three: frogs, salamanders, and caecilians). None of these can be found in the fossil record before the Early Jurassic, and the earliest stem member is *Triadobatrachus*, an animal with both frog- and salamander-like features from the Late Triassic. To be called an amphibian in the strict sense, an animal would be more closely related phylogenetically to these modern forms than to any amniote group.

Amongst Palaeozoic forms were also many lineages that have left no modern descendants, and whose origins do not seem to fall into either the amphibian or the amniote lineages, and these simply have to be called 'early tetrapods'. (They may have had an 'amphibious' life style, laying eggs in water and emerging onto land only as adults, but this describes their mode of life, not their relationships, and in many cases there is no evidence of what their reproductive mode might have been.) This article looks mainly at these early tetrapod groups and the amphibian lineage; although the earliest amniotes were first found in the Late Carboniferous, they largely fall outside the scope of this article, but are dealt with in more detail in **Fossil Vertebrates:** Reptiles Other Than Dinosaurs.

'Tetrapod-Like Fish'

Tetrapods and their closest fish relatives belong to the 'lobe-finned vertebrates' or sarcopterygians. Other modern sarcopterygians are the lungfishes (dipnoans) and the coelacanths (actinistians), but during the Devonian there were many representatives of a lineage now known as the tetrapodomorphs (various subsets of these have been referred to as 'crossopterygians' or 'rhipidistians', but these terms can be misleading and are not now in general use). The best known tetrapodomorph is a fish called *Eusthenopteron*, mainly from the Frasnian of Canada and the Baltic states. It used to be pictured as the 'ancestor' of tetrapods, because it shares some fundamental features with tetrapods, including the form of its nose and the shape of some of the bones in its pectoral fin skeleton (Figure 1). More recently, other tetrapodomorph fish have been discovered which are more closely related to tetrapods than is *Eusthenopteron*, including the Frasnian genera *Panderichthys* (Figure 1) from the Baltics and Russia, and *Elpistostege* from Canada. Knowledge of their anatomy is helping to establish the timing and sequence of acquisition of some key tetrapod characters in skull, jaw, and limb construction. More recently still, tantalizing fragments, such as the lower jaw of an animal called *Livoniana* (Figure 1), have been found, hinting that, at the cusp of the origin of tetrapods, a wide range of animals existed that exploited different ecological niches. *Livoniana* sported at least seven rows of teeth on parts of its lower jaw, but also showed some subtle but key features marking it as close to tetrapods themselves. Figure 1 shows a diagram of the relationships of some of these early groups to one another, as well as the time around which they existed. The diagram implies that tetrapods originated around the Middle to Late Devonian boundary.

The Earliest Tetrapods

The earliest known limb or girdle elements belong to a tetrapod called *Elginerpeton* from the Frasnian of Scotland, associated with lower jaw and skull fragments. These establish a minimum date for the appearance of tetrapods. After that, during the Famennian, the fossil record expands, so that, at present, at least seven genera of tetrapods are recognized from that epoch. Three of them are known from articulated limb and digit material: *Acanthostega* (Figure 2), *Ichthyostega* (Figure 2), and *Tulerpeton*, the first two from East Greenland and the third from Russia. *Ichthyostega* was first discovered in 1929, and numerous skull and postcranial skeletons exist. Some parts of the anatomy have been described in detail, while other parts have only been sketchily portrayed, and the formal systematic nomenclature has been problematic until recently. Present studies show that there are two main species, associated with two distinct geological formations, an example of micro-evolution that is rarely possible to establish in early tetrapod material because of its rarity. *Ichthyostega* is, in many ways, an unlikely early tetrapod because some of its anatomy is highly modified and unlike that of any other tetrapod, ancient or modern. It has very broad overlapping ribs with a relatively huge shoulder girdle. Its recurved, keeled teeth are those of an undoubted predator, in a solidly constructed, massive skull. Its ear region and braincase suggest that it was specialized for underwater hearing, and its finned tail and paddle-like hind limbs support the notion of a largely aquatic animal.

Acanthostega, likewise, was an aquatic animal, but in a much more committed fashion. Its finned tail was deeper and longer than that of *Ichthyostega*, and both limbs appear to have been paddles rather than walking legs. Its ribs were short, and its shoulder girdle displayed some primitive fish-like features. Furthermore, grooved gill bars imply that functional internal gills were still present. Its generally primitive appearance suggests that it was primarily, rather than secondarily, aquatic, and that it gives us a good view of what a very primitive early tetrapod was like. *Acanthostega* is known from many specimens, although most of the best derive from a single small exposure that may represent a point bar against which a number of carcasses were swept in a flood.

Tulerpeton is known mainly from a single specimen that shows a complete fore and hind limb (Figure 3), some shoulder girdle elements, and part of a snout. Other fragments from the same locality may belong to it. The sediments from which it derives appear to have been marginal marine in origin, possibly a shallow, warm lagoon, but probably quite far from the nearest landmass. However, its limbs appear more like those of a terrestrial animal than those of the Greenland forms.

These three animals have something significant in common: they all possessed more than five digits on the limbs that are preserved (Figure 3). In *Ichthyostega*, the hind limb bore seven digits in a remarkable

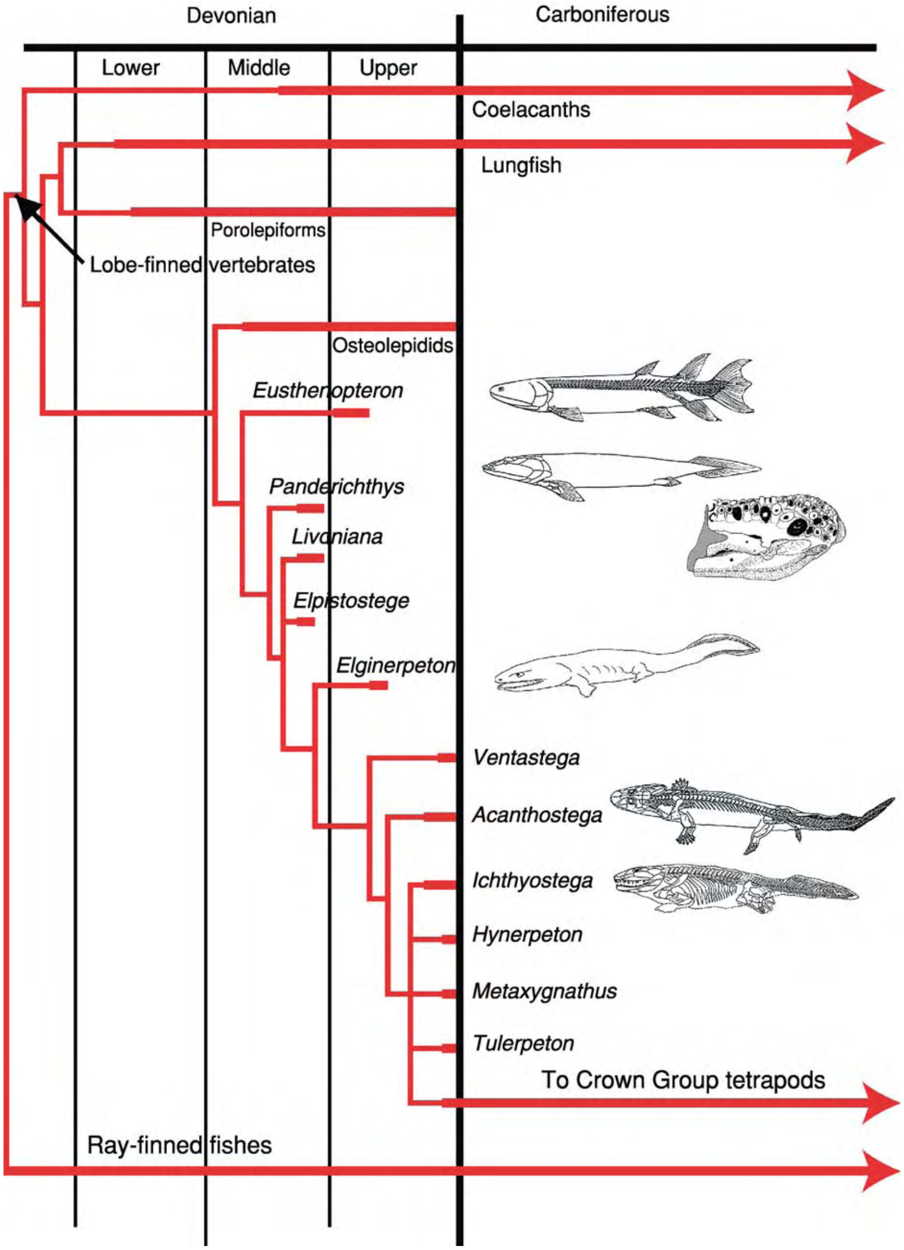


Figure 1 Cladogram of lobe finned fish and tetrapodomorphs, set on a time scale showing their occurrence. Note the loss of midline fins from *Panderichthys* onwards, and the retention of the tail finning in *Acanthostega* and *Ichthyostega*. *Livoniana* is known only from two jaw fragments, of which one is shown here. *Panderichthys* and *Ichthyostega* grew to over a metre in length, while *Eusthenopteron* and *Acanthostega* were slightly smaller, and *Elginerpeton* probably somewhat larger. The animals are not shown to scale. Arrows show that the group or its descendants survived to the present day.

pattern – at the front edge, two small followed by one extremely small digit formed a leading support for the paddle, with four stout digits forming the rest of the paddle (Figure 3). There are at least three specimens which show this. To be preserved in this way suggests that they were held in strong tissue in life. In *Acanthostega*, both the fore and hind limbs had at least eight digits, although there is only one specimen of each (Figure 3). In *Tulerpeton*, certainly the fore limb and probably the hind limb each had six digits. These discoveries have stimulated a rethink about how limbs with digits evolved and have provoked interchange with ideas coming from developmental biology and genetics. They also suggest that limbs with digits arose initially for swimming or paddling in water, not for walking on land.

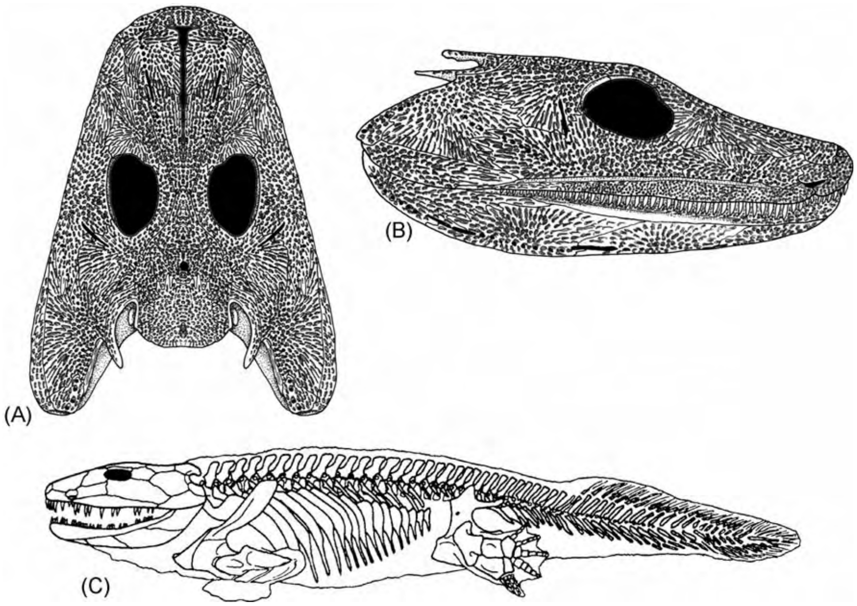


Figure 2 (A, B) Skull roof of *Acanthostega* in dorsal (A) and right lateral (B) views. The skull grew to about 200 mm in length. (C) Reconstruction of the skeleton of *Ichthyostega* based on recent information. However, study in progress indicates that further revisions will be necessary.

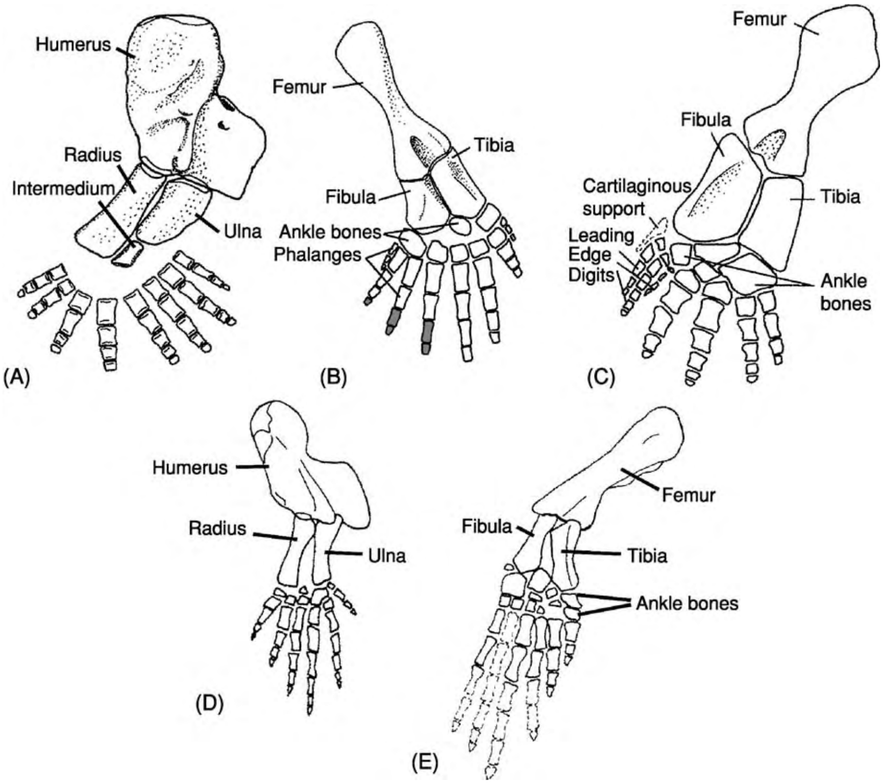


Figure 3 Limbs of Devonian tetrapods. (A) Fore limb of *Acanthostega*. (B) Hind limb of *Acanthostega*. (C) Hind limb of *Ichthyostega* (the manus is unknown in this form). (D) Fore limb of *Tulerpeton*. (E) Hind limb of *Tulerpeton*. Not to scale. Grey shading in *Acanthostega*, reconstructed elements. Broken lines in *Tulerpeton*, reconstructed elements. Adapted with permission from Clack JA (2002) *Gaining Ground: the Origin and Evolution of Tetrapods*. Bloomington, IN: Indiana University Press.

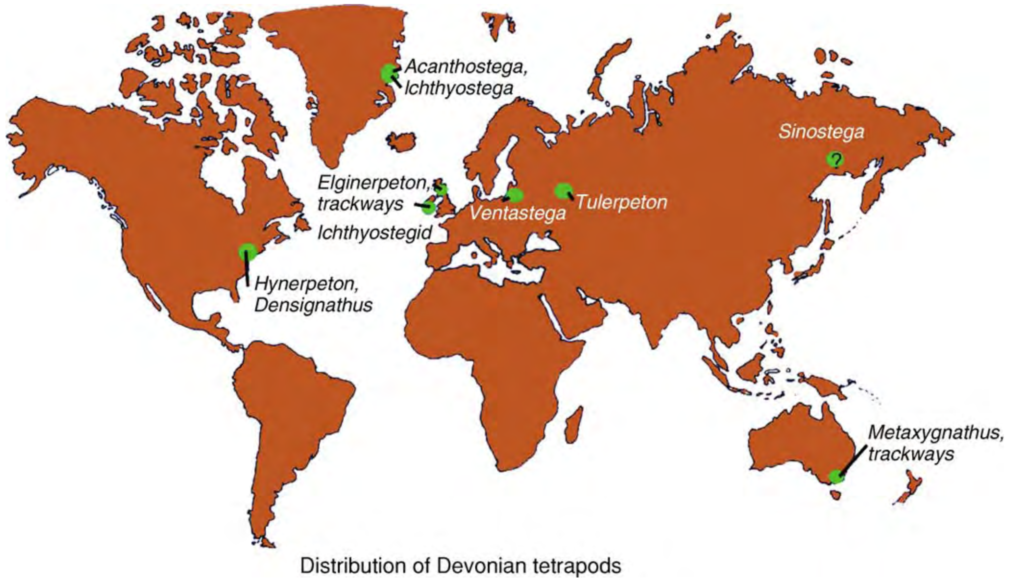


Figure 4 Map of the world showing the distribution of Devonian tetrapods worldwide.

Other Devonian tetrapods discovered over the last 10 years or so include *Hynerpeton*, known from a lower jaw and pectoral girdle material, and *Densignathus*, known from a lower jaw, both from Pennsylvania, USA. *Ventastega,* represented by excellent skull and pectoral girdle material, is from Latvia and, most recently, new taxa have been found in Greenland (as yet un-named), China (*Sinostega*) and Belgium (un-named). *Metaxygnathus*, from New South Wales, Australia, has been reconfirmed as a tetrapod in the light of work on *Acanthostega*. *Ventastega* is the best known of these and, interestingly, although it shows many similarities to *Acanthostega*, it also shows some more primitive and some unique features. Figure 4 shows the distribution of Devonian tetrapod finds worldwide.

Tetrapods of the Early Carboniferous

After the end of the Devonian, there is a long hiatus in the fossil record of tetrapods, informally known as 'Romer's Gap' (Figure 5). There may have been a mass extinction that correlates with the end-Devonian, but, whether sudden or gradual, there appears to have been a climate change affecting plants and animals alike for the next 15–20 million years. Unfortunately, it was during this period that tetrapods became fully terrestrial, but evidence to document this crucial period in the history of life on Earth is extremely scarce. Only a few specimens represent the Tournaisian and early Viséan epochs in tetrapod history, whereas the record of the late Viséan and Namurian is now much fuller than it was even two decades ago.

Pederpes finneyae is the only articulated tetrapod specimen currently known from Romer's Gap (Figures 5 and 6). It was found near Dumbarton in Scotland, in marginal marine cementstone sequences that have not otherwise yielded many fossils. It shows the earliest evidence of a hind limb that was used for walking and that was functionally pentadactyl, although what little is known of its fore limb suggests that it might have had relict accessory toes like those of the Devonian forms. Its closest relative, called *Whatcheeria deltae*, is from the Viséan of Iowa, USA. It is also possibly related to a mid-Viséan form, *Ossinodus*, from Queensland, Australia, and another possibly Tournaisian or Viséan form from Northern Ireland. They each show some primitive features, but share some more derived ones with later Carboniferous forms, and appear to form a clade (or possibly a grade) of Early Carboniferous forms that were widely distributed in time and space. Apart from these, a few isolated limb and girdle elements, and some large footprints from Horton Bluff in Nova Scotia, are all that represent the Tournaisian and early parts of the Viséan.

The tetrapods that have been discovered in the period following Romer's Gap show a range of sizes and body forms greater than that seen in almost any vertebrate group before (Figures 5 and 6). Certainly the tetrapods of the Late Devonian are quite conservative in shape and size – all are around a metre long with flattened and approximately spade-shaped heads. Those of the mid-Carboniferous range from mouse-sized to several metres in length with evidence of diversity in locomotory and feeding adaptations.

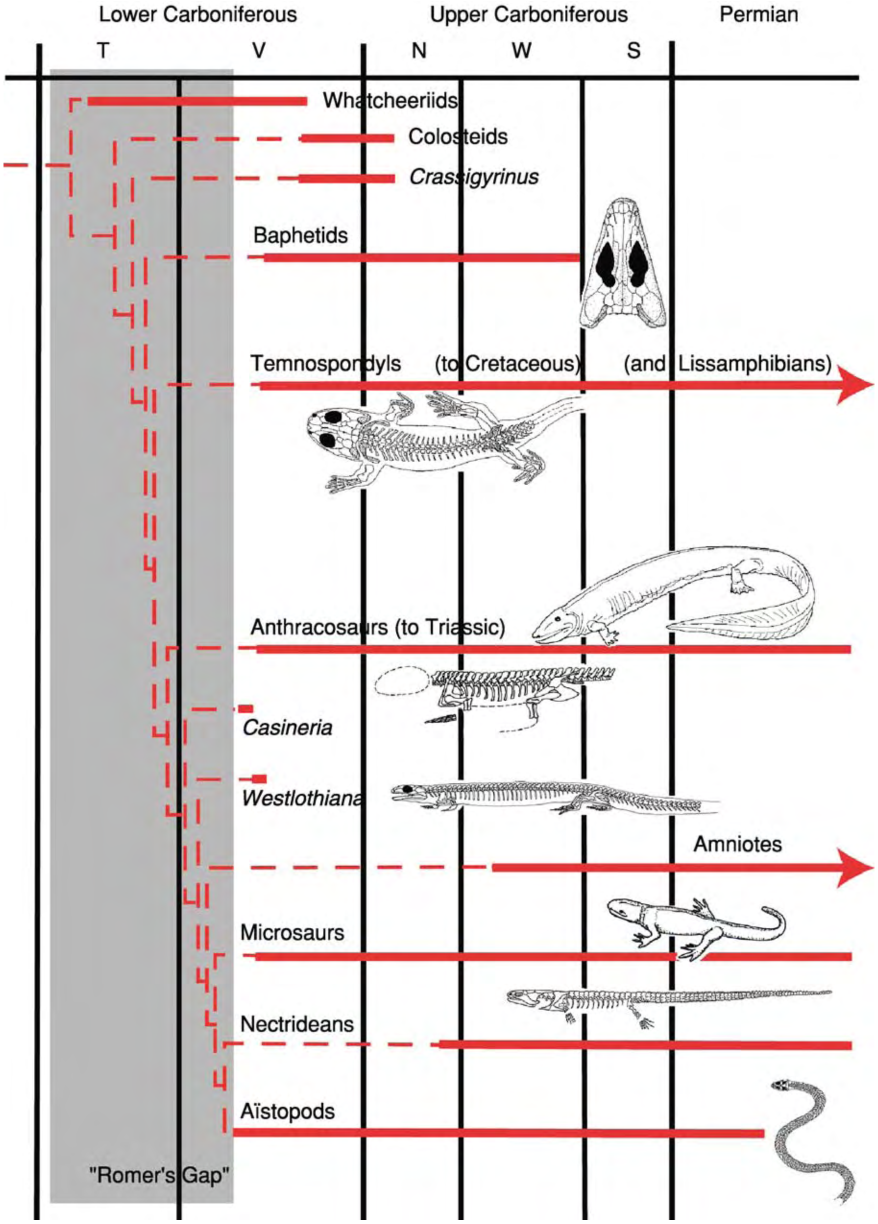


Figure 5 Diagram to show the distribution in time of some Late Palaeozoic tetrapod groups, with one view of their relationships. Romer's Gap is shown as a shaded band, and this time period probably represents that during which these major groups diversified. Representative animals are shown on some of the branches, but these are not to scale. Baphetid skulls, such as *Megalocephalus* figured here, reached over 300 mm in length. *Balanerpeton*, shown representing temnospondyls, grew to a total body length of between 300 and 400 mm. The embolomere, *Pholiderpeton*, shown representing anthracosaurs, probably reached a length of over 3 m. *Casineria* was about the size of a mouse, and *Westlothiana* that of a small lizard. The microsaur, *Pelodosotis*, featured here was only about 30 mm in length, but others grew to around 300 mm. *Keraterpeton*, representing the nectrideans, reached about 300 mm in total, although later ones grew much larger (see **Figure 8**). The aıstopod, *Lethiscus*, seen here had a skull length of about 30 mm, but the full length of its body is not known. Arrows show that the group or its descendants survived to the present day. T, Tournaisian; V, Viséan; N, Namurian; W, Westphalian; S, Stephanian.

In the later Viséan, three localities, all near Edinburgh, Scotland, have yielded significant tetrapod specimens. Two, close in both time and space, have each yielded a single but intriguing specimen, while the third has yielded a whole flora and fauna of terrestrial organisms. *Casineria kiddi* (Figure 5), from Cheese Bay on the Firth of Forth, is the earliest fully terrestrial tetrapod known. It is small, unusual in such an early tetrapod, and has a clearly pentadactyl hind foot that could grasp the substrate. Despite the specimen not preserving a head, this animal appears to be the earliest representative of the amniote stem

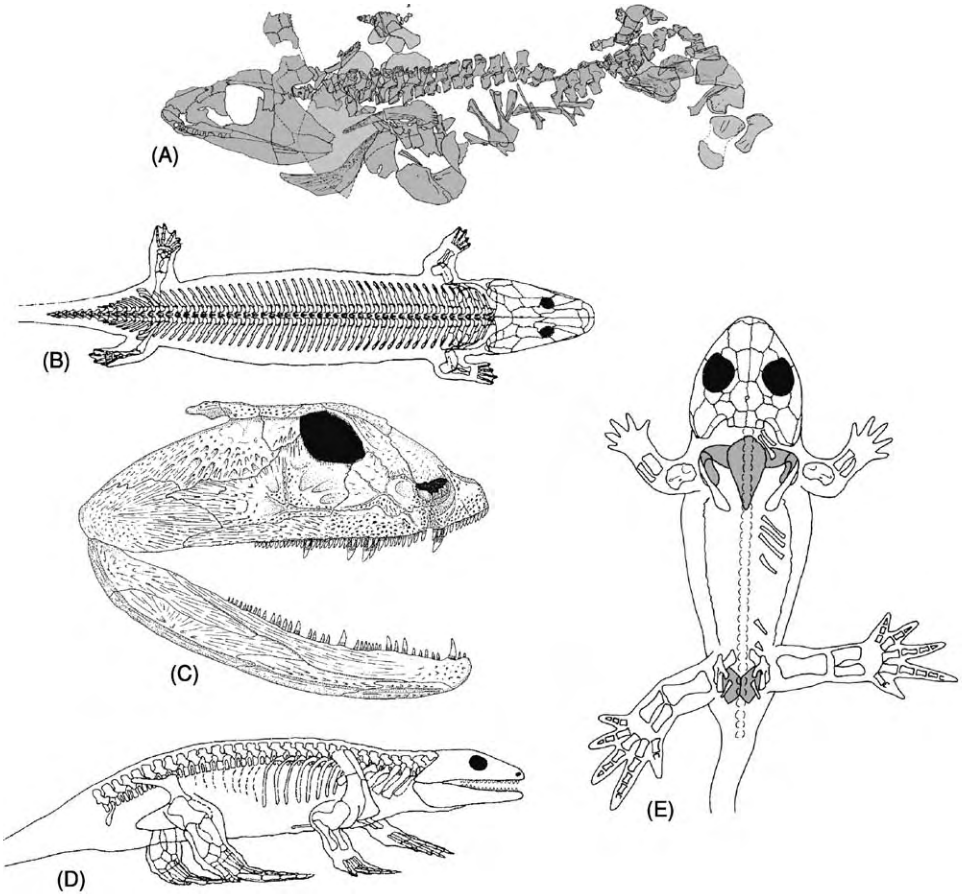


Figure 6 Early Carboniferous tetrapods. (A) Diagram of the (only) specimen of *Pederpes*, about 650 mm in length as preserved. (B) Reconstruction of the colosteid, *Greererpeton*, with a skull up to 180 mm in length. (C) Skull of *Crassigyrinus*. With its elongate body and tail, this animal probably reached between 2 and 3 m in length. (D) *Eldeceeon*, an early anthracosaur, about 300 400 mm in length. (E) *Eucritta*, probably a basal baphetid, with a maximum known skull size about 80 mm in length. (B) (E) adapted with permission from Clack JA (2002) *Gaining Ground: the Origin and Evolution of Tetrapods*. Bloomington, IN: Indiana University Press.

lineage. *Lethiscus stocki* (Figure 5), from Wardie, also on the Firth of Forth, is by contrast an aïstopod, an animal with a highly specialized skull structure and which has lost all its limbs and girdles, elongated its vertebral column and become a snake analogue. These two specimens hint at a radiation of tetrapods that took place soon after the end-Devonian, in which they began to explore and exploit a range of terrestrial niches, but of which we know virtually nothing.

East Kirkton, to the west of Edinburgh, has yielded at least six named genera of tetrapods, most of which seem to have been terrestrial. Its unusual volcanigenic geology has preserved plants, invertebrates, and tetrapods in some detail and diversity. The fauna includes the earliest representatives of at least four major groups of tetrapod that went on to radiate later in the Carboniferous, and some of which ultimately founded the lineages leading to modern amphibians and amniotes. The most common tetrapod from East Kirkton belongs to the group known as temnospondyls and is called *Balanerpeton woodi* (Figure 5). Temnospondyls, according to most analyses, are the group from which at least some modern amphibians (frogs and salamanders) arose. *Balanerpeton*, although the earliest, is apparently not the most primitive, and it shows many features characteristic of the group. *Silvanerpeton miripedes* and *Eldeceeon rolfei* (Figure 6) are the earliest members of the group known as the anthracosaurs, a group that traditionally has been seen as related to the amniotes, although the relationship may not be as close as suggested at one time. *Eucritta melanolimnetes* (Figure 6), an initially puzzling form that combines features of both temnospondyls and anthracosaurs, may be the earliest member of the baphetids, a group whose relationships to other tetrapods are still problematic. *Westlothiana lizziae* (Figure 5) is notable for having been considered the earliest stem amniote for a while, although it has recently been supplanted by *Casineria*. Its elongate body and short limbs may suggest, alternatively, that it belongs to a further group, the microsaurs, but its poor skull preservation does not yet allow the resolution of this question.

Most other late Viséan forms actually span the boundary into the early Namurian, but deserve to be treated as Early Carboniferous forms. A gap in the fossil record in the mid–late Namurian seems to represent some kind of faunal turnover, as there is only one genus that has a representative both before and after the interval. A distinctively Late Carboniferous fauna emerges during the later Namurian.

The mid-Carboniferous forms include the bizarre aquatic predator, *Crassigyrinus scoticus* (Figure 6), which features an enormous gape, huge palatal teeth, and a snout with a large midline foramen at the front, whose function is obscure. *Crassigyrinus* was an elongate animal with disproportionately small fore limbs, and was possibly an analogue of a modern moray eel. It is known from a few localities in Scotland. The colosteids form a group that had members in Scotland and the USA. They were permanently aquatic predators that retained a number of primitive features in their anatomy, and probably had persistent gillbars. Members include *Pholidogaster*, *Greererpeton* (Figure 6), and *Colosteus*. Their relationship with other Carboniferous groups is disputed, but that is true both of *Crassigyrinus* and another group that appears in the mid-Carboniferous, the baphetids (formerly known as loxommatids).

Tetrapods of the Late Carboniferous

Temnospondyls, anthracosaurs, and baphetids that originated in the mid-Carboniferous survived to radiate in the Late Carboniferous, producing large crocodile-like predatory forms characteristic of the coal swamps of the time. Baphetids are striking for their curious keyhole-shaped eye-sockets (Figure 5); the purpose of the forwardly directed extension is obscure. Very little is known about their postcranial skeletons, because, although their skulls are well known, only two have postcranial material associated with them. Anthracosaurs (Figures 5 and 7) retain many primitive characters of the skull, but these are shared with later amniote-like forms, and they also share their type of vertebral construction (called gastrocentrous) with more terrestrially adapted forms, including amniotes. By contrast, most early temnospondyls retained a primitive style of vertebral construction (called rhachitomous), associated with an aquatic life style and seen in *Acanthostega*. Their skulls, however, share features later seen in modern amphibians; large vacuities in the roof of the mouth (although covered with skin in life) might have been part of a breathing and hearing mechanism similar to that found in modern amphibians today. Temnospondyls also show other similarities to frogs and salamanders in their life histories. Several families are known to have had aquatic larvae with external gills and that metamorphosed into more terrestrial adults, and some, it is suspected, had members that remained as permanently aquatic larval-like forms, as in some modern salamanders, such as the axolotl. These larval temnospondyls are collectively known as ‘branchiosaurs’ (Figure 7), but the term embraces several different families. By contrast, some temnospondyls became much more fully terrestrial. One group, the dissorophoids, produced some of the most terrestrial of temnospondyls, while, at the same time, including forms that are apparently the most closely related to frogs (Figures 7 and 8).

The three groups described above (together with colosteids and a few others) were, at one time, lumped together as ‘labyrinthodonts’, but in fact these animals may not be particularly closely related to each other – indeed they may belong to opposing sides of the dichotomy leading to the two modern tetrapod groups. These groups filled the large-predator niche at the same time as several other groups filled the role of small aquatic or terrestrial nibblers of invertebrates. These forms, often collectively known as ‘lepospondyls’, include the microsaurs (both aquatic forms with elongate bodies and small short-bodied terrestrial forms that appear superficially amniote-like), the adelogyrinids and aïstopods (Figure 5) (elongate and almost or completely lacking limbs), and the nectrideans (united by specialized vertebrae in the tail that formed a vertically flattened and stiffened but laterally flexible and powerful organ for swimming). The skulls of all these groups are specialized and difficult to compare with each other or with non-lepospondyls. This is one reason why their relationships to each other and to non-lepospondyls are still highly controversial. Some studies have suggested that these forms are in fact closer to modern amphibians and amniotes than are temnospondyls or anthracosaurs.

Tetrapods of the Late Carboniferous are found almost universally distributed over a wide area of Central and Western Europe and eastern North America, representing the known extent of the coal forest. Although genera may differ, families produced their equivalents in widely spread localities, suggesting a uniformity of habitat and climate over many millions of years. Finds have usually been found in

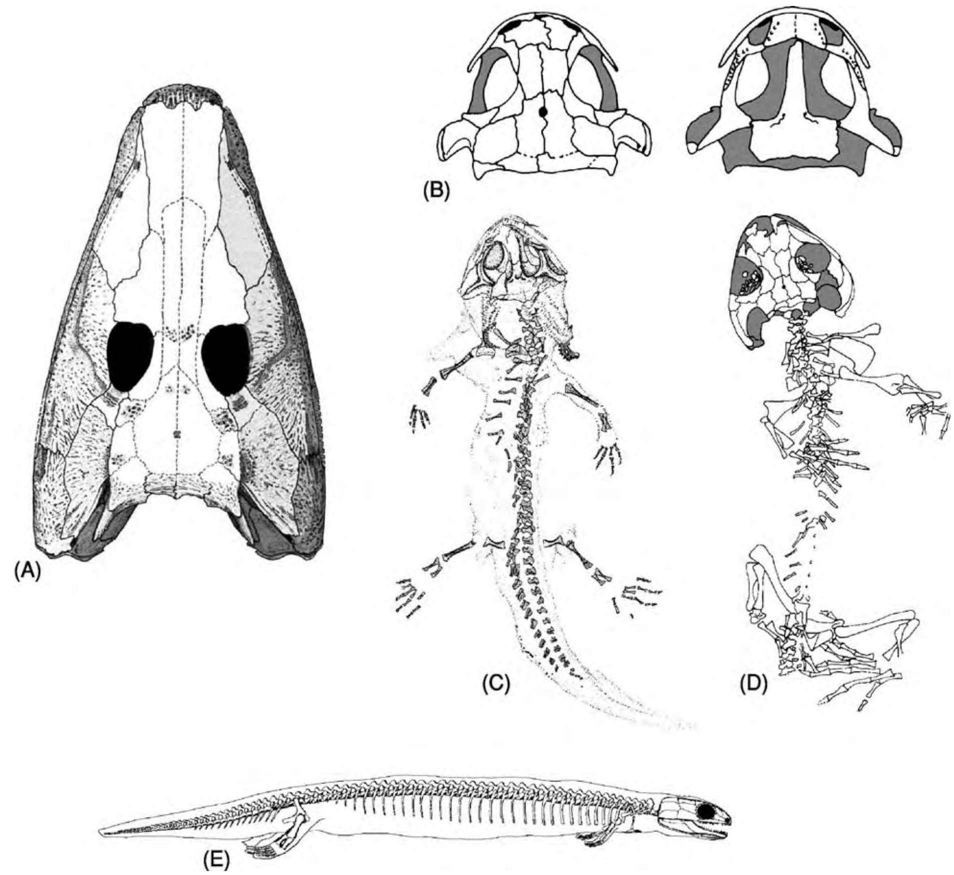


Figure 7 Late Carboniferous tetrapods. (A) Skull of the embolomerous anthracosaur, *Pholiderpeton*, in dorsal view, about 300 mm in length. (B) Skull of the branchiosaur, *Apateon*, in dorsal (left) and palatal (right) view. (C) Image of a branchiosaur fossil, about 50 mm in total length. (D) The dissorophoid, *Platyrhinops*, with a body length of about 120 mm. (E) The microsaur, *Microbrachis*, about 140 mm in length. (A) (D) adapted with permission from Clack JA (2002) *Gaining Ground: the Origin and Evolution of Tetrapods*. Bloomington, IN: Indiana University Press; (E) adapted with permission from Carroll RL, Bossy KA, Milner AC, Andrews SM, and Wellstead CF (1998) *Handbook of Palaeoherpetology*, Part 1, *Lepospondyli*. Munich: Verlag Friedrich Pfeil.

localities representing shallow channel-fills or deeper lake habitats, but few represent terrestrial environments for the Late Carboniferous. This may explain why the origins of amniotes and of modern amphibians are still poorly understood, in that their early members may have lived in environments that have not left a good fossil record. However, it is certain that the earliest true amniotes are found in the Late Carboniferous.

Tetrapods of the Early Permian and Their Descendants

Essentially, tetrapods of the Early Permian represent a continuity of lineages from the Late Carboniferous; this may be an artefact of stratigraphy in that many localities usually considered as earliest Permian may turn out to be latest Carboniferous after all. As the Carboniferous ended and the Permian began, the climate of the former coal swamp areas became gradually drier and the coal forests dwindled. Many localities from the Carboniferous–Permian boundary are formed from red-bed deposits characterized by evaporites, and there is a much greater representation of terrestrial environments. Even so, it has become clear recently that similar faunas existed over the areas of eastern North America and Western Europe.

Baphetids did not survive beyond the end of the Carboniferous, but anthracosaurs lived on into the Permian, with the long-bodied *Archeria crassidisca* of North America being the best known. The last members lingered on into the Late Permian and Early Triassic in Russia. Temnospondyls, by contrast, continued to dominate aquatic habitats, and produced a wide range of morphologies and life styles throughout the Permian. Although they were hit

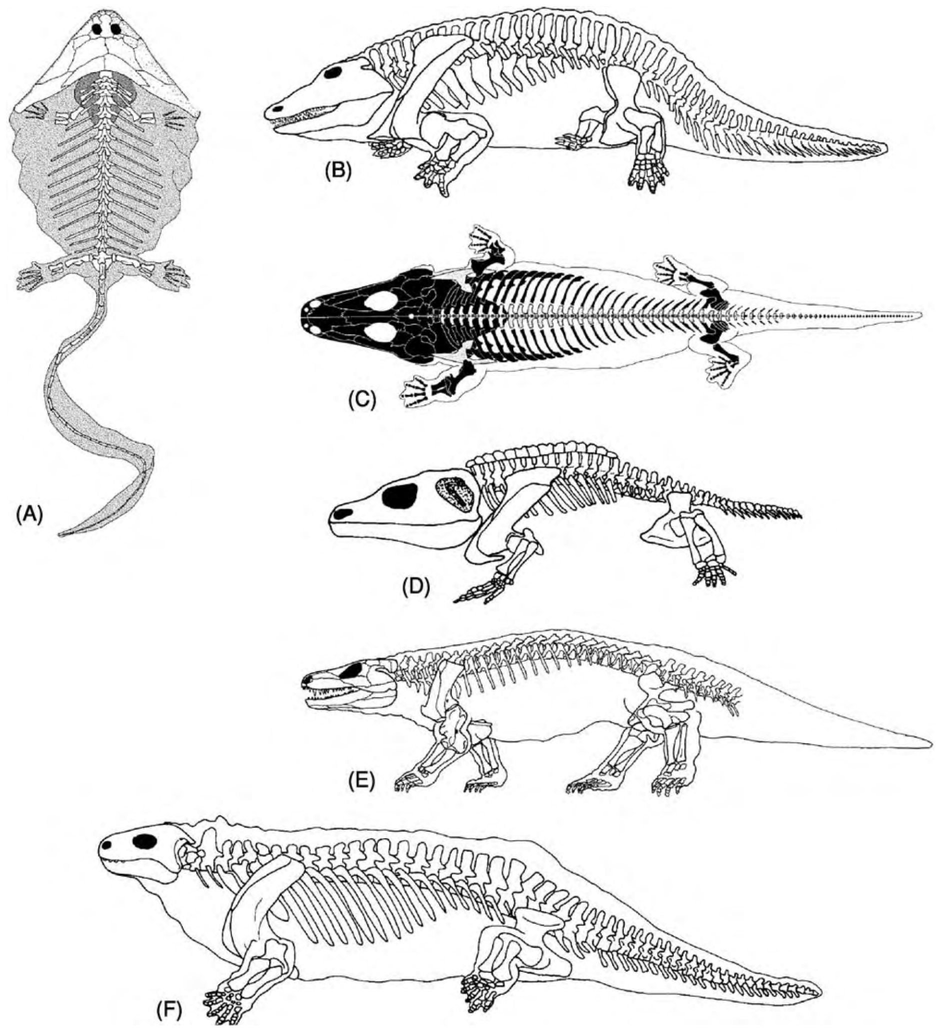


Figure 8 Early Permian tetrapods. (A) The nectridean, *Diplocaulus*, known from a range of sizes with skulls up to 147 mm in length. The body outline, suggested by trace fossils, is included in grey, although it is not clear whether the lateral skin flaps extended down the thighs. (B) The temnospondyl, *Eryops*, about 2 m in length. (C) The capitosauroid temnospondyl, *Mastodonsaurus*, an aquatic form up to 6 m in length. (D) The dissorophid temnospondyl, *Cacops*, a terrestrial form with a skull about 150 mm in length. (E) The seymouriamorph, *Seymouria*, about 1 m in length. (F) The diadectomorph, *Diadectes*, about 2 m in length. (B) and (D) (F) adapted with permission from Clack JA (2002) *Gaining Ground: the Origin and Evolution of Tetrapods*. Bloomington, IN: Indiana University Press; (A) adapted with permission from Carroll RL, Bossy KA, Milner AC, Andrews SM, and Wellstead CF (1998) *Encyclopaedia of Palaeoherpetology*, Part 1, *Lepospondyli*. Munich: Verlag Friedrich Pfeil; (C) adapted with permission from Schoch RR and Milner AR (2000) *Handbook of Palaeoherpetology*, Part 3B, *Stereospondyli*. Munich: Verlag Friedrich Pfeil.

heavily by the end-Permian extinction, after it they re-radiated into a range of aquatic forms, some extremely large, and at least one lineage survived into the Early Cretaceous. The best known of the Early Permian temnospondyls is *Eryops* (Figure 8) from south-eastern North America. It was thought of for many years as a 'model' early tetrapod, but this idea has turned out to be misleading. It is atypical for a temnospondyl in being a large semi-terrestrial form, when most were either large and fully aquatic or small and terrestrial. Dissorophids, for example, produced small terrestrial desert dwellers with nasal adaptations for retaining moisture (Figure 8), whereas capitosaurs were flat-headed aquatic giants (Figure 8).

Three groups that originated in the Late Carboniferous came to dominate the terrestrial habitats of the Early Permian. These are the seymouriamorphs, the diadectomorphs, and the synapsids. The latter were true amniotes and their lineage gave rise to mammals, but the other two, although fully terrestrial, are not thought to be fully amniote. Indeed, seymouriamorphs are known from some exquisite material from Slovakia that includes growth series, and this shows that they grew from aquatic larvae with external gills, like the branchiosaurs. As they grew, the land-going features of their skeletons, such as the attachment of the hip girdle to the vertebral column, became better developed. *Seymouria* (Figure 8), well known from Texas, Oklahoma, and German deposits, had a uniquely specialized ear region, suggesting terrestrial audition. *Diadectes* (Figure 8) was a large barrel-bodied form with crushing dentition, surely a herbivore eating tough terrestrial plants.

A few of the lepospondyl groups persisted into the Early Permian. Nectrideans produced some bizarre forms with heads shaped like advanced hang-gliders, such as *Diplocaulus* (Figure 8). The laterally pointing 'horns' are known from trace fossils to have been attached by broad flaps of skin to the body, and though nectrideans were undoubtedly aquatic, their precise mode of life is not known (Figure 8). Lysorophids, elongate and almost limbless derivatives of microsaurs, are known to have formed burrows in which to aestivate during dry seasons, as some lungfish do today. All these lepospondyl groups, however, had died out by the end of the Early Permian.

The Permian was the time when the land came to be dominated by the amniotes, as adaptations to land dwelling improved with changes to feeding and breathing mechanisms for dealing with terrestrial conditions. The story of non-amniotes is largely confined to temnospondyls until the mid-Mesozoic, when fossils representing modern amphibian groups are first picked up.

See Also

Fossil Vertebrates: Jawless Fish-Like Vertebrates; Fish; Reptiles Other Than Dinosaurs; Mesozoic Amphibians and Other Non-Amniote Tetrapods; Mesozoic Mammals. **Palaeozoic:** Devonian; Carboniferous; Permian; End Permian Extinctions

Further Reading

Ahlberg PE and Milner AR (1994) The origin and early diversification of tetrapods. *Nature* 368: 507 514.

Benton MJ (2000) *Vertebrate Palaeontology*, 2nd edn. Oxford: Blackwell Science.

Bolt JR and Lombard RE (2000) Palaeobiology of *Whatcheeria deltae*. In: Heatwole H and Carroll RL (eds.) *Amphibian Biology*, vol. 4, *Palaeontology*, pp. 1044 1052. Chipping Norton, NSW: Surrey Beatty.

Boy JA and Sues H D (2000) Branchiosaurs: larvae, metamorphosis and heterochrony in temnospondyls and seymouriamorphs. In: Heatwole H and Carroll RL (eds.) *Amphibian Biology*, vol. 4, *Palaeontology*, pp. 1150 1197. Chipping Norton, NSW: Surrey Beatty.

Carroll RL (1988) *Vertebrate Palaeontology and Evolution*. New York: W. H. Freeman and Co.

Carroll RL (2000) Lepospondyls. In: Heatwole H and Carroll RL (eds.) *Amphibian Biology*, vol. 4, *Palaeontology*, pp. 1198 1269. Chipping Norton, NSW: Surrey Beatty.

Carroll RL, Bossy KA, Milner AC, Andrews SM, and Wellstead CF (1998) *Handbook of Palaeoherpetology*. Part 1, *Lepospondyli*. Munich: Verlag Friedrich Pfeil.

Clack JA (2000) The origin of tetrapods. In: Heatwole H and Carroll RL (eds.) *Amphibian Biology*, vol. 4, *Palaeontology*, pp. 979 1029. Chipping Norton, NSW: Surrey Beatty.

Clack JA (2002) *Gaining Ground: the Origin and Evolution of Tetrapods*. Bloomington, IN: Indiana University Press.

Clack JA and Carroll RL (2000) Early Carboniferous tetrapods. In: Heatwole H and Carroll RL (eds.) *Amphibian Biology*, vol. 4, *Palaeontology*, pp. 1030 1043. Chipping Norton, NSW: Surrey Beatty.

Holmes R (2000) Palaeozoic temnospondyls. In: Heatwole H and Carroll RL (eds.) *Amphibian Biology*, vol. 4, *Palaeontology*, pp. 1081 1120. Chipping Norton, NSW: Surrey Beatty.

Laurin M (2000) Seymouriamorphs. In: Heatwole H and Carroll RL (eds.) *Amphibian Biology*, vol. 4, *Palaeontology*, pp. 1064 1080. Chipping Norton, NSW: Surrey Beatty.

Milner AR (1993) Biogeography of Palaeozoic tetrapods. In: Long JA (ed.) *Palaeozoic Vertebrate Biostratigraphy and Biogeography*, pp. 324 353. London: Belhaven Press.

Rocek Z and Rage J C (2000) Anatomical transformations in the transition from temnospondyl to proanuran stages. In: Heatwole H and Carroll RL (eds.) *Amphibian Biology*, vol. 4, *Palaeontology*, pp. 1274 1282. Chipping Norton, NSW: Surrey Beatty.

Ruta M, Coates MI, and Quicke DLJ (2003) Early tetrapod relationships revisited. *Biological Reviews* 78: 251 345.

Schoch RR and Milner AR (2000) *Handbook of Palaeoherpetology*. Part 3B, *Stereospondyli*. Munich: Verlag Friedrich Pfeil.

Smithson TR (2000) Anthracosaurs. In: Heatwole H and Carroll RL (eds.) *Amphibian Biology*, vol. 4, *Palaeontology*, pp. 1053 1063. Chipping Norton, NSW: Surrey Beatty.

Reptiles Other Than Dinosaurs

R R Reisz, University of Toronto at Mississauga, Mississauga, ON, Canada

Introduction

Who are the reptiles? The term 'reptile' can be readily associated with a series of extant animals, the turtles, crocodiles, lizards, and snakes, and the tuatara, but in a historical context this term also refers to a bewildering array of extinct forms that include the ancient Palaeozoic anapsids, flying reptiles, aquatic reptiles, mammal-like reptiles, and, of course, dinosaurs. In order to understand the evolutionary history of reptiles we must consider the wider implications of this name. In effect, we must start with the group Amniota, vertebrate tetrapods that are characterized by the presence of the embryological structure amnion and three other extraembryonic membranes that are critical to the survival and development of the embryo. All these animals lay an amniotic egg or did so primitively. Even the embryos of viviparous amniotes (therian mammals and some snakes) retain the extraembryonic membranes found in the amniotic egg (the allantois, yolk sac, chorion, and amnion). The allantois is a bladder storing metabolic wastes; it absorbs water from the environment because of the high osmotic pressure of its fluids and helps gas exchange. The yolk sac stores nutrients, the chorion provides a surface for gas exchange with the outside, while the amnion maintains a liquid environment around the embryo. This is in strong contrast to amphibians (frogs, salamanders, and apodans) which lack this complex set of embryological structures.

However, Amniota encompasses not only fossil and living reptiles but also mammals and birds, and this complicates the issue because the distant, ancient relatives of mammals were biologically reptile-like, and birds are technically reptiles because of their close association with dinosaurs. The early diversification of amniotes is probably the most exciting chapter in higher vertebrate evolution, and it is this section of the fossil record that I will discuss in some detail (Figure 1).

Early Amniotes

Amniotes can be divided into two large groups: synapsids and reptiles. Synapsida includes mammals and all fossil amniotes more closely related to mammals than to reptiles. The Palaeozoic synapsids are the so-called mammal-like reptiles that represent the most spectacular radiation of amniotes during the Carboniferous and Permian Periods. Reptilia includes extant lizards, snakes, crocodiles, collectively called diapsids because ancestrally they all had two temporal fenestrae behind the orbit, but also includes the anapsid testudines, and all the fossil amniotes more closely related to them than to mammals.

Many early reptiles were previously viewed as anapsids, referring to the absence of any temporal fenestrae behind the orbits. These include mesosaurids, parareptiles, captorhinids, and the so-called protorothyridids. The taxon Anapsida does not have any phylogenetic significance, being largely based on the absence of a series of derived characters. They are, however, part of the initial diversification of amniotes.

Mesosaurs were the first fully aquatic amniotes. They are known from the Permian of southern Africa and eastern South America. Their presence on both sides of the Atlantic has been used to support the theory of continental drift. Mesosaurs are highly specialised amniotes, characterized by numerous autapomorphies that are found in most parts of their skeletons. Their snout was very long, mainly composed of long, slender premaxillae and dentaries, and their unusually long teeth extended anteriorly (near the tip of the snout), laterally (on the anterior half of the sides of the snout), and ventrolaterally (on the posterior half of the sides of the snout) from the upper jaw. The dentition seems to have formed a basket for filtering water and capturing small invertebrates. The neck is long, an adaptation of several aquatic amniotes. The tail was long and laterally compressed, and was probably used for swimming. The trunk ribs are pachyostotic; they were very thick and heavy and served as ballasts. The scapular blade was very short, as in several other aquatic amniotes. Contrary to previous interpretations, there is no evidence for a lateral temporal fenestra in these forms.

Mesosaurs are now considered to be part of the Parareptilia, a group of Palaeozoic and Early Mesozoic reptiles that may include turtles and their close relatives, and also includes four main groups; millerettids, pareiasaurs, procolophonids (Figure 2), and turtles. They are united by the several synapomorphies, including a dorsally enlarged quadratojugal bordering the quadrate emargination. Millerettids appear at the beginning of the Upper Permian. They were small, insectivorous animals superficially similar to lizards. Several millerettids have either a lower temporal fenestra or a deep cheek emargination. They

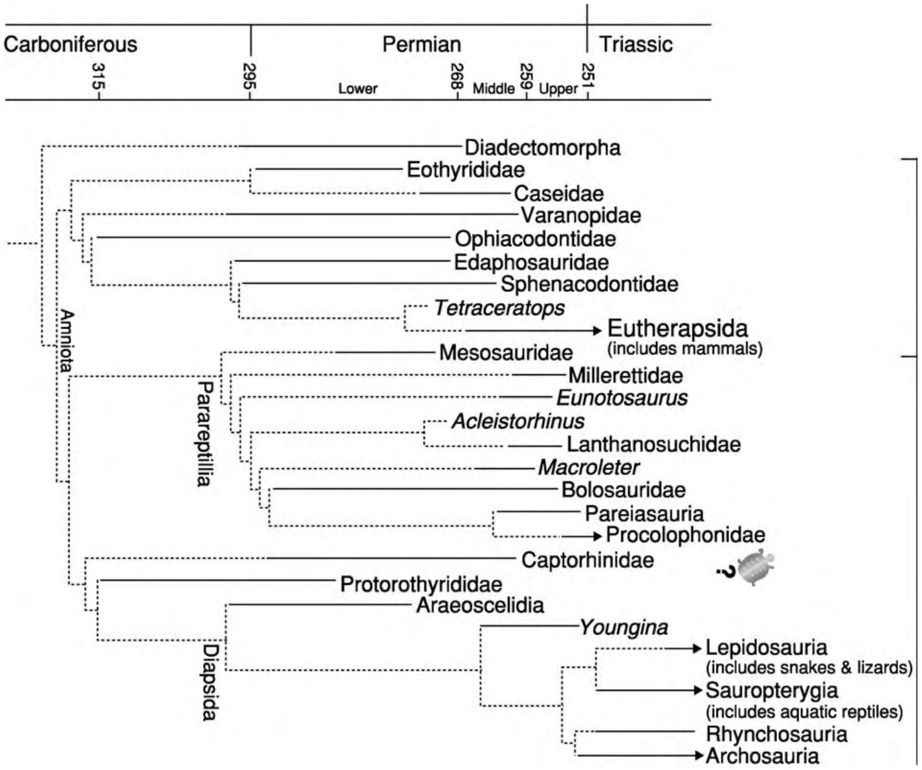


Figure 1 Early amniote phylogeny. The pattern of solid and dotted lines represent evolutionary relationships of the various taxa known from the fossil record. The solid lines with the taxonomic units to their right represent the known fossil record of those particular taxa. For example the fossil record of the Diadectomorpha extends from the Permo Carboniferous boundary throughout Lower Permian strata. The dotted lines represent reconstructed histories of particular taxa (also called ghost lineages). These are based on the combination of the known earliest appearance of certain lineages, and the proposed pattern of evolutionary relationships. For example, the known record of Eothyrididae extending to the Permo Carboniferous boundary automatically creates a long ghost lineage for Caseidae, whose record only starts towards the top or the Lower Permian.

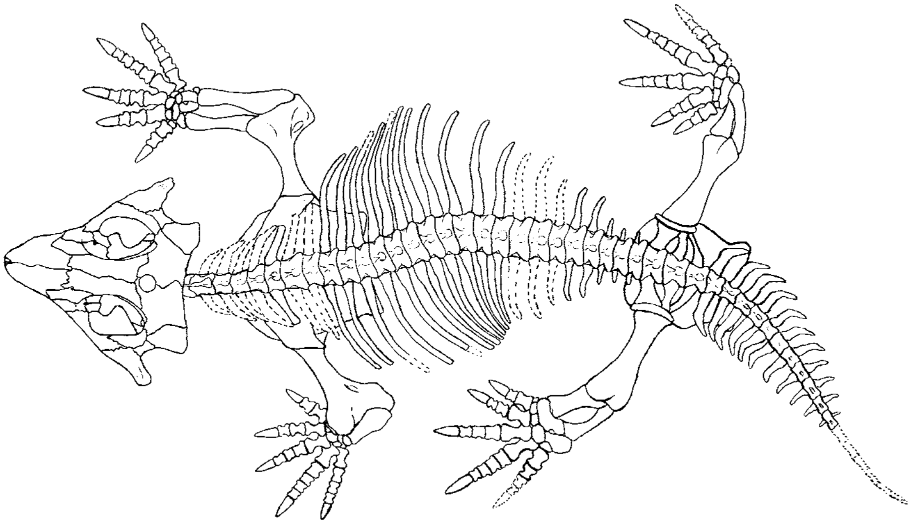


Figure 2 *Procolophon*. This small parareptile is known from the Early Mesozoic, Lower Triassic sediments of Gondwana.

have a quadrate emargination supporting a tympanum. Millerettids appear to be the first amniotes to acquire a tympanum. This tympanum may be homologous to the tympanum of turtles but it is not homologous to the tympanum found in diapsids, in mammals, or in frogs and temnospondyls. Pareiasaurs were large, herbivorous amniotes that lived in the Upper Permian in Africa, Western Europe, Russia, and China, one of the first group to achieve nearly global distribution. They were also among the largest Palaeozoic amniotes, reaching a length of 3 m. They were covered by an extensive armour of osteoderms (dermal bony scales). Their skull was massive, broad, and had strange bony processes in the cheek and lower jaw. They probably had a parasagittal, semi-erect posture. They had a reduced phalangeal formula in the foot, a character advantageous for an animal having a parasagittal posture. Procolophonids were small and primitively insectivorous, as millerettids were, but late procolophonids have transversely broad teeth that may indicate a herbivorous diet. Procolophonids appear in the fossil record at the very end of the Upper Permian and became extinct at the end of the Triassic. The orbit is expanded posteriorly, especially in late procolophonids. The last procolophonids had strange bony processes and their orbit was much longer posteriorly than the orbit of early procolophonids.

Testudines may be the only surviving parareptiles. They appear in the fossil record in the Upper Triassic, but they must have been present earlier, but are not preserved in the fossil record. Turtles have also been associated with diapsid reptiles, but their origins remain controversial.

Captorhinids were one of the most successful groups of early amniotes and are part of the group called Eureptilia. They appear in the fossil record in the Lower Permian and became extinct in the Upper Permian. Captorhinids are known mainly from sediments of North America, but a few individual taxa have also been recovered from North and Central Africa, India, and Russia, making it the earliest known group of reptiles to have a geographically widespread distribution. Early captorhinids were relatively small, and had a single row of marginal teeth, but soon after the appearance of the group, several captorhinids developed multiple marginal tooth rows in the upper and lower jaws. These tooth rows are restricted to the maxilla in the upper jaw, and the dentary in the lower jaw. Some of the more derived taxa, such as *Moradisaurus*, had up to 12 rows. In this derived, large captorhinid from the Late Permian of Africa, the cheeks are greatly expanded. The limbs were relatively short and massively built, with broad hands and feet.

The morphology of small captorhinids is known best from a vast amount of material collected from a Lower Permian fissure-fill locality near Fort Sill, Oklahoma. The overwhelming majority of the bones in these fissure fills are attributable to a single taxon, the small, multiple tooth-rowed *Captorhinus aguti* (Figure 3).

Protorothyridids are a group of small (100 mm snout-vent length), lizard-like eureptiles. We are not certain that they are a monophyletic group, because the interrelationships of members of this group have not been analysed, all previous studies having considered them only in terms of a generalized basal amniote pattern. They are closely related to diapsids because both groups have long and slender hands and feet (they are broad and robust in other amniotes).

The fossil record of protorothyridids is restricted to the Carboniferous and Early Permian of North America and Europe, and all the known taxa are represented either by a single, incomplete skeleton, or at most by a handful of partially preserved remains. *Hylonomus*, the oldest known amniote, has been recovered from the Middle Pennsylvanian (Westphalian A) of Joggins, Nova Scotia. All the specimens have been collected from inside upright *Sigillaria* stumps. Although known from several partial skeletons, all the specimens are completely disarticulated, and poorly preserved. A better known member of this group is *Paleothyris*, represented from three nearly complete skeletons from the Middle

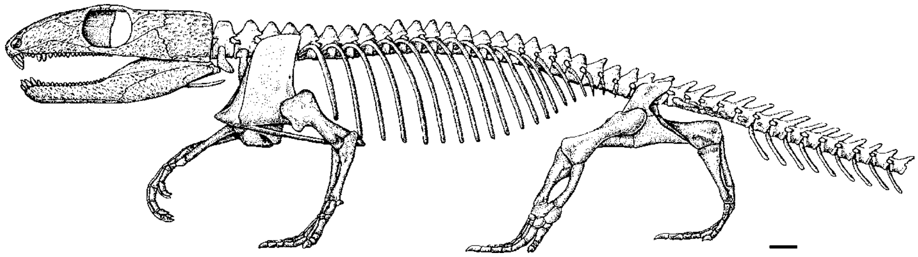


Figure 3 *Captorhinus*. This small eureptile is known from the Early Permian of North America. Members of this group, the Captorhinidae dispersed widely, and had a nearly cosmopolitan distribution in the Middle and Late Permian.

Pennsylvanian (Westphalian D) of Florence, Nova Scotia. Although these specimens were also collected from upright *Sigillaria* stumps, their preservation is superior to that of *Hylonomus*, allowing for a more detailed study. *Protorothyris* is a slightly larger form from the Early Permian of Texas. This form is known from a few skulls and a partial skeleton.

The meager evidence provided by the known protorothyridids give the general impression that they resembled early captorhinids, but had more slender limbs and narrow neural arches. This general impression is reinforced by the similarities between such taxa as *Protorothyris* and the recently described *Thuringothyris*, and captorhinids. Recent phylogenetic analyses indicate that protorothyridids are not an ancestral group of amniotes, as previously suggested, but are the sister-group, or closest known relative of diapsids.

Early Diapsid Reptiles

Diapsids are the most diverse and numerous eureptiles, and include all modern reptiles (except perhaps for turtles). They also include birds and several groups of extinct marine, flying, and terrestrial reptiles such as dinosaurs. The early history of this highly successful group extends into the Early Permian and the Late Pennsylvanian (latest Carboniferous), represented by the early diversification of the Araeoscelidia. Three well-known members of this group are now known, *Petrolacosaurus, Araeoscelis*, and *Spinoaequalis*.

The oldest known diapsid is *Petrolacosaurus* from the Upper Pennsylvanian of Kansas (Figure 4). Numerous specimens have been recovered from sediments that have filled in a Late Carboniferous tidal channel and preserved the most diverse amniote assemblage known for that period. *Petrolacosaurus* is the most commonly found amniote at this small site, other skeletons representing a wide variety of endemic synapsids. This small diapsid is characterised by an exceedingly delicate skull, which carries numerous, delicately built marginal teeth. *Petrolacosaurus* is unique among Late Pennsylvanian terrestrial vertebrates in having an unusually elongate neck, a long, slender tail, and slender, elongate limbs. Particularly striking among these is the delicately constructed forelimb, with long, slender fingers. The well preserved skeletal anatomy provides clear evidence that the oldest known diapsid was the most agile amniote of its time.

Araeoscelis from the Lower Permian of Texas is also a small diapsid, with a lightly built, slender skeleton, but its skull is much more massively built than that of *Petrolacosaurus*, and its marginal dentition is bulbous, and transversely expanded. *Araeoscelis* is a particularly fascinating early diapsid because there is strong evidence to indicate that it exhibits a secondarily closed lower temporal fenestra. Both of these diapsids are represented by several good skeletons, but the third, more recently described araeoscelidian, *Spinoaequalis*, from the Late Pennsylvanian of Kansas, is known from a single partial skeleton. This small reptile is only slightly younger than *Petrolacosaurus*, lacks the elongated neck, but has elongated limbs, as seen in the other two members of this clade. The appearance of *Spinoaequalis* in the fossil record, so soon after *Petrolacosaurus*, provides direct evidence for the diversification of diapsids within the Carboniferous. Of particular interest is the discovery

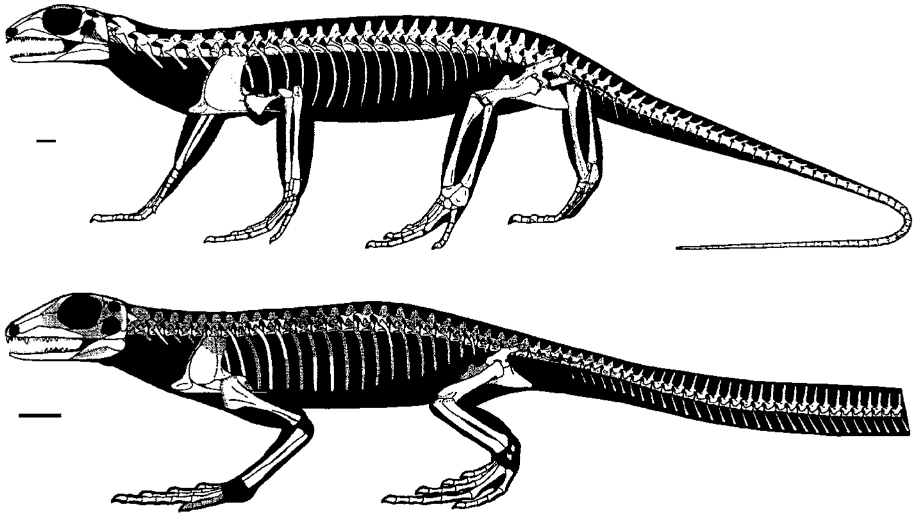


Figure 4 *Petrolacosaurus* and *Spinoaequalis*, the oldest known diapsid reptiles.

of a number of skeletal features in *Spinoaequalis* which display evidence for aquatic specialisations. Most striking among these is the presence of a dorsoventrally expanded tail. Araeoscelidians also share the typical diapsid characters with more recent diapsids: two temporal fenestrae, an upper fenestra between postorbital, parietal, and squamosal, and a lower fenestra between the postorbital, squamosal, and jugal (the quadratojugal sometimes enters the edge of the fenestra); a suborbital fenestra between the maxilla, palatine, and ectopterygoid.

Younginiforms are an important group of fossil diapsids from the Upper Permian and Lower Triassic of South Africa, eastern Africa, and Madagascar. *Youngina* is the best-known member of this group. Younginiforms and modern diapsids form a monophyletic group called the Neodiapsida. They are united by the presence of a ventromedial flange of the parietal on which the jaw musculature originated.

All other, younger diapsids belong to two major groups: the Lepidosauromorpha and the Archosauromorpha. Together, these two taxa include the forms that epitomise the Age of Reptiles, the Mesozoic, and also include most extant reptiles, with the possible exception of turtles (if the latter are parareptiles).

Lepidosauromorpha include mainly small- to medium-sized reptiles, both fossil and extant, and are represented today by *Sphenodon*, lizards, and snakes. These forms are characterized by an incomplete lower temporal bar. The modern sphenodontid *Sphenodon* has reverted to the primitive diapsid condition and has a complete lower temporal bar. Sphenodontids are represented only by *Sphenodon* in the modern fauna, but they were once a moderately diverse group. They first appear in the fossil record in the Upper Triassic. *Sphenodon* lacks a tympanum, but its fossil relatives probably had one. Sphenodontids have a strong jaw musculature, so their temporal arches are bowed laterally to accommodate it. Their dentition is acrodont (the teeth are fused to the surface of the jaw).

The Squamata is one of the largest group of living reptiles. It includes 6000 species of lizards and snakes. Snakes appear in the Cretaceous and lizards are first found in the Upper Jurassic. Several Upper Permian and Triassic lizards have been identified, but these are probably not lizards. Lizards are not a natural group because they include the ancestors of snakes. Most squamates have a pleurodont tooth implantation. This means that the teeth are attached to the medial surface of the jaws. There are very shallow tooth sockets, when they are present, and the alveolar shelf is strongly inclined (far from the horizontal). Early tetrapods have a subpleurodont or subthecodont implantation, in which the teeth are set in shallow sockets and the alveolar shelf is gently inclined. A few large squamates such as the Cretaceous mosasaurs, have teeth set in deep sockets (thecodont implantation). In a few squamates, such as agamids and chamaeleontids, the dentition is acrodont (fused to the surface of the jaw), as in sphenodontids.

Among squamates, mosasaurs were a large group of large aquatic lizards from the Upper Cretaceous. Almost 20 genera are known; the largest one, *Hainosaurus*, reached a length of 45 feet. Mosasaurs had a long body (29 to 51 presacral vertebrae) and small limbs with a reduced number of elements. They swam by anguilliform motion, which is rather slow, and the limbs were used only for steering. Mosasaurs lived in shallow, coastal marine waters. Some mosasaurs seem to have fed on ammonites (large cephalopods resembling the nautilus), because we found several ammonite shells with tooth marks that look like they were made by a mosasaur.

Snakes are one of the most recent and most rapidly evolving groups of reptiles. Most modern genera belong to families that appeared at the beginning of the Miocene or in the Oligocene. Most snakes have no limbs and there is not even a trace of the pectoral girdle, but several genera retain vestiges of the pelvic girdle and posterior limbs. Recent evidence indicates that snakes are probably related to mosasaurs, which in turn are related to varanoid lizards. Snakes are, therefore, nested within the clade of lizards. *Dinilysia* is the oldest known snake, dating from the Upper Cretaceous.

Aquatic Reptiles

As exemplified by the ancient Permian mesosaurs and the Cretaceous mosasaurs, return to the water is a common phenomenon throughout reptilian history. It has been argued that because of the relatively low metabolic rates, tolerance to anoxia, and generally low body temperatures, reptiles can make the transition from a terrestrial lifestyle to an aquatic one without major difficulties. It has been shown experimentally, in the modern marine iguana that the metabolic cost of swimming is only one-quarter that of walking. However, a return to the water to catch an easy meal is not the same as life in the water. Many reptiles today are amphibious and at home in the water but they do not live their entire lives there. Back in the Palaeozoic and more notably in the Mesozoic, many reptiles did evolve a fully aquatic lifestyle.

The relationships between Palaeozoic and Mesozoic marine reptiles and their terrestrial relatives are difficult to assess, but the following groups are generally considered to be derived from diapsids, probably lepidosauromorphs, with the lower temporal fenestra being without a lower temporal bar, and only the upper temporal fenestra being retained in its original form.

Nothosaurs are medium-sized aquatic reptiles restricted to the Triassic. They are one of the two major groups of sauropterygians (the other one being plesiosaurs). They range in length from 20 cm to 4 m in body length. Their skull is relatively small, and the lower temporal bar has been lost. The slender stapes and quadrate emargination suggest that they may have had a tympanum. The interpterygoid vacuity is closed. This type of palate links nothosaurs with plesiosaurs. The ribs of nothosaurs were often pachyostotic. Their limbs and girdles are poorly ossified, but they have no striking aquatic adaptations. Nothosaurs probably swam relatively slowly by lateral undulation. The limbs were probably used for steering, although some people think that the forelimbs may also have been used for propulsion, because the coracoid is relatively large.

Plesiosaurs, a group of aquatic diapsids closely related to nothosaurs, appeared in the Jurassic and became extinct at the end of the Cretaceous. They had a short trunk and a short tail. Their coracoid, and to a certain extent their scapula, are greatly expanded ventrally. Young plesiosaurs had a shoulder similar to nothosaur's girdle. Plesiosaurs have massive ventral scales called gastralia, which may have strengthened the trunk. This would have been useful, because plesiosaurs used their limbs to swim. The exact way in which the limbs were used is still debated. They may have moved in the water much like the modern leatherback turtle (by subaqueous flight) or rowed like sea-lions. According to some, the short scapular and iliac blades implies that the muscles elevating the limbs were weak and poorly suited for subaqueous flight. On the other hand, perhaps the origin of these muscles had just shifted medially, closer to the vertebral column. The massive coracoids would, in any case, have provided a large surface for attachment of the forelimb retractor muscles. So, this question remains unsolved.

Placodonts are Triassic aquatic reptiles with robust tooth plates that strongly suggest that they had a durophageous diet; they may have fed on molluscs and arthropods. They had a short, stout body, and an extremely robust skull. The internal naris (choana) is a single, median opening. Their anterior dentary and premaxillary teeth are procumbent and spatulate; they were probably used to grab the prey. The interclavicle has a very short posterior stem and it is ventral to the clavicle, as in nothosaurs. The endochondral shoulder girdle is poorly ossified. Some genera had a carapace composed of polygonal dermal bones covered by epidermal scutes. In the most advanced members of this group the shoulder girdle has moved inside the armor and rib-cage, in a similar fashion to that seen in turtles. Some armored placodonts had lost their anterior teeth and may have had a horny beak instead.

Ichthyosaurs were similar to dolphins and sharks in size and shape. They appeared in the Lower Triassic, were numerous in the Jurassic, and became extinct before the end of the Cretaceous. Of all the Palaeozoic and Mesozoic marine diapsids, they were the most highly adapted to a marine existence. Typical ichthyosaurs (from the Jurassic and Cretaceous) had a relatively short, fusiform trunk, a dorsal fin, and a high, lunate caudal fin. We know about the dorsal fin because some specimens have the body outline preserved as a carbonaceous film. These features suggest that ichthyosaurs could swim rapidly. We know that ichthyosaurs were viviparious, because some of them died while giving birth and were preserved with the young ichthyosaur still in the pubic canal.

The orbit was large and the eye was protected by a well ossified sclerotic ring. The snout was long and bore numerous homodont, sharp teeth (in most genera). The quadrate was not notched posteriorly and the stapes was relatively massive and articulated with the quadrate distally, so there was probably no tympanum (it would not have been useful under water). The vertebrae of ichthyosaurs were short, wide amphicoelous disks. The neural arches never fused to the centra and there were no transverse processes.

Early Archosauromorphs and Crocodiles

Archosauromorphs include the most spectacular reptiles of the Mesozoic Era, were the dominant terrestrial and aerial vertebrates of that era, and include crocodiles and crocodile-like reptiles, dinosaurian relatives, flying reptiles, dinosaurs, and birds. The oldest known members of this large group show up near the end of the Permian, but the main radiation of this group starts during the Triassic.

Rhynchosaurs were the most common and widespread early archosauromorphs. They are found throughout the Triassic. They have been used for biostratigraphic correlations. They have a single, median external naris and a premaxillary beak. *Mesosuchus* is a Lower Triassic rhynchosaur, more primitive than later rhynchosaurs in retaining premaxillary teeth and in having a single row of maxillary and dentary teeth. Later rhynchosaurs have multiple tooth rows and a broad skull. Rhynchosaurs were among the most abundant large herbivores in the Middle and Upper Triassic. Their extinction at the end of the Triassic may support the existence of a mass extinction event at that time.

Perhaps the most spectacular feature of the archosaurian skull is the presence of a new opening in the skull in front of the orbit, the antorbital fenestra. This was probably just a pneumatic structure, as in modern birds, making the skull lighter, but this feature readily distinguishes these forms from other diapsids. With the exception of a few other basal members of this large group, like *Proterosuchus* and *Euparkeria*, all other archosauromorphs belong to either the Crurotarsi, which includes crocodiles and their fossil relatives, and the Ornithodira, which includes flying reptiles, dinosaurs, and their closest relatives, and birds.

The Crurotarsi include some very spectacular relatives of crocodiles, but the earliest crocodile-like forms belong to the Sphenosuchidae. Sphenosuchids are known primarily from the Upper Triassic of Europe, North and South America, and the Lower Jurassic of South Africa. In contrast to true crocodiles, Sphenosuchids were relatively lightly built and had an erect, bipedal posture. True crocodiles (crocodyliforms) first appeared in the Upper Triassic. They are divided into three major groups, each representing an important adaptive radiation. These three groups are the Protosuchia, Mesosuchia, and Eusuchia. The Protosuchia and Mesosuchia are probably paraphyletic, so they are not real taxa. The skull of true crocodiles is large, massively built, and low. The skull roof is flat and has heavy pitting on the surface. The antorbital fenestra is small in primitive crocodiles and absent in living crocodiles. Living crocodiles have a long secondary palate to separate the passage of air from the mouth, as in mammals.

Palaeozoic Synapsids, the Mammal-Like Reptiles

No evolutionary history of reptiles is complete without some consideration of the so-called mammal-like reptiles, the Palaeozoic synapsids. Palaeozoic synapsids (also called pelycosaurs and early therapsids) occupy a central position in early amniote evolution. Palaeozoic synapsids include some of the oldest known amniotes, and their fossil remains record the earliest successful adaptations of terrestrial vertebrates to herbivorous and active predatory modes of life. During the Palaeozoic, synapsids diversified greatly and became the most conspicuous terrestrial vertebrates of their time. Although both pelycosaurs and early therapsids are paraphyletic taxa at the base of the clade that includes advanced therapsids and mammals, these Palaeozoic reptiles have played a pivotal role in considerations of synapsid evolution, including the origin of mammals. The oldest known mammal-like reptiles have been found in sediments of Early and Middle Pennsylvanian (Late Carboniferous) age in Nova Scotia. The remains of *Protoclepsydrops haplous*, retrieved from upright *Sigillaria* stumps from Joggins, Nova Scotia are too fragmentary to provide any useful information; even the identification of these remains as a pelycosaur has been questioned. By the Late Pennsylvanian, pelycosaurs become the most common amniotes in Laurasian sediments. The fossil record of this group is most extensive during the Early Permian, representing nearly 70% of all amniote finds of this time; by the Late Permian, however, pelycosaurs disappear from the fossil record. Instead, the early therapsids become the most diverse and common amniotes. Only during the Triassic do true reptiles become the most diverse and common amniotes.

As primitive synapsids, pelycosaurs are characterized by the presence of an anteriorly inclined plate-like occiput, and a lateral temporal fenestra that is bordered by jugal, squamosal, and postorbital bones. This is in contrast to the condition in diapsids, where the lateral temporal fenestra is bordered by a fourth bone, the quadratojugal primitively, and the occiput has large postemporal fenestrae. Recent phylogenetic studies of pelycosaurs have proposed that within this group, six major clades (families) can be recognized: small insectivores (Eothyrididae), two different types of large, bulky herbivores (Caseidae and Edaphosauridae), and three different types of medium to large-sized carnivores (Varanopidae, Ophiacodontidae, and Sphenacodontidae).

Eothyrididae

This family was erected as a provisional group for the small pelycosaur *Eothyris*, known from a single skull from the Lower Permian of Texas, and fragmentary remains of three large pelycosaurs. Other fossils were subsequently placed within this family, but at present only two genera, *Eothyris*, from the Early Permian of Texas, and *Oedaleops*, from the Early Permian of New Mexico, can be assigned with confidence into the Eothyrididae. *Eothyris* is known from a single skull, whereas *Oedaleops* is represented by three partial skulls and some limb elements. These specimens, although fragmentary, have revealed that eothyridids were small carnivorous amniotes.

Caseidae

The casseids, closely related to the eothyridids, were a widespread and diverse group of herbivorous pelycosaurs, with a fossil record that extended from the Early Permian of North America and western Europe

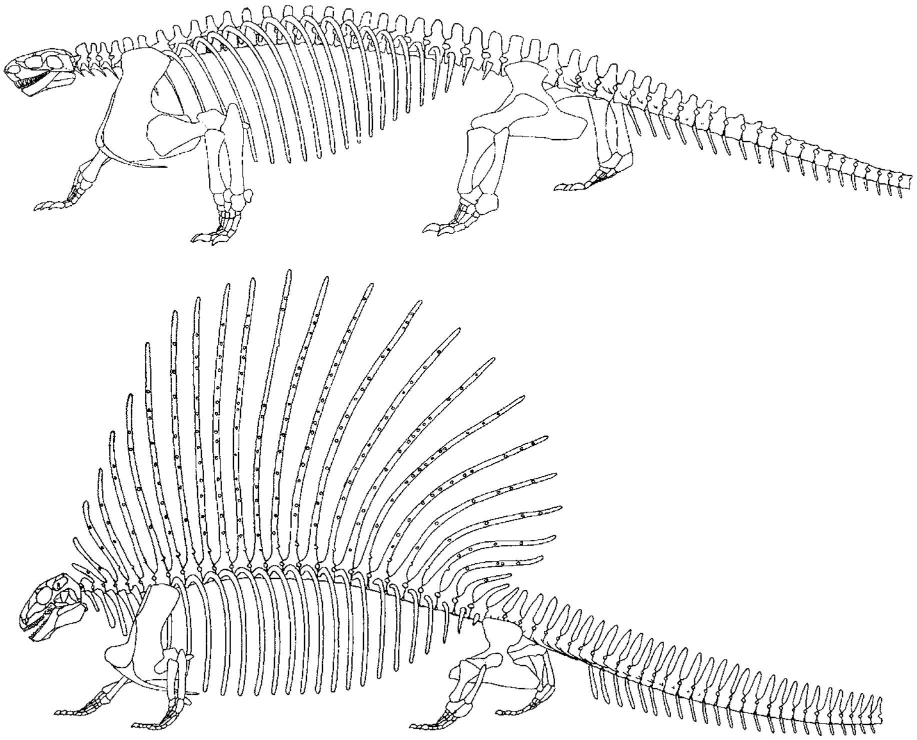


Figure 5 *Corylorhynchus* and *Edaphosaurus* two large Early and Middle Permian herbivorous synapsids (''mammal like reptiles'').

to the Late Permian of eastern Europe. Although ranging in size from 1 to nearly 4 meters in body length, caseids are surprisingly conservative in their skeletal anatomy and body proportions. The skulls are distinctive in the presence of relatively large temporal fenestrae, enormous external nares, and large pineal foramen. The snout overhangs dramatically the tooth row to form an anteriorly projecting rostrum, as in eothyridids. The external surface of the skull has unusual rounded deep pits and sometimes crevice-like depressions that form a distinctive sculpturing pattern. The marginal teeth have spatulate crowns and are serrated along the edge, in some respects similar to the teeth of pareiasaurs. The largest teeth in the marginal series is located anteriorly, on both the premaxilla and dentary. All caseids, whether modest in size, or enormous, are characterised by small cervical vertebrae, bulky, barrel-shaped bodies, and relatively massive limbs. There is reduction in the phalangeal formula, ranging from 2-3-4-4-3 to 2-2-2-3-2. Two distinct types of terminal phalanges can be found in caseids. In *Corylorhynchus*, the claw bearing element is very large and broad, with apparently sharp edges on the sides. In *Angelosaurus*, the claw bearing phalanges are also massive, but more conservative in morphology, roughly triangular in outline.

Caseids and eothyridids share a number of derived characters associated with the morphology of the snout and external naris that support the hypothesis that these primitive pelycosaurs form a clade, the Caseasauria. All other pelycosaurs can be placed in the Eupelycosauria, a clade that also includes therapsids and mammals (Figure 5).

Varanopidae

These faunivorous synapsids of small to moderate size are all characterized by slender, lightly built skulls with highly specialized marginal dentition (Figure 6). All varanopids have mediolaterally flattened teeth with anterior and posterior cutting edges, and are strongly recurved. Both the premaxilla and maxilla are also highly modified, the tooth bearing portion of the premaxilla being enlarged, and the maxilla being greatly elongated in association with the posterior extension of the marginal dentition. Other diagnostic fedatures of varanopseids includes enlargement of the temporal fenestra and the reduction of the subtemporal arch to a narrow bar. The lower jaw is slender, and the mandibular symphysis is reduced in size. The postcranial skeleton is also lightly built, giving varanopids a sleek appearance, superficially similar to modern monitor lizards. It is, therefore, reasonable to suggest that varanopid pelycosaurs were probably the most agile predators of their time. *Mycterosaurus longiceps* is the most basal member of the group. The larger varanopids, *Varanops, Aerosaurus*, and

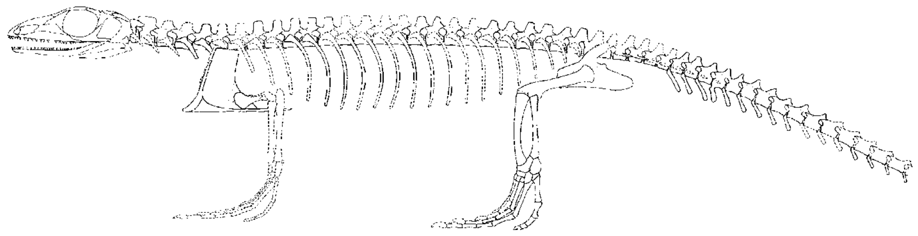


Figure 6 *Archaeovenator*, a Late Carboniferous varanopids synapsid (''mammal lime reptile'') from North America.

Varanodon, are all characterized by the presence of a massive external maxillary buttress above the caniniform region, a dramatic anterior tilt of the occiput, and anteroposterior enlargement of the temporal fenestra.

The hypotheses of phylogenetic relationships of this group of pelycosaurs has changed dramatically as our knowledge of primitive synapsids has increased. In 1940 varanopids were associated with the advanced sphenacodontids as a group of primitive carnivorous sphenacodonts, and it was even suggested that *Varanops* may have been close to the ancestry of sphenacodontids. The most recent hypothesis places varanopids within the Eupelycosauria, as its most primitive member. Although retaining many primitive cranial and postcranial features that appear in derived form in edaphosaurs, ophiacodontids and sphenacodontids, it is obvious that varanopids developed a number of autamorphies, and the advanced members of this group outlasted all other carnivorous pelycosaurs as highly specialised, agile predators. The large, predatory *Varanodon* is one of the youngest known pelycosaurs in North America, but two small varanopids, *Elliotsmithia* from the Upper Permian strata of South Africa, and *Mesenosaurus* from the Middle Permian of Russia extend significantly the fossil record of this group.

Ophiacodontidae

The oldest known pelycosaurs have been included in this family. *Protoclepsydrops* from the Middle Pennsylvanian of Joggins, Nova Scotia, has been placed within the Ophiacodontidae, but its identity and taxonomic status is uncertain, partly because the known skeletal remains are fragmentary.

Archaeothyris, from slightly younger sediments near Florence, Nova Scotia, is the oldest known diagnosable ophiacodontid.

Ophiacodontids include small to very large carnivorous pelycosaurs that have unusually long snouts formed by the elongation of the nasal, maxillary, and lacrimal bones. There is also a well developed lateral mandibular fenestra in a notch of the dentary. Among ophiacodontids there is a tendency towards increase in the skull to trunk ratio, larger taxa having unusually large, awkward looking, elongate skulls. In the postcranium, ophiacodontids have relatively large, massive shoulder girdles, with anteroposteriorly broad scapulocoracoids, but small headed interclavicles and clavicles. *Ophiacodon* is the best known member of the family, represented by an extensive fossil record in North America, that has been subdivided, somewhat arbitrarily (based largely on stratigraphic position and size) into six species. There appears to be extensive variation on morphology among the known specimens, even to the presence of two lateral temporal fenestrae on each side in larger specimens. In one particular specimen of *Ophiacodon retroversus*, the second opening is apparently present on one side of the skull but not on the other. A particularly puzzling characteristic of the postcranial skeleton is its reduced level of ossification, especially in the larges species, and the lack sharp claws as indicated by the truncated tip of the terminal phalanges. It is largely because of these features that *Ophiacodon* has been considered as being amphibious. This hypothesis is supported by the location of the orbit high on the side of the face, but it is nevertheless difficult to imagine how this animal, with a tall, slender skull, could be an effective aquatic predator.

Edaphosauridae

One of the most striking features exhibited by several pelycosaurs is the great elongation of the neural spines. The poorly known Middle Pennsylvanian *Echinerpeton*, the Lower Permian pelycosaur *Lupeosaurus*, at least three distinct genera of sphenacodontids, as well as all the edaphosaurs, have large sails supported by tall neural spines above the presacral portion of the vertebral column. Among these, edaphosaurs have the most spectacular sail because the greatly elongated neural spines have lateral projections or tubercles along each side. It is generally accepted that this great elongation of neural spines must have occurred independently at least three times among pelycosaurs. The presumed function of such elongation is for support of a membrane that spanned the space between successive spines. This membrane probably facilitated temperature control, but the sail and its associated spines may have also served in display behaviour, and in edaphosaurs it may have also served as protection against predators. This last hypothesis is supported by the presence of the lateral tubercles, as well as by the orientation of the neural spines; anteriorly tilted cervical and thoracic spines and posteriorly tilted lumbar spines probably provided protection in the neck and thigh region, respectively.

Edaphosaurus and *Ianthasaurus* are the only pelycosaurs that preserved the highly specialized lateral tubercles on the spines. Both are characterized by the presence of greatly elongated neural spines that are rounded in traverse section, and by the presence of remarkable lateral tubercles. The arrangement of these tubercles along the height of the spines is similar in the two taxa, the proximal lateral tubercles being not only the largest of the set, but also paired. *Ianthasaurus* is a small edaphosaur from the Upper Pennsylvanian that lacks many of the spectacular specialisations seen in *Edaphosaurus*. For example, the marginal dentition of *Ianthasaurus* is similar to that of insectivorous reptiles, with slender conical teeth which are slightly recurved at the tips, and there is a slight development of a caniniform region. The palatal and mandibular dentition is unspecialised, and there are no batteries of teeth for crushing of plant materials. This is in contrast to the cranial morphology of *Edaphosaurus*, where the palate has been greatly modified in order to carry a massive array of closely packed teeth that occluded against a similarly developed set of teeth on the medial surface of the mandible. The two edaphosaurs are also quite distinct from each other in their postcranial morphology. Whereas *Ianthasaurus* appears lightly built and was probably quite agile, *Edaphosaurus* possesses a combination of features that suggest that this animal was a heavy, relatively slow herbivore. For example, the cervical vertebrae are reduced in length in association with the reduced size of the skull, the dorsal vertebrae are massive, the tail is deep, the limbs are short, and robust and the ribs form a wide ribcage. Nine species of *Edaphosaurus* have been described, ranging in size from small to very large, bulky animals. The largest species, *Edaphosaurus cruciger* and *Edaphosaurus pogonias*, have modified their cervical and anterior thoracic spines into massive club-like processes. It is not unreasonable to suggest that the small, insectivorous *Ianthasaurus* represents the primitive edaphosaur pattern from which the larger herbivorous *Edaphosaurus* may have been derived.

Sphenacodontia

Haptodus is a relatively small pelycosaur that is considered to be a sister taxon to the sphenacodontians. *Haptodus* has been recovered from the Late Pennsylvanian and Early Permian of North America and Europe. Their morphology suggests that members of this genus were probably effective predators, being able to feed on both arthropods and small vertebrates. A number of haptodontine-like forms appear to represent progressively closer sister-taxa to the most spectacular pelycosaurs, the sphenacodonts. The Sphenacodontidae include the large carnivores *Dimetrodon*, *Sphenacodon*, *Ctenospondylus*, and *Secodontosaurus*. These forms were the dominant predators of their time, and their fossil record extends from the Late Pennsylvanian throughout the Early Permian. The significance of this group, however, lies with its phylogenetic relationships, because sphenacodontids are generally considered to be the nearest pelycosaurian relatives of therapsids. Sphenacodonts have a supracanine buttress on the medial surface of the maxilla, the first premaxillary and second dentary teeth are greatly increased in size, similar in length to the enlarged canine, the ventral edge of the premaxilla

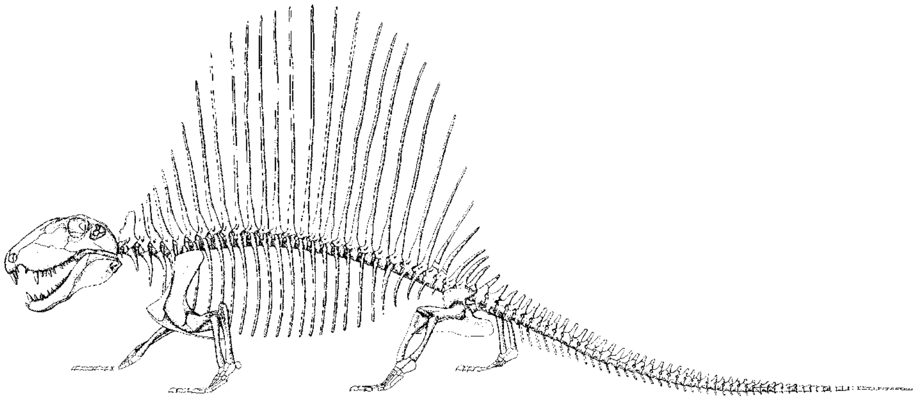


Figure 7 *Dimetrodon*, one of the earliest known top predators, an amniote synapsid (''mammal like reptile'') from the Permian of North America.

slopes anteroventrally, giving the snout a massive appearance, the supraorbital shelf forms a hood over the orbit and terminates posteriorly on the concave lateral surface of the postorbital, and the pineal foramen is located on a slightly raised area of the parietal, with a distinct ridge surrounding the foramen.

Sphenacodontids (*Dimetrodon*, *Sphenacodon*, *Secodontosaurus*, and *Ctenospondylus*) are all large to very large predators that have tall neural spines, strongly excavated centra and neural arches, and long limbs making them relatively agile, fast moving animals, especially when compared to their slower, bulkier herbivorous relatives. *Dimetrodon*, *Sphenacodon*, and *Ctenospondylus* all have similarly constructed, massive skulls with extremely large anterior incisors and canines (Figure 7). *Sphenacodon* and *Ctenospondylus* have blade-like neural spines, whereas *Dimetrodon* has greatly elongated spines that are rounded in transverse section. *Secodontosaurus* is greatly modified from the pattern seen in the other large sphenacodontids by reducing the width and height of the skull roof, while retaining the neural spine morphology of *Dimetrodon*. The cranial modifications of *Secodontosaurus* can only be interpreted as an adaptation to specialised feeding strategies, perhaps preying upon burrowing animals, or feeding on tetrapods that tried to evade capture by hiding in crevices.

Origin of Therapsids

A major morphological gap exists between pelycosaurs, which are mostly Pennsylvanian and Early Permian synapsids from North America and Europe, and therapsids, which are Late Permian and Triassic advanced synapsids from South Africa and Russia. Previous attempts at bridging this gap have not been successful. Among pelycosaurs the haptodontines and sphenacodontids have been proposed as the most likely candidates for sister-group status to therapsids. In 1940 it was proposed that of all the known sphencodontids *Haptodus* is the most likely ancestor of therapsids. More recently, in 1992, it was shown that the pattern of distribution of synapomorphies in haptodontines, sphenacodontids, and therapsids supports the hypothesis that sphenacodontids are the sister group of therapsids. The morphology of the reflected lamina of the angular provides the strongest evidence for the latter hypothesis: in both primitive therapsids and sphenacodontids, the reflected lamina is a prominent feature of the lower jaw, whereas in haptodontines the reflected lamina, if present at all, is a modest, slightly developed ventral projection with little or no lateral displacement.

Recent studies of *Tetraceratops* indicate that this enigmatic fossil from the Early Permian of Texas bridges the morphological and temporal gaps between the pelycosaurian and basal therapsid synapsids. As the oldest known therapsid, this form shares with other early therapsids a number of derived cranial features, but it also has numerous primitive features that appear in derived form in all other therapsids. *Tetraceratops* is, therefore, the sister taxon to all other therapsids.

Further Reading

Boy JA and Martens T (1991) Ein neues captorhinomorphes Reptil aus dem thüringischen Rotliegend (Unter Perm; Ost Deutschland). *Paläontologische Zeitschrift* 65: 363 389.

Carroll RL (1963) The earliest reptiles. *Zoological Journal of the Linnean Society* 45: 61 83.

Carroll RL (1988) *Vertebrate Paleontology and Evolution.* New York: W.H. Freeman and Company.

Carroll RL and Baird D (1972) Carboniferous stem reptiles of the Family Romeriidae. *Bulletin, Museum of Comparative Zoolology* 143: 321 364.

Clark J and Carroll RL (1973) Romeriid reptiles from the Lower Permian. *Bulletin, Museum of Comparative Zoolology* 144: 353 407.

deBraga M and Reisz RR (1995) A new diapsid reptile from the uppermost Carboniferous (Stephanian) of Kansas. *Palaeontology* 38: 199 212.

Dodick JT and Modesto SP (1995) The cranial anatomy of the captorhinid reptile *Labidosaurikos meachami* from the Lower Permian of Oklahoma. *Palaeontology* 38: 687 711.

Dilkes DW and Reisz RR (1996) First record of a basal synapsid ('mammal like reptile') in Gondwana. Proceedings of the Royal Society of London, *Series B* 263: 1165 1170.

Laurin M (1993) Anatomy and relationships of *Haptodus garnettensis*, a Pennsylvanian synapsid from Kansas. *Journal of Vertebrate Paleontology* 13: 200 229.

Laurin M and Reisz RR (1990) *Tetraceratops* is the oldest known therapsid. *Nature* 345: 249 250.

Laurin M and Reisz RR (1995) The osteology and relationships of *Tetraceratops insignis*, the oldest known therapsid. *Journal of Vertebrate Paleontology* 16: 95 102.

Laurin M and Reisz RR (1995) A reevaluation of early amniote phylogeny. *Zoological Journal of the Linnean Society* 113: 165 223.

Modesto SP and Reisz RR (1990) A new skeleton of *Ianthasaurus hardestii*, a primitive edaphosaur (Synapsida: Pelycosauria) from the Upper Pennsylvanian of Kansas. *Canadian Journal of Earth Sciences* 27: 834 844.

Modesto SP (1995) The skull of the herbivorous synapsid *Edaphosaurus boanerges* from the Lower Permian of Texas. *Palaeontology* 38: 213 239.

Modesto SP (1996) A basal captorhinid reptile from the Fort Sill fissures, Lower Permian of Oklahoma. *Oklahoma Geology Notes* 56: 4 14.

Olson EC (1991) An eryopoid (Amphibia: Labyrinthodontia) from the Fort Sill fissures, Lower Permian, Oklahoma. *Journal of Vertebrate Paleontology* 11: 130 132.

Reisz RR (1977) *Petrolacosaurus*, the oldest known diapsid reptile. *Science* 196: 1091 1093.

Reisz RR (1997) The origin and early evolutionary history of amniotes. *Trends in Ecology and Evolution* 12: 218 222.

Reisz RR, Heaton MJ, and Pynn BR (1982) Vertebrate fauna of Late Pennsylvanian Rock Lake Shale near Garnett, Kansas: Pelycosauria. *Journal of Paleontology* 56: 741 750.

Reisz R (1972) Pelycosaurian reptiles from the Middle Pennsylvanian of North America. *Bulletin of MCZ* 144: 27 61.

Reisz RR (1986) Pelycosauria. vol. 17 A. *Handbuch der Paläoherpetologie*. Wellenhofer P (ed.) Stuttgart, New York: Gustav Fischer Verlag, 1 102.

Reisz RR, Berman DS, and Scott D (1992) The cranial anatomy and relationships of *Secodontosaurus*, an unusual mammal like reptile (Synpasida: Sphenacodontidae) from the early Permian of Texas. *Zoological Journal of the Linnean Society* 104: 127 184.

Reisz RR and Modesto SP (1996) *Archerpeton anthracos* from the Joggins Formation of Nova Scotia: a microsaur, not a reptile. *Canadian Journal of Earth Sciences* 33: 703 709.

Romer AS and Price LW (1940) Review of the Pelycosauria. *Geologica Society of America Special Papers* 28: 1 538.

Dinosaurs

A M Yates, University of the Witwatersrand, Johannesburg, South Africa

Introduction

Dinosaurs were the dominant terrestrial vertebrates of the Mesozoic Era. All large terrestrial animals from the end of the Norian Stage (Late Triassic) to the close of the Cretaceous were dinosaurs. This is not to say that any large Mesozoic animal is automatically referred to as a dinosaur. Membership of the group can be diagnosed by the presence of numerous evolutionary novelties (Figure 1), which indicate that dinosaurs are a natural evolutionary group or clade.

On the evolutionary tree of terrestrial vertebrates, dinosaurs nest deeply within the Reptilia. Within the reptiles they belong to the Diapsida (reptiles with two openings, or fenestrae, for jaw muscles in the temporal regions of their skulls), within the diapsids they belong to the Archosauria (diapsids with an additional fenestra in the skull behind the nostril and in front of the orbit and that lay calcite-shelled eggs), and within the archosaurs they belong to the Avemetatarsalia (archosaurs with a simple hinge-like, or mesotarsal, ankle joint and elongate bunched metatarsal bones in the foot).

The Dinosauria contains two subgroups (Figure 2), the Ornithischia ('bird-hipped' dinosaurs) and the Saurischia ('lizard-hipped' dinosaurs). These subgroups were long thought to have evolved from different reptiles, but modern cladistic work has shown that they almost certainly form a clade that excludes all other archosaurian groups.

Diagnostic Characters of the Dinosauria

A great number of derived characteristics have been proposed for the diagnosis of the Dinosauria. However, many have since been shown to be present in other archosaurs (e.g. an s-shaped neck) or to be independent acquisitions in different dinosaur groups (e.g. three or more sacral vertebrae). The following list includes the characters that have been well documented and are likely to remain diagnostic of the dinosaur clade.

1. Loss of the postfrontal bone from the skull roof. In other archosaurs there is a postfrontal bone between the frontal and postorbital bones.
2. A contribution of the frontal bone to the depressed area that surrounds the upper temporal fenestra. In life this fossa would have served as an attachment site for jaw closing muscles. In other archosaurs the frontal bone is excluded from this fossa.
3. Loss of the lateral sheet of the squamosal bone that covers the dorsal head of the quadrate bone. The dorsal end of the reptilian quadrate bone (which forms the upper part of the jaw joint) forms a smooth, rounded head that fits into a socket under the squamosal bone. In other archosaurs a lateral sheet of the squamosal bone obscures the quadrate head in lateral view but in dinosaurs the quadrate head is plainly visible.
4. The sternum consists of paired ossified plates. In early tetrapods the sternum was composed of a single median interclavicle situated behind the paired clavicles ('collar bones' in humans). The interclavicle was lost early in the history of archosaurs, while the clavicles were strongly reduced in size.

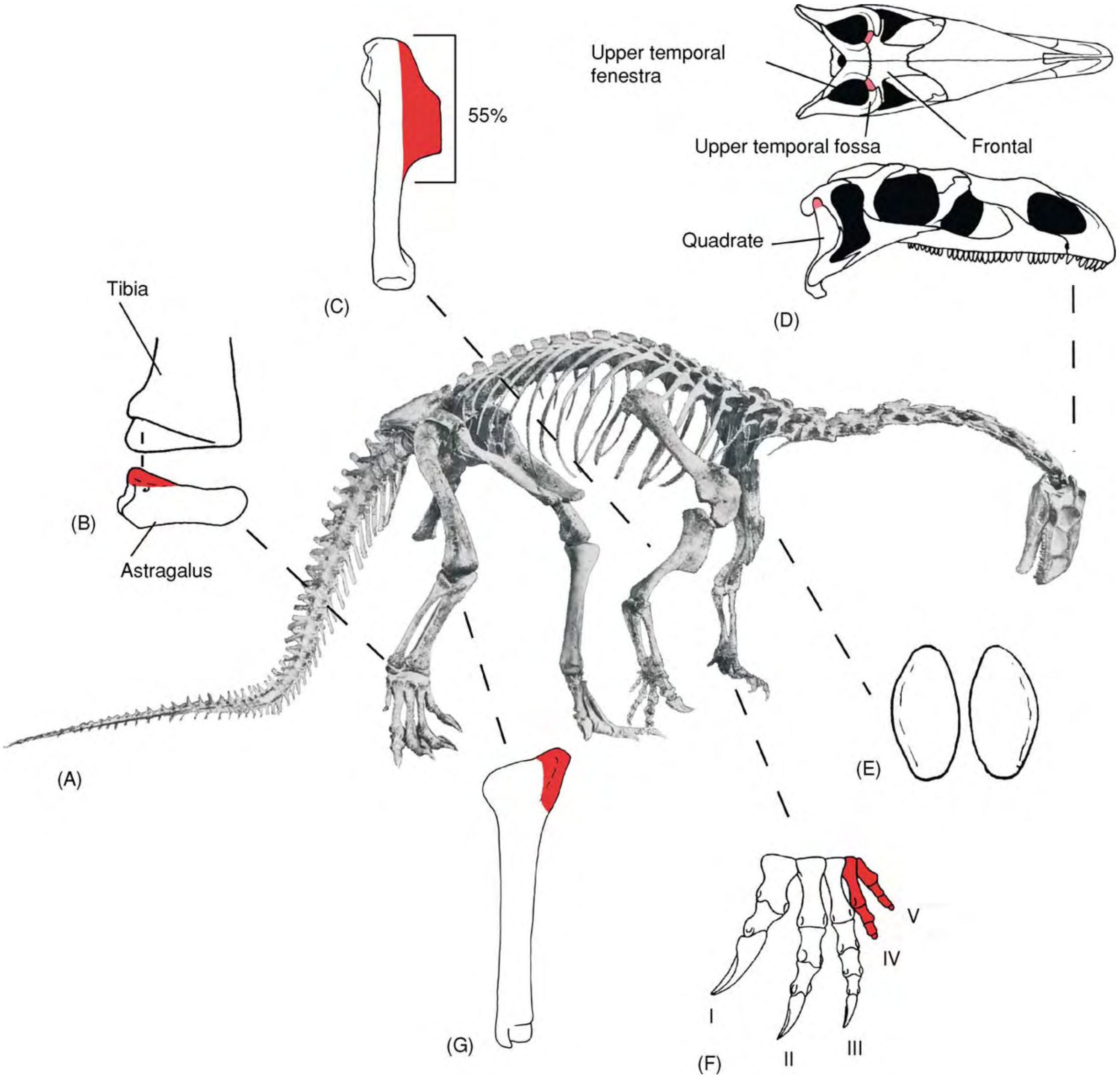


Figure 1 The diagnostic characters of the Dinosauria as shown by (A) the sauropodomorph *Plateosaurus.* (B) Distal part of the tibia and astragalus in anterior view, showing the ascending process (red) and the notch it fits into. (C) Right humerus in lateral view, showing the enlarged (over 30% of humeral length) and rectangular deltopectoral crest (red). (D) Skull in dorsal (top) and lateral (bottom) views, showing the exposure of the frontal bone in the upper temporal fossa (top) and the lateral exposure of the quadrate head (bottom) in red. (E) Paired sternal plates. (F) Left hand in dorsal view, showing the reduced and clawless digits IV and V (red). (G) Right tibia in lateral view showing the cnemial crest (red).

Dinosaurs evolved two new plates of bone to cover the anterior end of the chest. These plates are paired and lie on each side of the midline (Figure 1E).

5. The deltopectoral crest of the humerus is an elongated rectangular crest that extends for more than 35% of the length of the humerus (Figure 1C).
6. The outer (fourth and fifth) digits of the hand are short and clawless. This character is particularly unusual and is strong evidence that dinosaurs are indeed a natural group. Reduction or loss of digits is a common event in the evolutionary history of terrestrial vertebrates; however, it almost always occurs in a symmetrical manner about the central (third) digit. Furthermore, the first digit to be reduced is the innermost (in the hand this is the thumb or pollex) because it is the last digit to form during the development of the embryo. The asymmetrical hand of dinosaurs with the area of reduction centred upon the outermost digits (the first to form during

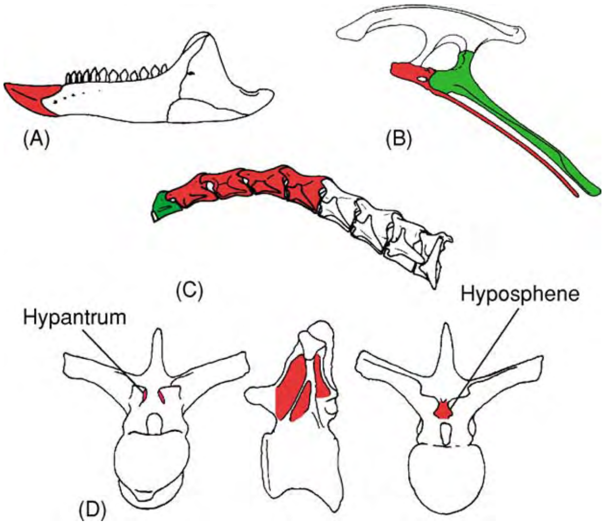


Figure 2 The diagnostic characters of (A, B) the Ornithischia and (C, D) the Saurischia. (A) Left lower jaw of an ornithischian in lateral view, showing predentary bone (red). (B) Pelvis of an ornithischian in left lateral view. The shaft of the pubis (red) has rotated backwards so that it lies parallel to the ischium (green). (C) The cervical series of a saurischian in left lateral view. The first four postaxial vertebrae (red) are elongate and exceed the axis (green) in length. (D) A trunk vertebra of a saurischian in (from left to right) anterior, left lateral, and posterior views. Red areas show the accessory articulations in the anterior and posterior views, while the lateral view shows the large pneumatic fossae separated by thin laminae. Figures not to scale.

development) is unique among digit-bearing vertebrates (Figure 1F).

7. There is a cnemial crest. This is a vertical crest of bone that grows out of the anterior face of the proximal end of the tibia, just below the knee joint. It supported various lower-leg muscles and seems to be correlated with enhanced running ability. No other archosaurs have a cnemial crest, but other terrestrial vertebrates, such as cursorial mammals, do (Figure 1G).
8. The astragalus has an ascending process on its anterior edge. The astragalus is the larger of the two bones in the proximal row of the tarsus (ankle joint). Dinosaurs have a mesotarsal ankle, where there is a simple straight hinge running between the proximal and distal rows of bones in the tarsus. The ascending process is a triangular block of bone that fits into a corresponding notch in the distal tibia, thus preventing any movement in this part of the ankle (Figure 1B).

Origin of the Dinosauria

The closest known relatives of the dinosaurs are small slender avemetatarsalians from the Middle Triassic of South America (Figure 3A). These were relatively unspecialized and form good models for a dinosaurian precursor. *Marasuchus* is the best represented and most thoroughly studied of these animals. It was exceptionally small compared with its dinosaurian cousins, standing no more than 150 mm high at the hip and weighing no more than a few hundred grams. It was an agile long-legged carnivorous biped.

By the end of the Carnian Stage of the Late Triassic we have the first known remains of true dinosaurs. They are scattered around the world, but the best specimens come, once again, from South America. Though much larger than the earlier avemetatarsalians (*Herrerasaurus* weighed around 200 kg), the early dinosaurs were neither the largest terrestrial animals of their time nor particularly abundant in terms of numbers of species or individuals. Nevertheless, a modest radiation had occurred with carnivores (*Eoraptor, Staurikosaurus*, and *Herrerasaurus*), omnivores (*Saturnalia*), and herbivores (*Pisanosaurus*) all represented. Some palaeontologists have proposed that a small but significant extinction event hit many of the dominant groups of terrestrial vertebrates at the end of the Carnian, leaving many of the large animal niches vacant in the succeeding Norian Stage, but this hypothesis remains controversial. For whatever reason, dinosaurs dominated the large-herbivore niches around the globe in the Norian, and smaller carnivorous dinosaurs were common. Nevertheless, other tetrapod groups were still significant components of the large terrestrial faunas. These groups included the last of the dicynodont synapsids and various non-dinosaurian archosaurs, including the rauisuchians, which were the top predators of the time. However, by the beginning of the Jurassic Period these groups had become extinct and the ascendancy of the dinosaurs was complete. There may have been another mass extinction at the Triassic–Jurassic boundary, and there is some evidence, from eastern North America, for an extraterrestrial bolide impact at that time.

Dinosaur Subgroups

Ornithischia

The Ornithischia are a distinctive group, and there is no doubt that they are clade. Even the earliest ornithischians were specialized towards herbivory (although some of the earliest members may well have been somewhat omnivorous), and all subsequent species seem to have been herbivorous. Two of the most distinctive derived characters are the arrangement of the pelvic bones (Figure 2B) and the presence of an extra bone, the predentary (Figure 2A), in the lower jaw. In ornithischians the symphysis between

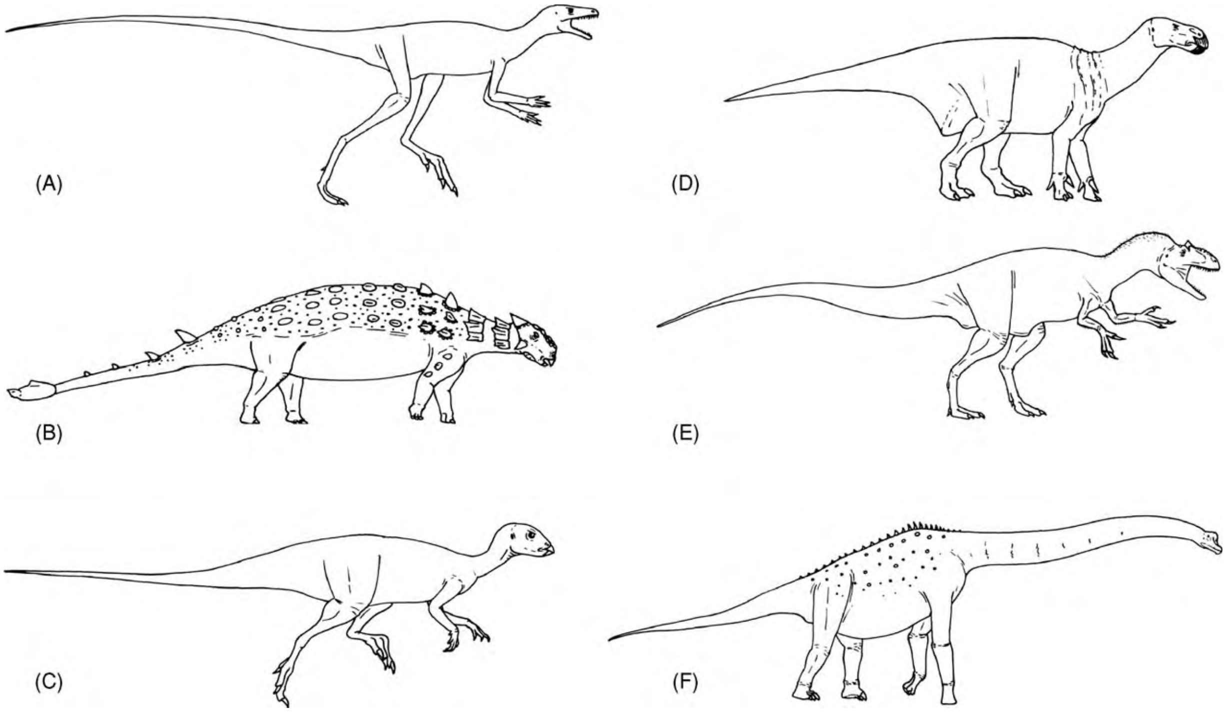


Figure 3 A sample of the diversity of dinosaurs and their closest relatives. (A) *Marasuchus*, a dinosaurian precursor (total length of 0.5 m). (B) *Saichania*, an ankylosaurid thyreophoran (total length of 6.5 m). (C) *Hypsilophodon*, a neornithischian that retained the primitive ornithischian body plan (total length of 1.5 m). (D) *Iguanodon*, a derived neornithischian (total length of 11 m). (E) *Allosaurus*, a typical theropod (total length of 8 m). (F) *Brachiosaurus*, a sauropod (total length of 21 m). Figures not to scale.

the pubic bones is interrupted and each pubis is rotated backwards to extend parallel to the ischia. This may have allowed the gut to expand backwards into the space between the legs, thus increasing its length (to deal with poorly digestible plant material) without shifting the centre of gravity forwards. The position of the centre of gravity is important as the early ornithischians retained the bipedal locomotion of their ancestors. Another feeding adaptation was the development of a predentary bone at the front of the mandible. This acted as a stable single-piece platform to support a horny beak while freeing the two mandibular rami to rotate slightly around their long axes, thus increasing the amount of tooth-on-tooth contact during chewing.

Ornithischians were a rare group during the Triassic. It was only in the Jurassic that they became abundant and diverse. These more successful ornithischians comprise two large clades, the Thyreophora and the Neornithischia.

Thyreophora Thyreophorans abandoned a bipedal cursorial lifestyle for a slower quadrupedal existence. They retained a primitive simple set of teeth and jaws and clearly relied heavily on their capacious guts to deal with their plant diet. Instead of using speed to escape predation, thyreophorans evolved a protective coat of keratin-covered dermal bones, or osteoderms, arranged in longitudinal rows. In early forms the osteoderms were simple studs, but they became highly elaborated in the two main thyreophoran groups: the stegosaurs and the ankylosaurs. In the stegosaurs the rows of osteoderms on each side of the dorsal midline were enlarged into a series of large plates and spines at the expense of the rest of their armour. In contrast, ankylosaurs increased the number of osteoderms to create a protective pavement that covered their dorsal surface (Figure 3B). One Late Cretaceous group, the Ankylosauridae, developed a large bony club at the end of the tail, which could be wielded like a mace.

Neornithischia Unlike the thryreophorans, the neornithischians evolved a number of specializations of their jaws and teeth for improved oral processing of plant matter. These included a larger cropping beak at the front of the jaw (indicated by an enlarged predentary bone), a diastema between the beak and the cheek teeth, and a thick layer of enamel covering one side of the ‘cheek’ teeth. This last feature creates a ridge of hard enamel that juts above the dentine when the tooth is worn down. Thus, a rougher file-like chewing surface is developed for efficient processing of plant matter. Apart from these features, the neornithischians display a great diversity of body plans. Early forms were small and bipedal, superficially resembling basal ornithischians (Figure 3C). Some primitive taxa (e.g. *Thescelosaurus*) kept this conservative body plan right through to the end of the Cretaceous. Two separate clades, the Neoceratopsia and the Iguanodontia (Figure 3D), continued to improve the oral processing of their food and culminated in the Ceratopsidae and the Hadrosauridae, respectively. Members of these clades became important herd-forming herbivores in the Late Cretaceous. Both developed batteries of hundreds to thousands of continuously replaced teeth to grind their food efficiently. They also developed large apparently species-specific display structures (frills and horns in the ceratopsids, expanded nasal crests in the hadrosaurids), which may have been significant for maintaining social structure within their herds.

Saurischia

At first sight saurischians are less obviously a natural group than the ornithischians. The 'lizard-like' hip is a primitive feature that is also found in many other reptiles and is not evidence of a close relationship. Nevertheless, there are several derived characters that unite the saurischians together as a clade. Many of them are modifications of the vertebral column. These include a lengthening of the anterior cervical (neck) vertebrae (Figure 2C), the presence of extra articulations between the vertebrae of the trunk (called the hyposphene–hypantrum system; Figure 2D), and the excavation of the vertebrae by large hollows that are bounded by thin webs or struts of bone (Figure 2D). The additional articulations in the trunk vertebrae stiffen the vertebral column and reduce flexibility, while the longer cervical vertebrae lengthen the neck. Air-filled outpockets of the lungs occupy the vertebral excavations in the only surviving saurischian group (birds), and it is reasonable to assume that this was also the case in extinct saurischians.

Two large clades arose from within the Saurischia: the largely carnivorous Theropoda and the predominantly herbivorous Sauropodomorpha. There is much debate about whether some of the very earliest saurischians (Herrerasauridae and *Eoraptor*) were early examples of the Theropoda or represent early branches of the saurischian tree that diverged before the split between Theropoda and Sauropodomorpha.

Sauropodomorpha Sauropodomorphs were saurischians characterized by long necks, small heads, and leaf-shaped teeth. These adaptations indicate that they were eating significant amounts of plant matter. However, early examples (which are sometimes placed in a clade of their own, the Prosauropoda) lack specialized chewing mechanisms, tooth-on-tooth contact, and large guts to deal with a strict diet of foliage. Thus, it seems likely that they fed on softer highly nutritious parts of plants (new shoots, fleshy reproductive parts, etc.), perhaps supplemented with small prey items and carrion. These early sauropodomorphs also had a generalized skeleton that could accommodate both bipedal and quadrupedal locomotion. In contrast, the Sauropoda, which first appeared in the Late Triassic but did not rise to prominence until the Middle Jurassic, consisted of strictly herbivorous obligate quadrupeds (Figure 3F). The Sauropoda are justly famous for producing the largest land-dwelling animals of all time. Even the smallest species weighed several tons and reached 10 m in length. The largest, for which we have only a few bones, may have reached 40 m in length and weighed up to 100 tons. They had massively constructed pillar-like limbs and an exceptionally small head relative to body size, which was perched atop an extremely elongated neck. They are often characterized as having very weak dentition, but their oral processing was more extensive than that of prosauropods, with broad jaws, precise tooth-on-tooth contact, and improved jaw musculature (shown by the development of a tall coronoid process in the lower jaw). Sauropod diversity in the northern continents declined in the Early Cretaceous, whereas they appear to have remained the dominant large herbivores in South America and Africa. By the Late Cretaceous titanosaurian sauropods were the dominant large herbivores in most parts of the world except North America and Asia. These were robust, and sometimes armoured, sauropods with distinctive articulations between the tail vertebrae. The group includes both the largest and some of the smallest sauropods.

Theropoda Most Mesozoic species retained the predatory bipedal way of life that they inherited from the earliest dinosaurs. Most were long-legged hunters with large skulls and serrated blade-like teeth (Figure 3E). Despite their outwardly conservative body form, the group shares a great number of anatomical specializations that make it clear that they are a clade. These include a deep, probably pneumatic, recess on the ventral surface of the braincase, a distally stiffened tail (forming a dynamic stabilizer), fusion of the clavicles into a furcula ('wishbone'), complete loss of the fifth finger, and a splint-like first metatarsal that carries a functional toe but does not reach the ankle joint. Although it is clear that almost all theropods were carnivorous, it is difficult to find conclusive evidence that they were predators as opposed to scavengers of carrion. Nevertheless, all modern terrestrial carnivores include the hunting and killing of prey in their behavioural repertoire, and it is most unlikely that theropods did not. The strong build of their skulls and their grasping hands seem adapted for grappling with powerful struggling prey, supporting this idea. One clade of derived theropods, the Coelurosauria, produced a range of divergent body plans. This radiation probably began in the Middle Jurassic, but coelurosaurs did not become abundant until the Cretaceous. The group included gigantic carnivores (tyrannosaurids), toothless presumed herbivores (ornithomimosaurs and oviraptorosaurs), and long-necked herbivores (therizinosaurs). The radiation also produced the sickle-clawed deinonychosaurs and their close relatives the flying dinosaurs or birds.

Bird Origins

Dinosaurs did not become entirely extinct at the end of the Cretaceous. One highly specialized group of theropods, the birds, survived and are still a major part of modern vertebrate faunas (*see* **Fossil Vertebrates:** Birds). The great preponderance of evidence from skeletal anatomy, eggshell microstructure, and integumentary structures indicates that the origin of birds lies deep within the theropod clade, with the Deinonychosauria being their closest relatives. These two groups share a number of significant similarities, including a decoupling of the tail from the muscular system driving the hind legs (shown by the extreme mobility of the articulations of the proximal tail vertebrae and the reduction of the bony attachment sites for the caudofemoralis muscle), retroversion of the pubis (similar to the condition in Ornithischia), and a laterally facing shoulder socket that allows a flapping motion of the forelimb. The caudofemoralis muscle is the main leg-retracting muscle in modern reptiles and extends from the posterior surface of the femur to the side of the tail, keeping the two organs tightly coupled during locomotion.

The most compelling evidence for the dinosaur ancestry of birds is the discovery of numerous species of non-avian coelurosaur that are preserved with a covering of feathers in the Early Cretaceous rocks of the Liaoning Province, China. These range from simple filaments in the primitive *Sinosauropteryx* to complex vaned feathers (with a central rachis and lateral barbs) in more derived species such as *Caudipteryx*, *Protarchaeopteryx*, *Sinornithosaurus*, and *Microraptor*.

The environment in which birds evolved and the possible behaviours and selection pressures that led to the origin of flight are more difficult to elucidate. Critics of the dinosaur hypothesis for the origin of birds point to the improbability of flight evolving in a ground-dwelling runner. Nevertheless there is no logical reason why some dinosaurs could not have adopted a scansorial (climbing) existence. Indeed, some of the newly discovered small deinonychosaurs from Liaoning seem to show some scansorial adaptations.

Physiology

No other area of investigation into dinosaur biology has produced as much controversy as the issue of their metabolic and thermal physiology. Although the debate is often portrayed as a simple dichotomy between ‘warm-blooded’ and ‘cold-blooded’ dinosaurs, the issue is more complicated.

Few doubt that dinosaurs were capable of sustained aerobic exercise and that they maintained at least some degree of homeothermy (stable body temperatures); these are not typical characteristics of modern ectotherms (animals that require an external heat source to reach their optimum body temperature). Evidence for this comes from their erect gaits, their frequent cursorial adaptations, and the fibrolamellar microstructure of their bones (indicating fast growth). The question remains, did they achieve homeothermy through an elevated resting metabolic rate (tachymetabolism) and the internal heat that it generates (endothermy), or did they use other means, such as the thermal inertia resulting from their large size and the fairly equable climate of the Mesozoic? Most modern ectotherms cannot maintain homeothermy (a condition known as poikilothermy) but there are exceptions, such as the leatherback turtle (*Dermochelys*).

Most of the evidence that has been used to support endothermy in non-avian dinosaurs either implies homeothermy, but not necessarily endothermy, or simply cannot be trusted owing to the distortions caused by the processes of taphonomy and diagenesis (e.g. predator–prey ratios and stable isotope signatures in bone). The evidence for ectothermy is also weak, although it has been claimed by some researchers that the absence of nasal passages large enough to house respiratory turbinate bones is the ‘Rosetta Stone’ that demonstrates dinosaurian ectothermy. However, the recent discovery of small coelurosaurs that were insulated by feathers means that it is almost certain that at least these dinosaurs were endotherms. Since coelurosaurs are included amongst the dinosaurs that have been claimed to be ectotherms by virtue of their small nasal passages, it would now seem that nasal-passage size and thermal physiology are not as tightly correlated as the proponents of this hypothesis contend.

Reproduction and Growth

Like all other archosaurs, dinosaurs laid calcitic-shelled eggs. Eggs have been found that (owing to the presence of embryos) can be attributed to theropods, neornithischians, titanosaurian sauropods, and prosauropods. Many other egg types have been found, and these probably represent all the major dinosaur groups. However, matching an egg type to a producer without direct evidence from embryonic remains is fraught with difficulty. An intriguing feature of non-avian dinosaur eggs is their relatively small size. The largest non-avian dinosaur eggs are about 4 l (laid by gigantic sauropods) and are actually smaller than the largest bird eggs, which are about 9 l (laid by the recently extinct elephant bird). A further difference is the average number of eggs laid in a single clutch. Fossilized non-avian dinosaur clutches usually contain tens of eggs (up to 40), but the actual average was probably higher, as fossil clutches are almost always partly destroyed by erosion when discovered. Avian clutch sizes can be as low as one and are usually less than ten. Further differences are that non-avian dinosaurs partly buried their eggs and apparently did not turn them once laid. However, oviraptorosaur skeletons found on top of their nests show that they did assume a bird-like brooding posture over their eggs. Furthermore, hadrosaurids show evidence of post-hatching parental care. The degree of ossification of the articular ends of the baby limb bones found inside nest structures indicates that they were nest-bound and must have relied on food bought back to them by their parents.

Owing to the great size difference between adult and hatchling dinosaurs (three, perhaps as much as four, orders of magnitude), young dinosaurs must have grown rapidly, which is borne out by the fast-growth-style microstructure of their bones. Estimates of actual growth rates have been obtained by matching the histological type of fossil taxa to modern taxa of known growth rate or by counting lines of arrested growth (LAGs), which are assumed to be seasonal. The range of growth rates that can produce each particular type of bone histology has not been adequately explored, and recent evidence suggests that the range is broader than previously suspected. Thus, estimates based on the first method are suspect. LAGs have proved even more unreliable, with the number of LAGs varying widely between different bones of a single individual. Experimental work has shown that LAG formation is not necessarily correlated with seasonal hardship. Until the biological meaning of LAGs is better understood, they should not be used to determine growth rates in dinosaurs.

See Also

Fossil Vertebrates: Reptiles Other Than Dinosaurs; Birds; Flying Reptiles. **Mesozoic:** Triassic; Jurassic; Cretaceous; End Cretaceous Extinctions.

Further Reading

Benton MJ (1983) Dinosaur success in the Triassic: a non competitive ecological model. *Quarterly Review of Biology* 58: 29 55.

Farlow JO and Brett Surman MK (1997) *The Complete Dinosaur.* Bloomington: Indiana University Press.

Farlow JO, Dodson P, and Chinsamy A (1995) Dinosaur biology. *Annual Reviews of Ecology and Systematics* 26: 445 471.

Horner JR (2000) Dinosaur reproduction and parenting. *Annual Reviews of Earth and Planetary Science* 28: 19 45.

Ji Q, Currie PJ, Norell MA, and Ji S A (1998) Two feathered dinosaurs from northeastern China. *Nature* 393: 753 761.

Olsen PE, Kent DV, Sues H D, *et al.* (2002) Ascent of dinosaurs linked to an iridium anomaly at the Triassic Jurassic boundary. *Science* 296: 1305 1307.

Padian K, de Ricqlès AJ, and Horner JR (2001) Dinosaurian growth rates and bird origins. *Nature* 412: 405 408.

Sereno PC (1997) The origin and evolution of dinosaurs. *Annual Reviews of Earth and Planetary Science* 25: 435 489.

Sereno PC (1999) The evolution of dinosaurs. *Science* 284: 2137 2147.

Starck JM and Chinsamy A (2002) Bone microstructure and developmental plasticity in birds and other dinosaurs. *Journal of Morphology* 254: 232 246.

Weishampel DB, Dodson P, and Osmolska H (1990) *The Dinosauria.* Berkeley: University of California Press.

Birds

G J Dyke, University College Dublin, Dublin, Ireland
L M Chiappe, Natural History Museum of Los Angeles County, Los Angeles, CA, USA

Introduction

The evolutionary history of birds – class Aves – and their relationships with other groups of fossil and living vertebrates have been debated for more than 100 years. Now, however, the overwhelming consensus among palaeontologists and zoologists is that all living birds and their fossil cousins are the descendants of carnivorous theropod dinosaurs; indeed, evolutionary hypotheses place Aves as a nested group within this subdivision of the dinosaurian family tree (Figure 1). As a group, birds are characterized by their active flapping flight and have an almost uninterrupted fossil history from the Late Jurassic (140 Ma) to the Holocene. The earliest phases of their evolution occurred in the Cretaceous, possibly including the origination of modern birds (Neornithes), whose extant orders and families are first recorded in the latest Cretaceous–earliest Tertiary, some 60–70 Ma ago. Throughout the Cenozoic birds radiated (diverged and diversified) to produce the approximately 10 000 species alive today. In this article, we discuss many of the key fossil taxa that have aided our understanding of avian evolution throughout the Mesozoic, outline the geological environments in which they are preserved, and present arguments for the pattern of the evolutionary radiation of modern birds. Although birds are among the most familiar and immediately recognizable of the major living vertebrate groups, very little was known about their early evolutionary history until just a few years ago – recent times have been very exciting for Mesozoic avian palaeontologists: in just the last five years, more fossil bird taxa have been discovered and described from the Cretaceous alone than were known for much of the preceding century.

Birds from the Mesozoic: Not just *Archaeopteryx*

Although it was described nearly 150 years ago from the lagoonal limestones of Solnhofen in Bavaria (Germany), *Archaeopteryx* is still the oldest recognized bird. Eight skeletal specimens and a feather are all that is known of this primitive taxon; all have been found in the 140 Ma old Solnhofen limestone deposits, which were formed in a poorly oxygenated hypersaline shallow tropical lagoon, thus facilitating the exquisite preservation of these specimens. A great deal has been written about *Archaeopteryx*. – How well was it able to fly? Was it an arboreal or a terrestrial animal? Do all known specimens come from one or several closely related species? Since its initial description in the 1860s, one thing has remained clear: *Archaeopteryx* provides a tantalizing glimpse of the earliest stages of avian evolution. Having a long bony tail, sharply clawed forelimbs, and a primitive pelvis, this toothed bird is in many ways anatomically more similar to theropod dinosaurs than to today's birds. At one time *Archaeopteryx* was one of very few fossils from the Mesozoic upon which solid evolutionary inferences could be based; however, over the course of the last two decades, more and more fossil information has come to light, revealing an unexpectedly large diversity of primitive birds that existed throughout the last half of the Mesozoic Era. These fossil birds vary vastly in age, degree of preservation, and the environments they inhabited. Although none of these newly discovered lineages have left living descendants, the many exquisite fossils allow us to reconstruct patterns of morphological change, including the origin and development of modern avian wing anatomy. They also furnish evidence documenting the pattern of genealogical relationships of early birds and help us better understand the radiation of extant species.

In addition to *Archaeopteryx*, new fossils that have come to light over the last few years include several of the other most primitive birds known, *Jeholornis* and *Zhenzhoraptor* (which are perhaps synonymous taxa) and *Rahonavis*, from the Early and Late Cretaceous of China and Madagascar, respectively. Although both these taxa retain the long bony tails of *Archaeopteryx* and other non-avian theropods, they show that more advanced shoulder girdles and wings (which characterize more derived birds) had begun to develop by the earliest Cretaceous (*ca*. 125 Ma ago) (Figure 2). *Rahonavis* is remarkably theropod-like in its anatomy, even retaining the enlarged sickle claw seen on the feet of theropods such as dromaeosaurids. Even though *Rahonavis* is much younger than *Archaeopteryx* (more than 60 Ma younger), the two birds may be closely related.

Fossil discoveries of birds at the base of the avian evolutionary tree, close in their anatomy to *Archaeopteryx*, have characterized some, but not all, of the

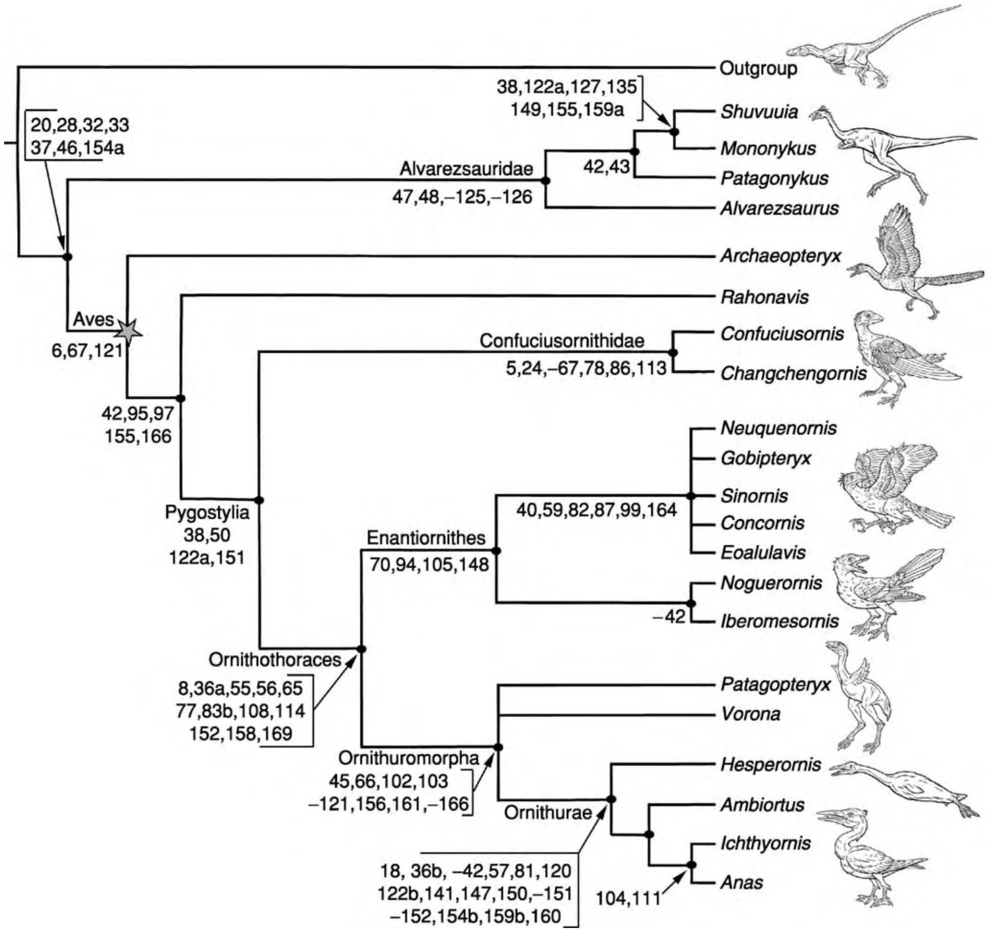


Figure 1 Cladogram of Mesozoic bird relationships showing theropod dinosaurs at the base of the tree. Many of the taxa seen at the tips of this tree have been described since 1990. Character numbers and names of nodes are taken from Chiappe (2001).

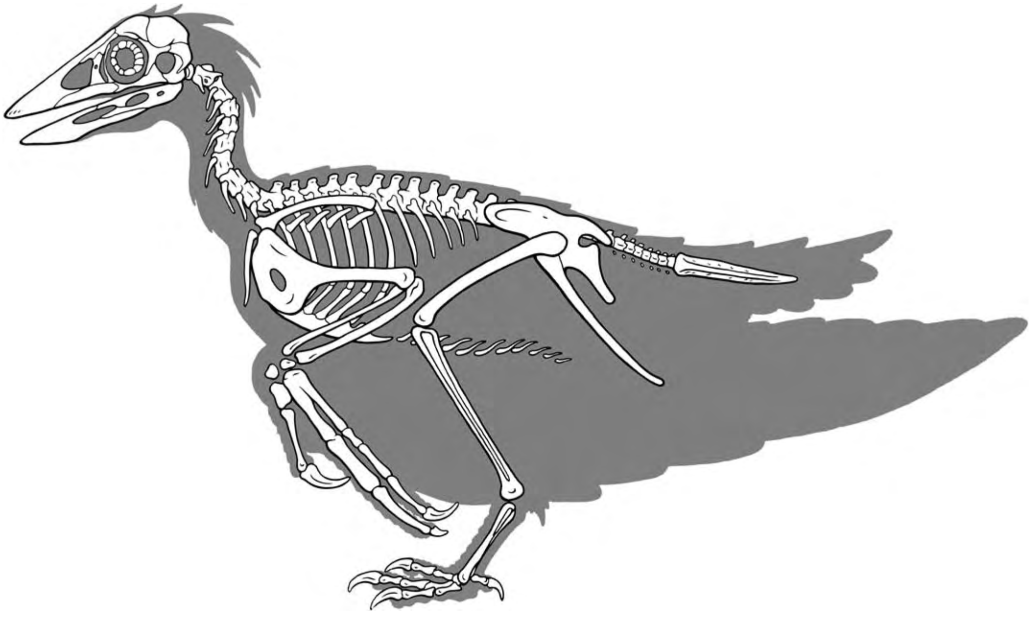


Figure 2 Artist's rendering of *Confuciusornis*, a primitive fossil bird from the Cretaceous of China.

advances in the study of avian evolution. Arguably the most significant development in this area came in the early 1980s with the discovery of an entirely new group of Cretaceous birds, the subclass Enantiornithes. More than 20 genera of Enantiornithes have now been discovered and described from all kinds of environments and from throughout the Cretaceous (Table 1). It is important to remember, however, that these birds were recognized only in 1981, from the Late Cretaceous of Argentina. The Enantiornithes were by far the most diverse group of Mesozoic birds, showing a good deal of anatomical variation, as well as diversity in their feeding adaptations, flight apparatus, life styles, and patterns of skeletal development. Representatives of this lineage are now particularly well known from the Early Cretaceous of Spain and China. Although most were toothed (e.g. *Sinornis*, *Eocathayornis*, and *Logipteryx*), some did not have teeth (*Gobipteryx*), and they range in size from species that were the size of a sparrow (*Iberomesornis*, *Eoalulavis*) to taxa with wingspans of almost a metre (*Enantiornis*). Their anatomy demonstrates that Enantiornithes were proficient fliers (probably very similar to living birds) and, interestingly, provides some of the earliest evidence for perching (based on the morphology of their feet) that has been found in avian evolution. Although these birds have mostly been recovered from inland deposits, enantiornithines also occupied littoral and marine environments, even extending into the polar regions. Enantiornithes are an anatomically distinctive group (having unique morphological specializations in their forelimbs, hindlimbs, and pectoral girdles) but are not closely related to extant birds (Figure 1); these taxa comprise a basal series of lineages that did not survive the end-Cretaceous extinction event.

In addition to these diverse archaic birds, a number of other currently less well-known taxa (in terms of numbers of described fossil specimens) comprise a series of evolutionary intermediates that fits neatly in between the enantiornithine radiation and the divergence of a second major avian group, the Ornithuromorpha (Figure 1). Most importantly, this clade contains the immediate relatives of all living birds. Members of the Ornithuromorpha are much more modern in their morphology than are members of the other Mesozoic groups – their earliest records are from the Early Cretaceous of China. Just as in the Enantiornithes, there has been something of an explosion in fossil discoveries of ornithuromorphs in recent years. The exceptionally well-preserved *Apsaravis* from the Late Cretaceous of the Gobi Desert in Mongolia is one example (Figure 3). The bulk of the early fossil record of these birds was limited to the flightless loon-like *Hesperornis* and its kin (Hesperornithiformes) and the more modern looking *Ichthyornis* (Ichthyornithiformes) from Late Cretaceous marine sediments in the northern hemisphere. Although much more derived in their anatomy, these birds were described soon after *Archaeopteryx* in the 1870s from rocks formed in the Western Interior Seaway, a shallow tropical sea that bisected North America during the Late Cretaceous. Both are toothed, and *Hesperornis* is known to have been a specialized foot-propelled diver with extremely abbreviated forelimbs (about the size of an emperor penguin), whereas *Ichthyornis* was much smaller and able to fly.

The Radiation of Modern Birds: Bursting into the Cenozoic

Neornithes is the group that includes the 10 000 or so living species of bird. Today, they are a diverse and cosmopolitan group, but their early evolutionary history remains far from well understood. Although *Ichthyornis* is the best-known close relative of the Neornithes, several other recently described taxa might be closer. These taxa, however, are known only from extremely fragmentary remains, thus rendering problematic our understanding of their genealogical relationships to the Neornithes. In spite of this, debates have centred on the timing of the origination of the major lineages of Neornithes (the extant orders and families), and specifically on whether or not these taxa differentiated as a group prior to the end-Cretaceous extinction event 65 Ma ago and, if so, how deep does the history of modern birds extend into the Mesozoic?

Evidence to support or refute hypotheses for the time of divergence of the Neornithes is entirely geological, but the fossil record of putative modern birds from the Cretaceous remains scanty. Just a handful of fossil specimens have been described from before the Createaceous–Tertiary boundary and classified within modern groups; they are mostly single bones that lack clear diagnostic features of extant lineages. It is in the earliest Tertiary that the known fossil record of Neornithes improves dramatically: hundreds of fossils (in many cases complete skeletons, often with feathers and other soft-tissue impressions) are known from a series of localities of Paleocene and Eocene age (60–55 Ma), particularly in Europe and North America, and from a range of environments. This abundance of fossil birds that are modern in their anatomy has led some workers to propose that the bulk of their evolutionary radiation occurred rapidly, immediately after the Cretaceous–Tertiary extinction event (Figure 4). Related to this, there is a further hypothesis that perhaps the extinction itself allowed

Table 1 Fossil Mesozoic birds, their ages and environments (Chiappe and Dyke, 2002)

Taxon	*Stratigraphical age*	*Material*	*Depositional environment*	*Geographical distribution*	*Year described*
Alexornis antecedens	Late Cretaceous	Single specimen	Inland	Mexico	1974
Alvarezsaurus calvoi	Late Cretaceous	Single specimen	Inland	Argentina	1991
Ambiortus dementjevi	Early Cretaceous	Single specimen	Inland	Mongolia	1982
Apatornis celer	Late Cretaceous	Single specimen	Marine	USA	1876
Apsaravis ukaani	Early Cretaceous	Single specimen	Inland	Mongolia	2001
Archaeopteryx lithographica	Late Jurassic	Several specimens	Near shore	Germany	1861
Archaeovolans repatriatus	Early Cretaceous	Single specimen	Inland	China	2002
Avisaurus archibaldi	Late Cretaceous	Single bone	Inland	USA	1985
Avisaurus gloriae	Late Cretaceous	Single bone	Inland	USA	1995
Baptornis advenus	Late Cretaceous	Several specimens	Marine	USA	1876
Boluochia zhengi	Early Cretaceous	Single specimen	Inland	China	1994
Changchengornis hengdaoziensis	Early Cretaceous	Single specimen	Inland	China	1999
Chaoyangia beishanensis	Early Cretaceous	Single specimen	Inland	China	1995
Concornis lacustris	Early Cretaceous	Single specimen	Inland	Spain	1995
Confuciusornis sanctus	Early Cretaceous	Hundreds of specimens	Inland	China	1996
Enaliornis barretti	Early Cretaceous	Several specimens	Marine	UK	1876
Enantiornis leali	Late Cretaceous	Several specimens	Inland	Argentina	1976
Eoalulavis hoyasi	Early Cretaceous	Single specimen	Inland	Spain	1992
Eoenantiornis buhleri	Early Cretaceous	Single specimen	Inland	China	1999
Gansus yumensis	Early Cretaceous	Isolated bones	Near shore	Canada	1997
Gobipteryx minuta	Late Cretaceous	Several specimens	Inland	Mongolia	1996
Halimornis thompsoni	Late Cretaceous	Single specimen	Near shore	USA	2001
Hesperornis regalis	Late Cretaceous	Several specimens	Marine	USA	1876
Iberomesornis romerali	Early Cretaceous	Single specimen	Inland	Spain	1993
Ichthyornis dispar	Late Cretaceous	Several specimens	Marine	USA	1873
Lectavis bretincola	Late Cretaceous	Single bone	Inland	Argentina	1993
Liaoningornis longidigitus	Early Cretaceous	Single specimen	Inland	China	1996
Limenavis patagonica	Late Cretaceous	Single specimen	Inland	Argentina	2001
Longipteryx chaoyangensis	Early Cretaceous	Single specimen	Inland	China	2001
Mononykus olecranus	Late Cretaceous	Several specimens	Inland	Mongolia	1993
Nanantius eos	Early Cretaceous	Single bone	Inland	Australia	1986
Neuquenornis volans	Late Cretaceous	Single specimen	Inland	Argentina	1994
Noguerornis gonzalezi	Early Cretaceous	Single specimen	Inland	Spain	1992
Otogornis genhisi	Early Cretaceous	Single specimen	Inland	China	1992
Parahesperornis alexi	Late Cretaceous	Single specimen	Marine	USA	1989
Parvicursor remotus	Late Cretaceous	Single specimen	Inland	Mongolia	1996
Patagonykus puertai	Late Cretaceous	Single specimen	Inland	Argentina	1996
Patagopteryx deferrarisi	Late Cretaceous	Single specimen	Inland	Argentina	1992
Protopteryx fengningensis	Early Cretaceous	Single specimen	Inland	China	2000
Rahonavis ostromi	Late Cretaceous	Single specimen	Inland	Madagascar	1998
Shuvuuia deserti	Late Cretaceous	Several specimens	Inland	Mongolia	1998
Sinornis santensis	Early Cretaceous	Several specimens	Inland	China	1992
Soroavisaurus australis	Late Cretaceous	Single bone	Inland	Argentina	1993
Vorona berivotrensis	Late Cretaceous	Single specimen	Inland	Madagascar	1996
Yanornis martini	Early Cretaceous	Single specimen	Inland	China	2001
Yixianornis grabaui	Early Cretaceous	Single specimen	Inland	China	2001

the repatriation of avian ecological niches that were occupied by more archaic birds during the Cretaceous. These ideas are based on literal readings of the known fossil record of Neornithes – small numbers of taxa (perhaps misidentified) from the Cretaceous, compared with a much larger diversity of definitive Neornithes from the Early Tertiary. Can this approach alone resolve the age-old issue of modern-bird evolutionary dynamics?

Literal interpretations of the fossil record to explain the pattern of the radiation have, however, been challenged by a number of other lines of evidence, notably the genealogical interpretation of living birds based on genetic data. By considering the numbers of

Figure 3 The single known specimen of the well preserved ornithuromorph *Apsaravis* from the Late Cretaceous of the Gobi Desert, Mongolia (reproduced with the kind permission of Mark Norell and the American Museum of Natural History, New York).

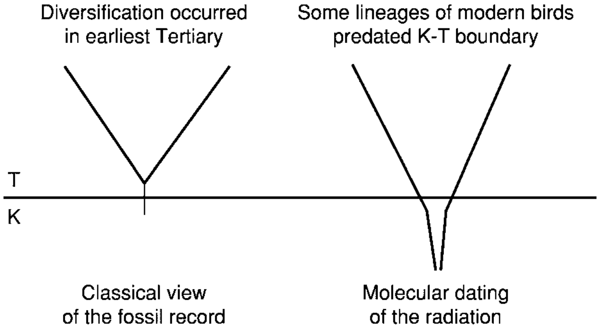


Figure 4 Two competing (alternative) hypotheses for the pat tern of the radiation of Neornithes (modern birds). The one on the left is based on fossil data; the one on the right is based on molecular data.

differences in the aligned DNA sequences (both nuclear and mitochrondrial) of living birds and by using either a fossil of known systematic position or a well-dated continental split as a calibration point, new ideas about the timing of the divergence of the major lineages of living birds have been developed, based on the so-called 'molecular clock'. Divergence estimates have been made for a number of groups of living birds and in all cases have indicated that the radiation of the Neornithes occurred much earlier than has been inferred from the known fossil record (Figure 4). 'Molecular clock' estimates vary, but have reached a consensus that the majority of the major lineages of living birds did originate sometime in the Cretaceous. It has been argued that these hypotheses, although currently receiving little support from the fossil record, agree with what is known about the pattern of breakup of the continental landmasses during the Mesozoic. A causal correlation between the breakup of landmasses and the differentiation of certain groups of extant birds (i.e. the occurrence of endemic southern-hemisphere taxa towards the base of the neornithine tree) has been proposed, and there is a general agreement among molecular systematists that the initial radiation of the Neornithes probably occurred in the southern hemisphere during the Mesozoic.

There is thus a discrepancy between the results of studies founded on molecular data and the apparent pattern seen in the fossil record (Figure 4). Despite a large number of exceptional fossil-bearing deposits from the Late Cretaceous (especially in the northern hemisphere), convincing remains of modern birds have yet to be found. Opinions about the apparent absence of these birds vary: while some have proposed that it probably reflects a real evolutionary pattern (fossils of other small vertebrates, such as mammals and lizards, are well-known from the Cretaceous), others have suggested that perhaps palaeontologists have not been looking in the right place (relatively little collecting effort has been made in the Cretaceous rocks of the southern hemisphere). What is clear is that the existing records of Cretaceous modern birds should be treated with caution – the earliest neornithine birds that are complete enough to be informative for cladistic analyses, and hence potentially informative for estimating the temporal divergence of the extant lineages, come from rocks that are roughly 55–60 Ma old, deposited some 10 Ma after the end of the Cretaceous.

Where From Here?

Thus, palaeontological evidence accumulated over the last decade has cemented the notion that birds are the living descendants of small carnivorous theropod dinosaurs. Exceptional new finds of Cretaceous fossils, especially from China, Spain, Argentina, and Mongolia, have documented a remarkable diversity of avian lineages that thrived and diversified during the latter stages of the Mesozoic. These discoveries have filled substantial anatomical and genealogical gaps in the early history of birds. Yet the evolutionary dynamics (time of divergence and radiation patterns) of the major lineages of extant birds remains an open question that can be resolved only by more well-preserved fossils and a better understanding of their evolutionary relationships.

See Also

Biological Radiations and Speciation. **Evolution**. **Fossil Vertebrates:** Dinosaurs; Flying Reptiles. **Mesozoic:** Cretaceous; End Cretaceous Extinctions.

Further Reading

Chatterjee S (1997) *The Rise of Birds*. Baltimore: Johns Hopkins University Press.

Chiappe LM (2001). Phylogenetic relationships among basal birds. In: Gauthier JD and Gall LF (eds.) *New Perspectives on the Origin and Early Evolution of Birds: Proceedings of the International Symposium in Honor of John H. Ostrom*. Yale Peabody Museum: New Haven.

Chiappe LM and Dyke GJ (2002) Mesozoic radiations of birds. *Annual Review of Ecology and Systematics* 33: 91 127.

Chiappe LM and Witmer LM (2002) *Mesozoic Birds: Above the Heads of Dinosaurs*. Berkeley: University of California Press.

Feduccia A (1999) *The Origin and Evolution of Birds*, 2nd edn. New Haven: Yale University Press.

Paul G (2002) *Dinosaurs of the Air: the Evolution and Loss of Flight in Dinosaurs and Birds*. Baltimore: Johns Hopkins University Press.

Swimming Reptiles

G W Storrs, Cincinnati Museum Center, Museum of Natural History and Science, Cincinnati, OH, USA

Introduction

Fossil reptile remains may be found in a variety of ancient depositional environments, but among the best preserved are those from marine settings. Consequently, fossil marine reptiles have occupied the interest of scientists, scholars, and naturalists from the earliest days of vertebrate palaeontological study. The first discoveries of such fossils were in the Mesozoic rocks of Europe, particularly in England, France, and Germany, in the late eighteenth and early nineteenth centuries. The occurrences of these fossil remains highlighted the relatively new concept of extinction and enabled the first palaeoenvironmental reconstructions, although the fossil discoveries predated the publication of both the theory of natural selection by Charles Darwin, in 1859, and the name Dinosauria, credited to Sir Richard Owen, in 1842. In spite of current popular belief, however, these marine-adapted reptiles were not dinosaurs. The major lineages of dinosaurs were, to the best of our knowledge, exclusively terrestrial in habit until the radiation of avian theropods (birds). Rather, marine reptiles are known to represent a variety of distinct adaptations to habitual life in the sea, and thus ocean-going lineages arose numerous times during the Mesozoic.

Marine Reptile Groups

A wide variety of fossil marine reptiles are known, most of which apparently originated in the Mesozoic. Indeed, most fossil marine lineages are restricted to the Mesozoic, excepting sea turtles and some minor faunal components such as sea snakes. However, a few swimming reptiles from freshwater and lagoonal environments are known from the Permian. *Claudiosaurus* of Madagascar and its relatives are notable for the controversial suggestion that they have an ancestral (or sister-group) relationship to the marine Sauropterygia. The Permian mesosaurs are famed for their role as palaeogeographical indicator fossils and evidence of 'continental drift' and a formerly united Gondwana (Figure 1). All marine reptiles, whatever their lineage, had terrestrial forebears, and thus their occupation of an aquatic environment is a return to the sea, or a secondary evolutionary adaptation. In returning to the sea, each species was faced with new physical constraints and functional needs requiring new adaptational solutions relative to their former life on land. Not the least of these adaptations was associated with the problems of aqueous locomotion, buoyancy control, food acquisition, and reproduction. How these problems were overcome often defines the disparate marine reptile groups.

Active subaqueous locomotion by vertebrate animals can be divided into two basic types: axial propulsion and paraxial propulsion. The former type is by far the most common and is characterized by sinuous flexion of the vertebral column. Constrained by evolutionary history, flexion is in the lateral plane in all reptiles that utilize this style of movement, as opposed to the vertical flexion of mammalian axial propulsors (e.g., cetaceans). Paraxial propulsion is accomplished through the use of the animal's limbs as the primary means of locomotion. Among Mesozoic axial propulsors are the ichthyosaurs, thallatosaurs, thalattosuchians, and mosasaurs. Paraxial swimmers include the sea turtles and the 'reptile wings' or Sauropterygia, i.e., the plesiosaurs and their kin, as defined by Richard Owen in

1860. The great majority of these animals functioned as 'habitual' swimmers; they were restricted to their role as oceanic creatures and few could return to the land, even for brief periods. As such, most probably gave birth to live young (i.e., were viviparous or ovoviviparous), having evolved beyond the need to lay a hard-shelled egg on land. Most of these animals ate fish or other reptiles, although some were apparently molluscivorous. Herbivory is an advanced dietary specialization and among marine reptiles is seen only in the turtles.

Figure 1 The Upper Permian aquatic reptile *Mesosaurus* of Brazil.

Axial Swimmers

The dominant marine group of the Triassic and Jurassic, at least in terms of numbers, was the Ichthyosauria, or 'fish lizards'. These species were also the most highly transformed from a terrestrial condition. Fossils dating to the Early Triassic indicate small, elongate, eel-like animals; by the Jurassic and Cretaceous, they had become powerful, streamlined swimmers with extremely derived, fish-like shapes (pisciform/thunniform). They ultimately developed a high dorsal fin and a symmetrically fluked tail supported by their down-turned, heterocercal vertebral column. Early discoveries in the English Liassic Series (Lower Jurassic) suggested to Owen in 1840 the presence of such caudal flukes. The flukes were later confirmed by discoveries of skeletons with surrounding organic films from the German Liassic (the famed Lagerstätte of Holzmaden and environs) (Figure 2). The limbs of ichthyosaurs were also much modified, ultimately becoming both hyperphalangic and polydactylous (i.e., with extra finger joints and digits). Used in manoeuvering, these paddles or fins became inflexible, and either long ('longipinnate') or broad ('latipinnate'). Ichthyosaurs had clearly lost their ability to leave the water early in their history.

Ichthyosaurs are first known from the Lower Triassic of Asia and survived until the Cenomanian (lower Upper Cretaceous). The Triassic mixosaurs were small and included animals with durophagous dentition (e.g., *Mixosaurus*), apparently for crushing nektonic cephalopods. The shastasaurids and cymbospondylids were giant, whale-like animals. As reported by Richard Ellis in 2003, a monstrous, 20-m-long ichthyosaur has been discovered in the Triassic of British Columbia, Canada. The Jurassic *Stenopterygius* and *Ophthalmosaurus* were among the most derived. The last records include fragmentary

Figure 2 The Lower Jurassic ichthyosaur *Stenopterygius* of Germany.

remains from Europe and North America when ichthyosaur abundance was seemingly a shadow of its former self.

Notable functional aspects of ichthyosaur anatomy include their obvious great speed and the possibility of an elevated physiology, and their well-developed eyes and vision, perhaps for finding prey at depth. Discoveries at Holzmaden of pregnant ichthyosaur (*Stenopterygius*) females, along with those killed in the act of giving birth, have demonstrated that these animals at least were viviparous. Ichthyosaurs in general are a classic example of convergent evolution, as is often cited in textbooks, whereby their habitus, or body shape, mimics that of fast teleost fish, sharks, and cetaceans.

The Thalattosauria ('sea reptiles') were a minor group of Triassic axial propulsors with rather more traditional reptilian proportions. *Askeptosaurus*, for example, from the Middle Triassic of the Monte San Giorgio region of Switzerland, had an elongate body and long, laterally compressed tail. Sinuous waves of the tail apparently propelled the animal through the water. The head had a pointed, elongate snout or rostrum, with numerous, needle-like teeth for capturing and holding slippery, struggling prey. The limbs were little modified from a terrestrial condition save for probable webbing of the feet. Similar animals are known from the west coast of North America. Aberrant, possibly bottom-dwelling forms such as *Clarazia* and *Heschelaria* may be related. All were near-shore, shallow-water animals, perhaps retaining some amphibious ability.

Thalattosuchians ('sea crocodiles') are derived Mesozoic crocodylomorphs (broadly, 'crocodilians', in older terminology). As catalogued by CW Andrews in 1913, like thalattosaurians, they also retained a broadly reptilian shape with elongate tail, body, and pointed rostrum. The powerful tail was the main propulsive organ and the limbs acted as directional controls. The Jurassic and Early Cretaceous teleosaurids (e.g., *Steneosaurus*) and metriorhynchids (*Metriorhynchus*, *Dakosaurus*) were extremely well adapted to life at sea. Indeed, at the height of their success, advanced metriorhynchids such as *Geosaurus* lost the bony armor of typical crocodyliforms and displayed modified, paddle-like limbs and heterocercal caudal tails, much in the manner of ichthyosaurs. Other crocodyliform groups such as the pholidosaurs (the Cretaceous *Teleorhinus* of North America) and the Cretaceous–Tertiary dyrosaurs (e.g., *Dyrosaurus*, *Hyposaurus*) dabbled with life in the ocean, but with less extreme morphological adaptation, retaining typical "crocodilian" limbs, for example. All possessed elongate, fish-catching rostra. The only crocodilian today that is able to tolerate fully oceanic waters, *Crocodylus porosus*, the estuarine or saltwater crocodile of Australasia, can hardly be called a true marine reptile because it is just as likely to be found well inland.

A very significant group of marine reptiles to utilize axial propulsion was the Mosasauroidea (Figure 3). Named for the Meuse River of north-western Europe, these animals were extremely numerous in the Late Cretaceous, to which they were restricted. Worldwide

Figure 3 The Upper Cretaceous mosasaur *Tylosaurus* of Kansas, as painted by Charles Knight.

in distribution, the mosasaurs were squamates related to today's lizards and snakes. Indeed, they appear to be super varanids – voracious predators that had abandoned life on land to dominate the seas. Mosasaurs, too, developed their fore- and hindlimbs into broad, webbed paddles to complement a tall, laterally compressed tail. Small, Early Cretaceous reptiles such as *Coniasaurus* and *Dolichosaurus* of Texas and England appear to be antecedents to the group. By the Late Cretaceous, giant animals such as the 10-m-long *Tylosaurus* and *Mosasaurus* devoured fish, cephalopods, other reptiles, and themselves. Mosasaurs are best known from the Smoky Hill Chalk beds (Niobrara Formation) of Kansas and adjacent states, but have been collected also from Europe, New Zealand, Africa, and Antarctica, for example.

An interesting functional characteristic of mosasaurs is their highly kinetic skull and unique jaw articulation. Hinged in the middle, the lower jaws were able to flex outwardly while rolling backwards on a movable skull/jaw joint (quadrate/articular). This apparently allowed large prey items to be systematically forced into the gullet even as they may have struggled to free themselves from the predator's grip. Escape of prey was generally prevented by a series of backwardly curving palatal teeth (pterygoid dentition). Interestingly, the V-shaped pattern of the marginal teeth and the curved lines of the palatal teeth have both been observed as a series of punctures on the fossilized shells of placenticerid ammonites, suggesting that these nektonic molluscs formed part of the mosasaurian diet. The crushing teeth of *Globidens* may have been a specialization for molluscivory. Mosasaurs probably bore live young.

A few living squamates may also be considered marine. The Galapagos Marine Iguana is physiologically well adapted to ocean waters, but is little transformed from terrestrial relatives and does not venture far from shore. It spends much of its time basking on coastal rocks but dives and forages underwater for its diet of sea grasses. Sea snakes (Hydrophiidae) do not come ashore and thus are marine reptiles in the truest sense. Extremely venomous and colourful, again with a vertically expanded, laterally compressed tail, they are a minor component of reptilian diversity and have no fossil record.

Paraxial Swimmers

Turning to paraxial swimmers, the familiar sea turtles have a long history from the Cretaceous to the Present. Clearly successful in their adaptation to oceanic life, they come ashore only to lay eggs in sandy beaches of the tropics. Ungainly and lumbering on land, they are graceful and far-travelled in water. Though most swimming reptiles have retained the sinuous lateral body movements of fish and early tetrapods, turtles have by necessity resorted to the dominant use of their forelimbs (paddles) for locomotion (Figure 4). Obviously, the stiffened thorax created by the presence of a hard, skeletal shell has constrained the direction of evolution in turtles. Prevented from flexing their bodies in any direction, sea turtles must swim with enlarged paddles, with which they essentially fly through the water. Their limb movements are akin to the lift-generating wing beats of birds, only being applied in a more viscous fluid medium, i.e., water, as opposed to air. Fossil sea turtles are relatively common in Cretaceous and Tertiary rocks, in particular. Continental deposits such as the shales and chalks of the North American Western Interior Seaway, and those of Europe and Africa, preserve large protostegids (e.g., *Protostega*, *Archelon*) and a variety of lesser sized cheloniids, osteopygids, toxochelyids, and others. They are all similar in appearance to living forms and are unmistakable as turtles. The living leatherback turtle (*Dermochelys*), at 2.5 m, is among the largest of all turtles. All living sea turtles are threatened or endangered.

Figure 4 Skeleton of the modern sea turtle *Dermochelys*.

The Mesozoic Sauropterygia contains a wide variety of paraxial propulsors. The earliest representatives are the Triassic pachypleurosaurus, 'nothosaurs' (i.e., stem-group Eusauropterygia) and placodonts. They share with the later plesiosaurus a 'euryapsid' skull condition and a reversed relationship of dermal and endochondral bones of the shoulder girdle, among other features. All typically also possess elongate necks, allowed by the development of limb-dominated swimming, whereas a short neck is a physical requirement for efficient axial propulsion. Pachypleurosaurs are exclusively known from the Alpine and Germanic basins of Europe and the east Tethyan province of China. They were small animals, generally much less than a metre in length, and had not advanced far beyond the terrestrial morphology of their ancestors. *Neusticosaurus*, *Keichousaurus*, and the others had long tails, limbs with obvious digits, and small heads with many, needle-sharp teeth. However, initial indications of paraxial propulsion are given by their lengthened necks, stiff thoraces, and broad and inflexible forelimbs. Pachypleurosaurs were near-shore (littoral) animals, but it is unknown whether they retained the ability to return to land. They probably ate small fish and invertebrates. Dense, pachyostotic vertebrae and ribs for use in buoyancy control were common features in pachypleurosaurs.

The Placodontia were bottom-dwelling, molluscivous forms of the Middle to latest Triassic, ranging in length up to 1.5 m. As far as is known, they were restricted to the Tethyan and adjacent provinces of Europe and the Mediterranean. Placodonts are generally separated into two subdivisions, the nonarmoured Placodontoidea (e.g., *Placodus* and *Paraplacodus*) and the armored Cyamodontoidea (e.g., *Cyamodus*, *Psephoderma*, and *Henodus*). All had short heads, a broad body, and some degree of limb-dominated propulsion. A variety of osteoderms or bony plates were present on the carapace and sometimes also fused to the back of the skull of cyamodontoids. The placodont skull was stout and akinetic, with a broad, strong cheek region in response to increased stresses related to molluscivory. Their diet is obvious from the durophagous, anvil-like crushing teeth of the palate and usually the maxilla and dentary. The Placodontia occupied shallow, near-shore, perhaps lagoonal, palaeoenvironments where the invertebrate benthos was abundant.

The remaining Triassic sauropterygians are relatively closely related to the later plesiosaurs but may not form a natural phylogenetic group. The Old World 'nothosaurs' were often large (up to 3 m) predators with a conservative, elongate body plan. Their swimming style was probably similar to that of the pachypleurosaurs. The flattened heads of 'nothosaurs' were often elongate and usually contained large teeth of unequal length (anisodont dentition). *Nothosaurus* of the Germanic Muschelkalk and *Lariosaurus* of the Alps were seemingly formidable fish eaters. Germany's *Simosaurus* was remarkable in having a rather shorter face with small, spatulate teeth. The behavioral significance of this morphology is unknown. Other taxa, including the pistosaurs (*Pistosaurus* of the Muschelkalk and *Augustasaurus* from the Middle Triassic Star Peak Group of Nevada), retained the primitive (plesiomorphic) nasal bone of 'nothosaurs', yet exhibited an open palate as in Plesiosauria. *Corosaurus* of the Wyoming Alcova Limestone may have had a sister-taxon relationship to pistosaurs.

The Plesiosauria (e.g., *Plesiosaurus*, *Pliosaurus*, and *Thalassiodracon*) were the most advanced and successful of the sauropterygians (**Figure 5**). They ranged from the latest Triassic to the Late Cretaceous when, like mosasaurs, they died off at the terminal Cretaceous mass extinction event. First known in England, they enjoyed a worldwide distribution by the Jurassic. Diverse, fully nektonic, and far ranging, they had no ability for terrestrial locomotion as evidenced by their wing-like limbs and separate ventral and dorsal skeletons. A viviparous or ovoviviparous reproductive style may be inferred. The adoption of paraxial locomotion using two symmetrical sets of hyperphalangic limbs remains unique to plesiosaurs. Also notable was their possession of elongate necks that, in the Cretaceous elasmosaurs (e.g., *Styxosaurus*), could include nearly 70 vertebrae. Once viewed as surface-going 'rowers' with flexible, swan-like necks, it is now known that plesiosaurs were subaqueous 'flyers' with relatively stiff, horizontally held necks. They were agile, mid-column swimmers, seemingly with well-developed underwater vision and perhaps olfaction. The development of sauropterygian 'underwater flight', perfected in the plesiosaurs, appears to have been constrained by stiff ventral baskets of dense gastralia, or ventral ribs. These heavy baskets were probably expanded to act as ballast in the skeletons of ancestral sauropterygians. Many plesiosaurs, particularly the elasmosaurs, swallowed 'stomach stones' (gastroliths) to aid in the production of neutral buoyancy.

Though probably not representing a natural phylogenetic division, two basic plesiosaur morphotypes paired very elongate necks with small heads ('plesiosauroids'), versus large heads with relatively short necks (pliosauroids). Intermediate combinations are also known. In general, small-headed forms with thin sharp teeth are thought to have been piscivorous, whereas those with large heads and teeth were carnivorous, rapid-pursuit predators. Known examples of bite marks indicate predation of the large on the

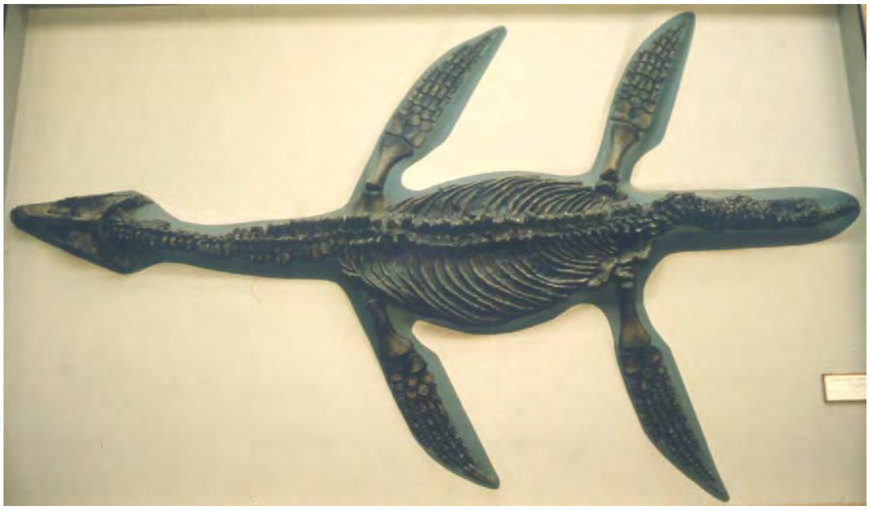


Figure 5 The Lower Jurassic plesiosaur *Rhomaleosaurus* of England.

small. Small, longirostrine pliosauroids, the polycotylids (e.g., *Trinacromerum*), are probably related to elasmosaurs.

Finally, a brief word may be said about the variety of ocean-living birds that might also be considered advanced marine reptiles. Notable among these is the Sphenisciformes, or penguins (Eocene to the Present). Advanced paraxial swimmers, they are far more at home in water than on land. As part of the radiation of archosaurs, the 'ruling' reptiles, penguins may legitimately be considered marine dinosaurs. Excepting birds, the extremely diverse dinosaurs never exploited the marine realm. Uniquely, dinosaurs adopted an upright stance or posture, as opposed to the more or less sprawling stance of typical reptiles. Most swimming reptiles adopted a sinuous undulation of their bodies and tails for locomotion. Others (turtles and plesiosaurs, for example) were prevented from doing so by a stiff shell or body. The upright stance of ground-dwelling dinosaurs may have prevented a viable swimming style. Because birds are unique in having two independent locomotory systems – legs for walking/running and wings for flight – the development of wings was essentially a pre-adaptation allowing underwater flight. Perhaps dinosaurs could effectively enter the sea only after abandoning the ground for the air.

See Also

Fossil Vertebrates: Reptiles Other Than Dinosaurs. **Lagerstätten**. **History of Geology From 1780 To 1835**. **Mesozoic:** Triassic; Jurassic; Cretaceous. **North America:** Continental Interior.

Further Reading

Andrews CW (1913) *A Descriptive Catalogue of the Marine Reptiles of the Oxford Clay. Part II*. London: British Museum (Natural History).

Bardet N (1994) Extinction events among Mesozoic marine reptiles. *Historical Biology* 7: 313 324.

Bell GL (1997) A phylogenetic revision of North American and Adriatic Mosasauroidea. In: Callaway J and Nicholls E (eds.) *Ancient Marine Reptiles,* pp. 293 332. San Diego: Academic Press.

Carpenter K (1997) Comparative cranial anatomy of two North American Cretaceous plesiosaurs. In: Callaway J and Nicholls E (eds.) *Ancient Marine Reptiles,* pp. 191 216. San Diego: Academic Press.

Cruickshank ARI, Small PG, and Taylor MA (1991) Dorsal nostrils and hydrodynamically driven underwater olfaction in plesiosaurs. *Nature* 352: 62 64.

Darwin C (1859) *The Origin of Species by Means of Natural Selection*. London: John Murray.

Ellis R (2003) *Sea Dragons: Predators of the Prehistoric Oceans*. Lawrence, KS: University of Kansas Press.

Hua S and Buffetaut E (1997) Crocodylia: introduction. In: Callaway J and Nicholls E (eds.) *Ancient Marine Reptiles,* pp. 357 374. San Diego: Academic Press.

Massare JA (1988) Swimming capabilities of Mesozoic marine reptiles: implications for method of predation. *Paleobiology* 14: 187 205.

Moody RTJ (1997) The paleogeography of marine and coastal turtles of the North Atlantic and Trans Saharan regions. In: Callaway J and Nicholls E (eds.) *Ancient Marine Reptiles,* pp. 259 278. San Diego: Academic Press.

Motani R, You H, and McGowan C (1996) Eel like swimming in the earliest ichthyosaurs. *Nature (London)* 382: 347 348.

Motani R (2002) Swimming speed estimation of extinct marine reptiles: energetic approach revisited. *Paleobiology* 28: 251 262.

Nosotti S and Pinna G (1996) Osteology of the skull of *Cyamodus kuhnschnyderi* Nosotti & Pinna 1993 (Reptilia, Placodontia). *Paleontologia Lombarda* 6: 1 42.

Nosotti S and Rieppel O (2002) The braincase of *Placodus* Agassiz, 1833 (Reptilia, Placodontia). *Memorie Società Italiana di Scienze Naturale Milano* 331: 1 18.

Owen R (1840) Note on the dislocation of the tail at a certain point observable in the skeleton of many Ichthyosauri. *Transactions of the Geological Society of London* 5: 511 514.

Owen R (1842) Report on British fossil reptiles. *Report of the British Association for the Advancement of Science* 1841: 60 204.

Owen R (1860) On the orders of fossil and Recent Reptilia, and their distribution in time. *Report of the British Association for the Advancement of Science* 29: 153 166.

Rieppel O (1994) Osteology of *Simosaurus gaillardoti* and the relationships of stem group Sauropterygia. *Fieldiana: Geology* 28: 1 85.

Rieppel O, Sander PM, and Storrs GW (2002) The skull of the pistosaur *Augustasaurus* from the Middle Triassic of north western Nevada. *Journal of Vertebrate Paleontology* 22: 577 592.

Schuchert C (1928) The continental displacement hypothesis as viewed by Du Toit. *American Journal of Science* 16: 266 274.

Storrs GW (1993) Function and phylogeny in sauropterygian (Diapsida) evolution. *American Journal of Science* 293A: 63 90.

Storrs GW (1997) Morphological and taxonomic clarification of the genus *Plesiosaurus*. In: Callaway J and Nicholls E (eds.) *Ancient Marine Reptiles,* pp. 145 190. San Diego: Academic Press.

Taylor MA (1997) Before the dinosaur: the historical significance of the fossil marine reptiles. In: Callaway J and Nicholls E (eds.) *Ancient Marine Reptiles,* pp. xix xlvi. San Diego: Academic Press.

Flying Reptiles

D Naish and D M Martill, University of Portsmouth, Portsmouth, UK

Introduction

Reptiles, the group of amniotes that encompasses turtles, lepidosaurs (lizards, snakes, and amphisbaenians), crocodilians, and all of their extinct relatives, appeared around 320 million years ago as small, quadrupedal terrestrial predators. During the Palaeozoic and Mesozoic, reptiles burgeoned into a phenomenal diversity, and a number of Permian and Triassic groups took to the air, often as superficially lizard-like gliders whose extendible wings were supported by ribs or rib-like structures. Some modern lizards and snakes glide using flaps of skin along the flanks and limbs, webbing between their toes, or expandable connected ribs. It should also be pointed out that, from the phylogenetic point of view, birds are reptiles because they are part of the Dinosauria. Flighted birds are thus also flying reptiles.

Why flight evolved in reptiles is an interesting question that may not have a single answer. Flying reptiles may have been able to pursue insects, avoid predators, and cover difficult terrain better than their grounded relatives, and wing membranes may have initially evolved as signalling devices.

Pterosaurs – Actively Flying Reptiles of the Mesozoic

Except in birds, the only reptiles to evolve true flapping flight were the pterosaurs ('wing lizards') or pterodactyls ('wing fingers'), a diverse and unique group from the Triassic, Jurassic and Cretaceous periods of the Mesozoic Era. Pterosaur anatomy, flight mechanics, and palaeobiology have proved controversial and, while scientific study of the group began as early as the late 1700s, the last three decades have seen a major increase in the study of the group.

The earliest known pterosaurs are from the Norian stage of the Late Triassic, but already they were true pterosaurs. The lack of a definite 'proto-pterosaur' has made it difficult to understand pterosaur origins, although some possible contenders have been described (see below). An array of primitive pterosaurs had mostly died out by the end of the Jurassic but, by this time, the advanced pterosaur group, the Pterodactyloidea, had evolved. These were abundant and widespread throughout the Early Cretaceous but, towards the end of the Late Cretaceous, their diversity waned and only one or two groups persisted until the end of the Maastrichtian. Why this decline occurred is unknown, but the Cretaceous diversification of birds may have decreased the ecological opportunities available to pterosaurs.

History of Discovery

The first pterosaur was discovered in the Late Jurassic Solnhofen Limestone of Eichstätt, Germany, and was described by Cosimo Collini (1727–1806) in 1784. Collini concluded that it was a possible sea creature of unknown affinity, although he did note bat-like features. In 1801, the great French anatomist Georges Cuvier (1769–1832) recognized that the creature was a reptile and that its elongated digits must have supported flight membranes. Cuvier was thus the first to recognize pterosaurs as flying reptiles and, in 1809, he coined the name 'Ptero-Dactyle'. This later became the generic name *Pterodactylus* (**Figures 1** and **4**).

In the decades that followed, a succession of further pterosaurs from the Solnhofen Limestone was announced, many in a spectacular state of preservation and some with their wing membranes intact. The first recognized British pterosaur, a specimen of the deep-skulled *Dimorphodon*, was discovered by Mary Anning (1799–1847) in 1827 in Lower Jurassic rocks of Lyme Regis, Dorset. We now know that Gideon Mantell (1790–1852), best known for the discovery of *Iguanodon*, found pterosaur remains before this in the Early Cretaceous Wealden strata of Sussex, but had thought that these were from birds. North America yielded its first pterosaur to the prolific palaeontologist O. C. Marsh (1831–1899) in 1871 and, by 1876, Marsh had recognized it as a new, distinctive genus he named *Pteranodon* (meaning 'winged and toothless'). With an estimated wingspan of 6 m, *Pteranodon* was huge compared to most earlier discoveries.

While these discoveries and others were being made, varied opinions on the nature and life style of pterosaurs were appearing, and they were variously depicted as swimming creatures, as bats, marsupials, or as kin of birds. By the early 1900s, it was generally agreed that pterosaurs were bat-like flying reptiles and, in 1901, Harry Seeley (1839–1909) published *Dragons of the Air*, the first book devoted to pterosaurs.

South American Cretaceous pterosaurs have proved to be among the most important in the world, but not until 1971 was the first pterosaur from the now famous Santana Formation of Brazil discovered. Since then a significant number of new kinds from around the world (around 70 genera are presently recognized) have revealed previously unimagined morphologies and maximum sizes. Until 1971, *Pteranodon sternbergi*

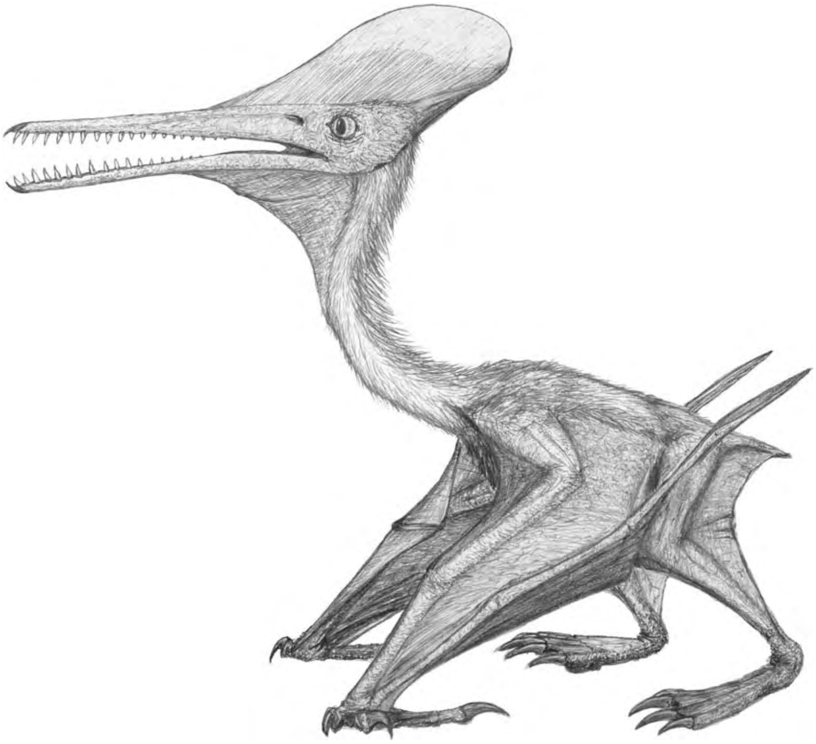


Figure 1 Life restoration of the Late Jurassic pterodactyloid, *Pterodactylus*, from the German Solnhofen Limestone in a quadrupedal stance. Note the presence of body hair and the soft tissue head crest. Reproduced with permission from Dino Frey. Buffetaut E and Mazin J M (2003) *Evolution and Palaeobiology of Pterosaurs. Geological Society Special Publication 217*. London: The Geological Society of London.

(wingspan, 9 m) was the largest known flying animal, but the discovery in Texas of *Quetzalcoatlus* revealed that the biggest pterosaurs achieved wingspans of 11 m. Related pterosaurs of similar or larger size were discovered in the 1990s in Spain and eastern Europe.

The Pterosaur Skeleton

The pterosaur skeleton was highly modified for flight, and the most obvious features are the huge size of the skull compared with the body and the extreme elongation of one of the fingers. Like birds, most pterosaurs had hollow bones with foramina (small openings), indicating that they contained air sacs connected to the lungs. Pterosaur bones were supported internally by struts, and the bone walls themselves, usually no thicker than 2 mm, are composed of multiple overlapping layers and thus combine lightness with strength.

Pterosaur skull morphology is varied, although the majority had long, slim, shallow jaws and all had large orbits (eye-sockets). In basal pterosaurs, the external nostril was separate from an opening in front of the orbit called the antorbital fenestra. In pterodactyloids, these two openings merged into a single one called the nasoantorbital fenestra. Pterosaur teeth were extremely variable. Widely spaced pointed teeth, were widespread and from ancestors with teeth like these evolved species with fang-like teeth at the jaw tips and the unique *Istiodactylus* with its short petal-shaped teeth. The Late Triassic *Eudimorphodon* and *Austriadactylus* possessed multicusped teeth while elongate, slender teeth numbering in the hundreds evolved in the ctenochasmatoids. Toothlessness evolved several times. Some pterosaurs skulls sport bony crests at the jaw tips, along the midline or at the back of the skull.

Unlike birds and bats, the main wing spar in pterosaurs was formed by a hypertrophied digit (**Figure 2**). This 'wing finger' is generally considered to be the fourth because the digital formulae of the pterosaur hand best matches that of digits one to four in the hands of other reptiles. However, a rod-shaped bone projecting from the pterosaur wrist, called the pteroid bone, has at times been argued to represent the first hand digit. This is a minority view today but, if it is correct, then pterosaurs have five hand digits and the wing finger is the fifth. Although most pterosaur fossils show the pteroid pointing towards the shoulder, some workers suggest that it pointed forwards parallel to the neck. Regardless, the pteroid was probably mobile and used to control the attitude of the propatagium (**Figure 3**).

The pterosaur pectoral girdle includes a (normally fused) scapula and coracoid that meet at an acute angle and, as expected for flying animals, the socket for the humerus faces sideways and slightly upwards. The coracoids attach to an enlarged keeled sternum that anchored most of the major flight muscles. The bones of the pterosaur pelvis were short, usually fused together, and with a closed hip socket. Pterosaurs have a pair of unique rod- or plate-like bones called the prepubes projecting forwards from the bottom of the pelvis. Like the gastralia ('belly ribs') that all pterosaurs possessed, they may have helped support the gut or keep the abdomen rigid.

The vertebral column in pterosaurs can clearly be differentiated into cervical, dorsal, sacral, and caudal portions. The number of vertebrae is variable and pterosaurs have 7–9 cervical, 11–16 dorsal, 3–10 sacral, and 11–40 caudal vertebrae. The many caudal vertebrae of basal pterosaurs are encased in long bony processes that make the tail stiff and rod-like. In derived Cretaceous pterosaurs, most of the dorsal

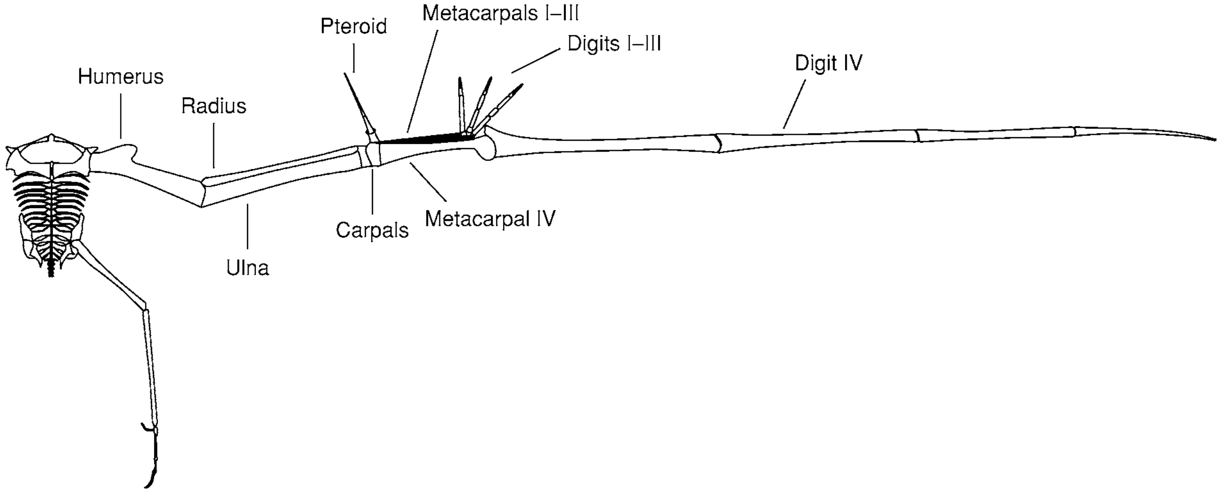


Figure 2 Wing skeleton of an ornithocheiroid pterodactyloid.

vertebrae are fused together forming a structure called the notarium.

Compared with the other wing segments, the pterosaur humerus is short, although generally with massive crests for muscle attachment. The ulna is always larger than the radius and both are attached distally to block-shaped carpal bones. Projecting from one of these is the unique pteroid. Pterosaurs had four metacarpals, the first three of which were slim and, with the exception of *Nyctosaurus* from Late Cretaceous North America, attached to short, clawed fingers. Why *Nyctosaurus* lacked clawed fingers is unknown, but in all other pterosaurs these digits may have served important functions. Trackways show that they were used in walking, and it is also possible that they were employed in grooming or climbing. The fourth metacarpal was robust and tipped with a twisted, roller-like distal end to which was attached the massive wing finger. This consists of four long straight bones, excepting a few genera where there were only three. Because of the twisted end of the fourth metacarpal, the wing finger would have lain parallel to the body's long axis when the wing was folded up.

The pterosaur hind limb is lightly built and the head of the femur is only slightly offset from the long axis of the shaft. Pterosaur hind limbs seem to have been quite flexible, but mostly sprawled to the sides. During flight, the hind limb was probably held in a bat-like orientation and could have been used to control the shape of the wing membranes. Basal pterosaurs are five-toed, with a prominent curving fifth digit that is hooked towards the tail. In pterodactyloids, the fifth digit is either absent or present as a tiny stub.

Because some articulated fossils indicate that the foot could assume a 90° angle relative to the tibia (and there is little evidence for much motion at the metatarsophalangeal joints), pterosaurs have generally been regarded as plantigrade (placing the whole length of the foot on the ground when walking). In 1983, Kevin Padian argued that this was not the case for *Dimorphodon* and that it may instead have been digitigrade (walking only on the toes). This was later inferred for all pterosaurs. An articulated *Dimorphodon* foot shows, however, that only limited motion was possible at the metatarsophalangeal joint, thus supporting a plantigrade posture. This is in agreement with probable pterosaur tracks preserved as trace fossils.

Soft Tissue, Integument, and Pterosaur Life Appearance

Many aspects of pterosaur life appearance remain unknown or controversial, although a number of exceptional fossils have provided some surprising details. Pterosaur body hair was reported as early as 1831 and described for various Jurassic pterosaurs between the 1920s and 1970s and today it is clear that pterosaurs had bristle-like hairs covering their necks and bodies (Figure 4). The active flapping flight and body hair of pterosaurs suggest that they had an elevated metabolism.

Other exceptional fossils show that some pterosaurs possessed a throat pouch, webbing between the toes, and scales on the soles of the feet. Soft

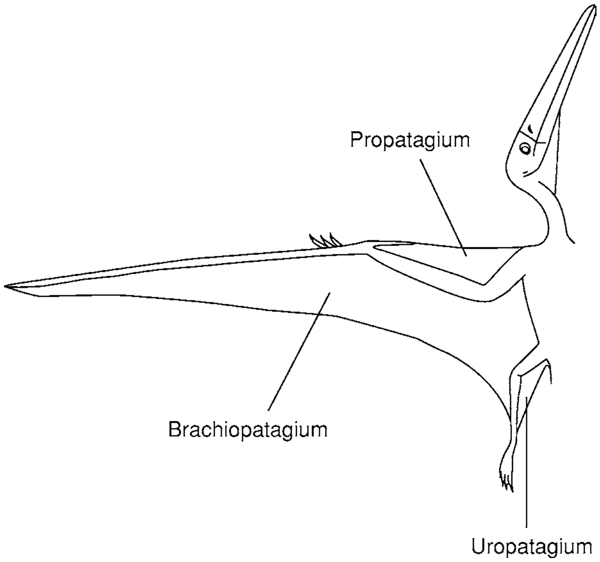


Figure 3 Schematic representation of the flight membranes in a generalized pterodactyloid pterosaur.

Figure 4 An exceptionally well preserved skeleton of the Late Jurassic pterodactyloid *Pterodactylus* from the German Solnhofen Limestone. This specimen preserves parts of the flight membranes, a throat pouch, and hairs on the neck and back.

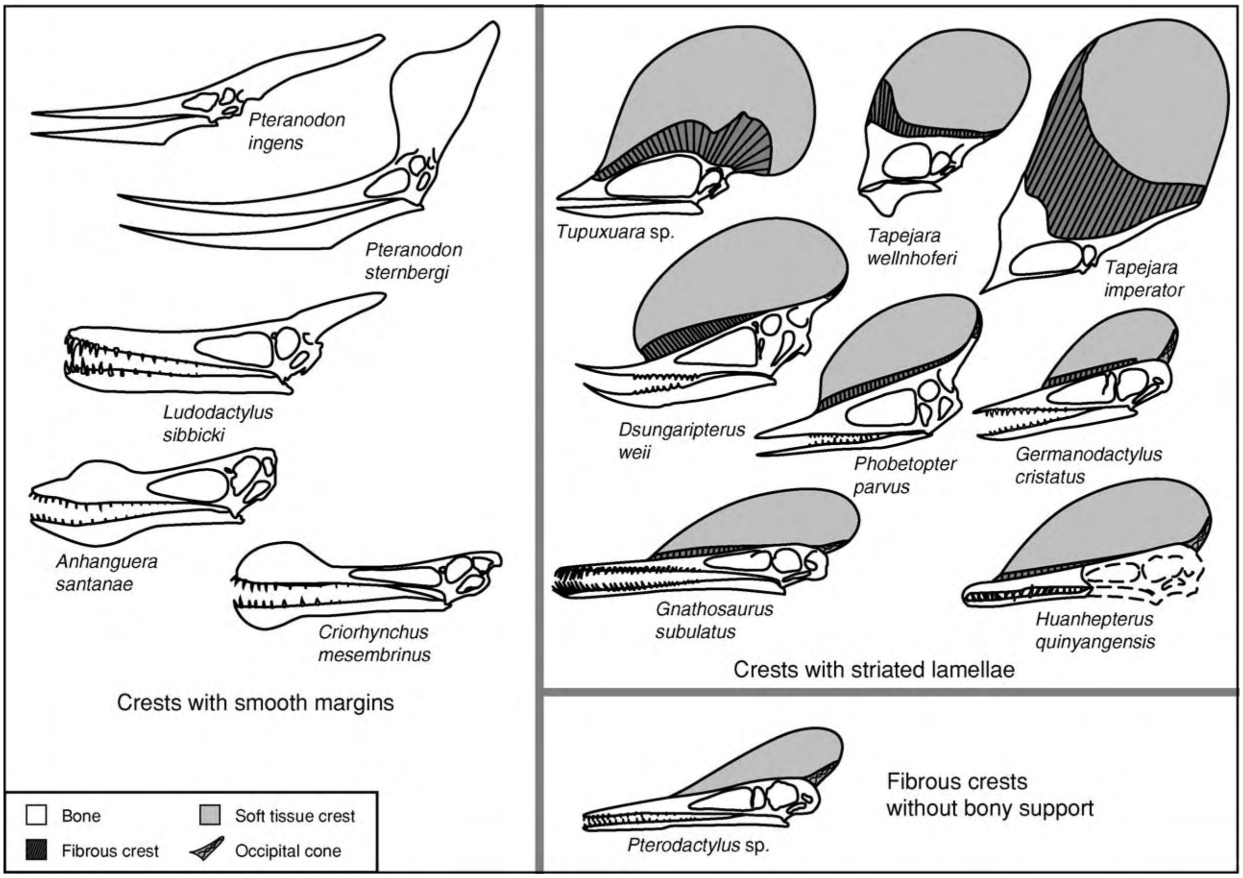


Figure 5 Variation in skull crest morphology in pterodactyloids. Soft tissue crests are now known for a wide diversity of pterodactyloids. Reproduced with permission from Dino Frey and Marie Celine Buchy. Buffetaut E and Mazin J M (2003) *Evolution and Palaeobiology of Pterosaurs. Geological Society Special Publication 217*. London: The Geological Society of London.

tissue skull crests connected to the underlying bony crests have proved to be widespread and appear to have doubled the size of the bony crests (Figures 5 and 6). An unexpected discovery is a soft tissue crest in *Pterodactylus*, a genus that lacks a bony crest (Figure 1). The presence of a distinctive bone texture on the pterosaur snout, jaw, and palate indicates that pterosaurs were beaked.

Pterosaur wing membranes are known from well-preserved specimens from the Solnhofen Limestone and the Early Cretaceous Brazilian Crato and Santana formations. A membrane called the propatagium extended from the shoulder to the pteroid and perhaps distally to encompass the first three fingers. The main flight membrane, the brachiopatagium (also called the cheiropatagium), extended from the tip of the wing finger to the hind limb, extending as far distally as the knee, shin, or ankle. Another membrane, the uropatagium, was present between the hind limbs (Figure 3). The wing membrane appears to have been complex, with a thin epidermis, a layer of vascular tissues, a layer of stiffening fibres called aktinofibrils, a thin sheet of muscle, and a

Figure 6 Skull of the tapejarid pterosaur *Topejara navigaus* from the Early Cretaceous Crato Formation of Brazil with bony and soft tissue skull crest. Buffetaut E and Mazin J M (2003) *Evolution and Palaeobiology of Pterosaurs. Geological Society Special Publication 217*. London: The Geological Society of London.

blood capillary network. *Rhamphorhynchus* and probably other long-tailed pterosaurs possessed a vertical diamond-shaped membrane at the tail tip.

With skin membranes connecting the wings, body, and legs, pterosaurs may have been superficially bat-like but, because bats are mostly dark-coloured nocturnal animals, it is doubtful that the similarities were strong. Pterosaurs mostly seem to have been ecological analogues of sea- and water-birds, and it might be that they were patterned in whites, blacks, and greys, although bright colours presumably decorated their crests.

The Affinities and Origin of Pterosaurs

Historically, pterosaurs have been allied with Mesozoic marine reptiles, bats, marsupials, and birds (*see* **Fossil Vertebrates:** Dinosaurs; Birds). However, major improvements in the understanding of vertebrate evolution allowed the palaeontologists of the nineteenth and twentieth centuries to realize that pterosaurs were related at least vaguely to dinosaurs (*see*) and their allies.

Although it is clear that pterosaurs are part of the major reptile assemblage known as the Diapsida, their affinities within this group are controversial. The presence of an antorbital fenestra has conventionally meant that pterosaurs have been regarded as archosaurs, the so-called ruling reptile group that incorporates crocodilians, dinosaurs, and kin. Among archosaurs, pterosaurs share a simple hinge-like ankle joint with dinosaurs and consequently have been regarded as close relatives of dinosaurs in most studies. This view was developed at a time when some workers thought that pterosaurs originated from terrestrial bipedal ancestors and that pterosaurs themselves were bipedal and digitigrade. A small bipedal, long-legged archosaur from Late Triassic Scotland, *Scleromochlus*, was argued to be a pterosaur ancestor, but recent studies refute this idea. The idea that pterosaurs might be close relatives of dinosaurs can certainly be regarded as the 'mainstream' view in vertebrate palaeontology today. However, several recent studies have questioned the evidence for this proposed affinity.

An alternative hypothesis argues that pterosaurs belong instead to a group of archosaur-like diapsids, the Prolacertiformes. Most prolacertiforms were superficially lizard-like, but *Sharovipteryx* from Late Triassic Kyrgyzstan appears to be intermediate between conventional prolacertiforms and pterosaurs. It has pterosaur-like hind limbs and vertebrae and membranes between its hind limbs and tail.

Some other models for pterosaur ancestry have been proposed. In 1996, S. Christopher Bennett argued that pterosaur hind limbs are only superficially similar to those of dinosaurs, and that re-analysis favoured a position for pterosaurs outside of the crocodilian–dinosaur group. Rather more heterodox recent ideas include the suggestion that pterosaurs are the closest relatives of birds and that pterosaurs are part of the Dinosauria.

Several different models have been proposed for the origin of pterosaurs, but the presence in basal pterosaurs of climbing features and of various details in the hind limb and pelvis indicative of a leaping ability suggest that pterosaurs first evolved as tree-climbing leapers.

Pterosaur Diversity and Phylogeny

It was recognized in 1901 that pterosaurs could be divided into two groups: the toothed, mostly long-tailed Rhamphorhynchoidea, and the short-tailed Pterodactyloidea (including both toothed and toothless kinds). Today, it is clear that rhamphorhynchoids include the ancestors of pterodactyloids and, consequently, Rhamphorhynchoidea is a grade and not a clade. Pterosaurs previously referred to as rhamphorhynchoids are nowadays termed basal pterosaurs or non-pterodactyloids. Although basal pterosaurs were diverse, it is notable that they were small compared with the majority of Cretaceous pterodactyloids.

The evolutionary relationships of pterosaurs are relatively understudied and only recently has pterosaur phylogeny been analysed. Although some areas of consensus have emerged, authors disagree on the details. We follow the phylogeny proposed by David Unwin of the Museum für Naturkunde in Berlin (Figure 7).

Perhaps the most basal pterosaur is *Preondactylus* from the Late Triassic of Italy. This form has a shorter coracoid and humerus and longer legs than other pterosaurs. Dimorphodontids, which include *Dimorphodon* from Early (and perhaps Middle) Jurassic

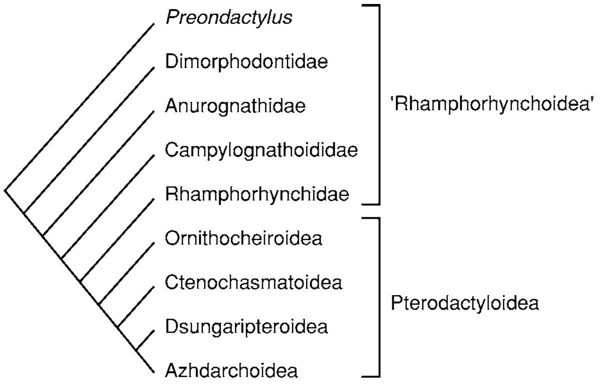


Figure 7 Cladogram depicting the relationships of all the major pterosaur groups. Reproduced from David Unwin.

England and Mexico, are basal pterosaurs with deep skulls superficially like those of puffins, whilst anurognathids were unusual in having short, broad snouts and abbreviated tails. A surprising recent discovery is the persistence of anurognathids into the Early Cretaceous. Two basal pterosaurs, *Eudimorphodon* from Late Triassic Italy and Greenland and *Campylognathoides* from Early Jurassic Germany and India, are united in the Campylognathoididae based on a distinctive lower jaw in which two pairs of large conical teeth are followed by multiple smaller ones.

Rhamphorhynchids were successful Jurassic pterosaurs known from Eurasia, North America, and Africa. *Rhamphorhynchus* from Late Jurassic Europe exhibits a laterally compressed, ventrally directed lower jaw tip and an array of forward-pointing teeth. It probably used these to grab fish and other small prey from the water. Another rhamphorhynchid lineage, the scaphognathines, had deeper skulls with teeth perpendicular to the jaw margins.

The Pterodactyloids

Pterodactyloids, the advanced short-tailed pterosaurs, originated from a rhamphorhynchid-like ancestor during the Middle Jurassic. The pterodactyloid radiation consisted of four major groups: the robust-jawed ornithocheiroids, the slim-jawed ctenochasmatoids, the low-crested dsungaripteroids, and the long-necked, crested azhdarchoids. A fifth group, the lonchodectids from Early Cretaceous England, are of uncertain affinity. Lonchodectids were small (wingspan, 1–2 m) with long, dorsoventrally flattened jaws with small teeth, each of which was supported by a low bony collar at its base.

Ornithocheiroids were large predatory pterosaurs (wingspan, 2–9 m) with robust beaks, often housing recurved, fang-like teeth at their tips (Figure 8). Their jaws frequently possessed keel-like dorsal and ventral crests, and some forms also possessed crests on the back of the skull. The toothless pteranodontids and nyctosaurids appear to be members of this group. The earliest known ornithocheiroids appear at the start of the Cretaceous, while nyctosaurids survived to the very end of the Cretaceous.

Ctenochasmatoids had needle-like meshes of teeth set in long, thin jaws. In *Pterodaustro* from late Early Cretaceous Argentina, the upturned lower jaw contains approximately 1000 bristle-like teeth. These were surely used for filtering small organisms from the water. Unlike ornithocheiroids, ctenochasmatoids had elongate cervical vertebrae and were generally small (wingspan, 50 cm to 2 m), although *Cearadactylus* from Early Cretaceous Brazil was a giant with a wingspan of 5.5 m.

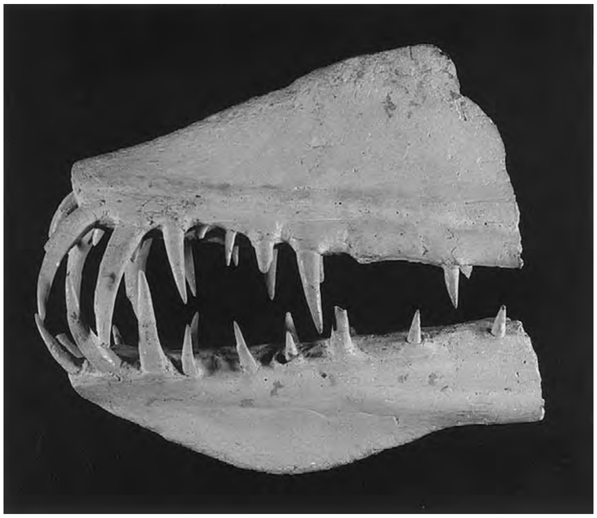


Figure 8 Jaw tips of the Cretaceous ornithocheiroid pterosaur *Coloborhynchus*. Note the massive fang like anterior teeth and the low keel like crests on both jaws. This specimen is from the Santana Formation of Brazil and would have belonged to an animal with a skull of approximately 1 m in length.

Members of the Dsungaripteroidea are known from the Late Jurassic and Early Cretaceous of Eurasia, Africa, and South America. Like some ctenochasmatoids, dsungaripteroids had a midline crest on the top of the skull. Their beaks often had toothless tips and they may have been predators of molluscs and crustaceans.

Finally, the strangest pterodactyloids must be the azhdarchoids. These include the long-necked azhdarchids and the crested tapejarids. Azhdarchids may have exceeded wingspans of 11 m and were widely distributed in the Late Cretaceous. They may have been ecological generalists akin to storks, and were probably not specialist carrion feeders or mud probers as has been proposed. Determining the life style of the tapejarids is more difficult. The vaguely parrot-like skull of *Tapejara* (Figure 6) from Lower Cretaceous Brazil led some workers to propose that it was a fruit eater, but it might better be imagined as an auk analogue. Recently, it has been suggested that *Thalassodromeus*, also from Lower Cretaceous Brazil, was a fish eater that trawled its blade-like lower jaw through the water.

Pterosaur Palaeobiology

Because pterosaurs are unique and extinct, reconstructing their palaeobiology is difficult and nothing is known about several aspects of their lives. Limited evidence does allow us, however, to reconstruct their sensory abilities, feeding behaviours, and styles of locomotion.

The large orbits of pterosaurs show that they had large eyes, and the abundance of visual display structures in the group implies that vision was important. Given that living reptiles see in colour, pterosaurs probably did as well, and pterosaur brain casts reveal enormous optic lobes suggesting excellent eyesight. The many skull crests seen in pterosaurs were therefore probably used in sexual display, and differing crest shapes seen in various genera, most notably *Pteranodon*, are suggestive of sexual dimorphism. Individuals with large, long crests are inferred males and those with small, short crests are inferred females. This is supported by the pelvic structure, as small-crested individuals have wider pelvic canals than large-crested individuals.

Pterosaur crests have also been suggested to have functioned in muscle attachment, thermoregulation, or aerodynamics. Keel-like ornithocheiroid crests may have functioned as cutwaters or stabilizers when the jaws were pulled through the water, and it is possible that the tall soft tissue crests of tapejarids and other pterodactyloids functioned as sails.

Little direct evidence for pterosaur diet is known. In general, pterosaurs seem to have been predators of fish and other aquatic prey, although insects and small terrestrial vertebrates could have been captured by some kinds. A Solnhofen Limestone dragonfly preserves what appears to be a pterosaur bite mark on one of its wings. Stomach contents are known for a handful of pterosaurs and show that *Eudimorphodon*, *Rhamphorhynchus*, *Pterodactylus*, and *Pteranodon* ate fish. Pterosaurs themselves were sometimes preyed on; the bones of a small Triassic pterosaur have been discovered in a gastric pellet probably produced by a fish, and pterosaur bones have been discovered among plesiosaur stomach contents. An azhdarchid wing bone gnawed on by a predatory dinosaur has been described from Late Cretaceous Canada.

Some pterosaur specimens reveal evidence of accidents and disease. Fractures have been reported for jaws, ribs, and wing finger and toe bones in *Rhamphorhynchus*, *Pterodactylus*, *Anhanguera*, and *Pteranodon*; evidence for secondary infection but an absence of evidence for healing in most of these cases indicates that the animals died shortly after receiving these injuries, although one *Pterodactylus* specimen lived long enough for its fractured femur to heal. Osteoarthritis and evidence for infected facial lesions have been reported for ornithocheiroids.

Virtually nothing is known about pterosaur reproduction, and eggs and nests are unknown. We infer that pterosaurs laid eggs on the basis of the egg-laying behaviour of their living relatives. Juveniles of a few species are known and suggest that at least some species were able to fly soon after hatching. Unlike birds, these pterosaurs therefore underwent much of their growth while they were capable fliers. According to some experts, large samples of the same species reveal all individuals ceasing growth at the same size and consequently it has been argued that some pterosaurs exhibited determinate growth like most birds and mammals. There are some indications, though, that basal pterosaurs and azhdarchids may have continued growing for their whole lives.

Pterosaur Locomotion

Pterosaur locomotion, both aerial and terrestrial, has been investigated intensively. Although pterosaurs were imagined in the past as weak gliders, largely unable to flap, the evidence for substantial wing muscles attached to the sternum, scapulocoracoid, and dorsal vertebrae indicates that even the biggest forms were powerful flappers. Their muscles mean that, although they were light for their size, the biggest pterosaurs were not as low in weight as often stated, and an azhdarchid with an 11 m wingspan probably weighed in excess of 100 kg. A comparison of pterosaur wing shapes with those of birds and bats indicates that, as a generalization, pterosaurs were slow, manoeuvrable fliers equivalent to large sea-birds. Large pterosaurs presumably exploited rising air and winds to soar, and various studies have shown that they had low gliding and stalling speeds. This would be ideal for animals that caught their prey from the water's surface. Pterosaurs probably controlled their rate of ascent and descent by using the propatagium to alter the flow of air over the wing's upper surface. How pterosaurs took off is mysterious given that they were probably not proficient runners. Small forms may have leapt from trees, rocks, or cliffs, while larger kinds may have relied on headwinds and updraughts.

Pterosaurs could probably swim and there is no good reason why some of them could not have dived beneath the water surface. The toe proportions and claw shapes of small basal pterosaurs suggest that they may have been good tree climbers.

The terrestrial ability of pterosaurs is one of the most controversial areas in pterosaur study. Trackways now thought to have been produced by pterosaurs were first reported from Late Jurassic rocks in the USA in 1957 and reveal a three-fingered hand and a long, four-toed foot, the proportions of which match the pterosaur foot skeleton. Dubbed *Pteraichnus*, these tracks support the view of pterosaurs as plantigrade quadrupeds. *Pteraichnus* tracks are now known from many Late Jurassic and Early Cretaceous North American and Eurasian deposits and were probably made by ctenochasmatoids. Quadrupedal trackways that appear to have been produced by large (estimated wingspan, 6 m) ornithocheiroids and basal pterosaurs have also been reported. However, although at least some pterosaurs seem to have been quadrupeds, it is still not certain that this was the case for all of them.

See Also

Fossil Vertebrates: Reptiles Other Than Dinosaurs; Dinosaurs; Birds; Swimming Reptiles; Mesozoic Amphibians and Other Non-Amniote Tetrapods.

Further Reading

Buffetaut E and Mazin J M (2003) *Evolution and Palaeo biology of Pterosaurs. Geological Society Special Publication 217*. London: The Geological Society of London.

Naish D and Martill DM (2003) Pterosaurs a success ful invasion of prehistoric skies. *Biologist* 50(5): 213 216.

Wellnhofer P (1991) *The Illustrated Encyclopedia of Pterosaurs*. London: Salamander.

Mesozoic Amphibians and Other Non-Amniote Tetrapods

A R Milner, Birkbeck College, London, UK

Introduction

In the Mesozoic, the various amniote groups (reptiles, birds, mammals, and their earlier relatives) came to dominate terrestrial niches; the fossil records of this era depict a general decline of the archaic amphibian groups and diversification of the modern amphibian groups. In this article, the effect of the Permo–Triassic extinction event on amphibians is considered and the archaic amphibians of the Triassic and the few relicts that survived into the later Mesozoic are reviewed. This is followed by a consideration of the modern amphibian groups – the frogs, salamanders, and caecilians – usually combined as the Lissamphibia, and what is known of their evolution through the Mesozoic.

Amphibians and the Permo–Triassic Event

The Permo–Triassic (P–T) extinction event (*see* **Palaeozoic:** End Permian Extinctions) appears to have had a dramatic effect on amphibian turnover, although the Late Permian record is too patchy for there to be certainty that many of the lineages died out during the event and not before. It is clear that an immense diversity of new types of amphibians appeared in the Early Triassic, most belonging to the Stereospondyli. This has been interpreted as a direct response to the extinction event, but it has recently been suggested that it represents a Late Permian diversification in Gondwana that appeared suddenly in Laurasia as the result of dispersal from the south after the P–T extinction. The Lissamphibia (modern groups such as frogs and salamanders) also make their first appearance in the fossil record after the P–T event. They may represent a major diversification following the P–T extinctions, but the Triassic record is too poor for such a pattern to be seen and it can only be inferred. As well as the Stereospondyli and Lissamphibia, a few other lineages of temnospondyls and of chroniosuchian anthracosaurs survived the P–T event and persisted as relicts during the Early and Middle Triassic, without further diversification.

Mesozoic Assemblages

Most Triassic amphibian assemblages are composed of large temnospondyls in red-bed floodplain or coastal lagoon situations. Much of the fossil material is disarticulated and many forms are known only from the massively constructed skulls. The temnospondyls appear to have filled the niche later occupied by parasuchians and crocodiles and were gradually replaced by them during the Triassic. Only a very few Triassic localities have produced evidence of small amphibians, either juvenile temnospondyls or early representatives of the modern amphibian groups.

In the Jurassic and Cretaceous, fossils of the few remaining large amphibians occur in lake-bed or floodplain deposits. The increasingly abundant frog and salamander fossils occur either as intact skeletons in lake deposits (localities, or 'lagerstätten', where prevailing conditions have produced highly conserved or diverse fossils, termed 'konservat-lagerstätten') (*see* **Lagerstätten**) or as isolated elements in microvertebrate bone accumulations. The latter are common and produce abundant fossil material, although large samples are needed to make associations between isolated bones. Notable Cretaceous konservat-lagerstätten for frogs and salamanders are Las Hoyas, Spain and Liaoning, China.

Temnospondyls in the Triassic

Stereospondyli

The stereospondyls comprise one of the largest subgroups of the amphibian group Temnospondyli. They were mostly 1- to 2-m-long aquatic and amphibious carnivores filling a broadly crocodile-like niche. The 'stereospondylous' type of vertebra, after which the group is named, is one in which the anterior central element, the intercentrum, enlarges to form the entire central body (Figure 1A). In fact, this structure characterizes only a few genera (including some of the first to be discovered, hence its original use as a characteristic). One of the most obvious, though not unique, characteristics of the group is that the braincase is broadly sutured to the lateral palatal bones (pterygoids) on either side.

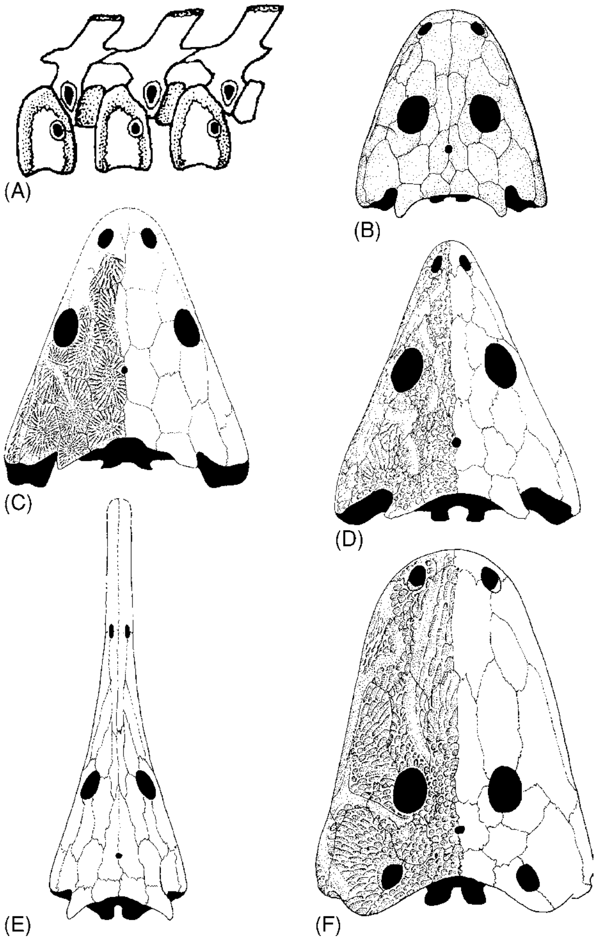


Figure 1 Stereospondyli. (A) *Mastodonsaurus* vertebra, demonstrating the stereospondyl condition, with an enlarged intercentrum and rudimentary pleurocentra; (B) skull of the lydekkerinid *Lydekkerina huxleyi*; (C) skull of the rhytidosteoid *Peltostega erici*; (D) skull of the trematosauroid *Lyrocephaliscus euri*; (E) skull of the trematosauroid *Aphaneramma rostratum*; (F) skull of the capitosauroid *Cyclotosaurus robustus*. All reproduced with permission from Schoch RR and Milner AR (2000) Stereospondyli. In: Wellnhofer P (ed.) *Handbuch der Paläoherpetologie,* part 3B. Munich: Pfeil.

The stereospondyl radiation had commenced diversification in the Late Permian but is one of the characteristic elements of continental and coastal Triassic faunas. There is not full agreement on the classification of the stereospondyls, but three major groups, Rhytidosteoidea, Trematosauroidea, and Capitosauroidea, have recognized by several workers, together with three families of controversial relationships.

Rhytidosteoidea

All of the members included with certainty in the Rhytidosteoidea are restricted to the Late Permian and Early Triassic. The basal forms are in the family Lydekkerinidae; these small forms are less than 1 m long and have orbits set in the middle of a blunt alligator-like skull (Figure 1B). They appear to have been amphibious animals. The more advanced Rhytidosteidae were more specialized aquatic forms, with small laterally placed orbits and either very triangular skulls (Figure 1C) or flattened rounded skulls. This appears to be a fundamentally Gondwanan group, with a few lineages penetrating into Laurasia. Most lydekkerinids and rhytidosteids are from the Lower Triassic of South Africa, Madagascar, India, Australia, and Antarctica, with fossil material recently reported from Brazil. Three genera are known from Greenland, Svalbard, and North Russia.

Trematosauroidea

The trematosauroids are an entirely Triassic clade of temnospondyls, first appearing after the P–T extinction and last appearing in the mid-Norian. The stem of the group is the Benthosuchidae, known largely from fossil material from Russia. Benthosuchids were a lineage of crocodile-like temnospondyls with elongate triangular skulls, and the most well-known taxon, *Benthosuchus* (Figure 2), is known from many hundreds of specimens preserved in three-dimensional form from the Sharzhenga River region. This taxon was described fully in 1940 by Bystrow and Efremov in a monograph that still serves as a standard for descriptions of this group.

The group adaptively radiated in the Early Triassic as the family Trematosauridae. The skull shapes within this family vary from an almost isosceles triangle shape (*Lyrocephaliscus*, from Svalbard; Figure 1D) to an extreme elongation of the snout that results in a gharial-like head (*Aphaneramma*, also from Svalbard; Figure 1E). All group members are characterized by a distinctive pattern of elongation of the back of the skull roof, small, laterally placed orbits, a braincase floor that underplates the occipital region, and a parasphenoid (central strut in the palate) that is extremely narrow. Most had poorly ossified postcranial skeletons and small limbs and can be assumed to have been aquatic predators. Early in the Late Triassic, the trematosaurids vanish from the fossil record, but two descendant groups appear, the Almasauridae and the Metoposauridae (Figure 3). Both have long skulls like the trematosaurids had, but the snouts are short and the orbits are quite near the front of the head. Almasaurids were small animals, no more than a metre in length, whereas metoposaurids grew to 2 m and had extremely flattened skulls.

Figure 2 Stereospondyli. Skull of the primitive trematosauroid *Benthosuchus sushkini*, from the Lower Triassic, Russia. Specimen at the Natural History Museum, London. © Andrew Milner.

The Trematosauroidea appears to have been a basically Laurasian group, with successful forms extending into Gondwana. Evidence of the basal Benthosuchidae and primitive trematosaurs such as *Thoosuchus* and *Angusaurus* is found only in continental deposits in the Lower Triassic of Russia. The more advanced trematosaurids are mostly Early Triassic, but one relict genus, *Hyperokynodon*, survived up to the Late Triassic in Germany. Records of the more advanced trematosaurids are also found in Laurasia from Arizona to Russia, but evidence of two subfamilies exists among all of the major Gondwanan temnospondyl faunas. Most are in freshwater deposits, but in Svalbard, a range of trematosaurid fossils is found in unambiguously marine deposits, providing one of the few examples of a marine amphibian fauna. The Late Triassic almasaurids and metoposaurids appear to have been restricted to the floodplains associated with the developing rift valley systems across north and central Pangaea, and most fossil materials are found in North America, Europe, North Africa, and India.

Capitosauroidea

The capitosauroids (also known as mastodonsauroids) were a large, successful, and comparatively structurally uniform group occurring globally throughout the Triassic, but not surviving into the Jurassic. Most adults were 2–3 m long (Figure 4), but individuals in a few taxa (e.g., *Mastodonsaurus*) grew to a total length of 5–6 m. The skull was superficially alligator-like in most genera, with a long and broad snout. One capitosauroid family of note is the Cyclotosauridae, a group characterized by having the tympanic (eardrum) region of the skull surrounded by outgrowths of the neighbouring skull bones (Figure 1F). Cyclotosaurids first appear as a minor group in the early Triassic but are seen to be the predominant capitosauroid group in the Late Triassic. *Mastodonsaurus* from the Middle Triassic of Germany is noteworthy as the first

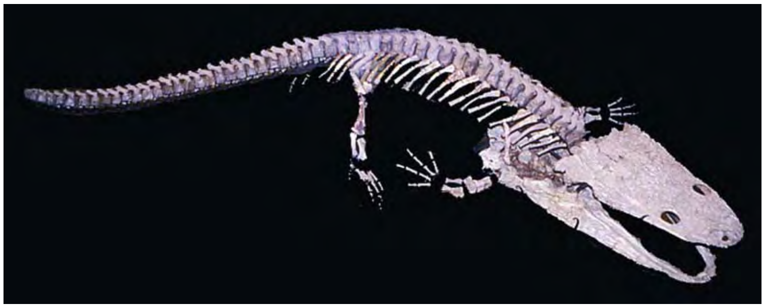


Figure 3 Stereospondyli. Skeleton of the metoposaurid *Buettneria perfecta*, from the Upper Triassic, North America. Specimen at the American Museum of Natural History, New York. © Andrew Milner.

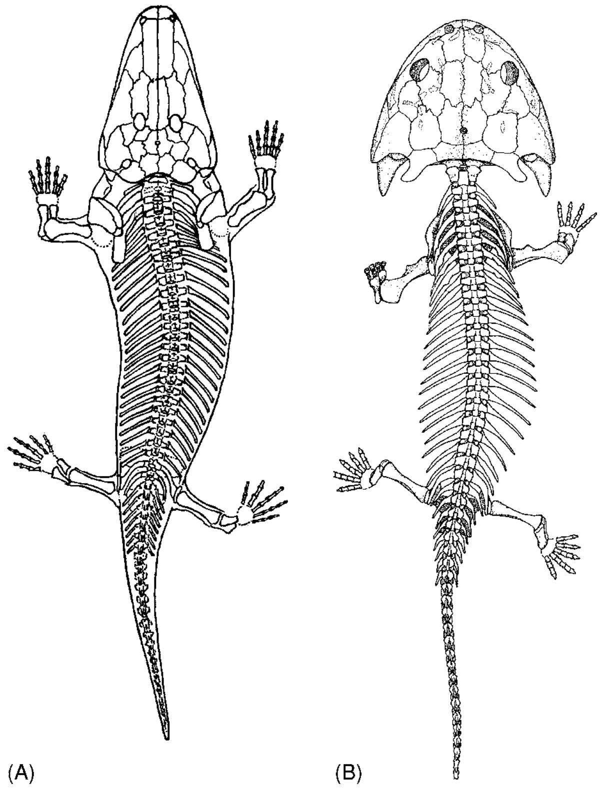


Figure 4 Stereospondyli. (A) Reconstruction of the capitosaurid *Paracyclotosaurus davidi*, from the Middle Triassic, New South Wales, Australia; (B) reconstruction of the chigutisaurid *Siderops kehli*, from the Lower Jurassic, Queensland, Australia. Reproduced with permission from (A) Watson, DMS (1958) A new labyrinthodont (*Paracyclotosaurus*) from the Upper Trias of New South Wales. *Bulletin of the British Museum (Natural History) Geological Series* 3, 233–263; (B) Warren AA and Hutchinson MN (1983) The last labyrinthodont? A new brachyopoid (Amphibia, Temnospondyli) fron the early Jurassic Evergreen Formation of Queensland, Australia. *Philosophical Transactions of the Royal Society, London Series B* 303, 1–62.

archaic fossil amphibian to be recognized and described (in 1828). Capitosauroidea may have originated in Gondwana, where the most primitive genus, *Watsonisuchus*, is known to have existed, but much of the record of this group suggests a uniform presence across Pangaea. The only subgroup that may be local in distribution is the Mastodonsauridae (*sensu stricto*), which has not yet been recognized outside Eurasia.

Enigmatic Paedomorphic Families

Three other families of temnospondyls represent very specialized aquatic forms that have apparently evolved by paedomorphosis (maintaining the larval body form throughout growth to a large size). The relationships between these three groups (Plagiosauridae, Chigutisauridae, and Brachyopidae) are controversial and they are treated here as three distinct groups of uncertain affinities.

Figure 5 Stereospondyli. Skull of the plagiosaurid *Gerrothorax rhaeticus*, from the Upper Triassic, Germany. Specimen at the Staatliches Museum fur Naturkunde in Stuttgart. © Andrew Milner.

The Plagiosauridae were one of the strangest-looking groups of temnospondyls; they were extremely flat, having a very wide and shallow gape, massive orbits (Figure 5), and vertebrae suggestive of a vertically undulating back, and they were covered in a chain-mail-like armour both dorsally and ventrally. The branchial (gill) arches were large and well-ossified and they clearly had a large internal gill apparatus. Most were less than 50 cm long, but one large fossil specimen from Svalbard is at least 2 m long. They are believed to be a paedomorphic lineage evolved from a more terrestrial stereospondyl group, and two quite different terrestrial genera, *Peltobatrachus* and *Laidleria*, have both been proposed as the terrestrial relatives of the plagiosaurs. Almost all plagiosaurids are known from Eurasia and Greenland with some fragments having been found in Thailand and Australia.

The Chigutisauridae were large stereospondyls with semicircular skulls bearing a suction–gulping feeding apparatus and small, anteriorly placed orbits (Figure 4B). They may be paedomorphic relatives of the Rhytidosteidae. All Triassic chigutisaurids are from the Upper Triassic of Gondwana. The Brachyopidae are similar (Figure 6), but much of the resemblance may be convergence and they have some similarities to the Tupilakosauridae (see later), which are certainly not stereospondyls. Most Triassic brachyopids are from Gondwana, with one problematic genus (*Virgilius*) from Arizona. The chigutisaurids and brachyopids occur at low taxonomic diversity through their history.

Other Temnospondyls

A few non-stereospondyl lineages of temnospondyls are known to have survived the P–T event and persisted as relicts for a short period in the Lower

Triassic. Best known is the genus *Micropholis* (Figure 7), which is locally common in the Karoo of South Africa. It is a late survivor of the Carboniferous–Permian family Amphibamidae, resembling long-legged salamanders and believed to be relatives of the modern amphibians by some workers. The genus *Tungussogyrinus* from the Tungus Basin in Siberia is a tiny gill-bearing temnospondyl, first described as a larva but recently argued to be a late member of the Branchiosauridae, small paedomorphic temnospondyls best known from Permo–Carboniferous lake deposits in central Europe. Finally the Tupilakosauridae known, from Greenland, Russia, and South Africa, were probably found throughout Pangaea and were long-bodied, almost eel-like, forms (Figure 8).

Figure 6 Skull of the brachyopid *Batrachosuchus haughtoni*, from the Middle Triassic, South Africa. Specimen at the Natural History Museum, London. © Andrew Milner.

Figure 7 An aggregation of five individuals of the temnospon dyl *Micropholis stowii*, from the Lower Triassic, South Africa. Spe cimen at the Bayerische Staatsammlung fur Geologie und Palaontologie, Munich. © Andrew Milner.

Chroniosuchians

Anthracosaurs, a significant group of Carboniferous and Early Permian large carnivores, are widely believed to be stem-amniotes, i.e., amphibian-grade organisms closer to the origin of reptiles, birds, and mammals than to modern amphibians. One relict anthracosaur taxon, the chroniosuchians, survived in the Late Permian of Russia (*see* **Palaeozoic:** End Permian Extinctions). They were small (50 cm–1 m) terrestrial carnivores superficially resembling monitor lizards. Until the early 1990s, they were assumed to have been wiped out by the P–T event. Recently, however, isolated characteristic armour-bearing vertebrae of chroniosuchians dating to the Middle Triassic have been found in Russia and Germany.

Post-Triassic Temnospondyls

Until 1977, it was generally assumed that all remaining temnospondyl lineages became extinct in the Norian during the extensive faunal turnover that occurred in the Late Triassic. This still may be true for central and western Pangaea, but two families of aquatic temnospondyl survived in the eastern peripheries of the continent until much later. The Brachyopidae are now known from the Middle–Upper Jurassic of Mongolia (*Gobiops*) and China (*Sinobrachyops*), and the similar Chigutisauridae have been described from the Lower Jurassic (*Siderops*; Figure 4B) and mid-Cretaceous of Australia. The latter record, *Koolasuchus* from the Aptian Stretzlecki Formation, is the latest known temnospondyl.

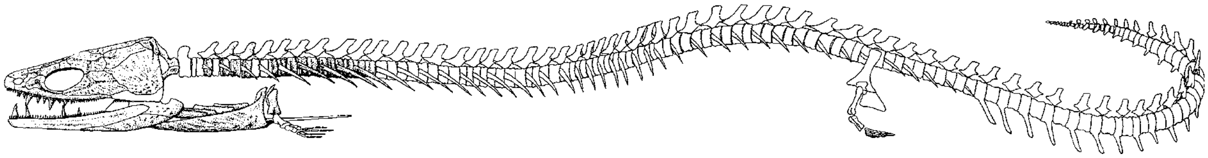


Figure 8 Reconstruction of the temnospondyl *Thabanchuia oomie*, from the Lower Triassic, South Africa. Reproduced with permission from Warren AA (1999) Karroo tupilskosaurid: a relict from Gondwana. *Transactions of the Royal Society of Edinburgh: Earth Sciences 89, 145 160.*

The Lissamphibian Groups

The taxon Lissamphibia applies to the living frogs, salamanders, and caecilians and their immediate relatives. They are believed by most workers to be a single adaptive radiation that started to diversify in the Permian, with the frogs demonstrably present in the Lower Triassic. The other groups first appear in the Jurassic and are presumed to be as yet undiscovered in Triassic rocks.

Frogs

Frogs occur on all non-polar continents and on many islands, particularly those that are relicts (New Zealand, Madagascar, and the Seychelles) of larger

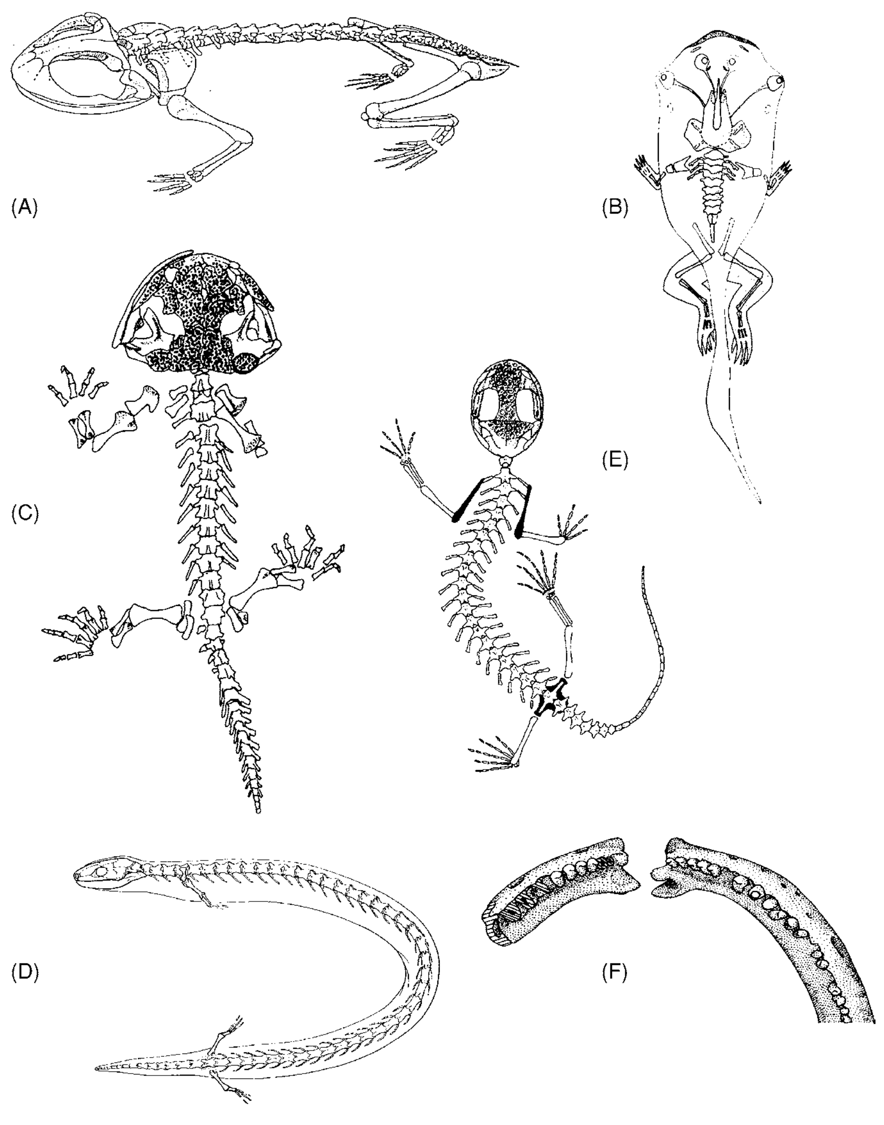


Figure 9 Lissamphibia. (A) Reconstruction of the protofrog *Triadobatrachus massinoti*, from the Lower Triassic, Madagascar; (B) the pipid tadpole *Shomronella jordanica*, from the Lower Cretaceous, Israel; (C), reconstruction of the primitive salamander *Karaurus sharovi*, from the Upper Jurassic, Kazakhstan, (D) reconstruction of the primitive caecilian *Eocaecilia micropodia*, from the Lower Jurassic, Kayenta Formation, Arizona; (E) reconstruction of the albanerpetontid *Celtedens ibericus*, from the Lower Cretaceous, Las Hoyas, Spain; (F) mandibles of the albanerpetontid *Celtedens*, showing the unique asymmetrical symphysial articulation. Reproduced with permission from (A) Rage, J C and Roček, Z (1989) Redescription of *Triadobatrachus massinoti* (Piveteau, 1936) an anuran amphibian from the early Triassic. *Palaeontographica A* 206: 1 16; (B) Estes, R, Špinar, ZV and Nevo, E (1978) Early Createceous pipid tadpoles from Israel (Amphibia: Anura) *Herpetologica* 34: 374 393; (C) Milner, AR (1994) Chapter 1. Upper Triassic andJurassic amphibians, fossil record and phylogeny. 5 22. In: Fraser, N and Sues, H D (eds), *In the Shadow of the Dinosaurs: Early Mesozoic Tetrapods*. Cambridge: Cambridge University Press; (D) Jenkins, FA and Walsh, D (1993) An early Jurassic caecilian with limbs. *Nature* 365: 246 250; (E) McGowan, G and Evans, SE (1995) Albanerpetontid amphibians from the Cretaceous of Spain. *Nature* 373: 143 145; (F) Fox, RC and Naylor, BG (1982) A reconsideration of the relationships of the fossil amphibian *Albanerpeton*. *Canadian Journal of Earth Science*. 19: 118 128.

continents. Frog skeletons are readily recognizable; the skeleton is highly modified for jumping, the backbone is very short (seven to nine vertebrae), and the tail vertebrae are fused as a rodlike urostyle. In the pelvis, the ilium is elongate, and in the hindlimb, the tibia and fibula are fused and the ankle bones are elongated. Fossils with some of the characteristics of frogs appear in the Early Triassic as *Triadobatrachus* (Madagascar) (Figure 9A) and *Czatkobatrachus* (Poland). *Triadobatrachus*, known from a near-complete skeleton, had the long ilium and modified ankles but had not yet acquired the other features. There are no other Triassic frogs or frog relatives, and the first frogs with the full suite of jumping adaptations are the Early Jurassic genera *Prosalirus* from Arizona and *Vieraella* from Argentina. By the Upper Jurassic, there are frogs attributable to the primitive modern family Discoglossidae in Laurasia, and it appears that the diversification to the modern families was under way.

Cretaceous frogs are not well documented, but the limited record suggests that much of the diversification to extant families had taken place. Discoglossids, pelobatids, rhinophrynids, and palaeobatrachids are known from Laurasia; leptodactylids and pipids are known from South America; and pipids are known from Africa. The genus *Gobiates*, sometimes elevated to the family Gobiatidae, is abundant in the mid-Late Cretaceous of central Asia but disappears in the Campanian and may be the only major type of Cretaceous frog not to survive into the Cenozoic. Rich assemblages of tadpoles of pipid frogs are known from the Lower Cretaceous of Makhtesh Ramon in Israel (Figure 9B).

Salamanders

The major diversity of living salamanders occurs in North America, central America, and Eurasia. A few genera are found in South America and North Africa, and these are believed to represent Neogene range extensions. Most of the Mesozoic record of salamanders is also restricted to Laurasia and the group appears to have been endemic to this region for much of its history. Diagnostic remains of salamanders first appear in the Bathonian of England and central Asia, and by the Late Jurassic they were widespread throughout Eurasia and North America. The known Jurassic salamanders are more primitive than most, if not all, living forms. A single articulated skeleton (*Karaurus*) is known from the Kimmeridgian of Kazakhstan (Figure 9C).

During the Cretaceous, the group diversified and basal members of some of the modern lineages appeared, and by the Maastrichtian, most modern families are represented in the record or can be inferred to have been present. Articulated skeletons are known from the konservat-lagerstätten at Las Hoyas, Spain (Figure 10), and Liaoning, China, and there are large numbers of salamander-producing microvertebrate assemblages in Eurasia and North America. Almost all salamanders were still restricted to the northern continents, but some reached the African region of Gondwana and are known from Israel (*Ramonellus*), Sudan, and Niger (*Kababisha*).

Figure 10 The salamander *Valdotriton gracilis*, from the Lower Cretaceous, Las Hoyas, Spain. Specimen at Las Hoyas Museum, Spain. © Andrew Milner.

Caecilians

The caecilians are tiny, burrowing amphibians now found in moist soils and leaf litter in South America, Africa, the Seychelles, India, and South-east Asia. They are superficially earthworm-like in shape, with no limbs or tail and with rudimentary eyes, and a scale distribution that gives them a segmented appearance. They are poorly represented in the Mesozoic, with just three records. The Lower Jurassic *Eocaecilia* from Arizona is known from articulated material and retained small limbs (Figure 9D). The two Cretaceous records from Morocco and Sudan are both based on isolated bones.

Albanerpetontids

The albanerpetontids are the extinct fourth group of lissamphibians; first recognized in 1976, subsequent discoveries have proved that they were widespread in the northern continents during the mid- to Late Mesozoic. They were small, superficially salamander-like forms (**Figure 9E**), uniquely characterized by the two halves of the lower jaw articulating anteriorly by means of an asymmetrical ball-and-socket joint (**Figure 9F**). Even small fragments of such jaws are readily recognized in microvertebrate assemblages. Articulated albanerpetontids are known from the konservat-lagerstätte at Las Hoyas, Spain, and this material is the key to understanding their morphology. Albanerpetontids first appear with the full suite of characteristics in the Bathonian, and they appear to have been present throughout Eurasia and North America in the Upper Jurassic and Cretaceous, with a recent report from the Lower Cretaceous of Morocco representing the first Gondwanan record.

See Also

Fossil Vertebrates: Palaeozoic Non-Amniote Tetrapods; Cenozoic Amphibians. **Lagerstätten**. **Palaeozoic:** End Permian Extinctions.

Further Reading

Bystrow AP and Efremov IA (1940) *Benthosuchus sushkini* Efr. A labyrinthodont from the Eotriassic of Sharjenga River. *Trudy Paleozoologicheskogo Instituta Akademii Nauk SSSR* 10: 1 152.

Holman JA (2003) *Fossil frogs and toads of North America*. Bloomington and Minneapolis: Indiana Press.

Milner AR (1990) The radiations of temnospondyl amphibians. In: Taylor PD and Larwood GP (eds.) *Systematics Association Special Volume 42. Major Evolutionary Radiations*, ch. 15, pp. 321 349. Oxford: Clarendon Press.

Milner AR (2000) Mesozoic and Tertiary Caudata and Albanerpetontidae. In: Heatwole H and Carroll RL (eds.) *Amphibian Biology, Volume 4, Palaeontology: The Evolutionary History of Amphibians*, ch. 18, pp. 1412 1444. Chipping Norton, NSW: Surrey Beatty.

Roček Z (2000) Mesozoic anurans. In: Heatwole H and Carroll RL (eds.) *Amphibian Biology, Volume 4, Palaeontology: The Evolutionary History of Amphibians*, ch. 14, pp. 1295 1331. Chipping Norton, NSW: Surrey Beatty.

Sanchíz B (1998) Salientia. In: Wellnhofer P. (ed.) *Handbuch der Paläoherpetologie*. part 4. Munich: Pfeil.

Schoch RR and Milner AR (2000) Stereospondyli. In: Wellnhofer P (ed.) *Handbuch der Paläoherpetologie*. part 3B. Munich: Pfeil.

Warren AA (2000) Secondarily aquatic temnospondyls of the Upper Permian and Mesozoic. In: Heatwole H and Carroll RL (eds.) *Amphibian Biology, Volume 4, Palaeontology: The Evolutionary History of Amphibians*, ch. 8, pp. 1122 1149. Chipping Norton, NSW: Surrey Beatty.

Cenozoic Amphibians

A R Milner, Birkbeck College, London, UK

Introduction

The fossil record suggests that archaic amphibian groups had entirely vanished by the Cenozoic, and that at least the majority of extant families were already established, many in the regions in which they still occur. In this article, the Cenozoic fossil record of each of the modern groups and of the extinct albanerpetontids is reviewed.

Amphibians and the Cretaceous and Tertiary Boundary

By the Late Cretaceous, the fossil record indicates that the only amphibians were the three modern lissamphibian groups (frogs, salamanders, and caecilians) and the enigmatic albanerpetontids. Such evidence as exists suggests that the Cretaceous and Tertiary (K–T) boundary event (*see* **Mesozoic:** End Cretaceous Extinctions) had no effect on the diversity of amphibians. In Montana, United States, and Alberta, Canada, there are freshwater faunas bracketing the Maastrichtian to Paleocene transition, and the frog and salamander assemblages are unchanged in diversity and taxonomic content through this time. Evidence from the rest of the world is consistent with this, but at too poor a resolution to be convincing in its own right.

Cenozoic Assemblages

Cenozoic amphibians are represented either as complete skeletons in assemblages in a few konservat-lagerstätten (*see* **Lagerstätten**) (localities, or ‘lagerstätten’, where prevailing conditions have produced highly conserved or diverse fossils) or, more commonly, as isolated elements in microvertebrate accumulations. Notable lagerstätte assemblages are the Eocene brown coals from Messel and Geiseltal in Germany, and the Oligocene–Miocene freshwater limestones from Oeningen in Switzerland and from Bechlejovice in the Czech Republic. These produce large numbers of complete, articulated frogs and tadpoles and a smaller number of salamanders. Modified body tissues may be preserved, but the skeletons are generally crushed flat. In contrast, microvertebrate assemblages may contain few to abundant amphibian bones that are three-dimensionally preserved but are not associated, resulting in problems of association unless there is significant similarity to living taxa.

The Cenozoic record of caecilians is negligible and that of albanerpetontids is too patchy to evaluate in the absence of living descendants. The record of frogs and salamanders predominantly derives from North America and Eurasia, with frogs also well represented in South America. For salamanders, this 'Laurasian' range may represent a genuine pattern, because most living genera of salamander are found in the same area, but for frogs it is undoubtedly an artefact of the distribution of microvertebrate collecting. There is ample morphological and molecular evidence that the Neobatrachia or higher frogs originated in Gondwana and that they diversified during the breakup of that supercontinent. It can be anticipated that a substantial record will eventually be found in Africa, India, and Australia as well as in South America.

Aquatic amphibians living in lakes and ponds are preserved most readily and so are disproportionately well represented in the fossil record. The frog families Pipidae and Palaeobatrachidae (Figure 1) and the salamander family Cryptobranchidae fall into this category. At the other extreme, stream-dwelling amphibians and those living in damp terrestrial environments are hardly represented at all. The salamander family Plethodontidae, for example, comprised mainly of such forms, is the largest family of living salamanders but has one of the poorest fossil records.

Frogs

Frogs occur on all non-polar continents and on many islands, particularly those such as New Zealand, Madagascar, and the Seychelles, which are relicts of larger continents. Early in Mesozoic frog evolution, there appears to have been a major dichotomy between Laurasian and Gondwanan frogs, with the discoglossids, pelobatids, pelodytids, and palaeobatrachids evolving in Laurasia, and the pipids and neobatrachians (~20 families) evolving in Gondwana. By the Cenozoic and up to the present day, this division is still recognizable in the distribution of frog families. However, during the Cenozoic, several successful Gondwanan families, the ranid frogs (Figure 2), bufonid toads, and hylid tree frogs, extended their ranges into Laurasia, diluting the original pattern. First appearances of fossils of these families in the northern continents show this to have happened in the Paleocene and Eocene, for ranids and bufonids.

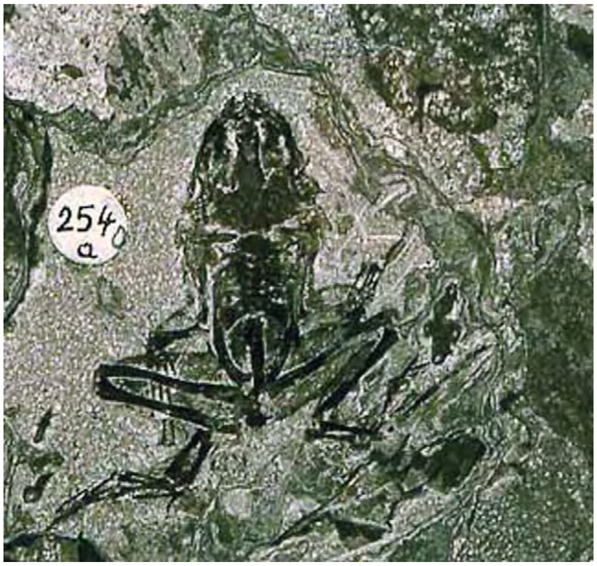


Figure 1 The palaeobatrachid frog *Palaeobatrachus grandipes*, from the Oligocene, Bechlejovice, Czech Republic. Specimen at the National Museum, Prague. © Andrew Milner.

Much of the Cenozoic fossil record of frogs comprises earlier representatives of the extant frog faunas of each continent, e.g., leptodactylids in South America, pipids and ranids in Africa, myobatrachids in Australia, and discoglossids and pelobatids in Europe. Occasionally, however, frog fossils reveal that past faunas had distinctive components. *Latonia gigantea*, for example, a discoglossid from the Oligocene–Pliocene of Europe, was twice as large as any modern discoglossid and had a heavily sculptured skull.

Salamanders

The major diversity of living salamanders occurs in North America, central America, and Eurasia. A few genera are found in South America and North Africa, and these are believed to represent Neogene range extensions. The Cenozoic record is consistent with this, with much of the fossil material belonging to the same families and genera that occur in each continent today. Most of the European fossil record is of salamandrids and proteids; the Asian record comprises salamandrids, hynobiids, and cryptobranchids; and the North American record is of cryptobranchids, sirenids, amphiumids, proteids, ambystomatids, and

Figure 2 The ranid frog *Rana ridibunda* (*'Rana pueyoi'*), from the Miocene, Libros, Spain. Specimen at the Natural History Museum, London. © Andrew Milner.

plethodontids. Some records fall outside this pattern. A large Paleocene–Eocene salamander, *Piceoerpeton* (of uncertain relationships), not only occurs within the modern range of salamanders in North America but has been found as far north as Ellesmere Island in Eocene beds, demonstrating a major Paleogene range extension for the group. The large aquatic cryptobranchid salamanders appear to have originated in Asia, where they occur as Paleocene fossils and the living *Andrias*. *Andrias* (Figure 3) first appears in Europe in the Late Oligocene, after the 'Grand Coupure' event, a combination of climate change and faunal dispersal between Europe and Asia. *Andrias* fossils are abundant from the Late Oligocene to the Pliocene, after which the species became extinct in Europe.

Caecilians

Caecilians are now found in South America, Africa, the Seychelles, India, and south-east Asia. Phylogenetic evidence from living forms indicates that the first four areas represent relicts of an original west–central Gondwana range, whereas the south-east Asian taxa represent a range extension out of India after its contact with Asia. Cenozoic fossils might be expected in any of these areas, but, in practice, caecilian bones are so rare and so tiny that the only discoveries to date are two single vertebrae from the Paleocene, one from Brazil and one from Bolivia.

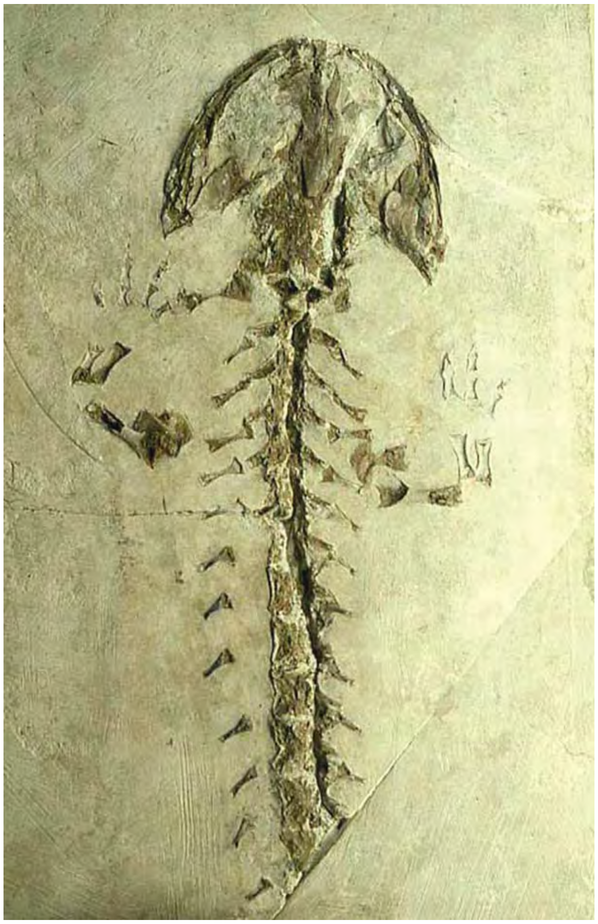


Figure 3 The cryptobranchid salamander *Andrias scheuchzeri,* from the Oligocene, Oeningen, Switzerland. Specimen at the Museum fur Naturkunde, Humboldt University, Berlin. © Andrew Milner.

Albanerpetontids

The Albanerpetontidae are a group of Jurassic–Pliocene amphibians; they were small salamander-like animals but are structurally distinct from any of the living groups. A unique feature is that the two halves of the lower jaw articulate anteriorly by means of an asymmetrical ball-and-socket joint. Albanerpetontid bones had been misassociated with those of salamanders for many years, but were finally recognized in 1976 from Miocene microvertebrate material from La Grive St Alban in France, this material forming the original *Albanerpeton inexpectatum* (Figure 4). Most subsequent discoveries have

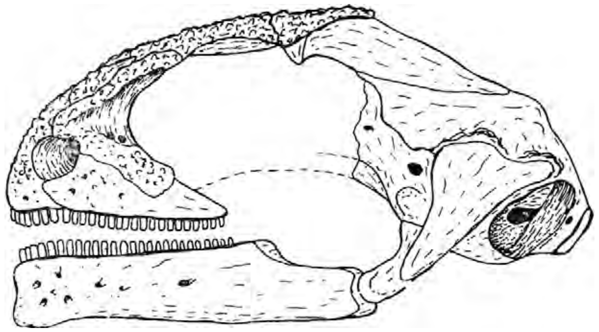


Figure 4 The skull of the Miocene amphibian *Albanerpeton inexpectatum*, from La Grive St Alban, France. Reproduced with permission from Estes R and Hoffstetter R (1976) Les urodèles du Miocène de la Grive Saint Alban (Isère, France). *Bulletin du Muséum National d'Histoire Naturelle, Sciences de la Terre* 57: 297 343.

been Mesozoic, but further finds have been made in Paleocene material of North America, Miocene material of Germany, and Pliocene material of Romania. At present, it appears that a few lineages survived the K–T event in the northern continents, making the last appearance of these species in North America in the Paleogene, but lingering on in Eurasia until the Pliocene, when they appear to have died out during the glaciations.

Amphibians and the Pleistocene Glaciations

The sequence of Pleistocene glaciations appears to have had a dramatic effect on the amphibian fauna of Europe. In part, this is undoubtedly because the southward movement of aquatic vertebrates was constrained by the Pyrenees, the Alps, the Carpathians, and, more profoundly, the Mediterranean Sea. Several genera of amphibians, widespread as fossils over Early Neogene Europe, are now restricted to various permutations of the Iberian Peninsula, Italy, and the Balkans (the salamander genera *Pleurodeles* and *Chioglossa* and the frog genus *Discoglossus*). More dramatically, four families became extinct in Europe, the hynobiid and cryptobranchid salamanders surviving elsewhere, but the palaeobatrachid frogs and albanerpetontids dying out entirely, Europe apparently being their last refugium.

See Also

Lagerstätten. **Mesozoic:** End Cretaceous Extinctions.

Further Reading

Báez AM (2000) Tertiary Anura of South America. In: Heatwole H and Carroll RL (eds.) *Amphibian Biology, Volume 4, Palaeontology: The Evolutionary History of Amphibians* ch. 15, pp. 1388 1401. Chipping Norton, NSW: Surrey Beatty.

Holman JA (1995) *Pleistocene Amphibians and Reptiles in North America. Oxford Monographs on Geology and Geophysics, 32*. New York, Oxford: Oxford University Press.

Holman JA (1998) *Pleistocene Amphibians and Reptiles in Britain and Europe. Oxford Monographs on Geology and Geophysics, 38*. New York, Oxford: Oxford University Press.

Holman JA (2003) *Fossil Frogs and Toads of North America*. Bloomington and Minneapolis: Indiana Press.

Milner AR (2000) Mesozoic and Tertiary Caudata and Albanerpetontidae. In: Heatwole H and Carroll RL (eds.) *Amphibian Biology, Volume 4, Palaeontology: The Evolutionary History of Amphibians* ch. 18, pp. 1412 1444. Chipping Norton, NSW: Surrey Beatty.

Rage J C and Roček Z (2003) Evolution of anuran assemblages in the Tertiary and Quaternary of Europe, in the context of palaeoclimate and palaeogeography. *Amphibia Reptilia* 24: 133 167.

Roček Z and Rage J C (2000) Tertiary Anura of Africa, Asia, Europe, North America and Australia. In: Heatwole H and Carroll RL (eds.) *Amphibian Biology, Volume 4, Palaeontology: The Evolutionary History of Amphibians*. ch. 15, pp. 1334 1389. Chipping Norton, NSW: Surrey Beatty.

Sanchíz B (1998) Salientia. In: Wellnhofer P (ed.) *Handbuch der Paläoherpetologie,* part 4. Munich: Pfeil.

Mesozoic Mammals

Z-X Luo, Carnegie Museum of Natural History,
Pittsburgh, PA, USA

Introduction

Two-thirds of mammalian history occurred during the Mesozoic, when dinosaurs dominated the land. During their long history in the Mesozoic, mammals diversified into many branches of a prolific evolutionary tree and underwent an enormous morphologic evolution. To date, more than 280 genera of Mesozoic mammals and mammaliaforms are known to science, and these belong to more than 25 distinctive family-level lineages, each of which has unique dental characteristics. Only a tiny fraction of these diverse Mesozoic mammalian groups survived to give rise to the three modern mammalian groups: the egg-laying monotremes, the pouched marsupials, and the placentals, with an elaborated placental reproductive structure.

Mesozoic mammals represent the trunk and the basal branches of the entire mammalian family tree, and they provide evidence of the ancestral condition from which modern mammals have evolved. Humans are primates, primates are placentals, placentals are eutherians, and eutherians are rooted in the evolutionary past of all Mesozoic mammals. The fossil record of Mesozoic mammals, therefore, is indispensable for understanding the deep history that gave rise to extant mammals, including the placental lineage to which humans belong.

Mammals are part of a more inclusive evolutionary lineage known as the synapsids, which are distinguishable from other non-synapsid vertebrates by the presence of a lower temporal fenestra in the skull. The earliest known fossils of the synapsid lineage are from the Late Carboniferous (300 million years ago) in the Palaeozoic. Cynodonts are a derived subgroup of synapsids; they emerged in the Late Permian, around 250 million years ago, and thrived in the Triassic, with some relicts surviving into the Early Cretaceous. Cynodonts are distinguished from precynodont synapsids in having a better developed secondary bony palate that separates the nasal passage (for breathing) from the mouth cavity (for feeding), an enlarged temporal fenestra, and a larger coronoid process on the mandible for the better developed jaw adductor muscles, plus a functional middle ear in the posterior part of the mandible.

Mammaliaforms are a subgroup of cynodonts (Figure 1), and they are distinguishable from premammaliaform cynodonts in three ways: by the presence of a derived jaw hinge, formed by the dentary bone of the mandible and the squamosal bone of the cranium; by the more derived features of the ears, such as the promontorium, which encloses the inner ear cochlea; and by precise dental occlusion. Mammaliaforms are close relatives of modern mammals, with many very mammal-like features; but they are not advanced enough to be placed within the evolutionary lineage of modern mammals. Many mammaliaforms are known from the fossil record of the Late Triassic through the Late Jurassic.

Modern Mammalia are a subgroup of mammaliaforms, and they are defined as an evolutionary lineage consisting of the ancestor common to all three living mammalian groups, plus those fossil taxa that can be placed within these living mammal groups by comparative morphological evidence in the bone and dental structures.

Origins of Mammalian Features

Modern mammals are more derived than most non-mammalian vertebrates are, in five aspects: (1) the jaw hinge formed by the dentary condyle and squamosal glenoid (instead of the quadrate and articular bones), (2) the middle ear suspended in the base of the cranium (and detached from the mandible), (3) the inner ear with elaborated cochlear canal, a part of the auditory bony labyrinth structure, (4) the enlarged brain endocast and a series of braincase structures to support and protect the brain, and (5) a diphyodont (two generations) dental replacement (instead of multiple dental replacements as in toothed modern reptiles). Some precursor conditions to these derived evolutionary features of modern Mammalia can be traced to the mammaliaforms in the Late Triassic to Early Jurassic (Figure 1); other mammal-like precursor features can be seen in the advanced cynodonts. By mapping the patterns of distribution of various precursor conditions of mammalian features on the successive hierarchies of mammals, mammaliaforms, and premammaliaform cynodonts on the evolutionary tree (Figure 2), it is possible to infer the historical sequence through which the derived mammalian features arose during the transition from premammaliaform cynodonts to the earliest mammaliaforms, and then from mammaliaforms to mammals. Mammals underwent significant anatomical evolution in their early history; the fossil evidence of this evolution is so extensive that the origins of mammals

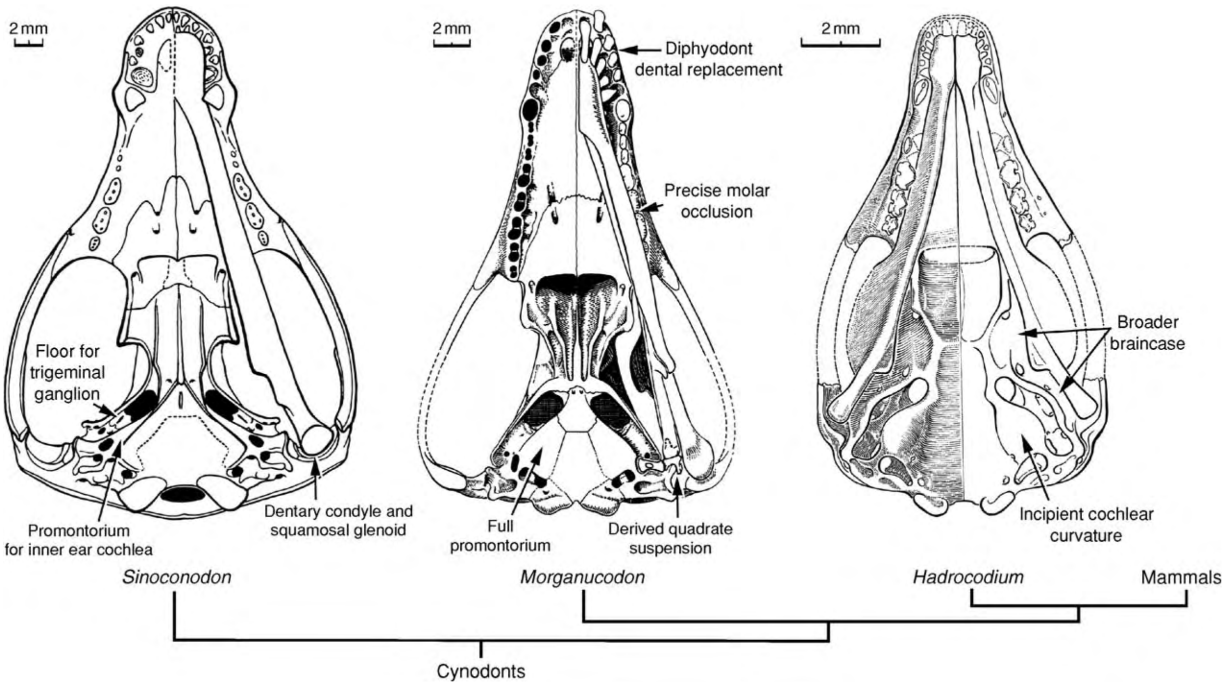


Figure 1 Mammaliaforms (skulls in ventral view) of the Late Triassic and Early Jurassic, showing the derived (diagnostic) mammalian characteristics that evolved with these transitional mammaliaforms. *Sinoconodon*, from the Lower Jurassic, is one of the most primitive mammaliaforms known; it has a jaw hinge of dentary condyle and squamosal glenoid, a petrosal promontorium for the cochlear canal, and a braincase floor for the trigeminal cranial nerve ganglion. *Morganucodon*, from the Late Triassic and Early Jurassic, is the earliest known mammaliaform with precise occlusion of the upper and lower molars and a diphyodont dental replacement pattern. *Hadrocodium* from the Early Jurassic, is the most derived among mammaliaforms in having a more enlarged braincase, an enlarged promontorium with curved cochlea, and the absence of primitive mandibular structures for attaching the middle ear. (Morganucodon modified from KA Kermack *et al.* (1981). *Zoological Journal of Linnean Society*. London 71: 1 158.)

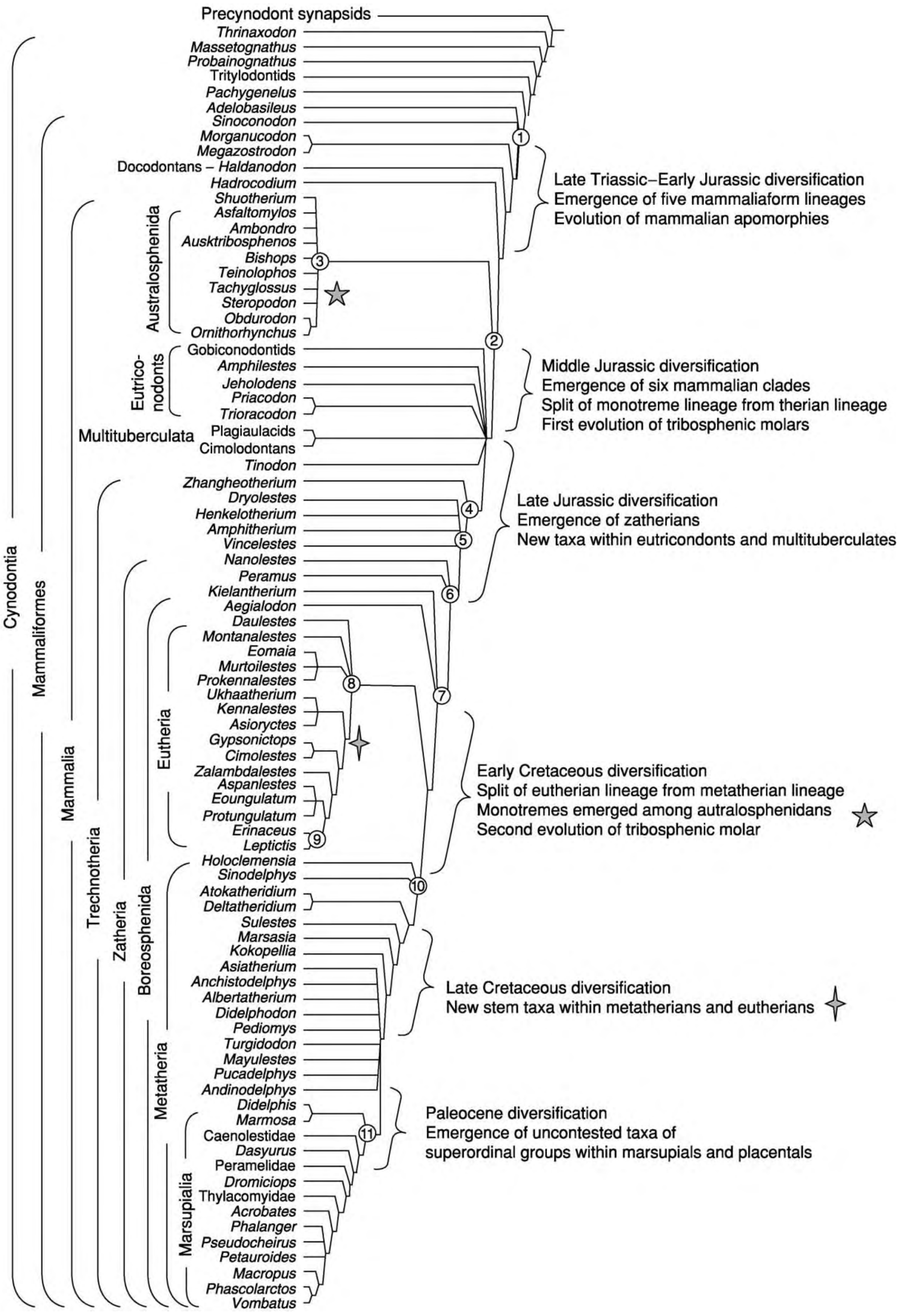
from premammalian cynodonts is one of the best documented cases for the origin of a major vertebrate group.

Jaw Hinge and Middle Ear

The earliest mammaliaforms are characterized by a 'double joint' in their jaw hinge (Figure 2: *Morganucodon*). They developed the derived (mammalian) dentary condyle to squamosal glenoid in the jaw joint, but also retained the primitive (cynodont) features of an articular-quadrate jaw joint. The earliest mammaliaforms are similar to premammaliaform cynodonts in that the middle bones (also known as the postdentary bones) and the ossified Meckel's cartilage are attached anteriorly to the mandible, in contrast to extant mammals, in which the embryonic Meckel's cartilage disappeared and the middle ear bones are separated from the mandible in adults. The quadrate (homologue to the incus of the mammalian middle ear) has developed contact with the petrosal bone in some mammaliaforms, as in modern mammals. In other words, the earliest mammaliaforms have achieved the diagnostic mammalian features (the dentary/squamosal jaw joint and the way in which the middle ear is suspended by the cranium via the quadrate) (Figure 2).

Other features of the jaw hinge and the middle ear in these Jurassic mammaliaforms are primitive features shared by more distant premammaliaform cynodonts. In the more derived eutriconodont mammals, the middle ear elements, such as the angular (homologue to the tympanic in mammals), achieved some

Figure 2 Evolutionary relationships of Mesozoic mammals, mammaliaforms, and cynodonts. The phylogenetic ranks of Mesozoic mammals and mammaliaforms correspond approximately to the temporal sequence of their diversifications. Taxic macroevolution of Mesozoic mammals is characterized by several successive episodes of diversifications and emergence of new clades with new dental features. Numbers represent nodes for successive ranks or hiearchies of Mesozoic mammalian phylogenetic relationships: 1, Mammaliaformes; 2, Mammalia; 3, Australosphenida; 4, Trechnotheria; 5, Cladotheria; 6, Zatheria; 7, Boreosphenida; 8, Eutheria; 9, Placentalia; 10, Metatheria; 11, Marsupialia. (Cladogram modified from Z X Luo *et al.* (2003) *Science* 302: 1934 1940.)



Late Triassic–Early Jurassic diversification
Emergence of five mammaliaform lineages
Evolution of mammalian apomorphies

Middle Jurassic diversification
Emergence of six mammalian clades
Split of monotreme lineage from therian lineage
First evolution of tribosphenic molars

Late Jurassic diversification
Emergence of zatherians
New taxa within eutricondonts and multituberculates

Early Cretaceous diversification
Split of eutherian lineage from metatherian lineage
Monotremes emerged among autralosphenidans ☆
Second evolution of tribosphenic molar

Late Cretaceous diversification
New stem taxa within metatherians and eutherians ✦

Paleocene diversification
Emergence of uncontested taxa of
superordinal groups within marsupials and placentals

degree of medio-lateral separation from the mandible, whereas the middle ear bones are connected anteriorly to the mandible via the Meckel's cartilage, even in adults. The best available evidence suggests that loss of the Meckel's cartilage may have occurred separately in modern monotremes and in the lineage of placentals plus marsupials, after the ancestor of the latter groups had split from eutriconodontans.

Elaborated Structure of Inner Ear Cochlea

Mammals are most derived among living vertebrates in their complex inner ear cochlea and related hearing adaptation. The petrosal bone has developed an enlarged bony housing for the cochlear part of the inner ear; this is crucial for sensitive hearing function, especially for high-frequency sound. Development of the derived inner ear and its bony housing occurred in several evolutionary steps in the transition from cynodonts and early mammals, as evidenced by the precursor conditions of these structures in mammaliaforms.

Premammalian cynodonts have a cochlear cavity for the auditory function in the inner ear. This cochlear cavity is small and globular; it does not extend anterior to the fenestra vestibuli in the reconstructed inner ear endocast. The bony housing for the inner ear is formed by a mosaic of multiple bones. Two of these bones, the prootic and the opisthotic bones, correspond to the homologous bones in extant nonmammalian vertebrates, which are developed from ossification of the embryonic auditory capsule in modern vertebrates. None of the bones surrounding the inner ear are fused. The bony housing of the inner ear in adults of the cynodont *Thrinaxodon* is formed by a mosaic of several bones, which can be either endochondral or intramembranous in embryonic origin among extant vertebrates.

By comparison, the inner ear of Early Jurassic mammaliaforms has an elongate bony cochlear canal that is much better developed compared to the cochlear cavity of premammaliaform cynodonts. The elongate bony cochlear canal may indicate a greater sensitivity to high-frequency sound, which is very important in the hearing function of all extant mammals and was probably important for at least some of the earliest mammals. The bony housing of the inner ear in stem taxa of mammals is formed exclusively by the petrosal, which is the single bone homologous to the fused prootic and opisthotic elements in premammaliaform cynodonts. The petrosal is not only much larger than the prootic and opisthotic, but it also forms a distinctive structure of the promontorium. The enlarged petrosal excludes the other cranial bones, such as the basisphenoid complex, the basioccipital, and the exoccipital, from the bony housing for the inner ear. The mosaic of multiple bones for the inner ear housing of cynodonts is replaced by a single bone in the derived mammaliaforms, including modern mammals.

The differences between mammals and cynodonts in the inner ear and its bony housing are the result of their correlated structural transformation. Enlargement of the promontorium is correlated with elongation of the cochlear canal. The enlarged promontorium displaced the neighbouring sphenoid complex and basioccipital bone. Inflation of the bulbous promontorium in the mammalian crown group is associated with coiling of the cochlear canal. If the precursor condition in mammaliaforms is mapped on the cynodont-mammal evolutionary tree, there is a clear pattern of incremental evolution of mammalian characteristics in the inner ear cochlear canal and the inner ear bony housing.

The Larger Brain

In early mammalian evolution, there is a clear tendency towards a larger volume of brain endocasts, and their morphological features are also better differentiated in the more derived taxa. The relative brain sizes of the transitional mammaliaforms of the Early Jurassic are intermediate between the larger brain size of the derived Cretaceous mammals and the smaller brain size of the Triassic cynodonts. Larger brain capacity of mammals compared to that of nonmammalian vertebrates indicates a larger volume of metabolically expensive neural tissues of the mammalian brain. This can be further correlated to the development of a more elevated metabolism in early mammaliaforms than in premammaliaform cynodonts. It could also be correlated with better sensory perception and elaborated neural control of the skeleto-muscular system for mastication and locomotion.

In the larger brain endocasts in the successively more derived mammaliaforms and mammals, the cerebral hemispheres are much better developed in the successively more derived groups. The posterior part of the brain endocast tends to show a better differentiation of the mid-brain from the cerebellar structure in the mammalian crown groups: metatherians, eutherians, and monotremes. Related to the tendency of increasing brain size, new structures in the braincase in the derived mammaliaforms are absent in more primitive cynodonts. For example, the braincase of mammaliaforms has a bony floor to enclose the trigeminal ganglion of the cranial nerve V into the braincase (Figure 1: *Sinoconodon*). The enlarged brain is correlated with a posterior shift of the braincase relative to the jaw hinge and other skull structures. In most premammaliaform cynodonts, the floor to the narrow anterior braincase was not ossified and the related orbital structure of the skull is also absent. The floor for the anterior part was presumably formed by the cartilaginous structure of the cranium in life. In mammaliaforms, by comparison, the floor of the anterior braincase is fully ossified, as are most of the orbital structures.

Diphyodont Dental Replacement

The diphyodont dental replacement in modern placental mammals is characterized by a single tooth replacement for incisors, canines, and premolars, but the molars have only one generation and are never replaced. Marsupial mammals show a more reduced dental replacement compared to the typical diphyodont replacement of placentals. The number of replacements per tooth locus in modern mammals is far smaller than in the polyphyodont ('multiple generations') replacement of nonmammalian vertebrates with teeth, as in crocodiles, dinosaurs, and premammaliaform cynodonts.

Evolution of reduced dental replacement rates in mammals is correlated with lactation. The reduction of dental replacement is feasible because lactation after birth makes it possible for mammalian neonates to achieve fast cranial growth without teeth, resulting in a delay in eruption of the deciduous teeth. Also, because relatively fast early cranial growth of the mammalian skull during lactation will slow after weaning, skull growth terminates relatively early in the life of individuals, accompanied by an early termination of dental replacement. The end of skull growth (determinate pattern) usually coincides with the eruption of the last molar. Therefore, diphyodont dental replacement is a crucial and derived feature that indicates the life history characteristics of modern mammals.

Some precursor conditions to modern mammalian dental replacement are seen in *Sinoconodon,* one of the earliest mammaliaform lineages to undergo a split. *Sinoconodon* has mammalian characteristics in the jaw hinge and in the ear region (Figure 1), but has retained some cynodont-like characteristics of the dental replacement, and a 'reptilian' pattern of skull growth. The incisors and canines in *Sinoconodon* were replaced at least three times in an alternating pattern, as seen in many cynodonts and other toothed non-mammalian vertebrates. The posterior molars of *Sinoconodon* have one replacement in the larger (presumably older) individuals, in a manner that is similar to the replacement pattern of some cynodonts, such as diademodontids.

Replacement of the postcanines (premolars and molars) in *Sinoconodon* is sequential in the anteroposterior direction, a mammal-like feature. But after the replacement of the premolars, the premolars and anterior molars are lost, resulting in a dental diastema (gap in the dental series) that becomes increasingly larger in older individuals. Coupled with successive addition of the newly erupted molariforms at the posterior end of the tooth row, there is a posterior shift of the functional tooth rows in the jaws. These primitive features are present in some cynodonts and also in other mammaliaforms to a lesser extent.

The currently available sample of *Sinoconodon* specimens shows a large range of growth, from the smallest individual, with an estimated body mass of about 13 g and already showing the canine replacement, to the largest, with an estimated body mass of more than 500 g. During this growth, the posterior molariforms were being replaced while the upper and lower jaws continued to lengthen in the successively older individuals. This suggests that *Sinoconodon* had indeterminate growth at least in its skull, associated with continuous tooth replacement, as in premammaliaform cynodonts. Given its basal position in the mammalian family tree, the characteristics of the dental replacement and skull growth of *Sinoconodon* can be regarded as an intermediate stage in the evolution from the primitive pattern of polyphyodont replacement seen in most cynodonts to the derived diphyodont replacement of mammals. Reduction in premolar and molariform replacement in the posterior dentition preceded reduction of the replacement of incisors and canines in the anterior dentition.

Interestingly, the mammaliaforms (e.g., *Morganucodon* and docodontans) that are closer to modern mammals all have the typical diphyodont replacement of modern mammals and a narrower range of skull growth similar to that of modern mammals. When dental replacement characteristics are mapped on the mammalian evolutionary tree, it helps to show that the transition from reptile-like skull growth and dental replacement to the typical mammalian determinate skull growth and diphyodont dental replacement occurred in the Late Triassic through Early Jurassic. Related to diphyodont dental replacement, derived mammaliaforms *Morganucodon*, *Hadrocodium*, and docodontans also evolved precise occlusion of molars, in which the upper and lower molars developed wear facets associated with individual molar cusps. *Sinoconodon* lacked precise molar occlusion because it did not have a consistent pattern of opposition between upper and lower molars, as a consequence of the partial replacement of the posterior molariforms and the successive posterior shift of the functional tooth row, as part of the indeterminate growth pattern of the skull. However, *Morganucodon* and other taxa more closely related to modern mammals than *Sinoconodon* have developed the precise cusp occlusion between the upper and lower molars for better and effective mastication of food.

Successive Diversifications of Mesozoic Mammals

Compared to other small vertebrates in the terrestrial biota, Mesozoic mammals were never numerically abundant. Most taxonomic records of Mesozoic mammals are based on teeth and jaw fragments. Nonetheless, sufficient numbers of their fossils have been collected for palaeontologists to know that there were several episodes of diversification of Mesozoic mammals. The first burst of diversification occurred as several mammaliaforms emerged during the Late Triassic to Early Jurassic. Although these mammaliaforms have almost identical mandibular design, their dentitions are different. The five main groups of mammaliaforms in the Triassic and Early Jurassic have developed three distinctive types of specialized teeth. Haramiyidans have multirow, multicusped, or 'multituberculate-like' molars; *Sinoconodon*, morganucodontans, and *Hadrocodium* have 'triconodont-like' molars showing three main cusps in alignment. Kuehneotheriids have triangulated, 'symmetrodont-like' molars. The phyletic diversification is characterized most prominently by the differentiation in dental morphology.

The Middle Jurassic diversification saw the emergence of six more order- or family-level evolutionary lineages: amphilestids, with triconodont-like molars; eleutherodontans, with multituberculate-like molars; spalacotheriids, with 'symmetrical' and triangulated molars; and amphitheriids and peramurids, with a triangulate trigonid plus a talonid heel on the lower molars. These groups are known mostly from the Laurasian continents during the Middle Jurassic. A major morphological innovation in mammalian evolution of this time was the molar structure for grinding function, which evolved in at least three separate Middle Jurassic lineages: docodontans, with complex shearing and grinding surfaces on the molars; shuotheriids, with an anterior grinding basin ('pseudotalonid'); and australosphenidans from Gondwana continents, with the tribosphenic molars capable of the shearing ('sphen') and the mortar–pestle crushing and grinding ('tribo'). The tribosphenic lower molar has a posterior talonid basin, functioning as a mortar that receives the inner main cusp (protocone) of the upper molar that is analogous to a pestle. These advanced features of crushing and grinding of tribosphenid mammals are developed in addition to the primitive shearing features on the molars that are also shared by their pretribosphenic relatives. It is obvious that the adaptive molar structures for grinding functions are homoplastic among these groups.

Consistent with this theory of homoplastic evolution of the grinding functions of molars is the fact that the derived grinding features of multifunctional molars in docodontans, shuotheriids, and at least some Middle Jurassic australosphenidans co-existed with the primitive features of the postdentary trough on the mandibles of these groups. The combined analyses of these primitive mandibular features and the derived molar features in the Jurassic southern mammals suggest that the australosphenidans or the southern tribosphenic mammals are not closely related to boreosphenidans, or the northern tribosphenic mammals. The latter group evolved some 25 million years later in the Laurasian continents, as a separate lineage with a convergent molar design, but far more derived mandibular features.

The third episode of diversification among early mammals occurred in the Late Jurassic, with five newly evolved order-or family-level lineages: multituberculata (*sensu stricto*), triconodontids, spalacotheriids, tinodontids, and dryolestoids. The most significant apomorphic feature of these emergent groups of the Late Jurassic is the absence of the postdentary trough on the mandible, which is a primitive feature retained by almost all Early Jurassic and Middle Jurassic lineages (except for amphilestids and *Hadrocodium*). Other striking similarities are in the structure of the dentary. Despite the fact that eutriconodontans, multituberculates, and spalacotheriids have very different dentitions, these three groups all have fairly similar characteristics in the posterior part of the mandible – a rounded 'angular' region grades into the dentary condyle, and there is often a prominent medial pterygoid crest ('shelf') along the ventral border of the mandible.

The last major episode of diversification occurred in the Early Cretaceous when the stem taxa of metatherians, eutherians, and basal boreosphenidans appeared in Laurasia, while toothed monotremes and some of their australosphenidan relatives appeared in Gondwana. The only new major lineage to appear in the Late Cretaceous was that of gondwanatherians from several southern landmasses. The existing lineages of eutherians, metatherians, and multituberculates greatly increased both in generic diversity and in numerical abundance throughout the Cretaceous and eventually survived (albeit with varied success) the mass extinction at the end of the period. Dryolestoids survived into the Cenozoic in South America. However, various mammaliaform clades, spalacotheriids, eutriconodontans, and gondwanatherians had declined and were extinct by the end of Cretaceous.

Rise of Modern Monotremes, Marsupials, and Placentals

Mammals now thrive throughout the world, with about 4600 species. Only the three extant species of egg-laying monotremes survive today, and all are endemic to Australia. Although extant adult monotremes lack teeth, their extinct relatives from the Miocene of Australia have teeth, and their dental fossils have helped to trace the history of monotremes to three earliest known toothed monotremes in Australia, from 110 million years ago. One toothed monotreme was also discovered from the Paleocene of Argentina, suggesting that monotremes had a wider distribution on Gondwana landmasses during the Mesozoic.

The earliest monotremes shared some derived premolar and molar features with the southern tribosphenic mammals or australosphenidans that existed on the Gondwana landmasses from the Middle Jurassic to Early Cretaceous. The earliest monotremes and other australosphenidans also bear resemblance in their primitive mandibular features. It has been suggested that monotremes are a surviving lineage of the more diverse and more widely spread australosphenidans endemic to the Gondwanan landmasses, when the Gondwanan continents were more widely separated from Laurasian continents during the Middle Jurassic through the Early Cretaceous. This hypothesis of monotreme evolution is contingent on morphological evidence in support of a close relationship between monotremes and the southern tribosphenic mammals. Although corroborated by a recent discovery of an australosphenidan mammal from the Middle Jurassic of Argentina, the putative close relationship of monotremes to other australosphenidan mammals has been questioned. According to the hypothesis, the australosphenidan lineage including modern monotremes had split from the rest of Mesozoic mammals before 165 million years ago (Figure 2: Middle Jurassic diversification).

Placentals and marsupials dominated the world's terrestrial biotas for 65 million years in the Cenozoic, culminating in a vast array of species that make up 99.9% of all mammals today. Modern placentals have a worldwide distribution, but modern marsupials are mostly restricted to the Australian region and South America, with only one species in North America. In the Mesozoic, their geographic distribution was much different. The uncontested Cretaceous relatives to modern placentals and marsupials are all from Laurasia, as are the stem boreosphenidans, the nearest relatives of the common ancestor of modern marsupials and placentals.

The earliest fossils that can be reliably attributed to the placental lineage (eutherians) and the marsupial lineage (metatherians) are from the Yixian Formation of north-eastern China, dated to 125 million years. The earliest known eutherian, *Eomaia scansoria*, and the earliest known metatherian, *Sinodelphys szalayi* (Figure 3), have provided evidence on the earliest skeletal evolution, during the divergence of the placental and marsupial lineages. By the Late Cretaceous, there are dozens of taxa that can be unambiguously

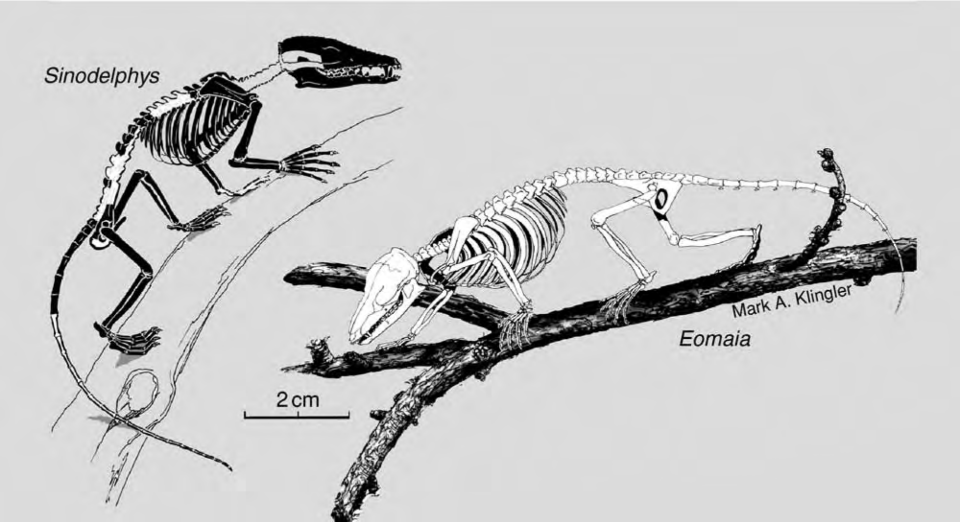


Figure 3 Restoration of the earliest known eutherian *Eomaia scansoria* and the earliest known metatherian *Sinodelphys szalayi*. *Eomaia* skeletal reconstruction is a composite from two specimens. *Sinodelphys* reconstruction is based on a partial skeleton, with preserved parts in black. Available skeletal evidence from these fossils shows adaptations that would allow climbing on uneven substrates and walking on branches. Illustration by M A Klingler, Carnegie Museum of Natural History, with permission.

assigned to metatherians and eutherians. Eutherians and metatherians can be distinguished by many characteristics of molars, anterior dentition, ear region, ankle bones, and wrist bones. The Late Cretaceous metatherians can be recognized by the presence of three premolars and four molars (seven positions of postcanines), with a single replacement at the last premolar, an inflected angular process of the mandible, and its posterior shelf, the masseteric muscle fossa. Other diagnostic features are from the wrist bones or carpals, such as enlargement of three carpal bones (the hamate, the triquetrum, and the scaphoid), enabling a more forceful hand grip. These marsupial-like characteristics are supplemented by features of ankle bones or tarsals. Metatherians can be recognized by features of the astragalar bone (or talus) and the heel bone (or calcaneus). The derived marsupial-like features in the ankle are correlated to a greater mobility between the tarsal bones and tibia and fibular at the upper ankle joint, and among the tarsals at the trans-tarsal joint. These features contributed to a greater range of the medial and lateral movement of the hindfoot. By contrast, the Early Cretaceous eutherians can be recognized by the presence of five premolars and three molars (eight positions of postcanines) and the presence of certain features related to the internal carotid artery and its branches in the petrosal bone of the ear region. The ankle bones of the Cretaceous eutherians (with a narrow head and pulley-like upper joint in the astragalar bone) are different from those of all non-eutherian mammals.

The newly discovered *Eomaia* and *Sinodelphys* help to establish the ancestral anatomical conditions from which modern marsupials and placentals could have evolved, and provide evidence for the sequence of evolutionary acquisition of the marsupial-like features in the metatherian lineage and the acquisition of placental-like characters in the eutherian lineage. Current evidence suggests that the foremost phylogenetic distinctions between metatherians (including marsupials) and eutherians (including placentals) reside in wrist and ankle anatomy, followed by the diagnostic characteristics of marsupial and placental dentitions, such as the reduced dental replacement related to specialized marsupial life history pattern, plus many characteristics in placental and marsupial molar crown morphology. Cretaceous metatherians and eutherians also show remarkably diverse locomotory adaptations in the features of hands and hindfeet. The phalangeal proportion of the finger bones and the shape of the claws suggest that both *Eomaia* and *Sinodelphys* were capable of branch walking and climbing on uneven substrates (Figure 3). The skeletons of many Late Cretaceous eutherians clearly show that these were terrestrial mammals. Locomotory skeletal structures of the Cretaceous eutherians and metatherians indicate that their locomotory adaptations were diverse, and this may have facilitated diversification of early eutherians and metatherians into different ecological niches.

See Also

Fossil Vertebrates: Palaeozoic Non-Amniote Tetrapods; Placental Mammals. **Mesozoic:** End Cretaceous Extinctions.

Further Reading

Allin EF and Hopson JA (1992) Evolution of the auditory system in Synapsida ("mammal like reptiles" and primi tive mammals) as seen in the fossil record. In: Webster DB, Fay RR, and Popper AN (eds.) *The Evolutionary Biology of Hearing,* pp. 587 614. New York: Springer Verlag.

Cifelli RL (2001) Early mammalian radiations. *Journal of Paleontology* 75: 1214 1226.

Hopson JA (1994) Synapsid evolution and the radiation of non eutherian mammals. In: Spencer RS (ed.) *Major Features of Vertebrate Evolution,* pp. 190 219. Knox ville, TN: The Paleontological Society.

Kemp TS (1982) *Mammal like Reptiles and the Origin of Mammals.* London: Academic Press.

Kielan Jaworowska Z, Cifelli RL, and Luo Z X (2004) *Mammals from the Age of Dinosaurs: Origins, Evolution and Structure.* New York: Columbia University Press.

Luo ZX, Cifelli RL, and Kielan Jaworowska Z (2001) Dual origin of tribosphenic mammals. *Nature* 409: 53 57.

Luo ZX, Crompton AW, and Sun AL (2001) A new mammal from the Early Jurassic and evolution of mammalian characteristics. *Science* 292: 1535 1540.

Luo ZX, Kielan Jaworowska Z, and Cifelli RL (2002) In quest for a phylogeny of Mesozoic mammals. *Acta Palaeontologica Polonica* 47: 1 78.

Luo ZX, Ji Q, Wible JR, and Yuan CX (2003) An Early Cretaceous tribosphenic mammal and metatherian evolution. *Science* 302: 1934 1940.

Novacek MJ (1992) Mammalian phylogeny: shaking the tree. *Nature* 356: 121 125.

McKenna MC and Bell SK (1997) *Classification of Mammals Above the Species Level.* New York: Columbia University Press.

Rowe TB (1993) Phylogenetic systematics and the early history of mammals. In: Szalay FS, Novacek MJ, and McKenna MC (eds.) *Mammal Phylogeny: Mesozoic Differ entiation, Multituberculates, Monotremes, Early Therians, and Marsupials,* pp. 129 145. New York: Springer Verlag.

Szalay FS (1994) *Evolutionary History of the Marsupials and an Analysis of Osteological Characters.* Cambridge: Cambridge University Press.

Wible JR, Novacek MJ, and Rougier GW (2004) New data on the skull and dentition in the Mongolia Late Cret aceous eutherian mammal *Zalambdalestes. Bulletin of the American Museum of Natural History* 281: 1 144.

Placental Mammals

D R Prothero, Occidental College, Los Angeles, CA, USA

Introduction

The Eutheria, or placental mammals, are the dominant group of vertebrates on the Earth today, and they have ruled the planet ever since the extinction of the dinosaurs at approximately 65 Ma. At the latest count, there are over 4400 living species of mammal, classified into over 1000 genera, 140 families, and 18 orders. However, the number of extinct mammals is at least five times as great. Most living mammals are terrestrial, including large beasts, such as elephants, rhinoceroses, hippopotamuses, and giraffes, as well as a great diversity of smaller land animals. The largest known land mammal is the extinct 20 tonne hornless rhinoceros *Paraceratherium*. Many groups of mammals evolved from land-dwelling ancestors to life in the water. These include manatees and dugongs (which are distantly related to elephants), otters (which are related to weasels), seals, sea lions, and walruses (which are distantly related to bears), and whales (which are distantly related to even-toed hoofed mammals), as well as numerous extinct groups. The living blue whale (at 30 m in length and 150 tonnes) is by far the largest animal that has ever lived, surpassing even the largest dinosaurs. Mammals have also taken to the air, with almost 1000 living species of bat, as well as numerous gliding forms, such as the 'flying squirrels' and 'flying lemurs' or colugos. Mammals are even more successful when their body sizes are small, with hundreds of small species of rodent, rabbit, and insectivore. The smallest living mammal, the 1.5 g Kitti's hog-nosed bat, is at the lower limit of possible body size for mammals, since the physiology and anatomy of mammals prevent them from thriving in the tiny-body-size niche inhabited by insects and other arthropods.

Anatomy and Physiology

Mammals are distinguished from all other animals by the possession of a number of unique characteristics. These include a body covered in hair or fur (secondarily reduced in some mammals, particularly aquatic forms); mammary glands in the females for nursing their young; a jaw composed of a single bone, the dentary; and three middle-ear bones, the incus, malleus, and stapes. All mammals maintain a constant body temperature through the production of metabolic heat. They have a four-chambered heart (two ventricles and two auricles), which keeps the circulation of the lungs separate from that of the rest of the body, resulting in more efficient oxygen transport to the body tissues. They have many other adaptations for their active lifestyle, including specialized teeth (incisors, canines, molars, and premolars) for biting, tearing, and grinding up their food for more efficient digestion. These teeth are replaced only once in the lifetime of the animal (rather than continuously as in other toothed vertebrates). Mammals have a unique set of jaw muscles, which allow the jaw to move in many directions for chewing and for stronger bite force. Their secondary palate encloses the internal nasal passage, and allows them to breathe while they have food in their mouth. Ribs (found only in the thoracic region) are firmly attached to the breastbone, so that expansion of the lung cavity is accomplished by movement of a muscular wall in the abdominal cavity called the diaphragm.

All living mammals have relatively large brains for their body size. Most mammals have excellent senses, and some have extraordinary senses of sight, smell, and hearing. To accommodate their larger brains and more sophisticated development, most mammals are born alive (rather than from eggs) and may require considerable parental care before they are ready to fend for themselves. Juvenile mammals have separate bony caps (epiphyses) on the long bones, which are separated from the shaft of the bone by a layer of cartilage. This allows the long bones to grow rapidly while still having a strong, bony articulation at the end. When a mammal reaches maturity, these epiphyses fuse to the shaft, and the mammal stops growing (in contrast to other vertebrates, which grow continuously throughout their lives).

Reproduction and Classification

Living mammals are divided into three major groups: the monotremes (platypus and echidna), which still lay eggs, retain a number of reptilian bones in their skeletons, and have other primitive features in their anatomy and physiology; the marsupials (opossums, kangaroos, koalas, wombats, and their relatives), which give birth to an immature embryo that must crawl into its mother's pouch (marsupium), where it finishes its development; and the placentals (the rest of the living mammals), which carry their young through a long gestation until they give birth to relatively well-developed progeny. This last group derives its name from the placenta, the membrane that surrounds the developing embryos within the uterus of the female. The placenta is formed by the fusion of the two other embryonic membranes, the chorion, which surrounds the embryo, and the allantois, which holds the embryonic wastes in egg-laying vertebrates. The placental membrane attaches the embryo to the wall of the uterus, and exchanges nutrients, gases, and waste products with the circulatory system of the mother, so that no hard-shelled egg (or its associated features, such as the allantoic waste sac or the large yolk sac) is needed. The placenta allows the embryos to undergo more development within the uterus, and to be born more mature than in other groups of mammals. By contrast, in marsupials the lack of a placenta protecting the embryo means that, when the immune system develops, the embryo must be born prematurely and move to the pouch, otherwise the mother's immune system would reject the embryo as a foreign object within the uterus.

In addition to these three living groups, there were many other major groups, such as the squirrel-like multituberculates, which are now extinct. The most recent classification of the mammals can be summarized as follows:

Class Mammalia
  Subclass Prototheria (monotremes)
  Subclass Theriiformes
    Infraclass Holotheria
      Cohort Marsupialia (marsupials or pouched mammals)
      Cohort Placentalia (placentals)
        Magnorder Xenarthra (sloths, anteaters, armadillos)
        Magnorder Epitheria
          Grandorder Anagalida (= Glires) (rodents, rabbits, elephant shrews)
          Grandorder Ferae (carnivores, pangolins, and many extinct groups)
          Grandorder Lipotyphla (hedgehogs, shrews, moles, tenrecs, and kin)
          Grandorder Archonta
            Order Chiroptera (bats)
            Order Primates (lemurs, monkeys, apes, humans)
            Order Scandentia (tree shrews)
          Grandorder Ungulata (hoofed mammals)
            Order Tubulidentata (aardvarks)
            Order Artiodactyla (even-toed hoofed mammals: pigs, hippopotamuses, camels, deer, antelopes, cattle, giraffes, pronghorns, and their relatives)
            Order Cete (whales and their extinct land relatives)
            Order Perissodactyla (odd-toed hoofed mammals: horses, rhinoceroses, tapirs, and extinct relatives)
            Order Hyracoidea (hyraxes)
            Order Tethytheria (elephants, manatees, and extinct relatives)

This classification does not list all the extinct groups, which include at least a dozen more ordinal-level taxa. It is a considerable improvement on previous mammalian classifications, which listed over 30 different orders with no indication of how they were related to one another. Although there are more ranks in this classification than is traditional, this nesting of groups within groups accurately reflects their evolutionary branching sequence (Figure 1).

In the past few years, molecular studies have suggested an alternative arrangement of placental relationships, which conflicts strongly with that indicated by the morphology and the fossil record. This classification groups the orders of placental mammals as follows:

Subclass Placentalia
  Superorder Xenarthra (edentates)
  Superorder Afrotheria (African insectivores, such as the golden moles and tenrecs, plus elephant shrews, aardvarks, and the tethytheres, including elephants, sirenians, hyraxes)
  Superorder Euarchontoglires
    Infraorder Archonta (primates, colugos, tree shrews)
    Infraorder Glires (rodents, rabbits)
  Superorder Laurasiatheria
    Infraorder Cetartiodactyla (whales plus artiodactyls)
    Infraorder Ferungulata (perissodactyls, carnivores plus pangolins)
    Infraorder Chiroptera (bats)
    Infraorder Eulipotyphla (shrews, moles, hedgehogs)

Advocates of this molecular classification argue that it makes biogeographical sense for the African mammal groups (the 'Afrotheria') to be closely related, since Africa was an island continent isolated from the rest of the world during the Early Cenozoic. Likewise, the only placentals from the island continent of South America, the edentates, are in a separate group. The rest of the placental mammals, which have lived on the northern ('Laurasian') continents through most of their history, also cluster together. Critics of this classification point to a huge number

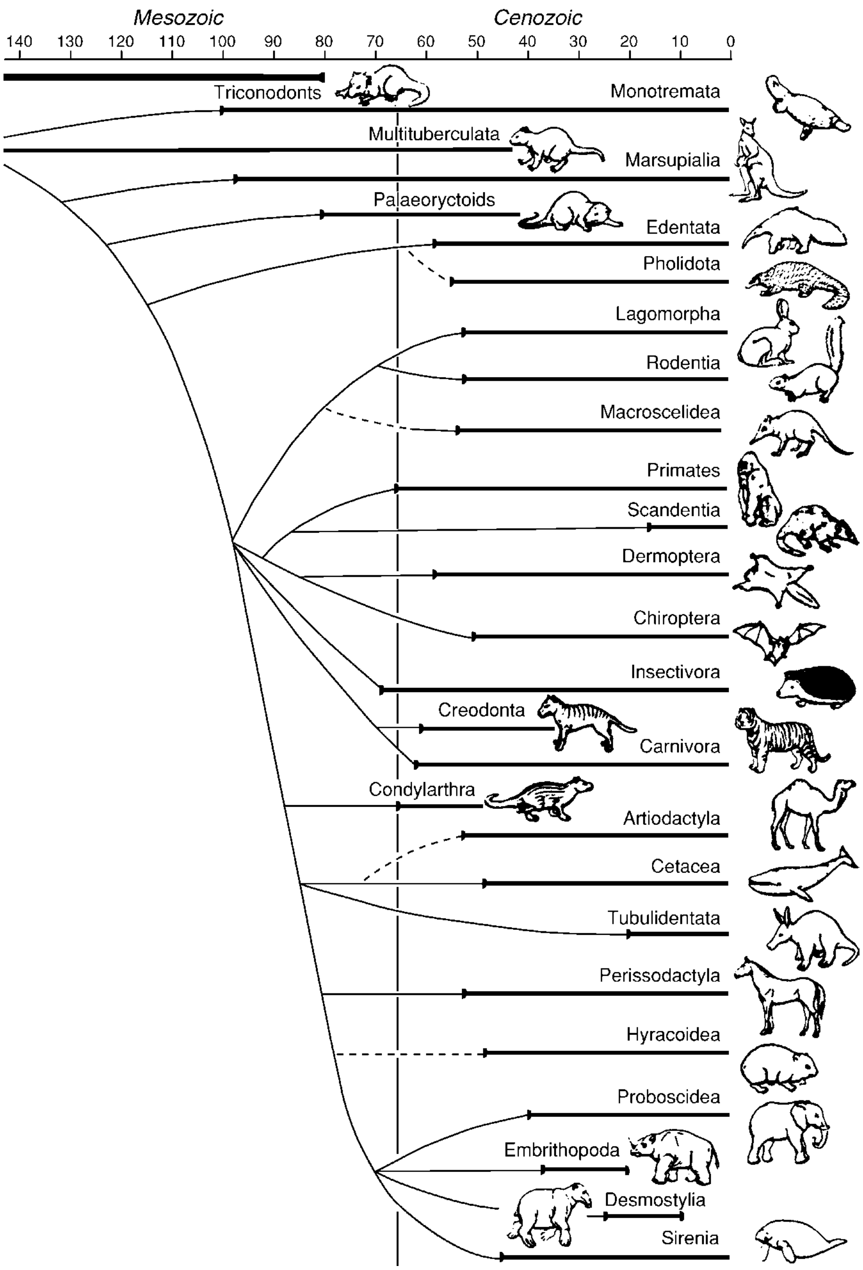


Figure 1 Evolutionary relationships and history of the major groups of mammals (modified from Novacek MJ (1994) The radiation of placental mammals. In: Prothero DR and Schoch RM (eds.) *Major Features of Vertebrate Evolution*, pp. 220 238. Paleontological Society Short Courses in Paleontology, no. 7. Paleontological Society, Lawrence, Allen Press).

of conflicts with well-established morphological evidence. For example, it breaks up the hoofed mammals by placing elephants and hyraxes in the Afrotheria, whereas the morphological evidence and the fossil record strongly suggest that they are related to the rest of the hoofed mammals (perissodactyls and artiodactyls plus whales). Likewise, good morphological evidence links at least some of the African insectivores with the true lipotyphlan insectivores, so most palaeomammalogists view the new molecular phylogenies with scepticism. They have good reason to do so, for molecular biologists have been claiming that they have 'the answer' to placental phylogeny in each new molecular system they examine, but, over the years, they have produced wildly varying results, so their classifications are not stable or reproducible enough yet to justify reordering the entire taxonomy of placental mammals.

Evolution

Mammals evolved from the Synapsida, an early branch of the terrestrial amniotes that has been erroneously called the 'mammal-like reptiles'. This name is inappropriate, because synapsids were never reptiles. Synapsids and reptiles originated independently about 320 Ma, and have evolved separately ever since. Early synapsids, such as the finback *Dimetrodon*, show relatively few mammalian characteristics, but, as their evolution progressed through the Late Palaeozoic, synapsids became progressively more and more mammal-like.

The first undoubted mammals appeared in the Late Triassic (about 210 Ma), and were tiny insectivorous forms much like living shrews. A number of different groups evolved over the next 145 Ma of the Jurassic and Cretaceous. Most remained tiny shrew-like animals, hiding from the dinosaurs in the underbrush and coming out mostly at night. The first two-thirds of mammalian history had passed before the dinosaurs became extinct 65 Ma, allowing mammals to emerge from their shadow.

Opossum-like marsupials are known from the Early Cretaceous (about 110 Ma), and they were more common than placentals just before the end of the age of dinosaurs. The first probable placental is the beautifully preserved skeleton (complete with hair impressions) from the Early Cretaceous of China known as *Eomaia scansoria*, which is about 125 Ma old. Although it is very primitive in most features, it had teeth for an insectivorous diet and limb proportions that suggest it was a tree dweller. By the Late Cretaceous (75 Ma), very primitive hoofed mammals (zhelestids) are known from Uzbekistan, and primitive members of the insectivore, carnivore, and primate lineages are known from the latest Cretaceous in both North America and Asia (65–70 Ma).

Although these early placentals are extremely primitive and hard to distinguish from early marsupials based on the limited evidence of their teeth and jaws, they can still be recognized by the trained palaeontologist. Early placentals had three upper and lower molars on each side of the jaw, and no more than five premolars (which reduces to four in later placentals). By contrast, early marsupials had four molars and usually only three premolars. In placentals, the last premolar has a tendency to be shaped like a molar (even though, as a premolar, it replaces a deciduous tooth, while a molar is never replaced). Primitive marsupial upper molars have a large expanded region on the outer edge of the tooth crown, whereas the upper molars of early placental mammals have little area outside the main cusps and instead are wider on the inside edge of the tooth. In the lower jaw of marsupials, the projection below the hinge (the angular process) flexes towards the midline, but in placentals it points straight back.

Between 65 and 55 Ma, a rapid adaptive radiation (Figure 1) yielded all the living orders of placental mammals and many extinct forms as well. The Xenarthra, or edentates, was the first group of placentals to branch off. Although the name 'edentate' means toothless, this is true only of anteaters; sloths and armadillos have simple peg-like teeth that lack enamel. The archaic nature of edentates among the placentals is shown by a variety of characters, including a uterus simplex (divided by a septum and lacking a cervix), a slower less well-regulated metabolism, retention of several reptilian bones lost in all other placentals, and a primitive rod-like stapes in the middle ear. Edentates evolved in isolation in South America throughout most of their history, developing into a variety of sloths (both tree sloths and huge ground sloths weighing up to 3 tonnes), anteaters, and armadillos (including the giant glyptodonts, which were 2 m long and had 400 kg of body armour and a spiked club on the tip of the tail). The long period of isolation in South America ended about 3 Ma ago, and ground sloths and glyptodonts migrated to Central America and parts of North America, disappearing at the end of the last Ice Age.

The remaining (non-edentate) placentals, or epitheres, diversified primarily in Eurasia and North America, and spread throughout the world in the Early Cenozoic. The true lipotyphlan insectivores (represented by shrews, moles, and hedgehogs) continued to diversify throughout the Cenozoic all over the northern continents. Most remained small in body size and ate insects, worms, and other small animals, although the extinct giant hedgehog *Dinogalerix* was the size of a large dog and killed sizable prey. The smaller lipotyphlan insectivores, such as the shrews, are among the smallest living placentals. Their body size is so tiny that they are constantly losing body heat because of their large surface area relative to their tiny mass. Consequently, they must eat almost continuously, or they will starve to death in a matter of days. This makes them extremely active and voracious predators, attacking not only worms and arthropods, but also animals much larger than themselves. They are fearless and will fight much larger animals when cornered.

The archontan radiation began with an enormous expansion of primitive lemur-like primates in the Early Cenozoic when the world had dense jungle vegetation all the way to the poles. Primates declined in the Oligocene, when their forest habitats disappeared, and became restricted to Africa (Old World monkeys and apes) and South America (New World monkeys). From the Old World monkeys evolved the great apes. About 15 Ma, apes were more diverse than monkeys in the Old World. Today only four groups of living great apes are known: the gibbons, the orang-utans, the gorillas, and the chimpanzees. They diverged from our own family, the Hominidae, about 7 Ma. This timing was originally established by molecular evidence, but recently a hominid fossil of that age, *Sahelanthropus tchadensis*, has been described from rocks 6–7 Ma in Chad in western Africa. In the past 5 million years, there have been dozens of species of hominid, divided into many genera: *Orrorin, Ardipithecus, Paranthropus, Australopithecus*, and our own genus, *Homo*, which appeared about 2.5 Ma. Fossils that are recognizably members of our own species, *Homo sapiens*, are known from South Africa in rocks about 100 000 years old.

The earliest bats (known from about 50 Ma) already had fully developed wings. They are the second most diverse group of mammals after the rodents, with almost 1000 living species and many more fossil species. They are divided into two main groups, the insectivorous Microchiroptera (which use echolocation to find their prey on the wing) and the fruit bats, or Megachiroptera (which fly during the day, eating fruit in the trees of the tropics, and do not echolocate).

The radiation of the Glires began in the Paleocene of Asia, where numerous primitive relatives of rodents and rabbits are found. In the Eocene both groups migrated to Europe and North America, where they soon took over the niche of small-bodied fruit, seed, and nut eaters that had been occupied by multituberculates and primitive primates. Rodents and rabbits are both characterized by chisel-like ever-growing front incisors that are used in gnawing. These incisors have a band of enamel only on the front edge; the rest of the tooth is made of softer dentin. The rodent keeps the teeth sharp by continuous gnawing, so that the more resistant enamel edge is worn and kept sharper than the dentin behind it. If the incisors are not worn down, but are misaligned by malocclusion, they will grow around in a curve until they pierce the skull.

Although rodents and rabbits are closely related, they form two different orders within the Glires and can be easily distinguished. Rodents have only a single pair of incisors, while rabbits have two. The enormous diversification of the Rodentia since the Eocene has given rise to over 1700 species (about 40% of the Mammalia), with forms ranging in size from the pig-sized capybara down to the many tiny mice and voles.

Predatory mammals (the Ferae) include the extinct creodonts (an archaic group that were the dominant predators and scavengers of the Early Cenozoic) and the living order Carnivora (cats, hyenas, mongooses, civets, dogs, bears, weasels and their kin, seals and sea lions, raccoons, and many extinct groups). All Carnivora are distinguished by their distinctive shearing teeth, the carnassials, developed between the last upper premolar and the first lower molar. True Carnivora began as weasel-like forms in the Early Eocene, but by the Oligocene they had taken over most of the predatory niches from the creodonts. By the Miocene, the ancestors of seals and sea lions had evolved from bear-like ancestors. Carnivorous mammals show remarkable convergence on a limited number of body forms. For example, sabertoothed forms evolved four times, once in the creodonts, once in the true cats, once in the extinct cat-like nimravids, which are related to dogs, and one extinct sabertoothed marsupial that has been found in South America. In North America, the borophagine dogs converged on hyenas, with similar bone-crushing teeth.

The hoofed mammals, or ungulates, are first known from about 85 Ma in central Asia. In the latest Cretaceous and the Paleocene, archaic hoofed mammals (‘condylarths’) were among the most common forms in North America and Asia. From these roots, numerous orders evolved. The first to branch off were the even-toed Artiodactyla, which have two or four toes on each foot and a distinctive ankle structure. First appearing in Pakistan in the earliest Eocene, artiodactyls quickly diversified into a number of different groups. Today there are over 190 living species of artiodactyl, and at least twice that number of fossil species are known. Artiodactyls include the suoids (pigs, peccaries, hippopotamuses), tylopods (camels and their extinct relatives), and ruminants (deer, giraffes, pronghorns, cattle, sheep, goats, antelopes). With their four-chambered stomachs for more efficient digestion, the ruminants became the dominant group of large herbivorous mammals as global climates became drier in the later Cenozoic and grasslands expanded.

The ancestors of whales were large hoofed predators known as mesonychids. Recently, transitional forms between mesonychids and primitive whales have been found in the Eocene of Pakistan. However, some of these fossils also show the ‘double pulley’ ankle bones found in all artiodactyls, which lends weight to the suggestion by molecular biologists that whales are descended directly from artiodactyls, rather than from mesonychids. By the Oligocene, whales had diversified into the predatory toothed whales (dolphins, orcas, sperm whales) and the filter-feeding baleen whales (blue, right, humpback, grey and many other whales).

The Perissodactyla, or odd-toed ungulates, have one or three toes on each foot. Today they include horses, rhinoceroses, and tapirs, but they were much more diverse in the past, with huge two-horned brontotheres and bizarre clawed chalicotheres. Closely related to perissodactyls are the hyraxes or conies and the tethytheres (elephants, sea cows, and their relatives). Tethytheres are so named because they originated from Late Paleocene ancestors that once lived along the Tethys seaway (which stretched from Gibraltar to Indonesia). Although sea cows had spread around the world by the Eocene, their earliest evolution was a mystery. Then a nearly complete specimen of *Pezosiren* was described from the Eocene of Jamaica, which showed what their earliest evolutionary transitions looked like. This fossil has a fully sirenian skull and teeth, and thick ribs for ballast, but retains fully functional legs and feet, rather than flippers.

The remaining tethytheres include the elephants, the huge two-horned extinct arsinoitheres, and the hippopotamus-like desmostylians from the Miocene of the Pacific Rim. The earliest relative of the elephant is *Paschatherium*, known from teeth from the Paleocene of Morocco. By the Eocene, the elephant family was represented by *Moeritherium*, which looked like a small pig or hippopotamus, although it had the beginning of a trunk or proboscis. The evolution of most of these groups was restricted to Africa until the Middle Miocene, when both groups (proboscideans and arsinoitheres) spread to Eurasia, and mastodonts even reached North America.

See Also

Evolution. **Fossil Vertebrates:** Mesozoic Mammals; Hominids. **Tertiary To Present:** Paleocene; Eocene; Miocene.

Further Reading

Benton MJ (2000) *Vertebrate Palaeontology,* 2nd edn. Oxford: Blackwell.

Corbet GB and Hill JE (1991) *A World List of Mammalian Species,* 3rd edn. Oxford: Oxford University Press.

Eisenberg JF (1981) *The Mammalian Radiations.* Chicago: University of Chicago Press.

Ji Q, Luo Z X, Yuan C X, *et al.* (2002) The earliest known eutherian mammal. *Nature* 416: 816 822.

Macdonald D. (1984) *The Encyclopedia of Mammals.* New York: Facts on File Publications.

McKenna MC and Bell SK (1997) *Classification of Mammals Above the Species Level.* New York: Columbia University Press.

Murphy WJ, Eizirik E, O'Brien S, *et al.* (2001) Resolution of early placental mammal radiation using Bayesian phylogenetics. *Science* 294: 2348 2351.

Novacek MJ (1992) Mammalian phylogeny: shaking the tree. *Nature* 356: 121 125.

Novacek MJ (1994) The radiation of placental mammals. In: Prothero DR and Schoch RM (eds.) *Major Features of Vertebrate Evolution,* pp. 220 238. Paleontological Society Short Courses in Paleontology, no. 7. Paleonto logical Society. Lawrence, Allen Press.

Nowak RM (1991) *Walker's Mammals of the World.* Balti more: Johns Hopkins University Press.

Pough FH, Janis CM, and Heiser JB (1999) *Vertebrate Life.* Upper Saddle River: Prentice Hall.

Prothero DR (1994) Mammalian evolution. In: Prothero DR and Schoch RM (eds.) *Major Features of Vertebrate Evolution,* pp. 238 270. Paleontological Society Short Courses in Paleontology, no. 7. Paleonto logical Society. Lawrence, Allen Press.

Prothero DR and Schoch RM (2002) *Horns, Hooves, and Flippers: The Evolution of Hoofed Mammals.* Baltimore: Johns Hopkins University Press.

Savage RJG and Long MR (1986) *Mammal Evolution: An Illustrated Guide.* New York: Facts on File Publications.

Szalay FS, Novacek MJ, and McKenna MC (eds.) (1993) *Mammal Phylogeny.* New York: Springer Verlag.

Vaughn TA, Ryan JM, and Czaplewski NJ (2000) *Mammalogy,* 4th edn. Fort Worth: Saunders College Publishing.

Wilson DE and Reeder DM (1993) *Mammal Species of the World.* Washington, DC: Smithsonian Institution Press.

Young JZ (1981) *The Life of Mammals,* 3rd edn. Oxford: Clarendon Press.

Hominids

L R M Cocks, The Natural History Museum, London, UK

Introduction

Humans and their history and activities are dealt with in a large number of enormous reference tomes, and thus the editors of this Encyclopedia have decided to refer to them only very briefly in this work. The studies of humans as organisms can be grouped into two, anthropology and medicine. The latter, dealing only with our ailments, receives no further mention here, but anthropology is divided into Physical Anthropology, which considers us as animals, and Social Anthropology, which examines how we interreact both with other humans and with our environments (a discipline which would be termed Ecology or Palaeoecology (*see* **Palaeoecology**) for other organisms elsewhere in this Encyclopedia).

Humans are placental mammals and are within the Primates, which also include such creatures as lemurs and lorises. Higher primates include monkeys and apes, and in particular gorillas, chimpanzees, and hominids (the latter systematically classified as the Family Hominidae). From genetic data, gorillas are thought to have diverged from the rest in Early Miocene times (at ~10 Ma), and hominids from chimpanzees at about 6.5 Ma. In the nineteenth century, much was made of the 'missing link' which apparently separated humans from their ancestors; however, that missing link has now vanished due to the large numbers of hominid specimens that have been found, particularly in East Africa since the Second World War. Because hominids are land animals, the preservation and fossilization of specimens was and is very chancy and they are best preserved in cave deposits and lake sediments. Although objective bone measurements can be and are made and used to biometrically distinguish between populations, the material is often sparse in numbers of specimens, and thus the resulting systematic classification somewhat subjective. The key features which physical anthropologists have found important are the braincase size and shape (including the development of brow ridges over the eye sockets), the teeth, the relative proportions of arms and legs, and the differing shapes of the ankle, knee, and elbow joints. Various DNA studies have been made, which have been differently interpreted in detail, but which in general support the story of hominid evolution and history outlined below.

Our own genus is termed *Homo*, and our species *sapiens*, and this brief article is divided into three: hominids other than *Homo*; *Homo* species other than *sapiens*; and *Homo sapiens*.

Hominids other than *Homo*

All hominids apart from *Homo* are known only from East and South Africa. A key feature of hominids is the development of bipedalism, which of course leaves the hands free for other activities, such as the gathering of food or the use of tools and other implements. Three genera with characters intermediate between chimpanzees and hominids, but known only from rather fragmentary fossils, are *Sahelanthropus*, found in Chad and dated to nearly seven million years ago, *Ardipithecus*, which lived in Ethiopia between about 6.0 and 4.4 million years ago, and *Orrorin* from Kenya, which lived at about 6 Ma. However, the earliest well-known hominid is their probable descendant *Australopithecus*, which lived between 4.2 and 2.4 million years ago (Figure 1). Possibly the best-known relatively complete specimen of *Australopithecus* is the one named 'Lucy', which was found at Hadar in Ethiopia and dated to 3.2 Ma. Modern chimpanzees have an average brain size of 390 c.c., Lucy's species a size of about 400 c.c. and modern humans about 1,300 c.c., and Lucy would have looked more like an ape than a human, walking upright, but with an ape-shaped body. Other hominid genera which have been named are *Paranthropus* (2.6 to 1.4 Ma) and *Kenyanthropus* (2.4 to 1.9 Ma).

Homo other than *sapiens*

The oldest representative of our genus is *Homo habilis* (and the similar and possibly synonymous *H. rudolfensis*), which lived in East and South Africa from 2.4 to 1.6 million years ago, with the best-known specimens, dated at about 2 Ma, being found by the Leakey family at Olduvai Gorge, Tanzania (Figure 2). *H. habilis* had a brain size of about 600 c.c. It made primitive stone tools from volcanic rock and chert, and was still somewhat chimpanzee-like, with long arms, short legs, and a thick waist. It was sometime around two million years ago that our ancestors made the important change from a largely vegetarian diet to one which also included meat. From then on they may be classified as hunter-gatherers, and that extension of activities into hunting might well have involved the combination into pack groups to make the killing of prey more effective.

Figure 1 Reconstruction by the late Maurice Wilson of *Australopithecus afarensis*, based on the 3.2 Ma specimens of 'Lucy' and other individuals from Ethiopia.

The invention of roasting food by fire appears to have been developed by about 1.6 Ma.

However, it was the successor of *habilis* which made much greater impact. This was *Homo erectus* (earlier versions of which are sometimes termed *H. ergaster*), which was relatively slim-waisted and, in some cases, just as tall as modern humans (Figure 3). It had a relatively larger cranium, but still had an extensive brow ridge. The brain case size had much increased; however, its average size at about 900 c.c. was only that of a one-year-old human child today. All these changes much improved mobility, and it was this species that expanded widely from the ancestral areas of East and South Africa for the first time. It has been postulated that these migrations would have been facilitated by climate changes related to the Plio-Pleistocene glaciation (*see* **Tertiary To Present:** Pleistocene and The Ice Age) near the Poles, which would have led to the expansion of more easily-traversed savannah grasslands and a reduction in the areas of tropical jungles. *H. erectus* existed from about 2.0 million to as recently as about 100 000 (100 Ka) years ago, and some of the best-known early material is from Koobi Fora and Lake Turkana, Kenya. By about 1.8 Ma the species (site names in brackets) had reached Georgia (Dmanisi) in Central Asia, soon afterwards Indonesia (Java) and subsequently China ('Peking' Man), and by 800 Ka it had spread over southern Europe to Italy (Ceprana) and Spain (Atapuerca). By 500 Ka years ago a descendant species, *H. heidelbergensis* (Figure 4), which lived from 800 to 300 Ka, had reached southern England (Boxgrove, Sussex).

A later species was *H. neanderthalensis*, commonly known as Neanderthal Man, which lived from around 400 000 years ago to as recently as about 30 000 years ago in Croatia (Figure 5). Neanderthals (named after their first-known nineteenth-century site in the Neander Valley in Germany), were relatively large-brained, big-faced but with a low skull and little chin, and most of their dietary protein was from animals. *Homo sapiens* lived in Europe together with Neanderthals between 40 and 30 thousand years ago: it is uncertain whether the latter became extinct through competition with humans or by a reduction of their habitats. It is also uncertain whether or not Neanderthals and modern humans could or did interbreed.

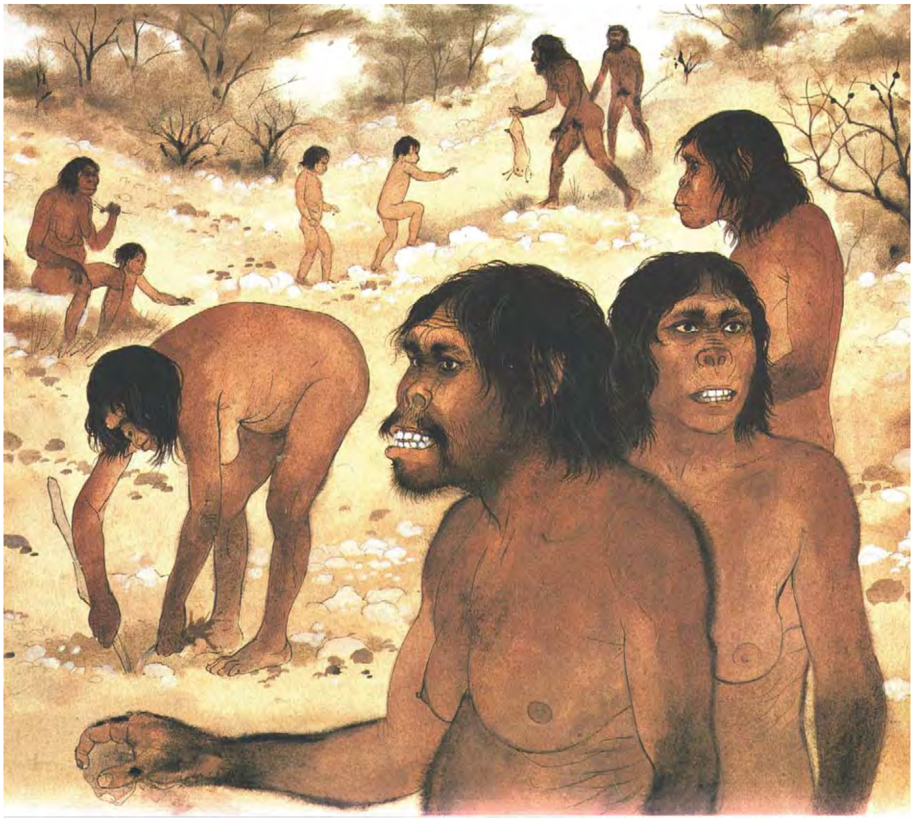


Figure 2 Reconstruction by the late Maurice Wilson of *Homo habilis*, based on 1.9 Ma individuals from Kenya and Tanzania.

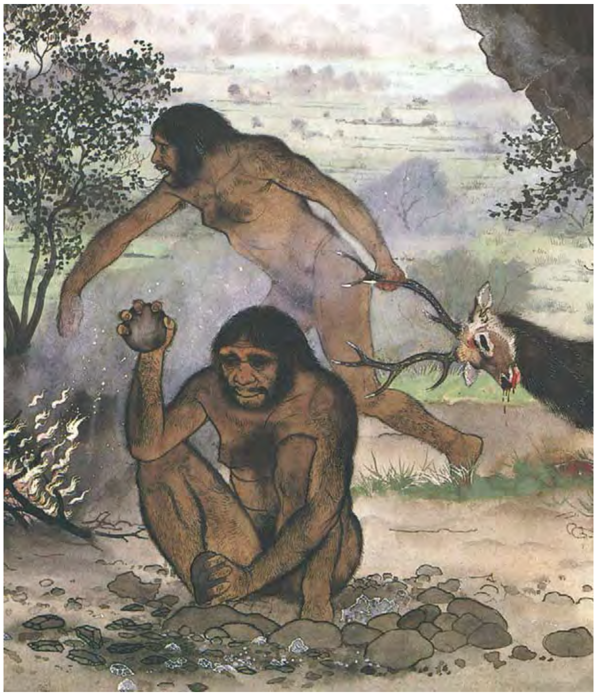


Figure 3 Reconstruction by the late Maurice Wilson of *Homo erectus* from 450 Ka individuals at Zhokoudian Cave, China.

Homo sapiens

Homo sapiens probably evolved from *H. heidelbergensis* at about 300 Ka, somewhere in eastern Africa. It is difficult to be sure whether the transition between *H. sapiens* and its ancestors was sudden or took place over a relatively long time period. They differ from their ancestors in having a much larger brain size and in the virtual absence of a brow ridge. They spread into the Middle East at about 100 Ka, and on into southern Europe relatively quickly after that. Migration to northern Asia appears to been have later (the earliest humans in China date from about 50 Ka). The oldest human remains in Australia also appear to date from about 50 Ka. Much later, humans colonised the Americas via the Siberian land bridge during a sea-level low stand before 13 Ka. Once in America they spread quickly, reaching Patagonia in southernmost South America at about 11 Ka. Differentiation between the various human racial types (African, Caucasian, etc.) occurred through separation of the breeding populations, but the process has not gone so far that we have ceased to be a single biological species.

A key change in the behaviour of humans was the transition, which was progressive over a long time

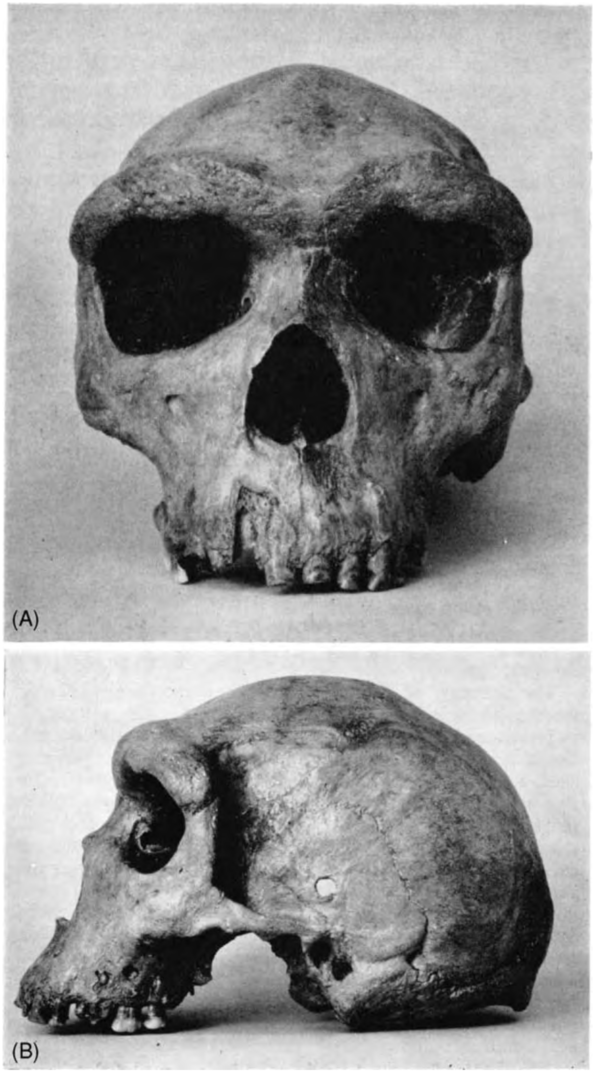


Figure 4 *Homo heidelbergensis*, a 300 Ka skull from Broken Hill, Zambia. Note the brow ridges.

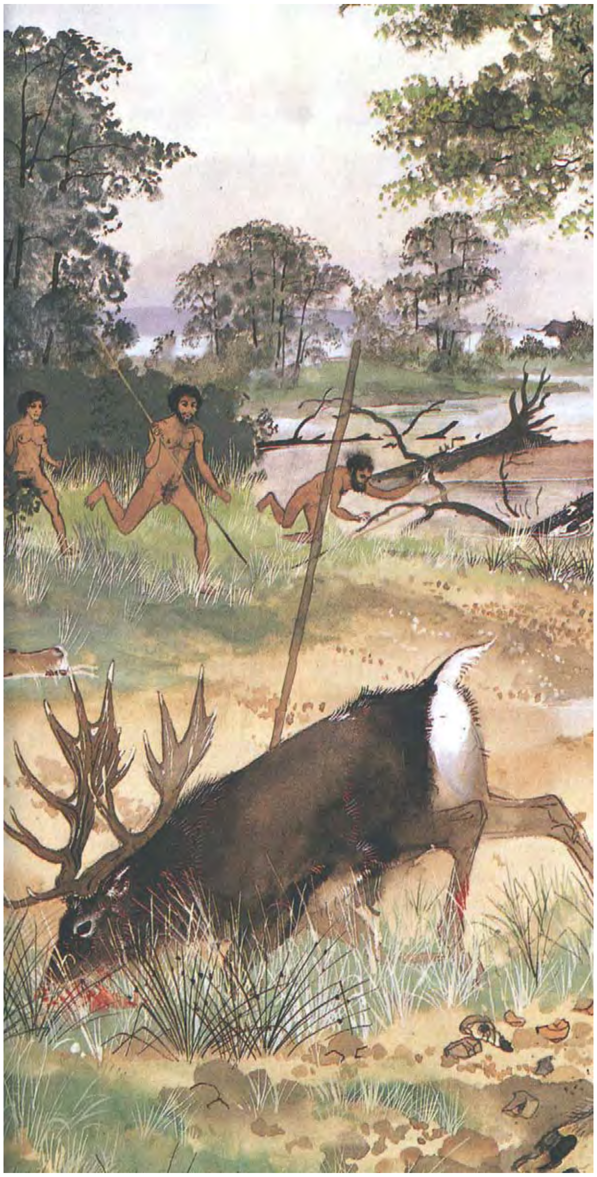


Figure 5 Reconstruction by the late Maurice Wilson of a hunting scene with *Homo neanderthalensis* at Swanscombe, Kent, England, at about 400 Ka.

period, from being hunter-gatherers to the planting of crops and the advent and development of agriculture. The latter implies a far more static existence, which led to more permanent dwellings and eventually into organized settlements, progressively villages, towns, and cities. The earliest organized human settlements yet known are those dating from about ten thousand years ago in Turkey and Syria. The more than minimal quantity of food produced by agriculture resulted in the introduction of both leisure time and also in the division of human activity into separate occupations; a process which led inexorably to the civilization and cultural divisions which we know today. It is not known when hominids developed speech and language, but the initiation of those facilities may have taken place as early as *H. habilis* times, with progressive development as the advantages of improved communications became self-evident.

Modern humans have much affected the geology on a global basis, not only in the extraction of many varied building stones, aggregates, minerals, hydrocarbons, and other resources, but also in the management of water, construction of buildings, and the varied activities which have led to the global warming which directly affects sedimentation and other geological processes both now and progressively more so in the future. Some anthropogenic details of these effects in historical times are outlined in the article on the Holocene of Europe (*see* **Europe:** Holocene).

See Also

Europe: Holocene. **Fossil Vertebrates:** Placental Mammals. **Palaeoecology**. **Tertiary To Present:** Pliocene; Pleistocene and The Ice Age.

Further Reading

Johanson D and Edgar B (1996) *From Lucy to language*. London: Weidenfeld and Nicholson.

Klein RG (1999) *The Human Career*. Chicago: University of Chicago Press.

Lewin R (2004) *Human Evolution: an Illustrated Guide*. Oxford: Blackwell Scientific Publications.

[no author] (2003) New look at Human evolution. *Scientific American* (special edition). June: 100.

Stringer CB (2002) Modern Human origins – progress and prospects. *Philosophical Transactions of the Royal Society, London B* 357: 563–579.

Stringer CB and Gamble C (1993) *In search of the Neanderthals*. London: Thames and Hudson.