

2

THE EVOLUTIONARY PROCESS AND THE FOSSIL RECORD



The Ordovician trilobite *Ogygiocarella*, $\times 1.35$. (Photograph courtesy of P.R. Sheldon.)

2.1 Molecular Palaeontology

G. B. CURRY

Introduction

In its widest sense, molecular palaeontology embraces the study of intact molecules in living organisms, as well as the investigation of the variably decayed remnants of ancient molecules which occur in great abundance in rocks and fossils (Runnegar 1986). In extant molecules such as DNA or proteins, the nucleotide or amino acid sequences are reproduced accurately from generation to generation with only minor changes caused by genetic drift or natural selection. Because such molecules are ubiquitous in all life forms, homologous molecules can be extracted from a wide taxonomic range of living organisms and their sequences compared, either directly or indirectly, to provide information on the systematic interrelationships of their host organisms.

Using such techniques, phylogenetic history can be investigated on a scale ranging from phyla to subspecies. The molecular clock hypothesis further extends this approach by attempting to use molecular 'distances' to date divergent events between taxa (Thorpe 1982). To do this the molecular clock must first be calibrated using sequences from organisms which have well dated divergence events — in effect organisms which have a good fossil record. The molecular clock hypothesis is hotly debated even by molecular biologists, and it is clear that the rates of sequence change are highly variable both in different molecules and in identical molecules within different species. Nevertheless, the interest in the molecular clock does mean that molecular biologists will increasingly be drawing conclusions and making predictions about topics which previously have been the exclusive preserve of palaeontologists.

Fossilization of organic molecules

Despite the palaeontological importance of recovering the prodigious 'historical' information stored in extant molecules, it is undoubtedly the investigation of molecular fossils which has the greatest potential for the development of molecular technology in palaeontology. The fact that resistant or-

ganic molecules can survive for many millions of years has been one of the most remarkable geological discoveries of recent years (see also Section 3.2). The presence of large quantities of organic debris in rocks has long been recognized, but previously it had been widely assumed that these compounds contained no palaeontological or biological information because of the extensive degradation they had experienced. It is now clear that such an assumption is wrong, and that certain robust molecules can survive virtually intact, or at least in recognizable form, for many millions of years. The key to such discoveries lies in the application of technological advances in subjects such as organic geochemistry and molecular biology which allow the recovery, purification, and characterization of organic materials with a precision never before attainable (Curry 1987).

The raw material for molecular palaeontology is the accumulation of the variably decayed remains of ancient animals, plants, and micro-organisms which has built up in rocks over many millions of years (Section 3.2). This organic debris occurs both interstitially in sedimentary rocks and as inclusions within fossil shells and skeletons. The vast scale of such accumulations is perhaps not generally appreciated — enormous as they are, the reservoirs of fossil fuels such as oil, gas, and coal represent only the tip of the iceberg. To a greater or lesser extent all sedimentary rocks contain less apparent, highly dispersed, and generally less degraded, organic debris. On average about 2% of the volume of sedimentary rocks is composed of organic compounds, and conservative estimates suggest that there is about 10 000 times more organic material in rocks than in the present-day global biomass. One group of ubiquitous molecular fossils, the bi-hopanoids, is present in quantities which equal or exceed the total mass of organic carbon in all living organisms.

Most of this reservoir is thought to be useless for molecular palaeontology, because it has been so intensely altered as to be totally unrecognizable. In addition to the complications imposed by hydrolyzation and other physical, chemical, and biological

processes of decay, investigations of molecular fossils have also to contend with the effects of a host of other geological phenomena such as diagenesis, vulcanism, and tectonism. As a result, fossil molecules can be exposed to an infinitely variable combination of heat, pressure, and percolating fluids, and can with time recombine into complex and intractable new structures by reacting with each other and with mobile components migrating from external sources. The possibility of contamination by molecules derived from extant organisms is an additional, and ever-present, complication.

In the face of all these problems, the extent to which molecules can survive fossilization is striking, most particularly in the well documented cases of molecules which are fully exposed to the ravages of geological processes. The commercial implications of accumulations of fossil molecules has provided a major stimulus for their study, and a wide range of molecular fossils have now been discovered in crude oils which have been intensely degraded. The great majority of fossil organic material has not been so intensely degraded, and hence should be particularly informative, although as yet such material has not been thoroughly investigated.

Analytical techniques

The major complications in investigating molecular palaeontology stem from the fact that raw samples contain extremely complex assemblages of molecular fossils, which occur in all stages of decay, from virtually unaltered to strongly degraded. Furthermore, most liberated fossil compounds readily combine into complex, heterogeneous agglutinations, which are often insoluble and hydrophobic and hence difficult to analyse. The first hurdle, therefore, is to separate the different constituents, and to recognize the various stages in the decay of a particular molecule, each of which will generally produce slightly different daughter products depending on its decay history. The development of chromatographic and electrophoretic techniques which can reliably separate and analyse samples of 1 µg or less provided a major stimulus to the study of organic debris within fossils, although the long-term survival of molecular fossils was first demonstrated using techniques from organic geochemistry.

Organic chemists now derive most of their information from computer-controlled combined gas chromatograph mass spectrometers (C-GC-MS) which first separate, and then analyse, the structure

of fossil molecules at a resolution measured in nanograms (10^{-9} gm). These procedures can identify geochemical fossils because the carbon skeleton of the original biological molecule is characteristically preserved either unaltered or with minor rearrangements, substitutions, or the removal of side chains (Fig. 1). A significant recent technological advance is the recognition that large, highly reactive molecules, which are usually particularly susceptible to degradation, can be stabilized over geological time when incorporated into inert polymeric agglutinations. These relatively intact and extremely informative fossil molecules can be released by the partial chemical dissolution of fossil polymers prior to C-GC-MS analysis, and have now been isolated from well preserved 50 Ma compounds. The main drawback to the organic geochemical approach is that the analytical techniques are complex and require specialist skills and expensive equipment.

The major preoccupation of molecular palaeontology has been to demonstrate unambiguous links between intact living molecules and their resistant fossil remains. A classic example is the petroporphyrins, which are common constituents of crude oils. As early as 1934 it was proposed that vanadyl petroporphyrin represented degraded chlorophyll *A* from plants; the fossil molecule has essentially the same structure, but has lost its side chain and the central magnesium ion has been substituted by a vanadyl ion (Fig. 1).

Subsequently the decay pathways of a wide range of compounds have been traced from living organisms through to sediments and rocks, providing information both on the origins of fossil fuels and on the conditions which prevailed during their alteration (because the state of decay is closely related to temperature, pressure, etc.). Molecular fossils can also provide information about ancient environmental conditions, e.g. when the chemical composition of a group of marine phytoplankton is known to vary directly with ocean temperature, and such variations can be detected in their fossil remains. Mapping out the distribution of such compounds in rocks therefore provides some indication of oceanic temperature variations in the past. Investigations of fossil molecules have also led to the discovery of previously unrecognized living molecules; biohopanoids, an important group of membrane-forming lipids in extant bacteria, were first recognized by their ubiquitous fossil derivatives.

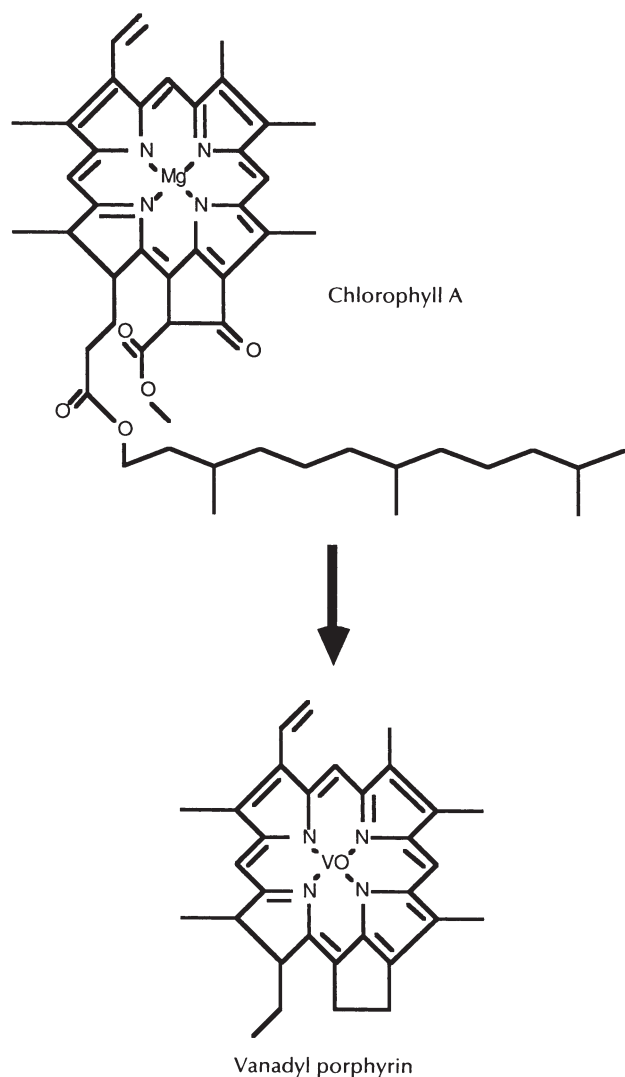


Fig. 1 A classic example of a geochemical fossil — vanadyl porphyrin, one of a suite of petroporphyrins (common constituents of crude oils), which is thought to be the breakdown product of the photosynthetic pigment chlorophyll *A* from plants. Over geological time the central magnesium ion has been replaced by a vanadyl ion, and the side chain has been stripped away. The side chains of chlorophyll are thought to give rise to other common geochemical fossils (phytanes and pristanes).

Organic compounds from biominerals

There is now considerable interest in the preservation of molecular fossils within the shells and skeletons of fossils (Wyckoff 1972; De Jong *et al.* 1974; Westbroek *et al.* 1979; Hare *et al.* 1980). Compared with free compounds in sediments and rocks, organic compounds from biominerals are relatively protected from decay and contamination — a kind of biological ‘fluid-inclusion’. It is widely believed that this important group of

molecular fossils represents the remains of original organic compounds which were incorporated into the skeletal fabric during shell growth as essential extra- or intracrystalline elements of ‘organic-matrix-mediated’ biomineralization. Although these entombed molecules do fragment over geological time, it is now clear that some of their breakdown products are often retained within skeletal fabrics (Figs 2, 3). The more porous mineralized matrices are obviously much more prone to contamination from percolating fluids, but the possibility of spurious results can be reduced by investigating the organic content of surrounding sediment, and by determining stable carbon isotope ratios which reveal the contamination of marine fossils by predominantly terrestrially-derived groundwater. Palaeontologists have long been interested in biomineralization processes, and an understanding of the intimate association between the organic matrix and the inorganic mineralized phase, which appears to be a major factor in the long-term stability of molecular fossils, is clearly crucial for the full exploitation of molecular palaeontology.

Organic compounds have now been isolated and characterized from a diverse taxonomic range of invertebrate and vertebrate fossils spanning the entire history of shelly faunas (i.e. 600 million years). The best results have been obtained with well preserved Mesozoic, Cenozoic, and Quaternary specimens (i.e. the past 250 million years). These data have mostly been presented in the form of amino acid mole percentages.

Detailed analyses of molecular remnants from living and fossil scallops spanning about 200 million years demonstrate the expected progressive decay in the total quantity, and percentage survival, of amino acids (Fig. 2). The rate of decay reaches a plateau at about the 1%–2% survival level during the Cretaceous (c. 100 million years), and thereafter the total quantities of molecular fossils remain relatively constant.

Free amino acids represent only one of three broad categories of molecular fossils recognized; insoluble organic residues and soluble peptides also display a similar decay profile (Fig. 3). Changes in the respective proportions of these components with time probably indicate some movement between categories. The consistency of amino acid profiles over geological time is remarkable, for example in nautiloids spanning almost 400 million years. Changes in relative abundance with time are generally thought to be a diagenetic effect, and such alteration phenomena are useful indicators of the

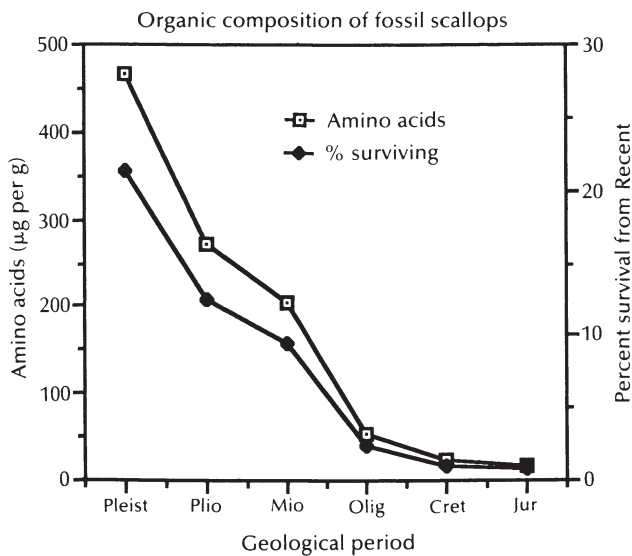


Fig. 2 The progressive decay of organic compounds from the shells of fossil scallops with time. Approximate ages, in millions of years, for the horizontal axes are Pleistocene (2), Pliocene (4), Miocene (15), Oligocene (30), Cretaceous (100), and Jurassic (175).

state of fossil preservation when coupled with isotopic data and laboratory experimentation on Recent shells.

Molecular fossils and systematics

Because of the ubiquitous distribution of amino acids in living tissues, the amino acid compositions

of fossils have as yet had only limited use as taxonomic indicators. However, consistent differences between the amino acid compositions of living and fossil brachiopods appear to distinguish between brachiopod orders, with the chitinophosphatic-shelled inarticulate *Lingula* being characterized by higher concentrations of alanine and lower glycine. While such variations are probably related to different shell mineralogy, amino acid profiles also distinguish between different orders of calcareous-shelled articulate brachiopods. It seems highly probable that the crude systematic application of such data will be restricted to the higher levels of taxonomic classification (i.e. superfamily, order).

A more recent innovation has been the use of immunological techniques to investigate molecular fossils (De Jong *et al.* 1974; Westbroek *et al.* 1979; Muyzer *et al.* 1984; Lowenstein 1986). Such an application makes use of the major attribute of the immune system, namely that antibodies recognize their target molecule by detecting a small diagnostic region or regions known as determinants. Antibodies prepared against living tissues or a particular fundamental molecule such as collagen, can therefore detect the presence of that molecule in fossils, provided that fragments containing the determinants have survived. As determinants are much smaller than intact molecules, their survival potential is much greater. Antigenic determinants are known to survive for at least 70 million years (De Jong *et al.* 1974; Lowenstein 1986). The advantages of immunology are the high specificity of antibodies

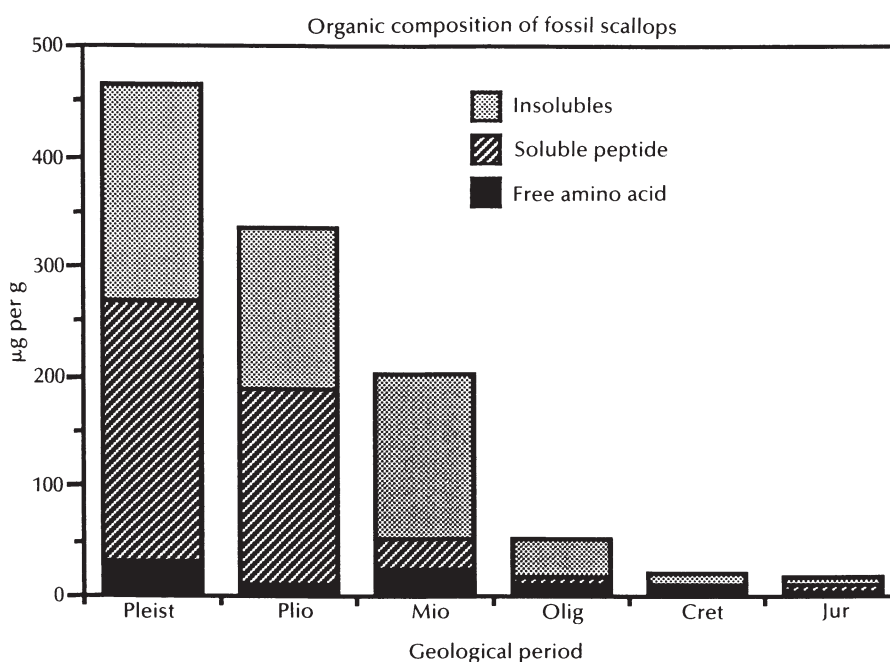


Fig. 3 Proportions of free amino acids, soluble peptide fragments, and insoluble organic residues in the shells of scallops of various ages. Horizontal axis as in Fig. 2.

(in particular monoclonal antibodies directed against a single molecule), and the ability to carry out large numbers of determinations once the antibody has been prepared, which allows rapid assessments of the extent of fossil organic preservation. The production of antibodies is, however, a complex and specialized field.

Immunological techniques potentially provide greater systematic resolution, at least to the generic and familial level, although species-specific reactions have been detected in relatively recent fossils (Lowenstein 1986). Such an application is in effect a variation of the widely used technique of reconstructing phylogeny from immunological distances, a procedure which can be justified on the grounds of the direct linear relationship between antigenicity and amino acid substitution rate. Experiments with living taxa have demonstrated the potential of immunology in this field, with antibodies against one bivalve species reacting with all but one of the other taxa in its family; the implication that the exceptional genus may be incorrectly assigned is apparently supported by other lines of evidence (Muyzer *et al.* 1984).

Full exploitation of the evolutionary and taxonomic applications of molecular palaeontology will require much more detailed information about the structure, composition, survival potential, and distribution of fossil molecules. Although there have been reports of the preservation of fragments of DNA in 2000 year old mummies and other relatively recent fossils, DNA molecules are relatively unstable, particularly susceptible to hydrolysis, and concentrated in vulnerable soft tissues rather than protective mineralized skeletons (Runnegar 1986). On present information, DNA is therefore likely to be very short-lived on a geological time-scale, although extant DNA is a potent source of information for palaeontologists. The search for molecular fossils must, at least for the present, concentrate on the more or less informative building blocks of organic molecules produced by DNA (e.g. protein residues) as a pathway to the partial understanding of the composition of ancient biochemical systems. Certainly the widespread preservation of amino acids, and the more restricted survival of appreciable portions of original molecular structure for at least 70 million years, has been demonstrated by immunology and organic geochemistry. The extent of such excellent preservation is unknown, however, and at the present time the field is one of considerable potential, tantalizingly glimpsed but with little hard data.

Reports of the preservation of characteristic original amino acid sequences of 80 Ma fossil shell proteins are particularly encouraging, as is the reported (but not published) sequencing of 15 amino acids at the N terminal end of a small Ca^{2+} binding protein from oyster shells of Recent, Middle Miocene (15 Ma), Middle Cretaceous (100 Ma), and Middle Jurassic (175 Ma) ages. However, it has so far proved extremely difficult to sequence segments of fossil molecules routinely, perhaps because of interference from co-existing dark polymeric compounds. Technological developments which are much more precisely tailored for the special conditions of the fossil record are clearly crucial, and palaeontologists may well have to become more familiar with the capabilities of existing equipment and possibly even involved in the design of new equipment.

Whatever the technique, it is clear that an integrated approach is necessary to avoid the many potential pitfalls in working with molecular fossils. Taxonomic studies should, in the first instance, concentrate on organisms which have a long and abundant fossil record and are still living today. This allows cross-checking of phylogenetic inferences against morphology, extant biochemical systematics, and geological history, and the tracing of organic preservation from living to fossil within single lineages. Large numbers of well preserved specimens of different ages are required, since the yield of organics per gram of fossil is low. A robust, non-porous, coarsely-crystalline, skeletal ultrastructure is also an obvious advantage because of the protection it provides for enclosed molecular fossils.

At the present state of knowledge only high level taxonomic indicators can realistically be anticipated from molecular fossils, and the most obvious and dramatic demonstration of such an application may well come from problematic groups whose morphology provides ambiguous clues as to their taxonomic affinities. In any event it is certainly now possible to begin utilizing fossil molecular data to augment or complement existing taxonomic methodology. All of the taxonomic tools available to the palaeontologist, including the study of morphology alone, have particular strengths and weaknesses; taken in combination, molecular and morphological, living and fossil, they will be a potent measure of taxonomic relationships.

Future developments

The assimilation of such complex and unfamiliar

technology into geological investigations is necessarily a slow procedure, but the investigation of molecular palaeontology is now a blossoming field. Although molecular data from the fossil record is still beyond the grasp of most scientists, the speed of development, and of automation, is such that analyses of this kind may soon be routine. Current studies have clearly demonstrated that such work can significantly augment or complement a wide diversity of geological and biological research. In addition to taxonomy, there are, for example, indications that the remains of ancient molecules also contain important information on geochronology (i.e. amino acid dating), the origins of fossil fuels, palaeoenvironment reconstructions, and the processes which operate during diagenesis. For the biologists such organic remnants could well provide valuable insights into evolutionary processes at the molecular level.

References

- Curry, G.B. 1987. Molecular palaeontology: new life for old molecules. *Trends in Ecology and Evolution* **2**, 161–165.
- De Jong, E.W., Westbroek, P., Westbroek, J.F. & Bruning, J.W. 1974. Preservation of antigenetic properties in macromolecules over 70 myr old. *Nature* **252**, 63–64.
- Hare, P.E., Hoering, T.C. & King, K. (eds) 1980. *Biogeochemistry of amino acids*. John Wiley & Sons, New York.
- Lowenstein, J.M. 1986. Molecular phylogenetics. *Annual Review of Earth and Planetary Sciences* **14**, 71–83.
- Muyzer, G., Westbroek, P., de Vrind, J.P.M., Tanke, J., Vrijheid, T., De Jong, E.W., Bruning, J.W. & Wehmiller, J.F. 1984. Immunology and organic geochemistry. *Organic Geochemistry* **6**, 847–855.
- Runnegar, B. 1986. Molecular palaeontology. *Palaeontology* **29**, 1–24.
- Thorpe, J. 1982. The molecular clock hypothesis: biochemical evolution, genetic differentiation and systematics. *Annual Review of Ecology and Systematics* **13**, 139–168.
- Westbroek, P., van der Meide, P.H., van der Wey-Kloppers, J.S., van der Sluis, R.J., de Leeuw, J.W. & De Jong, E.W. 1979. Fossil macromolecules from cephalopod shells: characterization, immunological response and diagenesis. *Paleobiology* **5**, 151–160.
- Wyckoff, R.W.G. 1972. *The biochemistry of animal fossils*. Sciencetechnica, Bristol.

2.2 Speciation

B. CHARLESWORTH

Species concepts

Modern evolutionary biologists are generally agreed that the *biological species concept* provides the most satisfactory basis for discussing the problem of the origin of new species (*speciation*). According to the biological species concept, species are 'groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups' (Mayr 1942). This concept was arrived at during the nineteen-thirties, as a result of the recognition by systematists and geneticists that purely morphological definitions of species are unworkable, in view of phenomena such as the occurrence of intergrading sets of morphologically distinct populations over a geographical range, and of morphologically nearly indistinguishable but reproductively isolated populations coexisting in the

same place. Under the biological species concept, the former situation is now treated as a case of a single *polytypic species* or *Rassenkreis*, and the latter as the existence of a number of *sibling species*.

Examples of polytypic species and sibling species abound (Mayr 1963), and demonstrate that there is no tight causal relation between the evolution of morphological differences of the kind that are detectable in the fossil record, and the evolution of reproductive isolation. A polytypic species is at least potentially capable of evolving as a unit, e.g. a selectively favourable mutation that arises at one end of its range is capable of diffusing throughout the whole species, as a result of migration, and replacing its alternative allele. Conversely, such a mutation has no prospect of spreading from one biological species to another under natural conditions.

In this sense, the biological species represents a natural unit of evolutionary change. Clearly, the concept applies only to sexually reproducing organisms. Furthermore, cases where it is difficult to apply are observed, and are indeed to be expected, since intermediate stages in the degree of reproductive isolation between geographically or ecologically separated populations must occur under almost any evolutionary hypothesis other than that of strictly saltatory evolution (Mayr 1963).

Modes of reproductive isolation

According to this view of the nature of biological species, the process of speciation is to be equated with the development of reproductive isolation between two populations that were formerly fully capable of interbreeding. This process is the ultimate source of the diversity of life on Earth, since sympatric populations that are not reproductively isolated will eventually lose their distinctness. Modes of reproductive isolation may be divided into the two broad categories of *prezygotic* and *postzygotic* barriers, referring to whether or not F1 hybrid individuals are produced by matings between members of the two species (Dobzhansky 1937; Mayr 1963). Prezygotic modes of isolation include differences in ecology, timing of breeding or flowering, and differences in mating behaviour or reproductive physiology, that prevent successful fertilization in interspecific matings. In many groups, such as *Drosophila*, behavioural prezygotic isolating barriers are the primary agents that prevent gene flow between sympatric species, which are often completely isolated genetically in spite of the absence of strong postzygotic barriers (Coyne & Orr 1989). Postzygotic isolating barriers include inviability or sterility of F1 individuals, or of subsequent hybrid generations.

Theoretical studies have shown that isolating barriers have to be extremely strong in order to prevent gene flow between two populations that are in contact. Neutral alleles will face no significant obstacles to diffusion from one population to the other unless there is nearly a 100% loss in fitness to F1 hybrids, or their probability of formation is near zero, except for loci which are very closely linked to genes involved in controlling the isolating mechanisms (Barton & Charlesworth 1984). The same applies to alleles that are selectively advantageous in both populations, but originally arise by mutation in only one of them. Studies of *hybrid zones*, where two genetically distinct geographical races come

into contact along a linear transect, resulting in a relatively narrow region where hybrids are formed, have demonstrated empirically that such gene flow occurs at enzyme and protein loci detected by electrophoresis (Barton & Charlesworth 1984). These are probably close to being selectively neutral. Loci that are under natural selection that favours different alleles in the two populations (because of differences in environment or genetic background) may remain differentiated between populations in contact; the extent of such divergence depends on the balance between the strength of the selection pressure concerned and the amount of gene flow. *Clines*, where populations vary along a linear transect in response to an environmental gradient in selection pressure, are the product of such a balance (Endler 1977). There are numerous examples of clines maintained over very short geographical gradients by intense selection pressures, the classic example being the evolution of metal tolerance by plants living on polluted sites such as old mine spoils (Endler 1977). There is thus no difficulty in understanding how morphological and physiological differences can be maintained between populations that are more or less freely exchanging genes, and which show little differentiation with respect to protein loci. These populations do not constitute separate species.

Knowledge of the genetic basis of isolating barriers between species is clearly of crucial importance in understanding the ways in which they may evolve. In flowering plants, it seems that ecological differences between related species often prevent interbreeding between them, even if they inhabit the same general area; habitat disturbances may result in the mingling of populations isolated in this way. There is clearly no direct genetic control of reproductive isolation in these cases, other than via the characters that lead to the ecological differences between the species. Such differences seem to be under the same type of genetic control as similar within-species differences (Stebbins 1950). This is generally true of morphological and physiological differences between related species (Mayr 1963; Barton & Charlesworth 1984). Genetic analysis of behavioural isolating barriers has largely been confined to *Drosophila*. What little information is available indicates that these are usually polygenic in nature, with different loci controlling male and female courtship behaviour (Barton & Charlesworth 1984).

The genetic basis of sterility or inviability of interspecies hybrids has been much more thoroughly

studied. Especially in plants, differences in chromosomal arrangements between related species often result in reduced pairing between chromosomes at meiosis in F1 hybrids, leading to the production of gametes containing abnormal numbers of chromosomes (Stebbins 1950). Zygotes resulting from these gametes suffer reduced viability, so that the effective fertility of the hybrid is low. Accidental doubling of the chromosome number in an interspecies hybrid can lead to restoration of fertility, in cases where failure of pairing of chromosomes in meiosis causes hybrid sterility in diploids, since tetraploidy allows pairing between homologous chromosomes derived from the same species (Fig. 1). Crosses between such tetraploid hybrids and the parental species result in the production of sterile triploid individuals, so that the former are effectively a new species. Polyploidy of this kind is an important mode of speciation in flowering plants (Stebbins 1950), and is the only known method of saltatory speciation other than by the spread of a parthenogenetic variant within an originally sexual population.

F1 sterility due to failure of proper development of the gonads or germ cells, rather than failure of chromosome pairing, is a common phenomenon in animals and plants. In animals with separate sexes and chromosomal sex determination, it has long been known that it is often the *heterogametic* sex (i.e. the sex that is heterozygous for the chromosome pair involved in sex determination) that is most severely affected (*Haldane's rule*). The same rule

also applies to hybrid inviability. This provides an opportunity to study the genetic basis of the sterility by means of crosses involving the fertile sex. This has been extensively exploited in *Drosophila* (Dobzhansky 1937; Muller 1940; Coyne & Orr 1989). The results of these studies show that sterility (or inviability) is caused by interactions between several genes, such that combinations of alleles derived from the two different species result in sterility. The simplest situation of this kind is when one species has a genetic constitution A_1B_1/A_1B_1 and the other A_2B_2/A_2B_2 , where A and B represent two different loci. Each species is, of course, perfectly fertile and viable; infertility or inviability results from adverse effects of interactions between the alleles A_2 and B_1 or A_1 and B_2 . It is frequently observed that loci on the sex chromosomes themselves often contribute disproportionately to the fitness breakdown of hybrids. This is probably the causal basis of Haldane's rule; X-linked or Y-linked alleles from one of the two parental species are fully expressed in F1 hybrids of the heterogametic sex, and have the potential to interact adversely with alleles from the other species at loci on non-homologous chromosomes (Muller 1940; Coyne & Orr 1989).

The origin of reproductive isolation

These data indicate that changes at several gene loci are usually required for the achievement of reproductive isolation, other than by polyploidy or parthenogenesis. This is virtually a logical necessity, since it is most unlikely that a mutant allele at a single locus could both confer a high degree of infertility with individuals carrying the original allele, and become fixed in the population in opposition to this intense pressure of selection. The only faintly plausible mode of speciation by a single genetic change is through the chance fixation in a small population of a chromosomal rearrangement that causes drastic fertility loss to its heterozygous carriers. However, the chance of such a fixation event is very small in a random-mating population when the fertility loss to heterozygotes is high. Furthermore, even the most infertile rearrangement heterozygotes rarely suffer a fitness loss of more than 50% or so, and it has been shown that this reduction in fertility does not create much of an isolating barrier. Thus, even chromosomal speciation is most likely to occur as a result of a number of steps, each of which has a small impact on fertility (Barton & Charlesworth 1984).

The empirical evidence from the genetic analysis

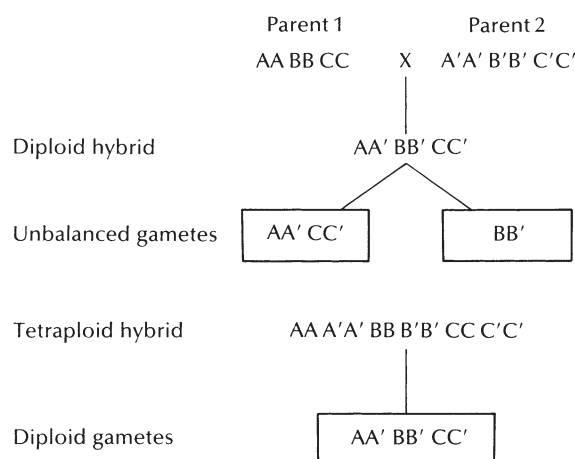


Fig. 1 The effect of polyploidy in restoring normal segregation in the F1 hybrid between two diploid species with three pairs of chromosomes. Note that a cross between the tetraploid hybrid (which produces diploid gametes) and either parental species will result in a triploid offspring. Such an individual will have low fertility, owing to the production of unbalanced gametes in meiosis.

of species crosses is in good general agreement with these population genetic considerations. It should be stressed, however, that we have little idea in any individual case as to what has caused the genetic divergence between two populations that results in pre- or postzygotic isolation. A variety of theoretical models that can generate the evolution of such isolation have been proposed. As first pointed out by Darwin (1859), when criticizing the idea that hybrid sterility has evolved in order to prevent the fusion of species, post-zygotic isolation cannot be selected for directly, since natural selection can never favour lowered fitness. Dobzhansky (1937) and Muller (1940) proposed that the accumulation of independent evolutionary changes in two totally geographically isolated populations will very probably eventually result in the evolution of reproductive isolation. This can occur even if the populations are subject to identical environments, since different mutant alleles at different loci will arise by chance in the two populations, and become fixed by genetic drift or natural selection. The process is facilitated by the existence of environmental differences between the populations, to which they become adapted.

Alleles that get fixed in one population are not selected to interact well with the alleles from the other population, but have to perform well on the background of their own population. Thus, a sufficiently long process of divergence will result in the establishment of gene combinations that function perfectly well within the population in which they occur, but produce a breakdown in fitness when alleles from one species are combined with those from the other. The loss in fitness to species hybrids is no more surprising than the fact that a carburettor from a car manufactured in the U.S.A. does not function in an engine made in Japan. This model is entirely consistent with the genetic evidence described above. Furthermore, the predominant role of the sex chromosomes in contributing to the lack of fitness of F1 hybrids between closely related species is predicted by this model, if the genetic changes concerned are due to the fixation by natural selection of alleles that are favourable on the background of their own species (Coyne & Orr 1989).

Similarly, prezygotic isolating barriers can be understood as a product of the gradual divergence in male and female courtship behaviour (in animals), or in the biochemistry of fertilization, between two geographically isolated populations. The male and female functions within any single population are, of course, always selected to en-

sure efficient mating and fertilization, but there is no such selection to preserve the ability to mate with individuals from a geographically separate population. Sexual selection, acting on mutations affecting male characteristics of relevance to success in competition for females, may promote the divergence of isolated populations with respect to mating behaviour. An alternative possibility, first suggested by Dobzhansky (1937), is that prezygotic isolating barriers are the product of selection for behaviour patterns that prevent the gamete wastage which occurs because of matings between members of two populations that are in contact, and kept separate by postzygotic barriers (the process of *reinforcement*). While reinforcement is a theoretical possibility, it seems unlikely to be the only cause of the evolution of prezygotic isolation, since many cases are known where this occurs between populations that have never been in contact. Furthermore, unless postzygotic barriers are very strong, it is probable that two populations in contact will merge before they evolve behavioural differences, except in systems where narrow hybrid zones are maintained or where sharp ecological gradients maintain differentiation (Fisher 1930). None the less, there is some evidence that reinforcement plays a role in the evolution of behavioural isolation in the genus *Drosophila*, since pairs of closely related *Drosophila* species that are sympatric tend to have stronger degrees of behavioural isolation than pairs of allopatric relatives (Coyne & Orr 1989).

The possibility that speciation may be triggered by random drift during periods of restricted population size associated with the foundation of new, geographically isolated populations (*founder effect speciation*) has been the subject of considerable debate (Mayr 1963; Barton & Charlesworth 1984; Carson & Templeton 1984). The basic idea is that two alternative stable equilibria may occur under natural selection. The ancestral population is located at one of these equilibria, and passage through a small founding population causes random sampling of genotype frequencies that can sometimes result in a chance transition from one equilibrium to the other. Under appropriate circumstances, a hybrid population formed from crossing populations located at the two alternative equilibria will suffer a substantial fitness loss, and so the two populations will be at least partially isolated reproductively.

The original motivation for this theory was the observation that populations of a species are often relatively uniform over a wide geographical range, but peripheral isolates may deviate sharply in their

characteristics from these. This led Mayr (1963) to propose that the potential for evolutionary change is restricted in large populations, because of the existence of genetic and developmental devices that prevent the manifestation of new variation on which selection could act. He claimed that these devices can be overcome by the random genetic changes that accompany the foundation of a new, isolated population. Later, Carson was stimulated to propose a related idea by the observation that inter-island migration of Hawaiian *Drosophila* is almost invariably accompanied by speciation (Carson & Templeton 1984). While theoretical models have shown that partial reproductive isolation can indeed evolve by this mechanism, the probability of producing anything approaching complete isolation by a single founder event is low. In addition, empirical and theoretical results of population genetics do not support the notion that evolutionary change is inhibited in large populations (Barton & Charlesworth 1984). Finally, alternative interpretations of the biogeographical data advanced in support of founder effect speciation have been proposed. For example, complete isolation permits two populations to differentiate with respect to favourable alleles, which would diffuse through both populations if they were connected by a chain of intermediate populations. This will produce an association between the foundation of new isolates and divergence or speciation. The role of founder events as causal agents of speciation thus remains controversial.

Sympatric and parapatric speciation

So far, the discussion of mechanisms of speciation has proceeded as though reproductive isolation takes place as a result of the genetic divergence of wholly isolated, geographically separate populations. This is the process of *allopatric speciation*, which can proceed by the mechanisms described above. There is little question that this is an important mode of speciation, perhaps the predominant one. This view has been vigorously championed by Mayr (1942), and there is indeed a wealth of distributional evidence which suggests that geographical isolation promotes genetic divergence and speciation. The most extreme alternative is *sympatric speciation*, the evolution of reproductive isolation between genotypes within a population that was originally mating randomly. Theoretical models of this process are reviewed by Seger (1985). The trigger for sympatric speciation is the maintenance of genetic polymorphism in a spatially heterogeneous

population, where different genotypes are favoured in different patches. In such a system, selection can favour the evolution of preferential mating between like genotypes, thus preventing the production of segregant offspring that may be ill-adapted to all types of patch. The existence of races of phytophagous insect species that are adapted to different host species is often quoted as an example of sympatric speciation, but it is not clear whether these are examples of true species (Futuyma & Mayer 1980). Since the theoretical conditions for the maintenance of genetic variation suitable for generating selection for preferential mating are rather severe (Seger 1985), the conclusion of Mayr (1963) that sympatric speciation is rare or non-existent seems likely to be essentially correct.

A more promising alternative is *parapatric speciation*, which involves the evolution of reproductive isolation between populations that are only partially isolated geographically. The classic model of this process is that of Fisher (1930), who suggested that a set of populations distributed along a geographical gradient of selection pressure would experience selection for mating preferences that would reduce the flow of genes between populations, and hence prevent the introduction of genotypes that are ill-adapted to the local environment. Later theoretical work has confirmed that this is, indeed, a mechanistically plausible process (Endler 1977; Barton & Charlesworth 1984). Of course, parapatric and allopatric speciation cannot be strictly distinguished, since populations at the extreme ends of a continental species range have very low probabilities of exchanging genes. There are several classic examples of reproductive isolation between populations located at the ends of such species ranges which have bent round on themselves, so that the extremes are now in contact (Mayr 1963). These represent cases in which reproductive isolation has evolved between populations that are connected by a series of other, adjacent populations between which gene flow may well be possible. If the intermediate populations were to become extinct, such cases would be open to misinterpretation as examples of strict allopatric speciation. Furthermore, the fact that the degree of geographical isolation tends to be correlated with divergence and speciation does not necessarily provide evidence for strict allopatric speciation, since (other things being equal) genetic divergence under the parapatric model will always be enhanced by restrictions on gene flow (Endler 1977). While it is always possible to interpret phenomena such as hybrid zones and

geographically disjunct species ranges in terms of secondary contact between species that have diverged in allopatry, such interpretations are not necessarily demanded by the data. Thus, the biogeographical evidence does not seem to permit a clear-cut conclusion to be drawn concerning whether or not speciation usually requires strict allopatry.

Ecological aspects of speciation

Up to now, the ecological significance of speciation has not been mentioned. For two related species to coexist stably in the same area, it is necessary that they become adapted to somewhat different ecological niches, otherwise the competitively superior one will cause the rapid extinction of the other. Especially in birds, there is extensive evidence for ecological differences between close relatives. The Galapagos finches provide a well documented example of the ecological divergence of related species, which appears to have been driven largely by divergent adaptations of geographically separated populations to different food sources (Grant 1986). Without the evolution of such different ecologies, speciation would not result in any increase in biological diversity within a given geographical area. Sibling species show that such ecological differences may arise without any gross morphological changes (Mayr 1963), although they are often associated with morphological differences, as in the case of the bills of the Galapagos finches. Ecological opportunities provided by the invasion of new habitats, unoccupied by competitors, must play a major role in stimulating the rapid increase in the number of species during adaptive radiations. Natural selection is, needless to say, the primary causal agent in this aspect of speciation (Grant 1986).

Speciation and the fossil record

Speciation, in the sense used here, is simply a population genetic process that results in the acquisition of reproductive isolation between two formerly interbreeding populations. It thus cannot be directly observed in the fossil record. Indeed, it is important to recognize that, in a sense, there is no such thing as a speciation event, since all the evidence suggests that the process of acquiring total reproductive isolation is a multistep process that requires numerous intermediate stages, examples of which can be studied in contemporary species (Mayr 1942, 1963). Of course, the whole process of acquisition of specific status may occupy only a few

thousand generations, as witnessed by examples of good species that are nearly identical at the molecular level (Barton & Charlesworth 1984; Coyne & Orr 1989). From the geological perspective, the time needed to develop complete reproductive isolation between two lineages may be effectively instantaneous. If no noticeable morphological differences evolve during the process, it will pass unnoticed in the fossil record, whereas morphological evolution in a single lineage might be counted as generating a new species, with the ancestral form A being replaced by a new form B. The fossil record thus provides a very incomplete picture of the process of speciation, particularly since subtle ecological, physiological and behavioural differences of the kind that frequently distinguish sibling species (Mayr 1963) will be missed. There are few cases in which it can be reasonably inferred that speciation has been observed in the record (Gingerich 1985; Section 2.3).

Of course, if form B appears in the record alongside A, it may be reasonable to infer that B originated in a speciation event elsewhere, and subsequently migrated into the range of A, which is unchanged morphologically. Patterns of this kind have been well documented by Cheetham (1987), for example. This kind of observation is the basis for the claim often made by supporters of the theory of punctuated equilibria, that morphological evolution usually only occurs in association with speciation (Gould & Eldredge 1977). As Turner (1986) has shown, it does not necessarily provide firm evidence for this claim. If speciation is often unaccompanied by detectable morphological change, then a progenitor of B with the same morphology as A could have coexisted alongside A, and only be distinguished by the palaeontologist as a result of evolutionary change that occurred well after speciation. Provided that evolutionary change in morphology is episodic, as is to be expected on most models of adaptive evolution, the punctuational mode of evolution can be explained without ascribing a special causal relation between speciation and morphological change. Such a causal relation does not seem to be consistent either with the evidence from present day organisms, where morphological change unaccompanied by reproductive isolation can be observed (as in polytypic species), or with population genetic theory (Turner 1986).

Nevertheless, the association of speciation with geographical isolation, and with new ecological opportunities, means that a correlation between episodes of rapid speciation and morphological evol-

ution and diversification is to be expected. Island radiations provide small-scale examples of this that have been intensively studied (Carson & Templeton 1984; Grant 1986). Larger-scale events, such as the mammalian radiations of the Eocene, may be inferred to have the same causes (Wright 1949), the successive occupation of new major modes of life by speciating lineages providing the basis for the origin of the diverse combinations of characteristics distinguishing higher taxa (see also Section 3.6).

References

- Barton, N.H. & Charlesworth, B. 1984. Genetic revolutions, founder effects, and speciation. *Annual Review of Ecology and Systematics* **15**, 133–164.
- Carson, H.A. & Templeton, A.R. 1984. Genetic revolutions in relation to speciation phenomena: the founding of new populations. *Annual Review of Ecology and Systematics* **15**, 97–131.
- Cheetham, A.H. 1987. Tempo of evolution in a Neogene bryozoan: are trends in single morphologic characters misleading? *Paleobiology* **13**, 286–296.
- Coyne, J.A. & Orr, H.A. 1989. Patterns of speciation in *Drosophila*. *Evolution* **43**, 362–381.
- Darwin, C. 1859. *The origin of species*. John Murray, London.
- Dobzhansky, T. 1937. *Genetics and the origin of species*. Columbia University Press, New York.
- Endler, J.A. 1977. *Geographic variation, speciation and clines*. Princeton University Press, New Jersey.
- Fisher, R.A. 1930. *The genetical theory of natural selection*. Oxford University Press, Oxford.
- Futuyma, D.J. & Mayer, G.C. 1980. Non-allopatric speciation in animals. *Systematic Zoology* **29**, 254–271.
- Gingerich, P.D. 1985. Species in the fossil record: concepts, trends and transitions. *Paleobiology* **11**, 27–41.
- Gould, S.J. & Eldredge, N. 1977. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* **3**, 115–151.
- Grant, P.R. 1986. *Ecology and evolution of Darwin's finches*. Princeton University Press, New Jersey.
- Mayr, E. 1942. *Systematics and the origin of species*, p. 120. Columbia University Press, New York.
- Mayr, E. 1963. *Animal species and evolution*. Columbia University Press, New York.
- Muller, H.J. 1940. Bearings of the *Drosophila* work on systematics. In: J.S. Huxley (ed.) *The new systematics*, pp. 185–268. Oxford University Press, Oxford.
- Seger, J. 1985. Intraspecific resource competition as a cause of sympatric speciation. In: P.J. Greenwood, P.H. Harvey & M. Slatkin (eds) *Evolution. Essays in honour of John Maynard Smith*, pp. 43–54. Cambridge University Press, Cambridge.
- Stebbins, G.L. 1950. *Variation and evolution in plants*. Columbia University Press, New York.
- Turner, J.R.G. 1986. The genetics of adaptive radiation: a neo-Darwinian view theory of punctuational evolution. In: D.M. Raup & D. Jablonski (eds) *Patterns and processes in the history of life*, pp. 183–207. Springer-Verlag, Berlin.
- Wright, S. 1949. Adaptation and selection. In: G.L. Jepsen, G.G. Simpson & E. Mayr (eds) *Genetics, paleontology and evolution*, pp. 365–389. Princeton University Press, New Jersey.

2.3 Microevolution and the Fossil Record

P. R. SHELDON

Introduction

Despite its many imperfections, the fossil record gives us a historical perspective on evolution that cannot be obtained from a study of living organisms alone. A lifetime's research on, say, fruitflies in a laboratory or peppered moths in a wood, though indispensable, can only span a fleeting moment in a species' history. A crucial task for evolutionary biologists is to integrate results from an increasingly detailed analysis of the fossil record into a comprehensive synthetic theory, one which bridges the

gap between the neontological and palaeontological scales of observation. This explains the growing attention paid by geneticists to high-resolution fossil data.

The term microevolution is taken here to mean all the evolutionary changes that occur within a species up to and including the formation of new species, either by lineage branching (i.e. cladogenesis or speciation) or by phyletic transformation (i.e. anagenesis).

Fossil species in practice

Although the fossil record of any species will always be deficient vertically (in time), laterally (in space), and morphologically, the hard-part record of some forms — especially shelly marine invertebrates — is rather better than sometimes asserted. Attempts to quantify completeness of stratigraphic sections have enabled palaeontologists to calibrate accessible levels of time resolution, and to define the kinds of evolutionary and palaeoecological questions that the fossil record is uniquely placed to answer (Section 3.12).

The acquisition of reproductive isolation can never be directly observed in the fossil record (Section 2.2). Nevertheless, many palaeontologists try to make the species they describe live up to a definition such as: 'Species are morphologically distinct groups within which variation is of the magnitude expected in interbreeding populations, and between which the differences are of the kind and degree expected to result from reproductive isolation of natural populations'. In practice, of course, such species can never be more than 'morpho-species' (units that embrace individuals of similar form), in which sibling species go undetected and from which truly conspecific variants and sexual dimorphs may be unwittingly excluded. Even greater conceptual problems arise if lineages undergo extensive anagenesis: some workers, especially biostratigraphers, divide such lineages into arbitrary chronospecies, whilst others, mostly theorists adopting a strict cladistic approach, would prefer to denote a single unbranching lineage with a single specific name, irrespective of the total change. The more continuous the record, the more problematic the nomenclature (e.g. Bown & Rose 1987; Sheldon 1987).

Patterns of evolution

The belief that many fossil species remained in morphological stasis throughout their existence prompted Eldredge and Gould (1972) to invoke a pattern of punctuated equilibrium as an alternative to the pervasive paradigm of phyletic gradualism. Previously, lack of intermediates between species had largely been accounted for by incompleteness of the record. Basing their proposition on Mayr's allopatric speciation model and on observations of Devonian trilobites and Pleistocene land snails, Eldredge and Gould argued that the rise of new species is expected to be episodic, local, and rapid

(as opposed to continuous, widespread, and slow) and so the chance of finding intermediates in the fossil record is bound to be low. A speciation event would normally span less than 1% of the species' later existence in stasis. Fossil sequences should show stasis with sharp morphological breaks marking the migration of the descendant form from the peripheral, isolated area in which it developed. The presumed ancestor is then expected to persist for a while alongside its putative descendant. Thus, according to punctuated equilibrium theory, significant evolutionary change occurs at events of branching speciation and not during the *in toto* transformation of lineages. These contrasting patterns have provided a very useful framework for discussion and stimulated a more rigorous analysis of the fossil record.

Perception of patterns

It has often proved impossible to establish the validity of an evolutionary pattern to universal satisfaction. Many relevant hypotheses, such as ancestor–descendant relationships or genetic versus ecophenotypic change, are never truly verifiable. But here, as usual in palaeobiology, we are not in the business of proof but in assessing the relative probabilities of competing hypotheses.

Fortey (1988) discussed the biases that influence perception of patterns, showing how it is comparatively easy for the convinced punctuationist to 'see' gradualistic change as a series of punctuation events. However, as intervals without data become restricted to smaller and smaller timespans, so one pattern rather than another can be shown to be more probable by applying the principle of parsimony (though, of course, the most parsimonious explanation is not necessarily the correct one).

Many of the standard textbook examples of gradualism dissolved in the wake of the punctuated equilibrium hypothesis. Most such cases were shown to lack sufficient documentation, whilst some, like the phylogeny of horses, were reinterpreted as showing punctuations, if anything. The ideal recipe for establishing evolutionary patterns includes some ingredients that are very difficult to obtain: many complete specimens from successive, small stratigraphic intervals whose relative age is unequivocal; a framework of well-constrained absolute ages; samples spanning the entire geographical and temporal range of all closely-related lineages; as many ontogenetic stages as possible (in order to recognize heterochronic relationships) and statistical data on

all available characters. To avoid generating artificial patterns, fossils should only be assigned to named species late in these procedures.

A knowledge of geographical variation is important because spurious vertical patterns of phyletic change could arise in local sections by waves of immigration and emigration of intraspecific variants tracking their favoured environment. Also, in theory, a peripheral isolate might evolve *gradually* to a new species which, on migration, appears abruptly alongside its parent species remaining in stasis. Geographical consistency of morphological change is not, however, a prerequisite for establishing the validity (i.e. genetic basis) of evolutionary trends. For example, dissimilar trends could occur in adjacent populations of a benthic species living in a tectonically unstable shelf area. The subpopulations might become isolated for a while in silled basins, each imposing different selection pressures, such as if one basin shallowed as another deepened. Eventually divergent strands of the lineage might be reunited if the physical barriers to gene exchange were removed whilst hybridization was still possible.

Examples from the fossil record

The case histories which follow have been selected to illustrate various aspects of the debate and for their implications for microevolutionary theory.

Williamson (1981) presented evidence for both stasis and punctuated speciation in many lineages of Cenozoic lacustrine molluscs from East Africa. The 'speciation events', which took 5000–50 000 years, were accompanied by an increase in morphological variance that Williamson interpreted as extreme developmental instability in transitional populations. The strong possibility remains, however, that the new short-lived forms were eco-phenotypic variants induced by intense environmental stress.

Hallam (1982) concluded that the Jurassic oyster *Gryphaea* showed a step-like pattern of punctuated change, allied with morphological trends, some of which were paedomorphic. He found no evidence of gradualism or cladogenesis. Recently, however, Johnson (*in* Fortey 1988) has reported gradual and continuous derivation of *Gryphaea* morphology in the Middle Jurassic from another oyster, *Catinula*.

Stanley and Yang (1987) documented stasis in shape for 19 lineages of Neogene bivalves, some spanning as much as 17 million years without taxonomically significant change. Populations millions of years old often resemble their Recent descend-

ants almost as closely as geographically separated conspecific living populations resemble each other. Stanley and Yang emphasized that shape and size should be kept separate in all calculations of evolutionary rates, arguing that most reported trends relate to variables representing only some measure of body size.

Cheetham (1987), examining 46 characteristics in nine species of Neogene bryozoans, found overwhelming evidence for stasis and, by inference, punctuated speciation. The few within-species trends present related to features not used in species diagnoses and he cautioned against judging patterns from single morphologic characters.

Planktic foraminifera and radiolarians yield some of the best known lineages in the fossil record, because of their widespread distribution, abundance in DSDP cores, and commercial use in biostratigraphy. These micro-organisms display a wide variety of evolutionary tempos and modes (see papers in *Paleobiology* 9 (4), 1983, and Banner & Lowry *in* Cope & Skelton 1985 for reviews). Gradual changes seem relatively common. Malmgren and Kennett (1981) demonstrated persistent gradualism in a lineage of temperate foraminifera that spanned four successive chronospecies in 8 million years. Planktic foraminifera also show a pattern best described as punctuated anagenesis (e.g. Malmgren *et al.* 1983), which is probably common in other groups too, including ammonites (e.g. Callomon *in* Cope & Skelton 1985).

Sheldon (1987) reported parallel gradualistic evolution of benthic Ordovician trilobites from central Wales. Over a period of *c.* 3 million years eight lineages underwent a net increase in the number of pygidial ribs, a character used in species diagnosis (Fig. 1). The end members of most lineages had previously been assigned to different species and, in one case, to different genera. In view of intermediate morphologies and temporary trend reversals, practical taxonomic subdivision of each lineage proved impossible. The apparent success of earlier Linnean nomenclature (with its implications of discrete species) could easily have been misinterpreted as evidence of punctuation and stasis. Perception of many other gradualistic patterns equally may have been hindered by conventional descriptive procedures, particularly the requirement to apply binominal taxonomy to fossils and the practice of lumping together specimens collected from different horizons in order to amass enough material for full 'species' description.

Although vertebrate data are sparse, gradual evo-

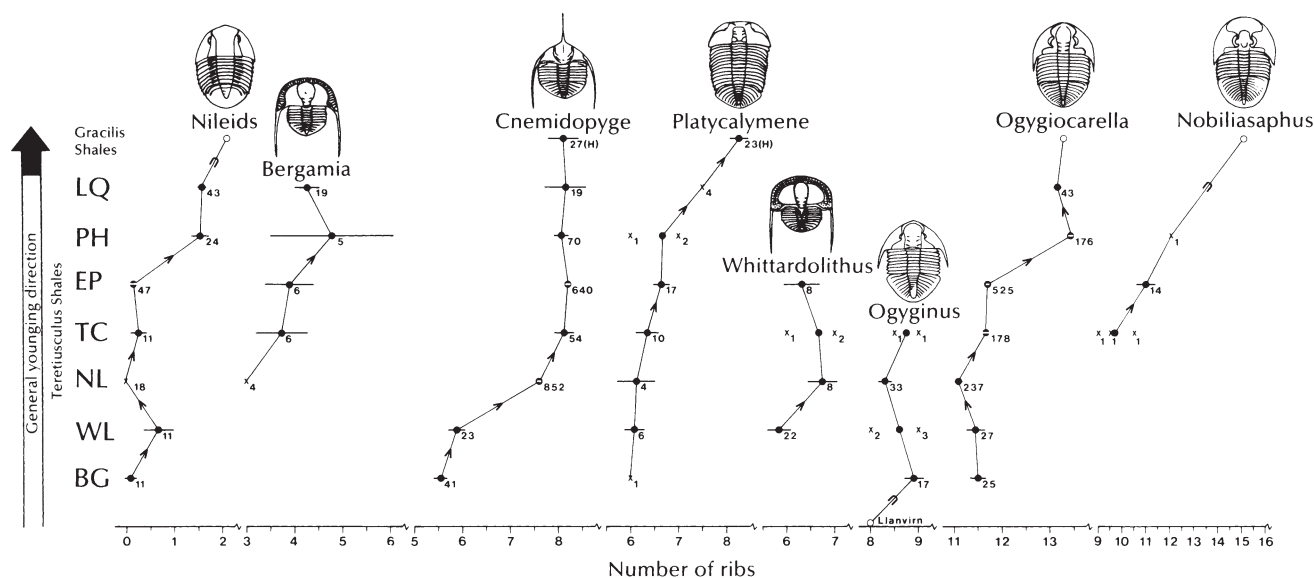


Fig. 1 Summary of changes in mean number of pygidial ribs for eight Ordovician trilobite lineages in the Builth inlier, central Wales. Means with 95% confidence interval and number of measurements are shown by \bullet_n and \circ_n . Approximate mean (\circ). Data from Hughes (see Sheldon 1987) (H). Individual measurements (X_n). Successive means (H) are significantly different at the 95% confidence level. Successive means (H) are certain to be significantly different but full data are unavailable. Vertical spacing between sections (BG, etc. see Sheldon 1987) is not to scale. Reversible shifts in means are common *within* sections (not shown subdivided). Sections in the Teretiusculus Shales span *c.* 3 million years. (From Sheldon 1987.)

lution appears to be fairly common, particularly in Tertiary mammals (e.g. Gingerich 1985; Godinot *in* Cope & Skelton 1985). Bown & Rose (1987) saw no sign of stasis in Eocene primates from Wyoming, reporting both gradual anagenesis and cladogenesis in sharp contradiction to the predictions of the punctuated equilibrium model. They highlighted the problems that gradualism causes for systematic palaeontology and biostratigraphy. Bell *et al.* (1985), in a multicharacter study of a Miocene stickleback lineage, found that taxonomically significant morphological change was accomplished by protracted trends *and* by rapid bursts of evolution, without tight synchrony of change among characters (mosaic evolution).

Although there are some well attested examples of gradual cladogenesis in the fossil record (Gingerich 1985), the great rarity of branching points where nodes are known is consistent with common patterns of change in which cladogenesis is rapid and/or involves small, isolated populations.

Random change and trend reversals

There has been much interest in the possibility that some morphological trends seen in the fossil record may be the result of processes other than natural selection. The genetic drift hypothesis predicts that

the morphology of selectively-neutral characters will vary through time as a random walk. Indeed, it has been argued that evolutionary rates exist only when the hypothesis of a symmetrical random walk can be refuted. However, Sheldon (1987) argued that temporary reversals of variable characters probably occur in all evolutionary lineages, and so many trends driven by selection may be indistinguishable from random walks. Reversals probably reflect times when some other attribute, genetically uncorrelated with the one under consideration, was selected. It would be unreasonable to expect that the one feature chosen for plotting was consistently the only one favoured by selection, or that it was always linked to every other favoured trait. In fact, long-sustained net trends in a single character may reflect genetic coupling to other characters having negative effects on fitness (see also Section 2.2).

The widespread tendency not to expect reversals, or to interpret them as ecophenotypic change or random drift, led to the unrealistic portrayal of phyletic gradualism as unidirectional change. Reversals have many consequences. For instance, they complicate the theoretical arguments (Fortey 1988) concerning differentiation between cases of gradualism and punctuated equilibria; they should not be automatically taken to indicate that the observed change is only ecophenotypic; and jumps in mor-

phological trends cannot be used to estimate the amount of time missing at diastems.

Patterns of evolution in different environments

It is still inappropriate to estimate the relative importance of particular patterns of evolution in different environments. Given the immense range of attributes of living organisms (e.g. complex life cycles and reproductive strategies) it would not be surprising to find different patterns emerging from broadly similar environments. Benthic invertebrates, for instance, have a wide diversity of larval dispersal modes and these early stages, although rarely preserved, might profoundly influence patterns. There is some evidence, as might be expected, that abrupt speciation and extinction are commonly associated with benthic species living in shallow marine settings. However, Sheldon (1987) suggested that, almost paradoxically, stasis seems to prevail in these more widely fluctuating, rapidly changing environments, whereas species living in, or able to track, narrowly fluctuating, slowly changing environments show persistent phyletic evolution rather than stasis.

Some of the perceived punctuations in shallow benthic settings may simply reflect higher rates of short-term deposition and more hiatuses (less completeness) than in offshore, pelagic environments. But, although the most reliable evolutionary patterns will come from the most complete sequences, the depositional conditions promoting completeness might in themselves encourage gradual phyletic evolution, especially of benthos.

Conclusion

Studies of the fossil record have revealed a wide spectrum of microevolutionary patterns, from which can be inferred a variety of evolutionary processes. Punctuated equilibrium and phyletic gradualism should be viewed as just two theoretical versions of many possible evolutionary patterns and the temptation to force poorly documented cases to fit one or other of these models must be resisted. Often there is simply too little data to assess patterns of change as, for example, with the genus *Homo* (Section 1.12). We are not yet in a position to assess accurately the relative frequency of particular patterns and the domain of their expected settings.

Individual taxa probably exhibit different patterns at different times, and different morphological characters in the same species may evolve at different rates. Episodic changes need not be associated with branching events and demonstrating stasis in a species is not the same as demonstrating punctuated speciation.

Most geneticists believe that a punctuated appearance of species is consistent with neo-Darwinian theory. In many ways it is explaining stasis — which is certainly more prevalent than would have been predicted from studies of living organisms — that represents the greater challenge.

References

- Bell, M.A., Baumgartner, J.V. & Olson, E.C. 1985. Patterns of temporal change in single morphological characters of a Miocene stickleback fish. *Paleobiology* **11**, 258–271.
- Bown, T.M. & Rose, K.D. 1987. Patterns of dental evolution in early Eocene anaptomorphine primates (Omomyidae) from the Bighorn Basin, Wyoming. *Memoir of the Paleontological Society* **23**, 162 pp.
- Cheetham, A.H. 1987. Tempo of evolution in a Neogene bryozoan: are trends in single morphologic characters misleading? *Paleobiology* **13**, 286–296.
- Cope, J.C.W. & Skelton, P.W. (eds) 1985. Evolutionary case histories from the fossil record. *Special Papers in Palaeontology* **33**, 1–203.
- Eldredge, N. & Gould, S.J. 1972. Punctuated equilibria; an alternative to phyletic gradualism. In: T.J.M. Schopf (ed.) *Models in paleobiology*, pp. 82–115. Freeman, San Francisco.
- Fortey, R.A. 1988. Seeing is believing: gradualism and punctuated equilibria in the fossil record. *Science Progress* **72**, 1–19.
- Gingerich, P.D. 1985. Species in the fossil record: concepts, trends, and transitions. *Paleobiology* **11**, 27–41.
- Hallam, A. 1982. Patterns of speciation in Jurassic *Gryphaea*. *Paleobiology* **8**, 354–366.
- Malmgren, B.A. & Kennett, J.P. 1981. Phyletic gradualism in a Late Cenozoic planktonic foraminiferal lineage; DSDP Site 284, southwest Pacific. *Paleobiology* **7**, 230–240.
- Malmgren, B.A., Berggren, W.A. & Lohmann, G.P. 1983. Evidence for punctuated gradualism in the Late Neogene *Globorotalia tumida* lineage of planktonic foraminifera. *Paleobiology* **9**, 377–389.
- Sheldon, P.R. 1987. Parallel gradualistic evolution of Ordovician trilobites. *Nature* **330**, 561–563.
- Stanley, S.M. & Yang, X. 1987. Approximate evolutionary stasis for bivalve morphology over millions of years: a multivariate, multilineage study. *Paleobiology* **13**, 113–139.
- Williamson, P.G. 1981. Palaeontological documentation of speciation in Cenozoic molluscs from Turkana Basin. *Nature* **293**, 437–443.

2.4 Heterochrony

K. J. McNAMARA

Introduction

Heterochrony is the phenomenon of changes through time in the appearance or rate of development of ancestral characters. While the recognition of a close relationship between ontogeny and phylogeny has a long history it was not until the late nineteenth century that it was formalized by E. Haeckel in his 'Biogenetic Law' (ontogeny recapitulates phylogeny). This involved a change in the timing of developmental events; but only in one direction — by terminal addition. Phylogenetically this meant that ancestral adult forms were encapsulated in the juvenile stages of their descendants. This became known as recapitulation.

Exceptions to this rule (known as 'degenerate' forms) were noted by the leading protagonists of recapitulation, particularly palaeontologists such as A. Hyatt (ammonites), R. Jackson (echinoids and bivalves), and C. E. Beecher (trilobites and brachiopods). With the awareness that these so-called degenerate forms were at least as common as examples of recapitulation, the Biogenetic Law began to slide into oblivion.

In the nineteen-twenties W. Garstang recognized that ontogeny did not always recapitulate phylogeny — it created it. Garstang believed that the retention of ancestral juvenile characters by descendant adults, which he termed *paedomorphosis*, was the key to understanding the evolution of many major groups of organisms, in particular the evolution of vertebrates from tunicate larvae. However, recent research has shown that both paedomorphosis and 'recapitulation' play important roles in evolution (Gould 1977; Alberch *et al.* 1979; McNamara 1986a; McKinney & McNamara 1991).

Nomenclature

Heterochrony involves the decoupling of the three fundamental elements of growth: size, shape, and time, or the extension or contraction of these elements. Temporal changes of size and shape relative to one another produce heterochrony, when either size or shape, or both, are affected by changes in their rate of ontogenetic development. Changes to

the timing of onset or cessation of morphological development and size change can also produce heterochrony. If size alone changes between ancestor and descendant, *dwarfs* or *giants* are produced.

If the rate of shape change is increased, or its period of operation is extended, the descendant adult passes morphologically 'beyond' the ancestor: this is *peramorphosis* (this equates, to some degree, with the Haeckelian 'recapitulation'). Conversely, if the rate of shape change is reduced, or its period of operation is contracted, the descendant adult passes through fewer growth stages, so resembling a juvenile stage of the ancestor: this is *paedomorphosis*.

These terms can be applied not only to the appearance of meristic characters (in other words, individual structures produced during an organism's ontogeny) but also to subsequent changes of shape of these structures during ontogeny. Thus not only may the rate of induction of structures vary, but the structures which are produced may show phylogenetic changes as they vary their rate of shape change. These two basic forms of heterochrony are known respectively as *differentiative heterochrony* and *growth heterochrony* (Figs 1, 2).

The relationship between size and shape is known as *allometry*. If the relative size and shape of a structure remain the same relative to overall body size during ontogeny, growth is *isometric*. However, if a particular structure increases in size relative to the whole organism, as well as changing its shape, growth shows *positive allometry*. Should a structure decrease in relative size, growth shows *negative allometry*. Increasing the degree of allometry is expressed phylogenetically as peramorphosis. Reducing it produces paedomorphosis. Similarly, extending or contracting the period of allometric growth produces peramorphic or paedomorphic descendants respectively.

Paedomorphosis and peramorphosis are morphological expressions of heterochronic processes. Paedomorphosis can occur by *progenesis*, *neoteny*, or *post-displacement* (Fig. 1). Peramorphosis can occur by *hypermorphosis*, *acceleration*, or *pre-displacement* (Fig. 2).

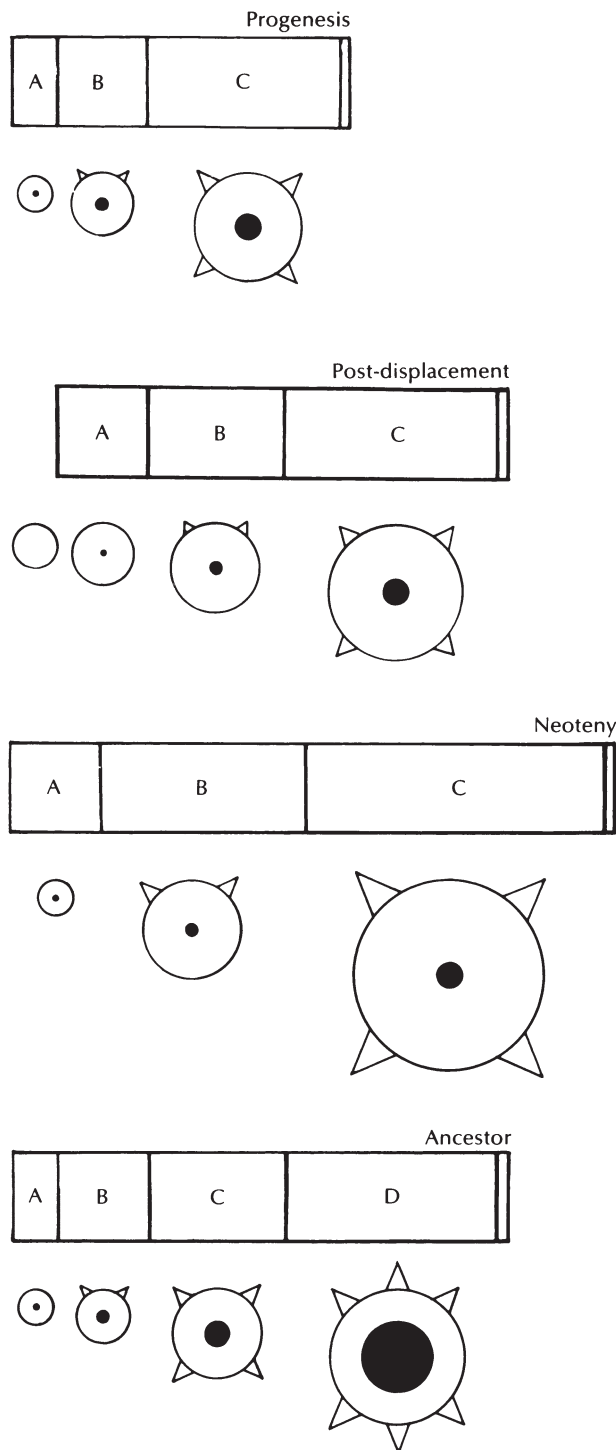


Fig. 1 The relationship of the three paedomorphic processes to the ancestor. *Progenesis* occurs by precocious maturation, *post-displacement* by the delayed onset of growth, and *neoteny* by reduced rate of morphological development. *Differentiative paedomorphosis* is shown by the spine production, *growth paedomorphosis* by the central spot. (From McNamara 1986a.)

Progenesis often occurs by precocious sexual maturation. Consequently morphological and size development is prematurely stopped, or severely retarded. The resultant adult paedomorph will be smaller than the ancestral adult. The prematuration morphological history of both the progenetic form and its ancestor will be identical. Progenesis is often global, affecting all structures, but it may also affect local growth fields. Some characters, however, are likely to have a more distinctly juvenile appearance than others. Thus, in the fossil record, it is generally possible to deduce the operation of progenesis: the morphotype is smaller than its presumed ancestor and resembles a juvenile stage of the ancestor. It will, however, be appreciably larger than the corresponding ancestral juvenile stage.

Neoteny is reduction in rate of morphological development during juvenile growth. If maturity occurs at the same age in both ancestor and descendant, they will be the same size. Often onset of sexual maturity is delayed in neotenic forms, consequently the neotenic forms attain a larger adult size. Neoteny may be global in its effects on the organism, or dissociated when it affects only certain morphological structures. Reduction in degree of allometry of specific structural elements will result in neoteny. Unlike progenesis, where the juvenile ontogenetic trajectories of ancestor and descendant are alike, juvenile growth trajectories are different between ancestors and descendants.

Post-displacement occurs by changes to the timing of onset of growth of particular morphological structures. Thus, by comparison with the ancestor, a structure commences development at a later stage, compared with other parts of the organism. Should subsequent development and cessation of growth be the same in the descendant as in the ancestor, the displaced structure will attain a shape at maturity resembling that found in a juvenile of the ancestral form. The displaced structure is also likely to be smaller than in the ancestor.

Hypermorphosis occurs by extending the juvenile growth period, by a delay in the onset of sexual maturation. Early juvenile development will progress at the same rate as in the ancestor. By extending growth allometries to a larger size, the hypermorphic adult can attain morphological characteristics quite distinct from those of the ancestral adult. Like progenesis, hypermorphosis is often global in its effects, but it too can affect only local growth fields.

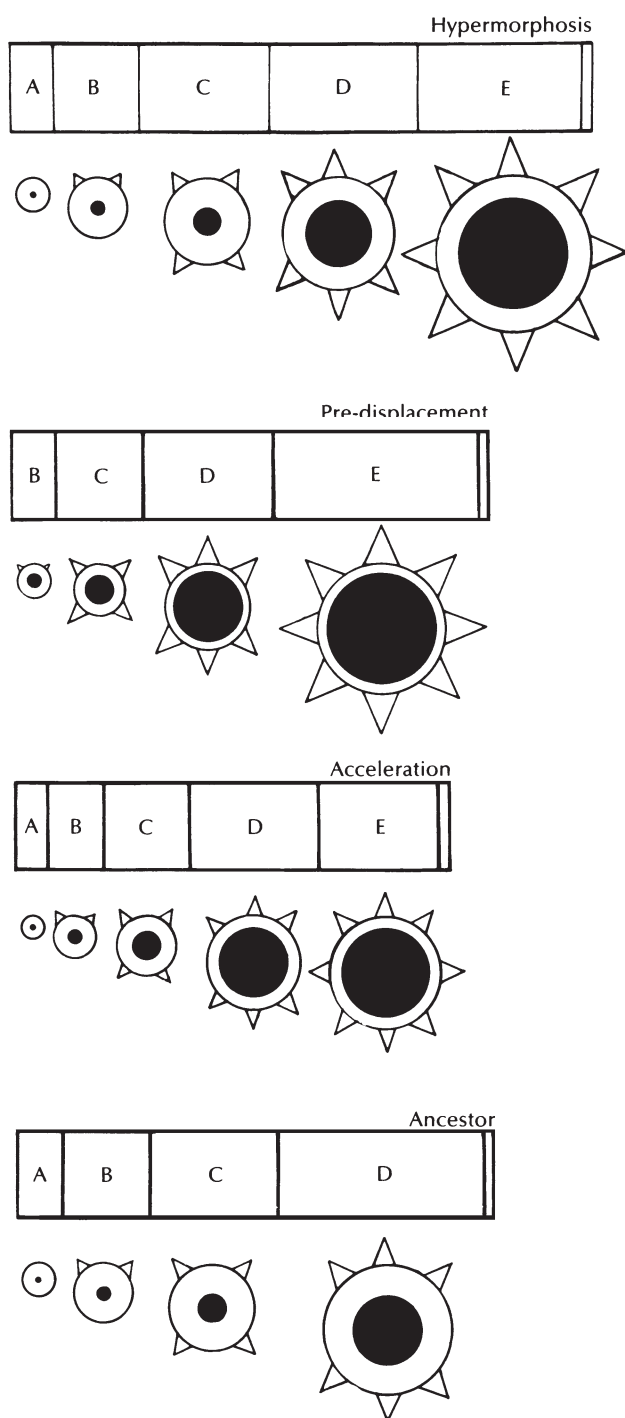


Fig. 2 The relationship of the three peramorphic processes to the ancestor. *Hypermorphosis* occurs by delayed sexual maturation, *pre-displacement* by earlier onset of growth, and *acceleration* by increasing the rate of morphological development. The spines and central spot demonstrate *differentiative* and *mitotic peramorphosis*, respectively. (From McNamara 1986a.)

Acceleration of rate of morphological development during ontogeny will produce a peramorphic descendant. In allometric terms, acceleration is an increase in the degree of allometry. For meristic characters it is an increase in the rate of production of structures. If acceleration is operating only on specific structures, then there need be no overall increase in body size. However, the particular structure is likely to be larger. As with neoteny, juvenile ontogenetic growth trajectories will be different in the ancestor and descendant.

Pre-displacement involves the earlier onset of growth of a specific structure. This allows a longer period of growth and development. Ancestral allometries will therefore, in effect, be extended. The resultant structure will be more advanced morphologically and larger than the equivalent structure in the ancestral adult, so long as cessation and rate of growth are identical in ancestor and descendant.

The identification of heterochronic processes in the fossil record is generally based on the precept that these processes can be characterized by the study of size and shape alone. The assumption is made that size is a proxy for time: the larger the organism, the longer period of time it took to reach that size. This assumption may not always be valid. Shea (1983) has suggested that two forms of progenesis and hypermorphosis can be recognized. In the first, time and size are not dissociated; thus smaller size correlates with shorter time, larger size with longer time. This Shea calls 'time hypomorphosis (= progenesis)'. The corresponding peramorphic process is time hypermorphosis. In the second case the progenetic form attained its reduced size and shape in the same amount of time that the ancestor took to attain maturity. This occurred because the rates of size and shape change were equally reduced through ontogeny compared with the ancestor. This Shea termed 'rate hypomorphosis (= progenesis)'. Time and rate progenesis or hypermorphosis can theoretically be distinguished in the fossil record. Early ancestral and descendant ontogenies will be the same when time progenesis has occurred, whereas they will differ in rate progenesis. Future emphasis on the study of growth lines in suitable invertebrate groups, such as molluscs, corals, and echinoids, will allow the true rates of growth of fossil organisms to be ascertained (McKinney 1988).

Heterochrony at different hierarchical levels

While most of the literature dealing with heterochrony as a factor in evolution concentrates on its role at the specific or supraspecific level, it needs to be stressed that much recognized intraspecific morphological variation in populations is, in fact, engendered by heterochronic processes. These act upon both meristic and allometric traits. For instance, intraspecific variation in ammonites often involves variation in the numbers of ribs or tubercles generated at a certain size. Similarly, in echinoids intraspecific variation often involves differences in the rate of production of meristic characters, such as the number of coronal plates and spines. Variation in numbers of these structures between two individuals of the same size may be accounted for either by variations in rates of development (neoteny or acceleration) or by onset and offset of growth (pre- or post-displacement and progenesis or hypermorphosis).

However, variations in rate of size increase may also produce such intraspecific differences. Thus if two individuals from a single population each 20 mm in length possess, in one case, six spines, and in the other eight, this may reflect a variation in rate of spine development (neoteny or acceleration), if both attained 20 mm in the same period of time. Alternatively, the individual with six spines may have increased in size at a faster rate through ontogeny, and thus only have had sufficient time to generate six spines. It is possible to test whether this latter mechanism has occurred by analysing the developmental patterns of other structures. For instance, if one of these organisms reached a length of 20 mm faster than the other, then all of its structures should appear relatively paedomorphic. However if, as is often the case, intrapopulation variation shows some characters to be paedomorphic and others peramorphic, then rates of structural development will have changed.

Selection of heterochronic morphotypes, and the resultant morphological evolution of a new species, is reflected in substantial shifts in the mean values of heritable phenotypic variation of shape or size of morphological structures. These occur by perturbations to the developmental programme. These may be under strong directional selection pressure (see below). Evolution of a substantial new heterochronic morphology may result in the evolution of new adaptive structures. These allow either geographical or ecological separation from the ancestral stock, and subsequent genetic isolation and estab-

lishment of a new species (see also Section 2.2).

In recent years documentation of heterochrony in the fossil record at the interspecific level has been undertaken in particular on ammonites (see McKinney 1988), echinoids (McNamara 1988), and trilobites (McNamara 1986b). It has been suggested (McNamara 1982) that heterochrony may be one of the factors responsible for rapid speciation events. This is particularly so where progenesis or hypermorphosis have occurred. However, gradual, phyletic changes may equally well be engendered by small modifications in growth rates between populations, resulting in subtle shifts in morphology through time.

Heterochrony has been proposed as a major factor in evolution at the supraspecific level. For instance, the orthodox view of the origin of vertebrates is that they may have arisen from the pelagic larva of a tunicate-like deuterostome invertebrate. This would have occurred by progenesis from an early larval stage. The free-swimming tunicate larva possesses all the fundamental chordate characters: a notochord, dorsal hollow nerve cord, gill slits, and post-anal propulsive tail. Attainment of precocious sexual maturation would have caused the retention of such ancestral larval characters into the adult phase and a consequent major adaptive breakthrough.

The earlier that perturbations to the embryological developmental system occur, the more profound the morphological consequence. Taxonomically, this is likely to be expressed at a high level. For instance, it has been suggested (McNamara *in* McKinney 1988) that progenesis at early developmental stages has been instrumental in the evolution of a number of higher taxa: saleniid, tiarechinid, neolampadoid, and clypeasteroid echinoids; edrioasteroids; baculitid ammonites; thecideidine and craniacean brachiopods; and branchiosaurid amphibians (Fig. 3).

Other heterochronic processes have also been instrumental in the evolution of higher taxa. For instance, it has been proposed that birds may have evolved from theropod dinosaurs. The very large orbits of birds, their inflated braincase, retarded dental development, and overall limb proportions indicate that early birds may have been paedomorphic theropods. Feathers are thought to have been present on juvenile theropods. The paedomorphic processes were probably neoteny and post-displacement.

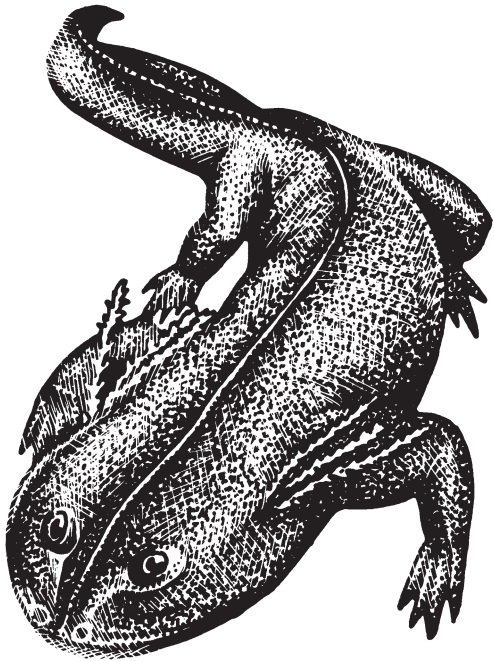


Fig. 3 Reconstruction of a paedomorphic branchiosaurid amphibian.

Heterochrony and directed speciation

The pattern that is emerging from studies of heterochrony in the fossil record is one of frequent directed heterochronic speciation. The direction of morphological evolution is strongly constrained by the nature of the organism's own ontogeny. Thus a number of characters in a lineage may show progressive paedomorphosis or peramorphosis. Provided that the descendant morphotypes are suitably adapted along an environmental gradient, a phylogenetic trend, in the form of a *paedomorphocline* or *peramorphocline*, may develop. The environmental and morphological directionality may be induced by the effects of either competition or predation. With induction of the heterochronic morphological gradient by competition, the persistence of the ancestral form constrains selection to one direction: along the environmental gradient away from the ancestral species. The phylogenetic pattern generated will be one of cladogenesis. Selective pressure from predation in one environment may induce the evolution of a paedo- or peramorphocline. In this case the phylogenetic pattern is one of anagenetic speciation.

Recent studies of echinoids, brachiopods, bivalves, ammonites, graptolites, and ammonites (see McNamara *in* McKinney 1988) indicate that the anagenetic pattern is common. Specific examples

include the evolution of a number of anagenetic paedo- and peramorphoclines in spatangoid echinoids, such as *Schizaster*, *Hemiaster*, *Lovenia*, *Pericosmus*, and *Protenaster* (Fig. 4). All show evolution from coarse to fine-grained sediments (probably shallow to deep water). Conversely, the Cenozoic brachiopod *Tegulorhynchia* evolved along a paedomorphocline from deep to shallow water into the genus *Notosaria* (Fig. 5). Similarly a number of trilobite lineages are thought to have evolved by heterochrony along the same environmental gradient (McNamara 1986b). In the marine environment changing water depth and sediment type are frequent environmental gradients along which paedo- and peramorphoclines develop.

Ecological causation of heterochrony

While many of the examples of directed heterochronic evolution have been interpreted as having arisen by selection of morphologically adaptive characters, it has also been argued that other factors, such as life history strategies, which affect elements such as size and time of maturation, may also be targets of selection. McKinney (1986) has suggested that for a suite of Tertiary echinoids selection favoured large forms along an environmental gradient from shallow to deep water (equating with unstable to stable environments) (Fig. 6). He argued that any subsequent morphological changes were incidental allometric by-products of the size change. The larger size was attained either by slower, neotenic growth or by extended, hypermorphic growth. This indicates that the target of selection was reproductive timing and/or body size. Such size increase along lineages (Cope's Rule; Section 2.10) may reflect K-selective pressure (large body size, delayed reproduction and development, and longer life spans in a stable environment).

While analyses of other echinoid lineages does not provide unequivocal corroboration of this pattern, there is ample evidence that many progenetic species are, conversely, *r*-selected (small body size, early maturation, high fecundity, and short life span in an unstable environment). Many so-called 'dwarfed' faunas may be *r*-strategists, inhabiting unstable, fluctuating environments. The small body size of progenetic Late Cretaceous oysters and ammonites may have been an adaptation to a soft, unstable substrate. The same is true for many progenetic brachiopods. High fecundity of progenetic species has been documented in edrioasteroids and trilobites (McNamara *in* McKinney

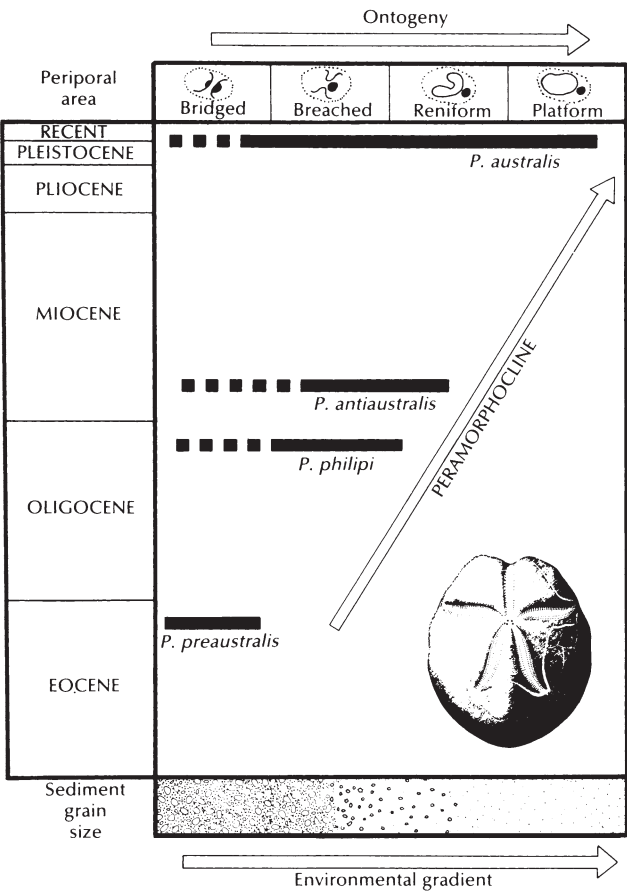


Fig. 4 Peramorphocline in the evolution of the periporal area in the Cenozoic spatangoid echinoid *Protaster*. Later species underwent greater morphological change in this character than earlier species. This led to the ability of later species to feed from progressively finer grained sediments. (From McNamara 1985.)

1988). Although in these cases small size and precocious maturation may have been the principal targets of selection, unless the resultant progenetic morphology was also adaptively significant in the new environment, selection would not have occurred. Most heterochronic changes occur as a result of changes to the internal developmental regulatory system. However, certain changes may actually be induced by environmental perturbations. The effect of changing environmental pressures may lead to facultative heterochrony within populations. For instance, the frequency of development of pae-domorphs in living populations of salamanders is directly influenced by the population density. When low, a large proportion of individuals develop as neotenic pae-domorphs, attaining maturity in their larval form, so remaining and reproducing in the aquatic environment. At high population densities

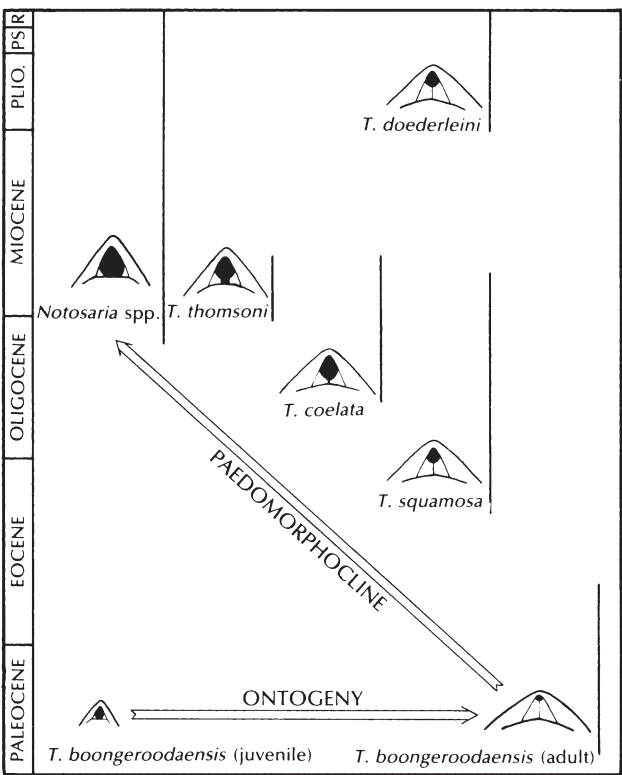


Fig. 5 Pae-domorphocline in the evolution of the umbonal region of the Cenozoic rhynchonellid brachiopods *Tegulorhynchia* and *Notosaria*, showing pae-domorphic retention of larger foramen (and hence thicker pedicle) and smaller umbonal angle. The environmental gradient along which the pae-domorphocline evolved was from deep to shallow water. (From McNamara 1983.)

few individuals are pae-domorphic, the high density levels inducing metamorphosis to the terrestrial form. There is also indirect evidence from the fossil record (McNamara 1986b) that changes in water temperature at different water depths in the marine environment may have been a factor in inducing progenesis in a number of lineages of Cambrian trilobites. Experimental work has demonstrated the effect of higher temperatures in inducing premature maturation in some living arthropods.

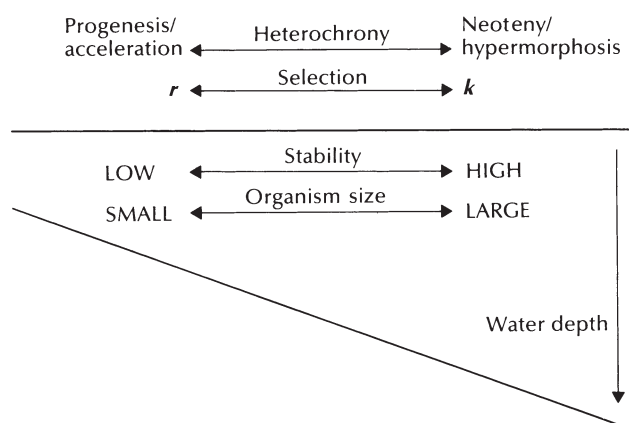


Fig. 6 Diagrammatic summary of McKinney's (1986) proposed relationship between water depth, ecological stability, selection, organism size and heterochronic process, based on a suite of Tertiary echinoids from southeastern U.S.A.

Frequency of heterochrony in the fossil record

Any attempt to assess the frequency of heterochrony or of the heterochronic processes is fraught with problems, not the least of which are historical prejudices. While the Haeckelian school were blinkered to the existence of paedomorphosis, the Garstang/de Beer school were equally contemptuous of peramorphosis. In a recent survey of palaeontological literature from 1976 to 1985, McNamara (in McKinney 1988) documented 272 examples of heterochrony; of these, 179 were of paedomorphosis, the remaining 93 were of peramorphosis.

The most comprehensive recent analyses of heterochrony in the fossil record have centred on trilobites, echinoids, ammonites, bryozoans, and graptolites. These studies have shown both paedomorphosis and peramorphosis to be important factors, but paedomorphosis still predominates. The greater frequency of paedomorphosis, if true, may occur because existing developmental programmes are utilized. Peramorphosis always requires the production of novel bauplans by extending the pre-existing developmental pathways.

Trilobites show a changing relative frequency of paedomorphosis and peramorphosis. During the Cambrian paedomorphosis, particularly that induced by progenesis, was predominant. Post-Cambrian forms, however, show a marked decline in the incidence of progenesis and a greater frequency of peramorphosis. It has been argued (McNamara 1986b) that this change may reflect an improvement in regulation of the developmental system in later trilobites.

Recent analysis of heterochrony in irregular echinoids (McNamara 1988) has highlighted the complex activity of heterochrony in single lineages, some characters being paedomorphic, others peramorphic. Furthermore, one paedomorphic structure might have evolved by neoteny, whilst another might have formed by post-displacement. The operation of a complex array of heterochronic processes has been termed *mosaic heterochrony*. With each structural element of an organism essentially following its own ontogenetic trajectory, and each being potentially subject to changes in developmental regulation, there is the possibility of the evolution of a multitude of heterochronic morphotypes. Any one of these may potentially form a new species, with the target of selection being the resultant morphotypes, size, or life history strategies.

Ammonites have featured prominently in studies of heterochrony for over 100 years. They were used initially as examples of 'recapitulation', by Hyatt and co-workers; while to O. Schindewolf and other workers in the first half of this century they showed evidence only of paedomorphosis. Recent research (see McKinney 1988) has shown the ubiquity of both phenomena, but peramorphosis, particularly of the septa, appears more common than paedomorphosis.

Colonial organisms, such as bryozoans and graptolites, show a two-tiered heterochronic pattern. Both the individual animals and the colony as a whole may be affected by heterochrony. The former is known as *ontogenetic heterochrony*, the latter as *astogenetic heterochrony*. This two-tiered structure is comparable with the two-tiered structure of differentiative and mitotic heterochrony present in non-colonial organisms. Astogenetic heterochrony has been reported more often than ontogenetic heterochrony (McKinney 1988), perhaps because astogenetic changes reflect developmental modification of ontogenetic characters, so reflecting the individuality of the colony as a whole. Heterochrony in colonial organisms may have been important in macroevolution. Ontogenetic heterochrony in highly integrated colonies may result in large morphological differences between ancestor and descendant. It would appear that periods of reef building correspond to periods of high integration in colonial animals. It is likely, therefore, that astogenetic heterochrony will predominate during periods of reef building.

Many of the examples of heterochrony involving vertebrates occur in amphibians. Most of these show paedomorphosis, akin to that seen in living

salamanders. Many of the interspecific and inter-generic differences in allometry of skull plates in Palaeozoic fishes are the result of mitotic heterochrony, though few studies have actually couched it in these terms. Similarly, phylogenetic changes in limb allometries in mammals are due to heterochrony. The limited evidence from studies of mammals seems to suggest a predominance of peramorphosis over paedomorphosis. This may occur because of the frequent operation of Cope's Rule in mammal lineages, suggesting size as being an important target of selection. For example, extension of ancestral allometries by increased size in horses through the Tertiary resulted in peramorphic descendants by hypermorphosis. However, some characters, such as development of the foot, show paedomorphic reduction in some digits.

Developmental processes underlying heterochrony

Changes to the onset, offset and rate of growth of morphological characters are essentially under three interactive levels of control: genetic, hormonal, and cellular. Perturbations to the genetic regulation of hormonal and cellular development, particularly at early embryological stages, are likely to be critical factors in heterochrony.

Developmental regulation is not simply a matter of discrete entities called 'regulatory genes' acting upon 'structural genes'. It involves a complex interaction between active sites or structural components of proteins, combined with cell-cell interactions (Campbell & Day 1987). Developmental processes are controlled by highly organized, dynamically structured multigene families. The manner in which the genome is encoded and expressed in development is far from clear, although it would seem that only a small area of the highly dynamic, constantly changing genome is occupied by genes for development.

The region involved in regulation in a typical eukaryote gene is the promoter region. This contains DNA binding proteins specific to the gene, and capable of controlling the level of transcription. The role of the promoter sequence in gene control, and its effect on growth, highlights the activity of hormones in growth, and how perturbations to the genetic control of hormone production can have a strong phenotypic expression.

Growth, moulting, and sexual reproduction in arthropods, for instance, are all under hormonal control. It has been suggested (Campbell & Day

1987) that hexapods evolved from a myriapodous ancestor by progenesis: a small change in the genetic control of the hormone responsible for the inception of maturation, and of the hormone controlling post-larval development, had a profound effect on the phenotype. Even within fossil lineages the activity of genes controlling hormone production can be inferred. Two forms of progenesis in trilobites have been identified: *sequential* and *terminal* (McNamara 1986b). Terminal progenesis is likely to have occurred by a premature cessation in production of a juvenile hormone. Sequential progenesis, where each intermoult period is shortened, is thought to have occurred by premature production of an ecdysone-like moulting hormone during each intermoult period. This premature hormonal activity will have been under direct genetic control.

The third factor in the developmental processes that cause heterochrony is activity at the cellular level. Hall (1984) has stressed the importance of the number and mitotic activity of the cells in the initial skeletal condensation in vertebrates. Thus onset of growth is determined by the number of stem cells that start condensation, the proportion that divides, rate of cell division, and amount of cell death. These parameters all act early in development and determine the time of onset of growth. The rate of growth of skeletal elements is influenced by adjacent tissues, hormones, and allometric factors. Muscle action, tendon insertion, blood flow, innervation, and growth of adjacent tissues modulate the growth rate. Cessation of growth is partially determined very early in development by the number of growth plate cells and the number of times they divide. Timing of development of secondary ossification centres also affects the offset signal. Metabolic inhibition by production of a growth inhibitor to suppress cell proliferation and protein synthesis also stops growth.

References

- Alberch, P., Gould, S.J., Oster, G.F. & Wake, D.B. 1979. Size and shape in ontogeny and phylogeny. *Paleobiology* 5, 296–317.
- Campbell, K.S.W. & Day, M.F. (eds) 1987. *Rates of evolution*. Allen & Unwin, London.
- Gould, S.J. 1977. *Ontogeny and phylogeny*. Belknap Press, Cambridge.
- Hall, B.K. 1984. Developmental processes underlying heterochrony as an evolutionary mechanism. *Canadian Journal of Zoology* 62, 1–7.
- McKinney, M.L. 1986. Ecological causation of heterochrony:

- test and implications for evolutionary theory. *Paleobiology* **12**, 282–289.
- McKinney, M.L. (ed.) 1988. *Heterochrony in evolution: a multi-disciplinary approach*. Plenum Press, New York.
- McKinney, M.L. & McNamara, K.J. 1991. *Heterochrony: the evolution of ontogeny*. Plenum Press, New York.
- McNamara, K.J. 1982. Heterochrony and phylogenetic trends. *Paleobiology* **8**, 130–142.
- McNamara, K.J. 1983. The earliest *Tegulorhynchia* (Brachiopoda: Rhynchonellida) and its evolutionary significance. *Journal of Paleontology* **57**, 461–473.
- McNamara, K.J. 1985. Taxonomy and evolution of the Cainozoic spatangoid echinoid *Protenaster*. *Palaeontology* **28**, 311–330.
- McNamara, K.J. 1986a. A guide to the nomenclature of heterochrony. *Journal of Paleontology* **60**, 4–13.
- McNamara, K.J. 1986b. The role of heterochrony in the evolution of Cambrian trilobites. *Biological Reviews* **61**, 121–156.
- McNamara, K.J. 1988. Heterochrony and the evolution of echinoids. In: C.R.C. Paul & A.B. Smith (eds) *Echinoderm phylogeny and evolutionary biology*. Oxford University Press, Oxford.
- Shea, B.T. 1983. Allometry and heterochrony in the African apes. *American Journal of Physical Anthropology* **62**, 275–289.

2.5 Red Queen Hypothesis

M. J. BENTON

Introduction

Palaeontologists have long argued that the distinctive features of the evolution of life were produced by changes in the physical environment. Changes in climate, or in sea-level, for instance, might explain why certain groups died out, or why an adaptive radiation took place at a particular time. This trend has continued in recent research into mass extinctions (Section 2.12), whether their cause is said to be changes in the earthbound physical environment, or the impact of asteroids.

On the other hand, many ecologists have viewed the large-scale aspects of evolution (macroevolution) as simply a scaled-up version of microevolution. Evolutionary change, they argue, can be produced by competition between organisms, and by interactions between predators and prey. This ecological view stresses the influence of the biotic environment, that is, other plants and animals, on evolution.

Van Valen's Law

The ecological view of macroevolution was codified by Van Valen (1973), who presented palaeontological and ecological evidence for a model of evolution that depended on the biotic environment, and termed the model the *Red Queen Hypothesis*. The palaeontological evidence was based on a study of

the rates at which different groups of plants and animals go extinct through time. Van Valen used plots of species survivorship (Fig. 1) which showed the proportions of an original sample of organisms that survive for various intervals. He found, contrary to his expectations, that the probability of extinction within any group remained constant through time — his *Law of Constant Extinction*. For example, families or species of modern mammals are just as likely to become extinct as were their Mesozoic ancestors living 200 Ma. A species might disappear at any time, irrespective of how long it has already existed. Evolutionary biologists might have intuitively expected species within any group to become longer-lived over time on average. Van Valen's startling discovery seemed to deny some basic assumptions of evolution. If evolution is taken to mean improvement in the adaptation of a species to its environment through time, why is it that modern mammals are not better at surviving than their Mesozoic forebears?

Van Valen's explanation for the Law of Constant Extinction was that the various species within a community maintain constant ecological relationships relative to each other, and that these interactions are themselves evolving. Thus, the antelope on an African savanna, for example, evolves greater speed in order to escape from the lion, but the lion

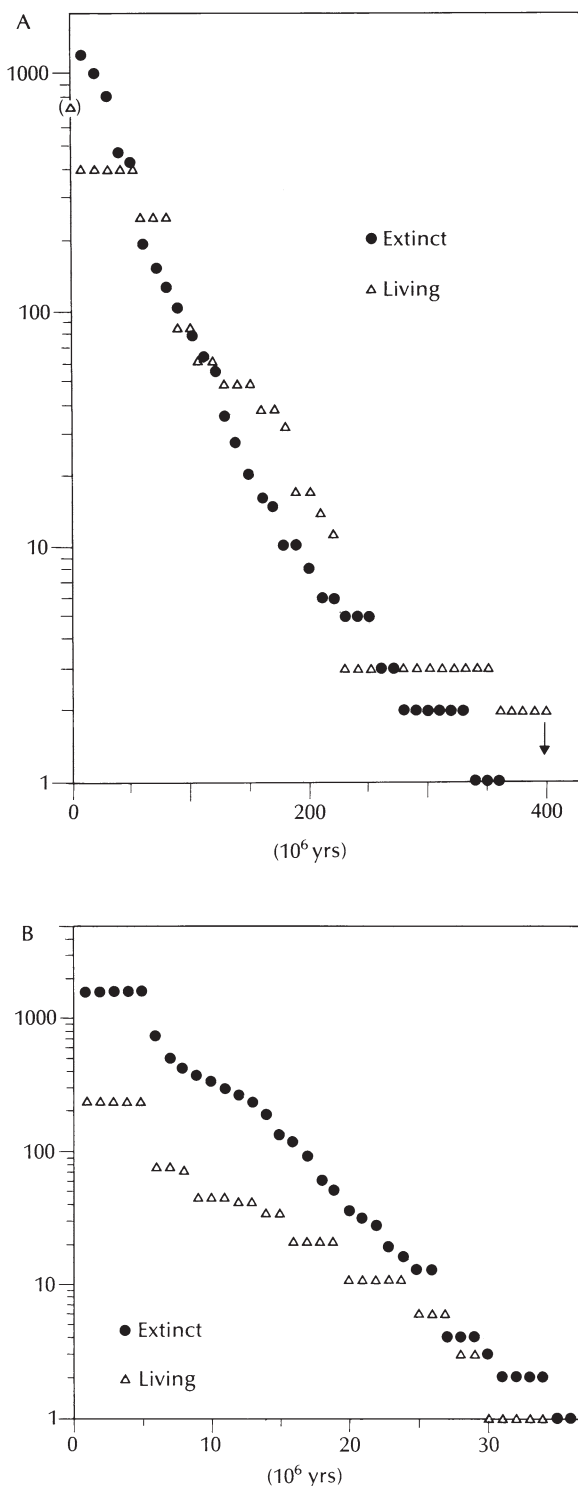


Fig. 1 Taxonomic survivorship curves for A, pelecypod (bivalve) and B, mammalian genera (major Therian orders). These graphs show roughly linear trends, and indicate constant probabilities of extinction for genera of any duration. (After Van Valen 1973.)

also evolves greater speed in order to catch its dinner. The status quo is maintained. If biotic interactions did not follow a pattern of ever-moving dynamic equilibrium, the community would be shattered. Were all antelopes to achieve a quantum improvement in their running speed, lions would starve and populations of antelope might outstrip the carrying capacity of the environment. This balance is the Red Queen Hypothesis. [In Lewis Carroll's *Through the Looking Glass*, the Red Queen told Alice, 'Now here, you see, it takes all the running you can do, to keep in the same place'.]

After 1973, many biologists accepted the Red Queen model, while others were critical. The main problem was simply the counter-intuitive claim that species do not improve their chances of survival through time. If organisms are continuously evolving and adapting, why do they not get any better, on average, at avoiding extinction? There were problems also with Van Valen's particular formulation of the Red Queen model. He explicitly made a zero-sum assumption: that there were fixed amounts of energy available to communities, and that any gain by one species was exactly offset by equal losses to others. It is not at all clear, however, that the amounts of energy, or resources in general, have remained constant through time. It is equally probable that the total global biomass has increased markedly many times as major new habitats were exploited (e.g. the move onto land (Section 1.8), the evolution of 'trees', the origin of flight (Section 1.9), and the evolution of deep-burrowing habits by various marine groups (Section 1.7.1)). The increase in biomass is possible by pulling more of the global carbon into the biotic part of the biogeochemical cycle, and/or by speeding up the rate at which carbon, and other essential elements, are cycled through the system (Benton 1987).

Problems with 'progress', and the Stationary Model

Although the Red Queen model does not predict improvements in the ability to avoid extinction, it does explicitly assume that, within any lineage, later members will be competitively superior to earlier ones: that a present-day antelope can run faster than its Pliocene forbear, that modern mammals are clearly competitively superior to their Palaeocene ancestors. This notion of progress is frequently assumed by biologists and palaeontologists, but is probably impossible to test directly. Nevertheless, simple assumptions of progress of this kind have

been criticized recently (Benton 1987). There seems to be no adequate way yet of demonstrating 'progress' in macroevolution, least of all competitive improvement. The evolution of horses can be taken as a well known example of an adaptive trend, or record of improvement through time. The early small leaf-eating horses of Eocene times gave way to larger animals with fewer toes (greater running speed) and deeper teeth (for grinding up the new silica-bearing grasses) in the Miocene (Section 1.11). However, if the fossil record were reversed, we could equally well demonstrate how the horses adapted to the diminishing grasslands by becoming smaller forest-dwellers, living a cryptic life and switching to a diet of tree leaves. Where is the progressive improvement of competitive ability? The whole question seems to hinge on how macroevolution is viewed. If organisms are generally very well adapted, finely tuned by natural selection, and if the physical environment has only minor effects, the Red Queen Hypothesis has to hold. If, on the other hand, organisms are viewed as only moderately well adapted, natural selection as only a sporadic force for evolutionary change, and the physical environment as an important influence through local and global extinction, and radiation events, then the Red Queen Hypothesis cannot be correct.

In 1984 Stenseth and Maynard Smith formalized an alternative to the Red Queen model, termed the *Stationary Model*. This model assumes that evolution is driven mainly by abiotic factors, and that it will cease in the absence of changes in the physical environment. The two models make very different predictions and, as Stenseth and Maynard Smith (1984) wrote, 'the choice between the Red Queen and Stationary Models will have to depend primarily on paleontological evidence'.

The Red Queen model predicts that the rates of speciation, extinction, and phyletic evolution will remain constant in ecosystems, even when the diversity of species has reached equilibrium so that the numbers of species do not change. The Stationary Model, however, predicts that at equilibrium no evolution will occur. Bursts of evolution, extinction, and speciation will happen only in response to changes in the physical environment. These two models can be visualized by plots of species survivorship over time, which gives a measure of the rate of extinction (Fig. 2A, B).

Testing the models

Hoffman and Kitchell (1984) applied a palaeonto-

logical test. The first problem they encountered was to find an example spanning several million years in which no environmental change had occurred. Such a case is highly unlikely, and it proved necessary to make allowances for episodic perturbations in the physical environment. The modified patterns are still distinctive (Fig. 2C, D). The Red Queen model predicts an approximately regular decline in the number of species surviving (that is, constant extinction), with occasional changes of slope that correspond to major environmental perturbations. The Stationary Model predicts a distinctly stepped pattern, with constant numbers of species at equilibrium, and sudden extinctions at times of environmental change.

Hoffman and Kitchell (1984) also examined the records of microfossils (coccoliths, foraminiferans, radiolarians, diatoms, and others) from 111 deep-sea boreholes through the past 50 million years of sediments of the Pacific Ocean floor. The species survivorship curves obtained from these data (Fig. 3) are more or less smooth, rather than stepped, and they seem to support the Red Queen model. An analysis of the cumulative appearance of new species also gave general support to the Red Queen model, although there was some evidence of stepping.

Further analysis shows there to be considerable variation in the probability of extinction over geological time: for example, there seem to have been particular periods in which all the microfossil groups had high extinction rates. These indicate plankton extinction events which would normally be attributed to sharp changes in the physical environment. When Hoffman & Kitchell (1984) made allowances for these events, the various analyses again pointed to the Red Queen model.

Another test, also using the plankton record, was carried out by Wei & Kennett (1983). Their study was based on the fossil record throughout the world of 149 species of foraminifera over the past 24 million years. They found that major changes in rates of extinction and speciation corresponded to palaeoceanographic perturbations (Fig. 4), and they regarded their data as consistent with the Stationary Model.

These two studies illustrate some of the practical difficulties involved in testing the Red Queen model. One serious problem is in separating biotic from abiotic factors in order to assess their relative significance: it is probably impossible to pigeon-hole both kinds of phenomena as independent factors. Secondly, in many real situations, and pos-

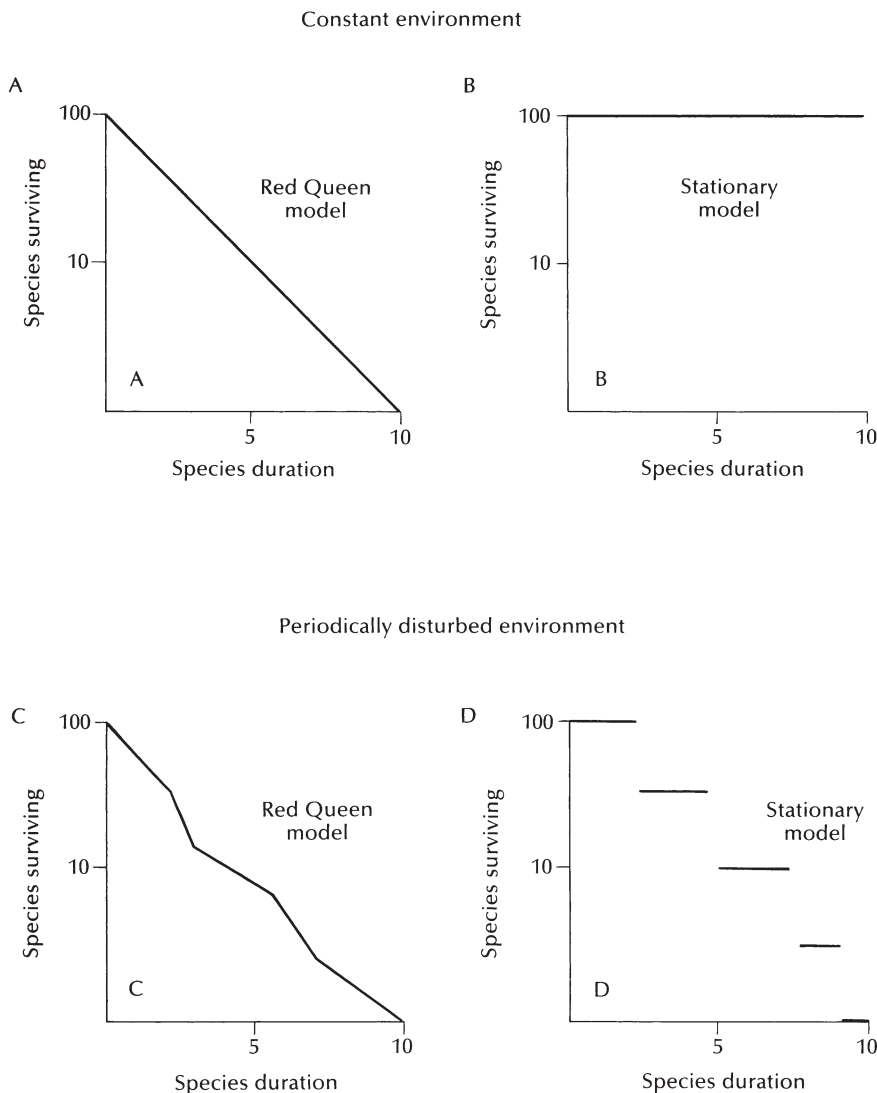


Fig. 2 Predictions of the Red Queen and Stationary models for species survivorship in constant (A, B), and periodically disturbed (C, D) environments. (After Hoffman & Kitchell 1984.)

sibly including the two described above, the test would be inconclusive. For example, it would be hard to distinguish between two predicted patterns if both curves were stepped as a result of rapid changes in the physical environment. At the other extreme, continuous small changes in environmental conditions might give similar gently sloping graphs for both models.

Furthermore, as Hoffman and Kitchell (1984) pointed out, both the Red Queen and Stationary models assume that ecosystems tend towards a diversity of species that is at equilibrium. Both hypotheses predict that the total number of species is constant: the addition of one species to the system causes the loss of another. Yet models that do not assume such an equilibrium could also account for the Law of Constant Extinction and for the other palaeontological data. In these models, there would

be no necessary limit to diversity, and the rates at which species arise or go extinct would not be correlated with each other, nor with total diversity. The data so far are equivocal on these points.

Other research seems to count against the Red Queen Hypothesis. Kitchell *et al.* (1989) studied a simple predator–prey relationship: naticid gastropods and bivalves. The naticid gastropods prey on bivalves by boring through their shells and extracting the flesh. Both groups are plentiful as fossils, and such predation has left identifiable borings in fossilized bivalve shells. Kitchell and her colleagues modelled the predator–prey system mathematically and found that, whatever the starting point, the system tended to a static position. The bivalves evolved either to reproduce early (before they were eaten) or to devote all of their energy to building a thick shell to minimize the chance of successful

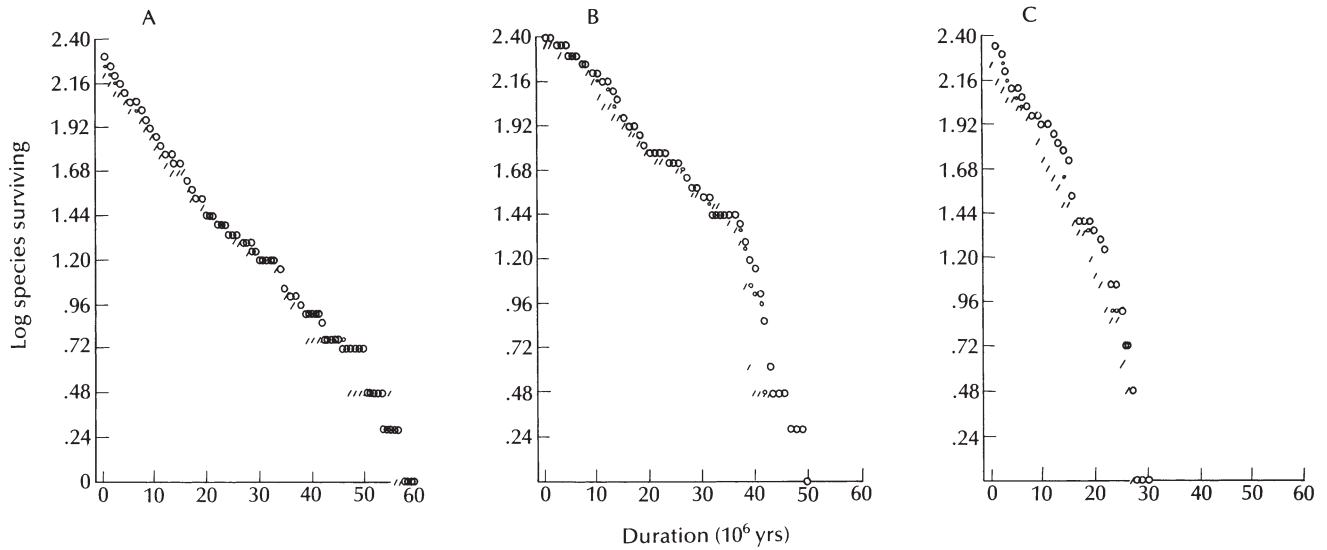


Fig. 3 Species survivorship curves for A, coccoliths; B, radiolarians; and C, planktic foraminifera. Each graph contains data for three cohorts — taxa arising in the earliest Pliocene, in the latest Pliocene, and in the Holocene, and each is represented by a different symbol. (After Hoffman & Kitchell 1984.)

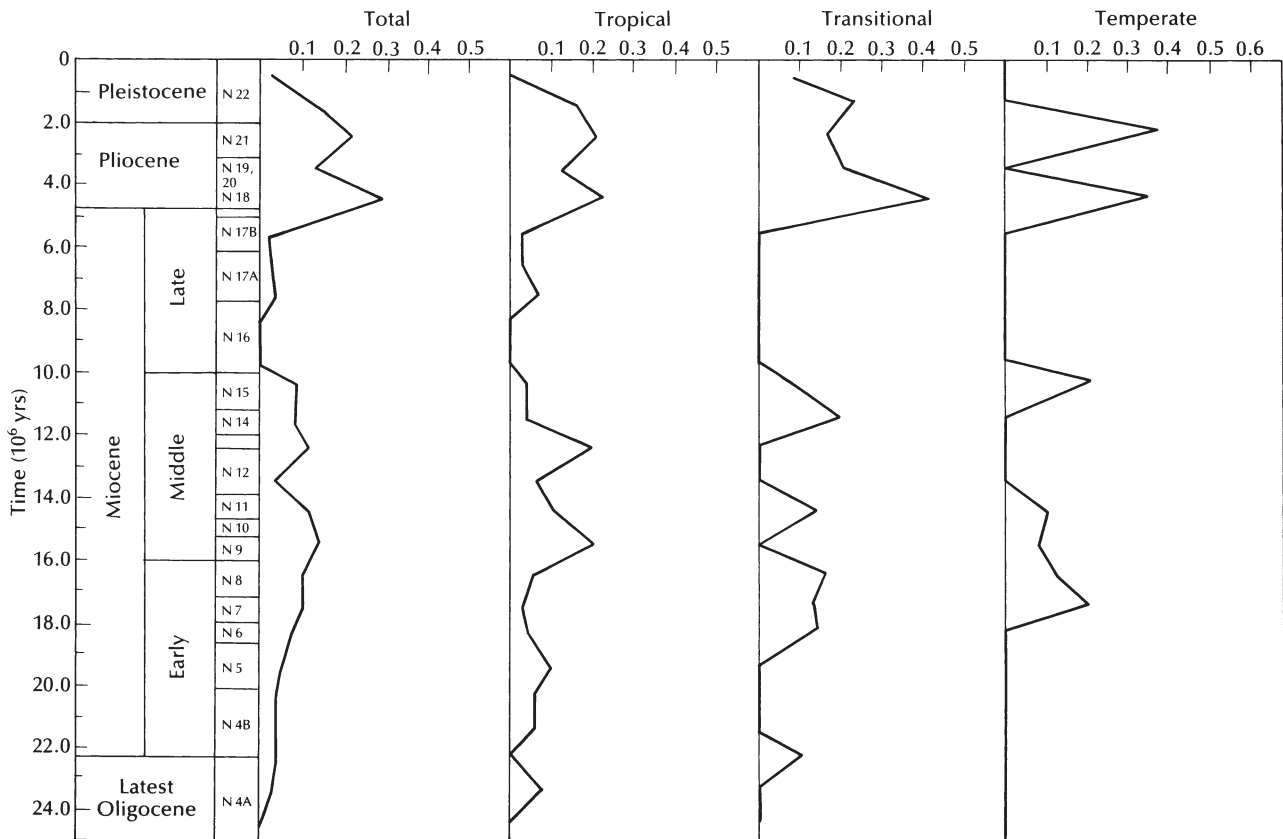


Fig. 4 Extinction rates for species of Neogene foraminifera for the whole of the oceans ('total'), and for tropical, transitional, and temperate assemblages, for the past 25 million years (latest Oligocene to Recent). The 'N' zones are standard Neogene planktic foraminiferal zones. Peaks in extinction are correlated globally, and linked to major changes in the physical environment. (After Wei & Kennett 1983.)

boring attacks. This result speaks against the Red Queen model, which would require constant evolution in a particular direction.

Other biologists have argued that species probably do not keep running towards unattainable goals, as the Red Queen Hypothesis predicts. Each species is faced with the need to make compromises. Many bivalves, for example, have to balance the need for a strong shell against the costs of a heavy shell. The compromise solution is to have a thin corrugated shell. 'Constant running' in one direction is often not possible in a lineage, and the simplistic view of the Red Queen model as continuous and endless evolution in one direction may be denied by the limitations of genetic variation, development, and mechanical design factors.

References

- Benton, M.J. 1987. Progress and competition in macroevolution. *Biological Reviews* **62**, 305–338.
- Hoffman, A. & Kitchell, J.A. 1984. Evolution in a pelagic planktic system: a paleobiologic test of models of multi-species evolution. *Paleobiology* **10**, 9–33.
- Kitchell, J.A., DeAngelis, D.L. & Post, W.D. 1989. Predator–prey interactions on the ecological and evolutionary time scale. In: N.C. Stenseth (ed.) *Coevolution in ecosystems and the Red Queen Hypothesis*. Cambridge University Press, Cambridge.
- Stenseth, N.C. & Maynard Smith, J. 1984. Coevolution in ecosystems: Red Queen evolution or stasis? *Evolution* **38**, 870–880.
- Van Valen, L.M. 1973. A new evolutionary law. *Evolutionary Theory* **1**, 1–30.
- Wei, K-Y. & Kennett, J.P. 1983. Nonconstant extinction rates of Neogene planktonic foraminifera. *Nature* **305**, 218–220.

2.6 Hierarchy and Macroevolution

N. ELDREDGE

Introduction

Evolution is the scientific explanation of the design apparent in organismic nature. Natural selection is generally seen to be the principal cause of deterministic modification of the phenotypic properties of organisms through time. Macroevolution most commonly connotes the degree of such modification, and thus in its most general sense is simply 'large-scale genotypic/phenotypic change'. Microevolution, in contrast, refers to the relatively slight amount of change that occurs on a generation-by-generation basis through natural selection and genetic drift. A connotation of elapsed time is often implicit in the distinction between micro- and macroevolution: microevolution takes place in relatively short amounts of time (e.g. 'ecological time,' over a few generations), while macroevolution is generally held to occur in geological time. Yet some theories of macroevolution (e.g. saltation theories of the geneticist R. Goldschmidt, or the palaeontologist O. Schindewolf) invoke brief (even single generation) genotypic and phenotypic transformation. A further distinction commonly drawn

between 'microevolution' and 'macroevolution' sees the former as a within-species or perhaps within-genus level phenomenon, in contrast with the degree of change associated more typically with the emergence of taxa of higher categorical rank (i.e. in the Linnaean hierarchy — families, orders, classes, etc.).

The 'evolutionary synthesis', dating from the mid nineteen-thirties, forms the core of modern evolutionary theory. The synthesis followed the successful fusion of Darwinian selection with an emerging understanding of the principles of heredity (achieved primarily through the efforts of geneticists R.A. Fisher, J.B.S. Haldane and S. Wright). This neo-Darwinian paradigm of drift- and selection-mediated dynamics of genetic stasis and change was then integrated with the data of systematics, palaeontology, and other biological subdisciplines to form what was widely heralded as a unified theory of evolution. It is the general position of the synthesis that 'macroevolution' is simply microevolution summed over geological time. Specifically, generation-by-generation stability and transformation, mediated by natural selection and genetic

drift, were held to be *both necessary and sufficient* to account for all aspects of the evolutionary history of life. Though the geneticist Dobzhansky (1937) and the systematist Mayr (1942) both sketched versions of macroevolutionary theory, it was left primarily to the palaeontologist Simpson (1944, 1953) to formulate the principles of macroevolution within the synthetic theory.

The assumption that microevolution yields a complete account of the evolutionary process when projected over evolutionary (geological) time restricts the study of evolutionary mechanics to laboratory and field investigations of living organisms. The role of palaeobiology in such a scheme, however, is by no means thereby rendered trivial: as Simpson (1944), for example, endeavoured to show, the integration of evolutionary theory with patterns of evolutionary events drawn from the fossil record is no simple matter. In particular, Simpson was concerned to show that it is the task of palaeontology to determine the relative intensities, and importance, of various microevolutionary processes (e.g. mutation rate, selection, population size, etc.) required to explain various evolutionary patterns of the fossil record. In that spirit, Simpson developed his model of 'quantum evolution' (rapid, 'all-or-nothing' modification of adaptive features of organisms in relatively small populations) to explain the relatively abrupt appearance so typical of many higher taxa.

Recent years have seen an alternative view emerge on the relationship between palaeontological data, geological time, and theories of the evolutionary process. In traditional evolutionary biology, it is the phenotypic (and underlying genetic) properties of organisms that are of central interest, and which 'evolve'. Organisms vary in these respects within local populations; populations are aggregated into species (Section 2.2). Natural selection 'sorts' the phenotypic attributes of organisms within populations to yield (1) stasis or change in phenotypes, and (2) the emergence of new species (and, by simple extension, higher taxa).

'Hierarchy theory' accepts the neo-Darwinian paradigm of within-population variation, selection, and drift, but seeks to extend the list of evolutionary entities beyond genes, organisms, and populations. Specifically, species, monophyletic (higher) taxa, and ecosystems have come to be viewed as having real existence, and are variously termed 'systems', 'entities', or even 'individuals'. The goal of hierarchy analysis is to elucidate the nature of each kind of large-scale entity, and thus to determine their pos-

sible role(s) in the evolutionary process.

If large-scale systems such as species, higher taxa, and ecosystems are real entities, they exist on a spatiotemporal scale which is too large to be encompassed in laboratory and field experimental studies of the Recent biota. It is the fossil record that reveals the actual dimensions of such systems, and thus it falls in large measure to palaeobiology to examine how they can be integrated with existing theories of the evolutionary process. Such work has two aspects: (1) the determination of any relevance of such large-scale systems to the original problem of evolution — that is, the origin, maintenance, and further transformation of adaptive phenotypic features of organisms; and (2) the recognition of other effects on the general history of life that may result from the existence of such larger-scale entities.

Specifically, the concept that large-scale systems such as species, taxa, and ecosystems are themselves entities, not merely epiphenomena or simple (and perhaps arbitrarily delineated) collectivities of organisms, has led to several palaeobiological theories that allege a degree of additional process to macroevolution, over and above — and in some instances 'decoupled' from — microevolutionary processes.

Hierarchies in evolutionary biology

Several meanings of the term 'hierarchy' are in general use in biology (Greene 1987). In the context of evolutionary theory, however, only two hierarchical systems are generally recognized: the *genealogical* and *ecological* (economic) hierarchies (Table 1). Both are thought to be implicated in the evolutionary process (Eldredge 1985, 1986, 1989; Salthe 1985), though some authors recognize one hierarchy and not the other. Both hierarchies consist of nested sets of entities forming distinct levels. Each level constitutes a class (or category — e.g. 'species'), specific examples of which are entities or 'individuals' (e.g. *Archaeopteryx lithographica*). The entities of any given level have as parts the entities of the adjacent lower level and form, in turn, parts of the adjacent higher level: demes have organisms as parts; in turn, demes are parts of species.

The entities at each level interact or behave in specific ways that unite them to form the entities of the next higher level. In the genealogical hierarchy the activity is 'reproduction' in the most general sense; thus, speciation is seen as the production of more entities (i.e. species) of like kind — an activity ultimately responsible for the ongoing existence of

Table 1. The genealogical and ecological hierarchies.

Genealogical hierarchy	Ecological hierarchy
Monophyletic taxa	Biosphere
Species	Ecosystems
Demes	Avatars (populations)
Organisms	Organisms
Germ line ^a	Soma ^b

^a Composed of hierarchically nested chromosomes, genes, codons and base pairs.

^b Composed of hierarchically nested organ systems, organs, tissues, cells and proteins.

higher taxa. In the economic hierarchy, direct interaction among entities of any given level cohere the entities of the adjacent higher level; thus it is the interaction among local populations of non-conspecifics (as in predator–prey interactions) that unites them into local ecosystems.

The two hierarchies arise out of the two types of organismic activity, that is reproduction, on the one hand, and processes related to matter–energy transfer on the other. Viewed in this light, Darwin's distinction between *sexual* and *natural selection* is clear. In sexual selection, relative reproductive success arises strictly from among-population variation in some aspect of reproductive behaviour, physiology, or anatomy. In natural selection, an organism's relative success in economic (matter–energy transfer) activities has an effect on that organism's probability of successful reproduction.

The two hierarchies are direct outgrowths of these two distinct categories of adaptation that arise under sexual and natural selection. In sexual organisms, reproduction implies a local pool of suitable partners — a 'deme'. In most instances, there will be pools of suitable partners elsewhere; thus local demes form regional 'species'. Most modern treatments of species recognize them as reproductive communities, within which mating occurs, outside which it does not. Paterson (e.g. 1985) recently suggested that species are reproductive communities composed of organisms sharing a particular set of reproductive adaptations, or 'specific mate recognition system' ('SMRS'). His concept obviates the ambiguity of disjunct distributions, where potential mates never meet. Moreover, because the SMRS is an adaptive system subject to (sexual) selection (favouring mate recognition in isolation) speciation minimally must entail (presumably allopatric) divergence of the SMRS. Speciation is seen

as an outgrowth simply of continued reproduction in isolation, leading to modification of the SMRS. Because new (sexual) species arise in this fashion as a matter of course, higher taxa are maintained (as long as speciation rate exceeds extinction rate). Higher taxa are seen strictly as lineages of species; they are recognized (just as are clones of strictly asexual organisms) only when new adaptations ('synapomorphies' of phylogenetic systematics; Section 5.2.2) arise and serve as markers for the lineage. As such, monophyletic taxa do not 'reproduce', that is, they do not produce additional entities of like kind. Genera do not give rise to new genera, the way that new species arise from old.

The economic activities of organisms of a species lead them to form local populations ('avatars') which may, but need not, be coextensive with local demes of the same species. But above this level (Table 1), a crucial distinction between the genealogical and economic hierarchies arises. Whereas the reproductive adaptations of organisms are shared by organisms in other demes elsewhere, the economic adaptations of organisms lead to *cross-genealogical* interactions between local populations belonging to different species. Local ecosystems interact with other such systems on a regional scale, but maps of genealogical systems and economic systems simply do not coincide. It is especially significant that *species are not parts of economic systems*. Thus, by sheer dint of the existence of two classes of organismic activity — hence adaptations — organisms are simultaneously parts of two separate, hierarchically arranged systems. And in particular, interaction within and between entities of the two different hierarchies is of the greatest importance in elucidating a full causal theory of the evolutionary process.

The evolutionary process: role of the genealogical and economic hierarchies

Discussions of macroevolution traditionally emphasize the origin of higher taxa in the context of large-scale adaptive change. Under this synthesis, linear trends are often said to be generated by 'orthoselection', i.e. long-term, predominantly directional natural selection, as distinct from 'orthogenesis', or linear phyletic change through unspecified causes internal to organisms. In general, the accumulation of significant amounts of adaptive transformation within a lineage has been termed *anagenesis*, which is commonly, if not invariably, held to be a process distinct from *cladogenesis*, or lineage splitting. Thus

much, if not all, macroevolutionary change has traditionally been considered to occur without any (or any significant) degree of speciation. A major exception to this generalization is the theme of *adaptive radiations*, in which morphological transformation proceeds rapidly and independently in several or many different directions, and lineage-splitting is directly invoked as part of the process. Simpson's (1944) earliest formulation of 'quantum evolution' also invoked lineage splitting (though not expressly termed 'speciation'); later (Simpson 1953) modified in favour of a purely phyletic conceptualization of quantum evolution.

The hypothesis of *punctuated equilibria* (Eldredge & Gould 1972) is based, in part, on the empirical claim that most species exhibit relative morphological stability throughout the bulk of their stratigraphic ranges (see also Section 2.3). Thus most anatomical change appears to occur along with speciation. Such species stability facilitates recognition of species as spatiotemporally-bounded entities; it further leads to the postulate that linear trends in macroevolution may reflect processes of *species sorting* in addition to directional natural selection. In general terms, such a model proposes that actual transformation of morphology occurs *via* directional natural selection (plus, perhaps, genetic drift) on a standard generation-by-generation basis. But the linearity of the trend through long periods of time — when the species remain morphologically stable and vary among themselves with respect to the evolving trait — arises through sorting of variation among species through a variety of potential causes.

The term 'species selection' itself embraces a number of variant conceptualizations. As developed as an outgrowth of *punctuated equilibria* (Eldredge & Gould 1972); the term itself was introduced by Stanley 1975; see also Stanley 1979), 'species selection' was virtually synonymous with the more general term 'species sorting' used here. Subsequent authors, seeking more precise parallel usage between organismic and higher-level selection, contend that 'species selection' is applicable only to species-level properties of species (*cf.* Vrba 1984). This argument holds that phenotypic (and underlying genotypic) properties of organisms are the focus of organismic selection. True species selection should be invoked only to explain species-level adaptations; it cannot logically be applied to the situation in which species differ merely in the frequencies of one or more *organismic phenotypic* traits. Williams (1966) was the first to argue this

point, claiming that 'group selection' can pertain only to group-level adaptations. Jablonski (1987) argued that geographical ranges are species-level properties, and show high heritability in his data on Cretaceous molluscs; he concluded that species ranges are therefore subject to true 'species selection'.

Hull (1980) has discussed two components that must be present for selection to occur at any level; these two components serve in addition as criteria for evaluating claims of species-level selection. According to Hull (1980), among entities involved in any instance of selection, there must be an *interactor* as well as a *replicator*. The relative success of interactors is recorded in the subsequent representation of their underlying replicators. Thus, in natural selection, relative economic success of organisms will affect their relative reproductive success, and hence the frequencies of the underlying genotypes. Organisms in this instance are both interactors and 'reproducers', with replicative fidelity supplied by their genes.

Hull's (1980) selection criteria imply that species selection cannot be directly analogous to natural selection. If species are genealogical entities (if, in other words, it is the reproductive activities of organisms that lead to the formation and continued existence of species), species are causally connected to the replicative activities of genes; but if it is further true that species, as whole entities, do not play direct roles in ecological systems, species cannot be said to be interactors, and Hull's criteria for selection are not met *by definition* for species. 'Species selection' appears to be most analogous to Darwin's *sexual selection* — because factors affecting rates of speciation and species extinction are involved.

Species sorting is a function of differential extinction and origination of species within a monophyletic clade. It is the goal of macroevolutionary theory to specify the causal processes underlying such sorting. In addition to processes at work *within* a given level, the entities above and below in the hierarchy provide constraints (initial, or boundary conditions) on processes occurring within any given level — the 'upward and downward causation' of many hierarchy theorists. Vrba's (1980) 'effect hypothesis' is an example of *upward* causation within the genealogical hierarchy. The effect hypothesis postulates that macroevolutionary patterns (for example, linear trends in one or more morphological attributes within a clade over geological time) may arise simply as an outgrowth (side effect) of the

biology of the organisms themselves. Nothing more — specifically, no selection at the species level — need be invoked as an explanation of such patterns.

Palaeontologists have sought links between characteristic speciation and extinction rates in lineages (macroevolutionary patterns), on the one hand, and aspects of organismic biology on the other — at least since Williams (1910) noted the apparent correlation between variation, niche width, and stratigraphic duration. Williams claimed that broadly niched (eurytopic) species, in addition to their characteristically wider geographical (habitat) occurrence, tend to display greater morphological variability (both within and certainly among populations) and longer stratigraphic durations than more narrowly niched stenotopes. Focusing especially on aspects of niche-width, macroevolutionary theorists have attempted to account for rates of both speciation and extinction. It seems, for example, that lineages comprised predominantly of eurytopic species show lower rates of species extinction and origination than lineages comprised of predominantly stenotopic species (see Eldredge & Cracraft 1980). The contrast is especially clear in sister-lineages. Indeed, Vrba (1980) used the Miocene–Recent sister lineages of Aepycerotini (impalas) and Alcelaphini (wildebeests, hartebeests, topis, etc.), the former species-poor and eurytopic, the latter speciose, with short-ranging stenotopic species, to illustrate one possible cause underlying the ‘effect hypothesis’ (see also Section 2.10). Vrba postulated that the trends in alcelaphine evolution were simply aggregates of higher rates of speciation and the accumulation of adaptive modification in the lineage of stenotopes — while little significant evolutionary transformation accumulated within the co-ordinate lineage of eurytopes, the aepycerotines.

Interhierarchic interaction and macroevolution

Darwin’s (1859) original formulation of natural selection (where relative economic success affects relative reproductive success within a local population of a species) serves as a model of the mechanics of interaction between the two hierarchies in the evolutionary process. Organisms, as members simultaneously in both the economic and genealogical hierarchies, patently stand as the prime causal link between the two (although some hierarchy theorists (notably Salthe 1985) see direct causal

interaction between entities at various levels of the two hierarchical systems).

Under the synthesis, species and higher taxa are generally depicted as having niches (or ‘adaptive zones’ in the case of higher taxa); further, in a widely used extension of Wright’s (1932) metaphor of ‘adaptive peaks’, species and higher taxa are generally depicted as occupying peaks, or series of adjacent peaks (i.e. in an ‘adaptive range’). Thus the most general approach to macroevolution under the synthesis holds that species and higher taxa are distinctly *economic* entities — effectively collapsing the dual hierarchy system into a single scheme.

Yet, following arguments outlined above, it has seemed to recent theorists that species and higher taxa are different sorts of entities from those that form complex biotic economic systems. Species are aggregates of local demes, all of which share a common fertilization system. From an ecological point of view, species are typically integrated into a variety of different ecosystems. Yet organisms within a species, as a rule, retain sufficient similarity in terms of economic adaptations that local populations are, to a great degree, redundant from one another. That is to say, the actual ecological role played by species is to serve as a reservoir of genetic information. Local populations are notoriously ephemeral; local extinction, on several geographical and temporal scales, is often counteracted by recruitment from neighbouring demes. An important consequence of the mere existence of species is that local parts of ecosystems are continually replenished from demes elsewhere. Recent studies of larval recruitment in intertidal communities — after events that range from slight to total disruption — amply bear out the role that species play as reservoirs of genetic information.

Darwin (1871) called species ‘permanent varieties’. The expression is apt in the context of macroevolution, because the complexion of ecosystems is forever modified upon the final extinction of a species; the possibility of replacing local populations with conspecifics is forever lost. In general, just as species display a within-species pattern of supply of organisms to replace local populations, following extinction events that result in the loss of many higher taxa, the identities of the surviving taxa determine the natures of the subsequently founded ecosystems. Disruption of ecosystems results in extinction — the more severe the disruption, the higher the characteristic level of disappearance of taxa, from species on up; and the higher the average level of taxonomic extinction,

the greater the change in economic systems.

Theories (e.g. the 'Red Queen Hypothesis', Section 2.5) often depict evolution as a process of inexorable adaptive change. Recent empirical and theoretical work in palaeobiology suggests rather a different picture: the ecological systems of which all organisms are parts, are formed from whatever organisms are extant at any given moment. With normal, small-scale fluctuations in composition and relative abundance of organisms, ecological systems appear to be quite stable. Speciation and extinction do occur, and so affect the composition of ecosystems. Some phyletic modification may accrue within species, but, because most demes are ephemeral, little net change typically accumulates within species throughout most of their histories. Little in the way of concerted evolutionary change, either within species or among species within lineages, tends to occur unless and until external perturbation disrupts ecosystems to the point where entire species — and higher taxa — become extinct, rendering impossible the resumption of ecosystems of the same composition as before. Thus, although the presence of genealogical entities (species and higher taxa — as packages of genetic information) are indispensable to the formation and ongoing existence of ecosystems, it appears that it is primarily the disruption of such economic systems that leads to significant amounts of change within entities of the genealogical hierarchy. Hence mass extinction appears to be an important causal cornerstone of macroevolution.

References

- Darwin, C. 1859. *On the origin of species*. J. Murray, London.
- Darwin, C. 1871. *The descent of man, and selection in relation to sex*. J. Murray, London.
- Dobzhansky, T. 1937. *Genetics and the origin of species*. Columbia University Press, New York.
- Eldredge, N. 1985. *Unfinished synthesis*. Oxford University Press, New York.
- Eldredge, N. 1986. Information, economics and evolution. *Annual Review of Ecology and Systematics* **17**, 351–369.
- Eldredge, N. 1989. *Macroevolutionary dynamics: species, niches and adaptive peaks*. McGraw-Hill, New York.
- Eldredge, N. & Cracraft, J. 1980. *Phylogenetic patterns and the evolutionary process*. Columbia University Press, New York.
- Eldredge, N. & Gould, S.J. 1972. Punctuated equilibria: an alternative to phyletic gradualism. In: T.J.M. Schopf (ed.) *Models in paleobiology*, pp. 82–115. Freeman, Cooper & Co., San Francisco.
- Greene, M. 1987. Hierarchies in biology. *American Scientist* **75**, 504–510.
- Hull, D.L. 1980. Individuality and selection. *Annual Review of Ecology and Systematics* **11**, 311–322.
- Jablonski, D. 1987. Heritability at the species level: analysis of geographic ranges of Cretaceous mollusks. *Science* **238**, 360–363.
- Mayr, E. 1942. *Systematics and the origin of species*. Columbia University Press, New York.
- Paterson, H.E.H. 1985. The recognition concept of species. In: E.S. Vrba (ed.) *Species and speciation*, pp. 21–29. Transvaal Museum Monograph No. 4. Transvaal Museum, Pretoria.
- Salthe, S.N. 1985. *Evolving hierarchical systems*. Columbia University Press, New York.
- Simpson, G.G. 1944. *Tempo and mode in evolution*. Columbia University Press, New York.
- Simpson, G.G. 1953. *The major features of evolution*. Columbia University Press, New York.
- Stanley, S.M. 1975. A theory of evolution above the species level. *Proceedings of the National Academy of Science* **72**, 646–650.
- Stanley, S.M. 1979. *Macroevolution: pattern and process*. W.H. Freeman, San Francisco.
- Vrba, E.S. 1980. Evolution, species and fossils: how does life evolve? *South African Journal of Science* **76**, 61–84.
- Vrba, E.S. 1984. What is species selection? *Systematic Zoology* **33**, 318–328.
- Williams, G.C. 1966. *Adaptation and natural selection*. Princeton University Press, New Jersey.
- Williams, H.S. 1910. The migration and shifting of Devonian faunas. *Popular Science Monthly* **77**, 70–77.
- Wright, S. 1932. The roles of mutation, inbreeding, cross-breeding and selection in evolution. *Proceedings of the VIth International Congress of Genetics* **1**, 356–366.

2.7 Patterns of Diversification

P. W. SIGNOR

Introduction

The past 3.5 billion years have witnessed substantial change in the numbers of protist, animal, and plant taxa on Earth. The magnitude of that net change is evident from comparison of the lush biological diversity present in so many modern habitats with Archaean sediments seemingly barren of fossils. But reconstructing the geological history of organic diversity has proved difficult. Biases in the preservation, collection, and study of fossils have combined to obscure patterns of change in diversity. Despite the difficulties, a variety of different patterns of diversification has now been documented at scales ranging from local communities to the entire biosphere. These patterns indicate that the net accumulation of taxa through time has been quite unsteady.

Biases in the fossil record

The geological history of taxonomic and ecological diversification is obscured by a variety of time-dependent and time-independent filters. Most of these are various sorts of sampling biases, which cause the observed fossil record to differ from the actual history of the biosphere (see also Section 3.12). The most severe of the time-dependent biases is the loss of sedimentary rock volume and area with increasing age (Raup 1976b). Both sedimentary rock area and rock volume correlate strongly with the numbers of animal species described from that stratigraphic interval (Raup 1972, 1976b). Rock volume and area affect apparent species richness by influencing the likelihood that a given species is preserved, discovered, and described (Raup 1976b). Similar biases have been documented in the fossil record of vascular plants on land (Knoll *et al.* 1979). The quality of preservation of fossils within sedimentary rock also tends to deteriorate with increasing age, because of extended exposure to diagenesis (Raup 1972). The kinds of sedimentary rock and, by implication, the environments preserved in the stratigraphic record have varied greatly through time. Variability in the representation of palaeoenvironments in the stratigraphic record must in-

fluence the composition of the fossil record. Sea-level, which largely controls epicontinental marine deposition and preservation of fossils therein, has varied throughout the geological past. Low sea stands are usually represented in the stratigraphic record as diastems, disconformities, or unconformities, and lack any fossil record of shelf faunas. Other time-dependent biases include monographic effects (Raup 1972) and the distribution of systematists' labour (Sheehan 1977).

There are significant time-independent biases. For example, terrestrial environments (and the organisms that inhabit them) are not well represented in the stratigraphic record, in comparison to marine habitats (e.g. Padian & Clemens *in* Valentine 1985). Among marine organisms, heavily skeletonized forms are preserved far more frequently than lightly or non-skeletonized forms. Palaeobiologists often presume that the ratio of heavily skeletonized to non-skeletonized species has been approximately constant, at least since the early Phanerozoic, but no data or arguments to support that contention have been advanced. On the contrary, there is some evidence that skeletons have become more robust in time in response to newly evolving predators (Section 4.13).

The net result of these biases is quite severe, amply justifying the ancient laments about the incompleteness of the fossil record. Only approximately 10% of the skeletonized marine species of the geological past and far fewer of the soft-bodied species are known (Sepkoski *et al.* 1981; Signor *in* Valentine 1985). No doubt whole clades and communities of the past remain to be discovered. More importantly, these biases continue to obscure all but the most fundamental patterns in the history of diversification.

A brief aside on the semantics of diversity might prevent confusion. The term *diversity* has been used in two senses. Unfortunately, the two usages are rather different, and treating the term carelessly confounds an important concept. In the palaeontological literature, diversity is often used to mean richness, or the number of taxa present. Diversity also has a second meaning, incorporating both rich-

ness and evenness of distribution. In this second sense, a community composed of three equally common species would be regarded as more diverse than a community of three species where one species far outnumbered the remaining two. The ecological literature generally restricts usage of diversity to the latter meaning. Most papers on the history of diversity, however, have treated diversity as synonymous with number of taxa present, and that is the approach used here.

Taxonomic diversity

Tabulations of classes, orders, or families, instead of species, are commonly employed to minimize sampling bias in palaeontological estimates of biological diversity. One need only find a single species to document the presence of a higher taxon, whereas every species must be discovered to provide complete documentation of species richness. Therefore, researchers have employed classes, orders, or families as surrogates for species in estimates of biological diversity in situations where our ability to sample species is hindered. Higher taxa have also been employed as metrics of morphological or ecological diversification (e.g. Erwin *et al.* 1987).

The utility of higher taxa as first order metrics of species richness is dubious (see Sepkoski 1978 for a contrary view). Tabulations of marine orders and families (Sepkoski 1978, 1979, 1982) and estimates of species richness (Sepkoski *et al.* 1981; Signor *in* Valentine 1985) are not congruent, indicating that numbers of higher taxa do not parallel changes in underlying species richness (Fig. 1A, B). Generic diversity is rather similar to estimated patterns of species richness, but patterns of the diversity of families, orders, or classes are increasingly dissimilar. Similarly, Raup's (1979) analysis of the Permo-Triassic mass extinction (Section 2.13.4) indicates that 17% and 52% reductions in the number of marine orders and families, respectively, represent approximately a 96% reduction in the number of species. Likewise, patterns in the numbers of terrestrial vertebrate orders are rather dissimilar from patterns in the numbers of genera (see Padian & Clemens *in* Valentine 1985). Higher taxa are buffered from fluctuations in numbers of species and consequently are poor metrics of changes in species richness.

Higher taxa are more or less artificial constructs that are not defined by species richness. Indeed, most higher taxa incorporate relatively few species (Sepkoski 1978). Therefore, the lack of concordance

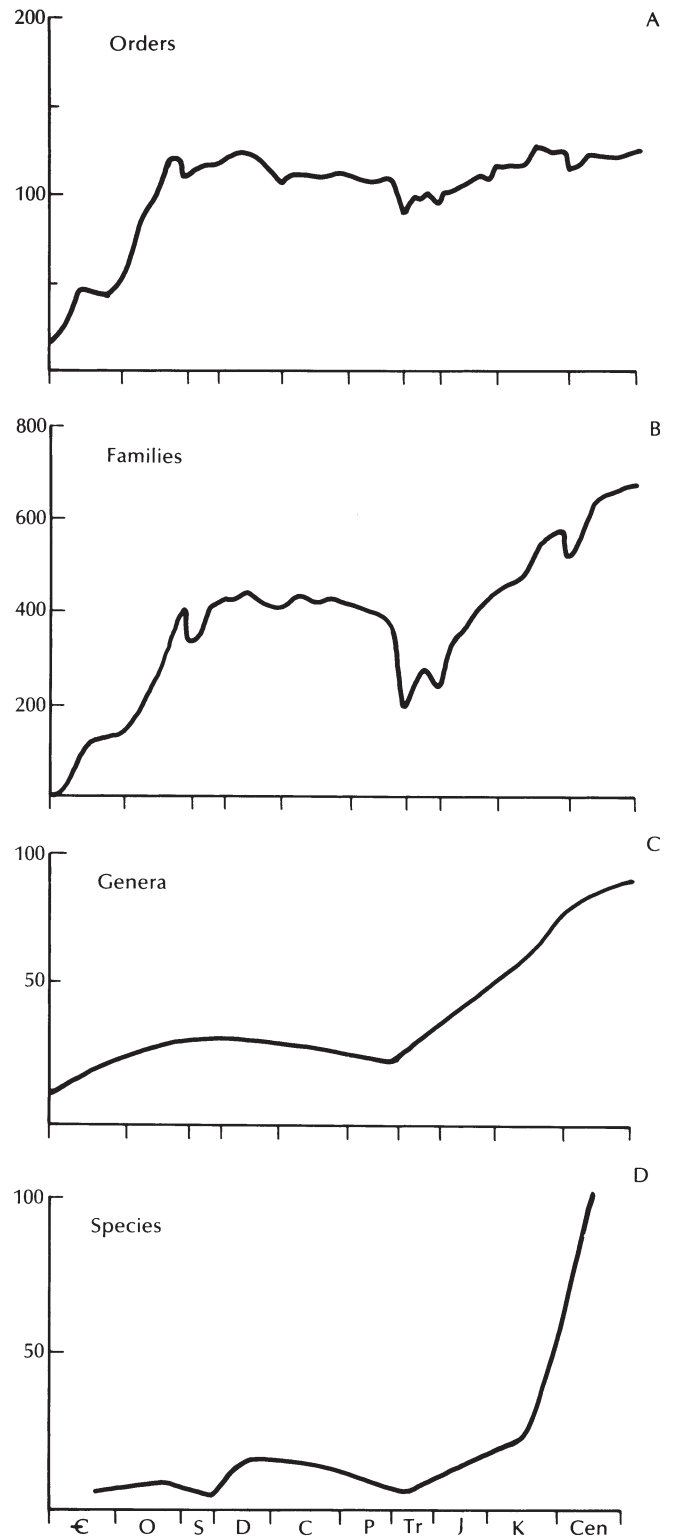


Fig. 1 Diversity of marine animals. A, numbers of marine orders through time. (Data from Sepkoski 1978.) B, numbers of marine families through time. (Data from Sepkoski 1979.) C, percentage change in the number of marine genera through time. (Data from Sepkoski *et al.* 1981.) D, estimated percentage variation in the number of skeletonized marine invertebrate species. (Data from Signor *in* Valentine 1985.)

between the number of species, families, orders, or classes through time is not surprising. In contrast, biological species can be defined and recognized through patterns of reproductive isolation: they are real biological units. As an entity, the species possesses biologically significant characteristics lacking in higher taxa. But change in the numbers of species through time is a more difficult problem to attack than change in higher taxa, because of the inherent deficiencies of the fossil record (Section 3.12).

The numbers of Phanerozoic marine orders and families have been tabulated by Sepkoski (1978, 1979, 1982) (Fig. 1A, B). The number of orders increased rapidly until the Late Ordovician, and then remained approximately constant for the remainder of the Phanerozoic. The number of families also increased rapidly in the Cambrian and Ordovician, reaching a plateau of about 400 families for the remainder of the Palaeozoic. Following the Permo-Triassic mass extinction (Section 2.13.4), the number of families has increased more or less continuously to the present. A preliminary tabulation of the number of genera shows a pattern generally similar to the change in numbers of families through time (Sepkoski *et al.* 1981) (Fig. 1C).

Compilations of the number of described marine species through time show low numbers through the Palaeozoic and Mesozoic, followed by a substantial increase in the Cenozoic (Raup 1976a). These tabulations are undoubtedly skewed by sampling biases, as discussed above (see Raup 1976b). Attempts to infer patterns of species richness from changes in the numbers of higher taxa have produced patterns generally similar to Raup's tabulation, but show lower numbers of species in the Palaeozoic and Mesozoic. Analytical calculations to remove the effects of sampling bias result in a similar pattern (Signor *in* Valentine 1985; Fig. 1D).

The history of diversification of terrestrial vertebrates produces a quite different pattern. Compilations of the numbers of tetrapod orders through time show no longstanding equilibrium (Padian & Clemens *in* Valentine 1985; Fig. 2A). There was a steady increase through the Middle Palaeozoic, reaching a Mesozoic plateau that began in the Late Triassic. Following the Cretaceous–Tertiary mass extinction (Section 2.13.6), the number of orders increased briefly and then began to decline (Fig. 2A). The Tertiary adaptive radiation of birds is superimposed upon this diversification, and nearly doubled the number of terrestrial vertebrate orders (Fig. 2A). The pattern at the generic level is similar,

but more exaggerated (Padian & Clemens *in* Valentine 1985). The number of genera rose quickly through the Palaeozoic to a peak in the Permian. Following a severe reduction in generic diversity at the end of the Permian, the number increased, regaining Permian levels in the Cretaceous. In the Cenozoic the number of genera increased nearly tenfold.

The history of the diversification of vascular plants forms still a third pattern (Fig. 2B). In the Northern Hemisphere, there was a gradual increase in species richness to a peak of over 40 species early in the Late Devonian (Niklas *et al.* *in* Valentine 1985). Following a slight decline in the Late Devonian, the number increased rapidly to over 200 species by the Middle Carboniferous. With the exception of a brief decline at the end of the Permian, the number of species increased gradually through the remainder

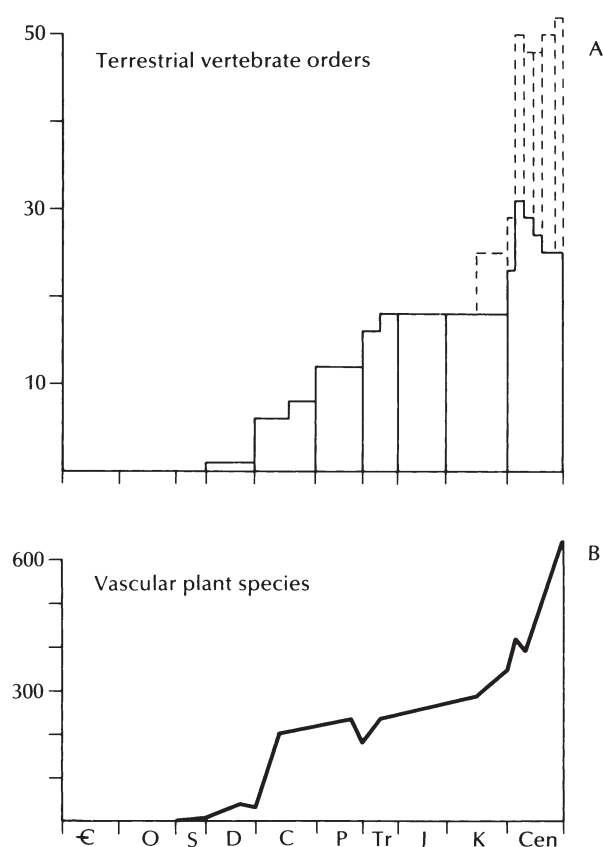


Fig. 2 Diversity of terrestrial organisms. A, numbers of terrestrial vertebrates through time. Solid lines indicate changes in the number of amphibian, reptile, and mammal orders; dotted lines indicate the number of avian orders. (Data from Padian & Clemens *in* Valentine 1985.) The Cenozoic is subdivided into the five epochs of the Tertiary plus the Pleistocene. B, numbers of terrestrial plant species (mostly in the Northern Hemisphere) through time. (Data from Niklas *et al.* *in* Valentine 1985.)

of the Palaeozoic, the Triassic, Jurassic and Early Cretaceous. Following the origin of angiosperms (Section 1.10) in the Late Cretaceous, species richness increased rapidly to over 600 in the Quaternary.

Patterns of global taxonomic richness differ at various levels of the taxonomic hierarchy. Unbiased species data would most accurately reflect changes in biological complexity through time, but species-level data are the most susceptible to sampling bias. The species richness of marine animals, vascular plants, and terrestrial vertebrates have quite different histories, but all indicate significant reductions in diversity at the Permo-Triassic, Norian, and Maastrichtian extinction events (Section 2.13). All three patterns also share a tremendous diversification of species beginning in the Cretaceous.

Local diversity

An important component of the history of diversity is the temporal pattern of species richness within individual communities. Bambach (1977) compiled counts of species present within 386 previously described ancient marine communities. He assigned the communities to one of three generalized habitats: nearshore high stress, nearshore variable, and open marine environments. *Alpha diversity*, or within-community diversity, remained constant in the high-stress environment communities throughout the Phanerozoic, but increased twofold in the variable nearshore and open marine environments during the Mesozoic (Bambach 1977; Fig. 3C).

The Mesozoic increase in alpha diversity apparently was accommodated through trophic diversification of the major clades of marine animals in shelf communities (Bambach 1983). In the Cambrian, there were relatively few clades and each clade had a limited range of roles. The number of clades increased in the Palaeozoic, an increase that was paralleled by a limited diversification of trophic roles. The Mesozoic increase in diversity was accompanied by a much larger diffusion of taxa into new trophic roles, especially into infaunal life-modes.

The expansion of marine animals into infaunal life-modes is one component of the pattern of increasing *tiering* (Section 1.7.1). Tiering, the spatial development of communities both above and below the sediment surface, had increased through the Phanerozoic. This increase has been attributed to a number of physical and biological processes, but

the net result is undoubtedly an increase in local habitat complexity and organic diversity.

Controls on diversity

The nature of the processes controlling species richness is the subject of considerable speculation. At the level of communities, such processes are not well understood, even in the modern world. Area, habitat complexity, environmental stability, physical disturbance, and other factors may well be important controls on alpha diversity. What transpired to bring about a twofold increase in within-habitat species richness in the late Mesozoic is equally unclear. A better understanding of the processes regulating species richness in the Recent is probably a prerequisite to resolving this question.

Area is a primary influence on diversity at local, regional, and global levels. Area appears to regulate diversity primarily through variation in rates of extinction. Reduction in habitable area decreases population size, which increases the chance of extinction. For organisms dwelling among the benthos on continental shelves, change in diversity appears to be related to variation in shelf area (Sepkoski 1976; but see Flessa & Sepkoski 1978). Severe reductions in shelf area have also been implicated as the cause of a mass extinction (Sections 2.12.1, 2.13.4).

Mass extinctions severely reduce the number of taxa present in the biosphere. Diversity generally rebounds following extinction events and often increases to surpass previous levels, but that rebound requires geologically significant intervals of time. The time necessary for recovery is, in part, proportional to the magnitude of the extinction (Sepkoski 1984). During the intervening time between extinction and recovery, diversity is reduced. If extinctions are spaced more closely than the necessary recovery time, the biosphere will remain relatively impoverished (e.g. Hansen 1988).

At the level of the biosphere, plate tectonics is undoubtedly the most potent control on the diversity of organisms (Valentine *et al.* 1978; Signor *in* Valentine 1985). Marine organisms often share common range boundaries, and geographical regions with relatively homogeneous faunas and distinct boundaries are termed *provinces* (see also Section 5.5). The boundaries of these provinces are defined by the joint limits of distribution of common species, and are controlled by patterns of climate and oceanic water circulation. In turn, climate and oceanic circulation are determined largely by the

distribution of continental land masses. Continuously changing continental configurations have thus regulated the number and distribution of provinces through time. Changing levels of provinciality have no apparent impact on biological complexity within individual communities, but alter global diversity through adding or subtracting whole provinces. The number of provinces has increased greatly since the late Mesozoic breakup of Gondwana and Laurasia (Valentine *et al.* 1978), a change undoubtedly responsible for much of the increase in species richness of marine animals.

The species richness of vascular plants and terrestrial vertebrates is also heavily influenced by plate tectonics (e.g. Padian & Clemens *in* Valentine 1985). Tertiary isolation of the terrestrial faunas of South America, Australia, Africa, and Madagascar permitted the evolution and persistence of unique faunas, while other clades dominated the Holarctic continents of Asia, Europe, and North America. Isolation and interspersed periods of faunal interchange have contributed greatly to the diversity and taxonomic composition of the terrestrial vertebrate faunas of the different continents.

In summary, local and regional patterns in the number of taxa vary semi-independently and are compounded at regional and global scales. Increasing alpha diversity within marine communities resulted in a comparable increase in global diversity in the Late Mesozoic. Similarly, changes in provinciality through time have altered global diversity. Changes at each level of the ecological hierarchy, from the community to the biosphere, influence trends in global diversity. Such trends therefore represent complex interactions of physical and biological processes operating on many scales.

Diversity within clades

A systematic pattern of temporal change in diversity within individual clades has recently been recognized (Gilinsky & Bambach 1987). Clades appear to contain more subtaxa early in their histories than later on. The primary cause of this trend appears to be a systematic decline in rates of origination within established clades through geological time, although rates of extinction also increase somewhat through time (Gilinsky & Bambach 1987). The obvious inference from this statistical generalization is that clades are established during brief adaptive

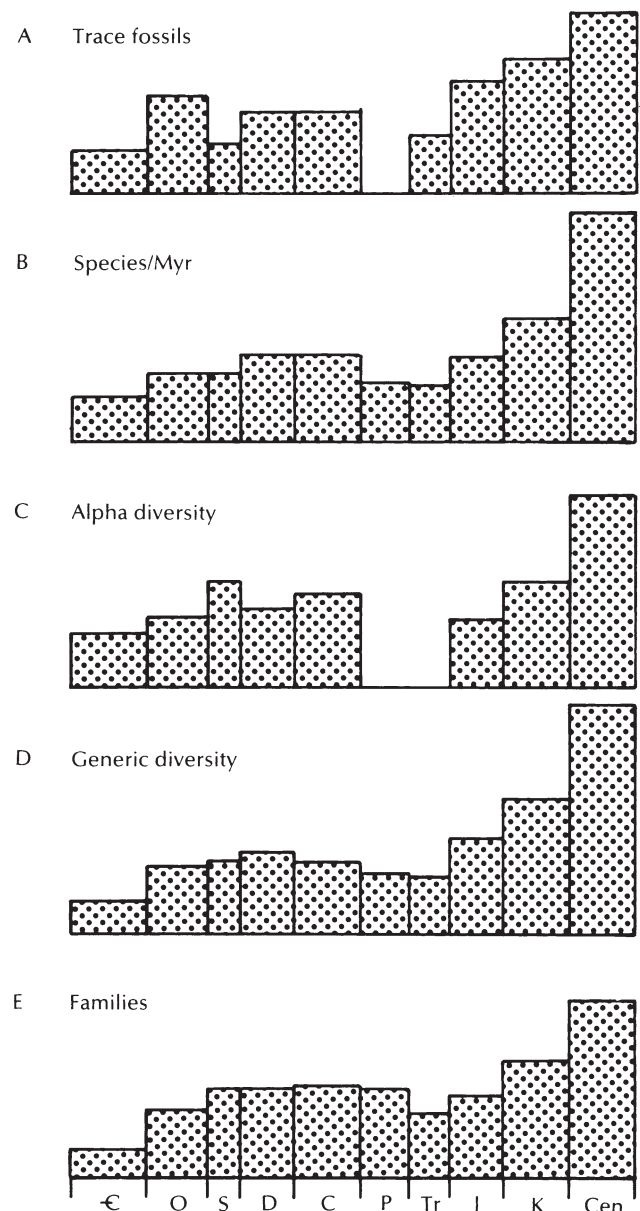


Fig. 3 Trends in marine animal diversity through time. (After Sepkoski *et al.* 1981.) The five trends in diversity accepted as evidence against an equilibrium model of Phanerozoic diversity. A, number of the trace fossil genera within individual habitats. B, numbers of described species, normalized for length of the geological period. C, alpha diversity of marine invertebrates. D, E, numbers of genera and families respectively. The original figure was presented without vertical axes, as shown here.

radiations and subsequently begin a long decline to eventual extinction. Speciation, not extinction, might be the dominant factor in clade diversity (Gilinsky & Bambach 1987).

Modelling change in diversity through time

Sepkoski (1978, 1979, 1984) applied quantitative models of population growth to the history of taxonomic diversity of marine organisms. These models were developed to describe and predict the increase in numbers of individuals within single populations. The mathematical assumptions of the simplest model, the logistic model of population growth, are: (1) there is a maximum number of individuals that can be supported in the environment (the *carrying capacity*); and (2) population growth is exponential and declines linearly as the population size approaches the carrying capacity. These assumptions may also reasonably apply to taxonomic diversification at the level of the biosphere. Sepkoski's (1978) successful application of the logistic model to describe the increase in numbers of marine orders suggests that the model describes the behaviour of change in the numbers of orders rather well. Similar results have been obtained in analyses of the radiation of angiosperms (Lidgard & Crane 1988). Further applications of more complex models from population biology treat different aggregates of clades as faunas, comparable to populations competing for resources (Sepkoski 1979, 1984). These models also produce good fits with the available data.

Faunal equilibrium?

Over the past 20 years, a number of theorists have questioned the empirical pattern of increasing species richness through time, suggesting instead that the diversity of marine animals has been at equilibrium through much of the Phanerozoic (see Signor *in* Valentine 1985 for review). In this view, changes in species richness reflect only biases and not biologically significant trends. The equilibrium itself could be absolute, with constant numbers of species through time, or dynamic, with an equilibrium shifting in response to the changing physical world. However, the similar patterns observed in a variety of separate metrics of diversity (Fig. 3) provide convincing evidence that the apparent pattern of species richness through time is not artifactual (Sepkoski *et al.* 1981). But important questions about historical patterns of local and global species richness, and the ultimate controls on those patterns, remain to be resolved.

References

- Bambach, R.K. 1977. Species richness in marine benthic environments through the Phanerozoic. *Paleobiology* **3**, 152–167.
- Bambach, R.K. 1983. Ecospace utilization and guilds in marine communities through the Phanerozoic. In: M.J.S. Tevesz & P.L. McCall (eds) *Biotic interactions in Recent and fossil benthic communities*, pp. 719–746. Plenum Press, New York.
- Erwin, D.H., Valentine, J.W. & Sepkoski, J.J., Jr. 1987. A comparative study of diversification events: the early Paleozoic versus the Mesozoic. *Evolution* **41**, 1177–1186.
- Flessa, K.W. & Sepkoski, J.J., Jr. 1978. On the relationship between Phanerozoic diversity and changes in habitable area. *Paleobiology* **4**, 359–366.
- Gilinsky, N.L. & Bambach, R.K. 1987. Asymmetrical patterns of origination and extinction. *Paleobiology* **13**, 427–445.
- Hansen, T.A. 1988. Early Tertiary radiation of marine molluscs and the long-term effects of the Cretaceous–Tertiary extinction. *Paleobiology* **14**, 37–51.
- Knoll, A.H., Niklas, K.J. & Tiffney, B.H. 1979. Phanerozoic land-plant diversity in North America. *Science* **206**, 1400–1402.
- Lidgard, S. & Crane, P.R. 1988. Quantitative analyses of the early angiosperm radiation. *Nature* **331**, 344–346.
- Raup, D.M. 1972. Taxonomic diversity during the Phanerozoic. *Science* **177**, 1065–1071.
- Raup, D.M. 1976a. Species richness in the Phanerozoic: a tabulation. *Paleobiology* **2**, 279–288.
- Raup, D.M. 1976b. Species richness in the Phanerozoic: an interpretation. *Paleobiology* **2**, 289–297.
- Raup, D.M. 1979. Size of the Permo-Triassic bottleneck and its evolutionary implications. *Science* **206**, 217–218.
- Sepkoski, J.J., Jr. 1976. Species diversity in the Phanerozoic: species–area effects. *Paleobiology* **2**, 298–303.
- Sepkoski, J.J., Jr. 1978. A kinetic model of Phanerozoic taxonomic diversity, I. Analysis of marine orders. *Paleobiology* **4**, 223–251.
- Sepkoski, J.J., Jr. 1979. A kinetic model of Phanerozoic taxonomic diversity, II. Early Phanerozoic families and multiple equilibria. *Paleobiology* **5**, 222–251.
- Sepkoski, J.J., Jr. 1982. *A compendium of fossil marine families*. Milwaukee Public Museum Contributions to Biology and Geology No. 51. Milwaukee Public Museum, Milwaukee.
- Sepkoski, J.J., Jr. 1984. A kinetic model of Phanerozoic taxonomic diversity, III. Post-Paleozoic families and mass extinctions. *Paleobiology* **10**, 246–267.
- Sepkoski, J.J., Jr., Bambach, R.K., Raup, D.M. & Valentine, J.W. 1981. Phanerozoic marine diversity and the fossil record. *Nature* **293**, 435–437.
- Sheehan, P.M. 1977. Species diversity in the Phanerozoic: a reflection of labor by systematists? *Paleobiology* **3**, 325–328.
- Valentine, J.W. 1985. *Phanerozoic diversity patterns*. Princeton University Press, Princeton.
- Valentine, J.W., Foin, T.C. & Peart, D. 1978. A provincial model of Phanerozoic diversity. *Paleobiology* **4**, 55–66.

2.8 Coevolution

S. CONWAY MORRIS

Introduction

Ever since the first species was joined by a second one, the potential for some sort of coevolution has existed. However, the possibility of documenting coevolution in the fossil record depends on the scope of the definition that is accepted. In a broad sense, coevolution has been taken to include almost any biological interaction, with emphasis often placed on mutualistic associations. Stricter definitions emphasize reciprocal responses between individuals of two species where each exerts, either sequentially or synchronously, an influence on the other's heritable characters. Whether such coevolutionary oscillations are stable over geological periods of time is not certain, and it seems questionable whether coevolution in a strict sense has ever been recognized in the fossil record. Accordingly, evidence for more broadly based interactions that seem in some sense to represent responses by one taxon or group to changes in another is presented here.

Gaia

On the grandest scale there has been considerable interest in the concept of Gaia, whereby a system of organically mediated feedbacks maintains the Earth's surface in a state of homeostasis that is largely independent of external vicissitudes that otherwise would imperil the continuation of life. However, while it is accepted that biological activities can mediate geochemical, and probably geophysical, cycles, there has been less enthusiasm for the notion that life *in toto* could act as the primary regulator of Gaia. This is because individual species, rather than the entire biosphere, must be accepted as the units for evolutionary selection, with life exploiting those opportunities offered by changing environments.

The Law of Constant Extinction

While not as grandiose in its scope, the so-called Law of Constant Extinction may have implications for interactions between members of entire ecological groupings of taxa. Analyses of numerous taxon

survivorship curves by L.M. Van Valen indicated that the rate of extinction is stochastically constant, i.e. the probability of extinction is constant irrespective of the duration of a particular taxon. Van Valen explained this pattern by the now well known Red Queen Hypothesis (Section 2.5), arguing that the environment in which evolution occurs is largely defined by biotic interactions that operate so that the improvement in fitness of any one species automatically reduces the fitness of all others (given that the sum of all fitnesses remains unchanged). In this sense ecological units consisting of numerous interacting species, which exhibit unceasing evolutionary change as species attempt to restore their fitness in the face of a constantly deteriorating biotic environment, may be said to show coevolution. Whether the Law of Constant Extinction is empirically demonstrable, and whether the Red Queen Hypothesis is the appropriate explanation, have both been extensively debated. Recent analyses of planktic species (mostly coccoliths, foraminifera, and radiolarians) from Cenozoic ocean deposits from mid-to-low latitudes give some support to the Red Queen Hypothesis, although the data on species survival have to be considered in the context of an environment that is not effectively constant.

There are three other areas in the fossil record that may be explained in the broad context of coevolution. These are claims for reciprocal patterns between: (1) predators and prey; (2) plants and animals, especially insects; and (3) phylogenetic congruence between symbiotic taxa, especially parasites and their hosts.

Predators and prey

Much interest has been expressed in the possibility of an arms race between predator and prey, with a spiralling escalation of attack and defence. There are, however, reasons to doubt that a long-term oscillation would persist. In particular, Vermeij (1982) pointed out that: (1) as most predators feed on at least several species, if confronted by an increasingly well-defended prey they will switch to one of greater vulnerability; and (2) the predator itself will be prey for other species, so that selective

factors that favour the predator's own survival, as against its ability to obtain a meal, will predominate. Indeed, the only cases where reciprocal evolution of prey and predator may show consistent trends are where a victim can on occasion maim or even kill its attacker.

Notwithstanding the potential problems in documenting predator–prey arms races, a number of attempts have been made to demonstrate reciprocity in the fossil record on the basis of broad-scale trends. One of the best known analyses concerns changes in the brain mass of Cenozoic ungulate herbivore and carnivore mammals. From a study of brain sizes Jerison (1973) concluded that the carnivores maintained proportionally larger brains than the ungulates, although both showed a persistent increase during the Cenozoic. This pattern was explained by a type of coevolutionary feedback whereby selection pressure exerted by the larger-brained carnivores forced a corresponding increase in the herbivores, which in turn fuelled further increases in the carnivores.

The validity of this analysis, however, has been questioned. Radinsky (1978) pointed out that: (1) many of the comparisons involve carnivores and ungulates of different stratigraphic age; (2) the estimates of body weight (needed as part of the calculation of the relative brain size) may require revision; and (3) some samples may be too small to provide reliable comparison. He concluded that, where the data are adequate, supposed differences between ungulates and carnivore brain sizes cannot be demonstrated. Moreover, other studies of mammalian evolution in the Cenozoic (Bakker *in* Futuyma & Slatkin 1983) have argued that while both carnivores and herbivores show trends towards greater efficiency (e.g. for running) the so-called adaptive gaps may widen, especially when replacement faunas arrive following a mass extinction.

Changes in predator–prey interactions have also been identified in the marine record (see also Section 4.13.1). The rise of predators in the Cambrian is followed by an episode of increasing predatory activity in the Middle Palaeozoic and finally a major reorganization of prey and predatory ecologies during the Jurassic and Cretaceous (the so-called Mesozoic Marine Revolution). However, apart from the parallel rise of offensive and defensive adaptations, it has not been possible to demonstrate specific series of reciprocal changes, and it seems that evolutionary responses may have been diffuse.

Amongst invertebrates, recent research has investigated possible coevolution between predat-

tory gastropods, especially naticids, and their prey of bivalve molluscs, which they attack by drilling through the shell (Kitchell *in* Nitecki & Kitchell 1986). While evidence for naticid attacks may extend back to the Triassic, it first became widespread in the Cretaceous. Study of drilling behaviour demonstrates a remarkable stereotypy over geological time in terms of both position on the prey and ability to resume attack after interruption. However, in terms of possible coevolutionary responses between predator and prey, the only persistent trend that can be documented in the fossil record is a mutual increase in size.

Plants and animals

The widespread inference of coevolution between plants and arthropods, especially insects, in modern biotas has often been extended into the geological past. However, despite some classic examples, such as between figs and fig-wasps, there is serious reason to doubt whether many Recent plant–animal interactions can be regarded as strictly coevolutionary. Nevertheless, there is a widespread assumption that the diversification of plants and insects in the fossil record has been governed by coevolutionary forces. At present only the growing evidence for plant–animal interactions can be documented, leaving open the question of whether any of the examples fall into the domain of strict coevolution.

Evidence for such interaction can be traced to the early Devonian, both in the form of direct associations (e.g. trigonotarbid arachnids lurking in sporangia of *Rhynia* from the Rhynie Chert), and more generally in plant morphology that ostensibly either promoted (e.g. spore sculpture) or hindered (e.g. stem spines) arthropod interactions. In these latter cases caution must be exercised, as specific features of plant anatomy may have had multiple functions including resistance to water loss, shielding from ultraviolet radiation, and so forth. By the Carboniferous there is considerable evidence for plant–animal interactions, both direct, such as spores in insect guts or various trace fossils (coprolites, borings, chew marks), and indirect, from insect mouth parts or plant anatomy, especially of spores and seeds. However, in no case has it been demonstrated that either partner was exerting reciprocal selective pressure over a period of geological time. Evidence of responses to arthropods in younger floras (Crepet 1979) includes study of reproductive structures, such as those of cycads,

whose cone anatomy appears to trend towards excluding insect attack. Particular attention has been given to the activities of insects in pollination, to which has been linked the rise of the bisporangiate condition. In the Jurassic gymnosperms, for example, coleopterans (beetles) may have been of particular importance, but with the rise of the angiosperms in the Cretaceous (Section 1.10) the role of dipterans (flies) and hymenopterans (bees and wasps) is regarded as crucial. In particular, links between insect group and flower or inflorescence anatomy may allow inferences on potential pollinators. However, in many examples the assumptions are based on uniformitarian premises and it is also important to realize that strict coevolution has not been demonstrated. Indeed, in many cases it is likely that evolution was sequential, the insects following plant diversification rather than acting as primary mediators.

Although the greatest interest in plant–animal coevolution has concerned the role of arthropods, speculation has extended also to vertebrates. Stebbins (1981) argued for coevolution between Cenozoic mammalian grazers (e.g. horses) and the grasses (see also Section 1.11). The development of hypsodont teeth to cope with the siliceous grasses (specifically the secretion of opalines in the plant epidermal cells), and of running ability in more open savanna, would seem to be linked intimately with the spread of grasslands. However, reciprocal connection between degree of hypsodonty and silica content or distribution, that could be taken as strict coevolution, has not been demonstrated.

Phylogenetic congruence

The final area where the fossil record may contribute to the documentation of coevolution is in the identification of congruent phylogenies where mutualistic associations, especially between parasites and their hosts, are reflected in their respective histories of cladogenesis (Mitter & Brooks in Futuyma & Slatkin 1983). This pattern is often referred to as Fahrenholz's Rule, but to date the evidence for congruence has varied widely and in very few clades of parasite and host is strict congruence evident. Where parasites possess limited abilities for dispersal then host–parasite congruence may occur. However, if a parasite species is pursuing a particular feature, in what is known as resource tracking, then typically it will occur in those taxa that happen to share the particular resource. Thus, in the Mallophaga (chewing lice) the limited dis-

persal of those infesting some mammals (e.g. pocket gophers) contrasts to the distribution of those parasitizing birds, where feather type seems to be of particular importance (Timm in Nitecki 1983). Where Fahrenholz's Rule appears to be applicable, then in principle the fossil record of a well skeletized host could give insight into the evolutionary history of the parasites, which are almost invariably soft-bodied and unknown as fossils. While the fossil record may throw light on times of divergence in such instances, the relatively few well documented lineages still only provide a broad indication of evolutionary events, and tightly constrained histories do not appear to be available. The numerous commensal associations that have been documented may prove a more fruitful area for establishing phylogenetic congruence between symbionts in the fossil record. These include host specific epizoans, e.g. cornulitids, spirobids and other 'worms', and more intimate associations such as those between stromatoporoids and corals. However, in no case does it appear that strict coevolution has occurred, and in at least some cases there is evidence that the host has evolved (at least morphologically) at a substantially faster rate than its partner.

Conclusion

While evidence of species interaction is manifest in the fossil record, examples of strict coevolution have yet to be documented. This may reflect problems of resolution and insufficient study, but it seems more likely that long term associations only rarely fall into the category of coevolution as it may be usefully understood.

References

- Crepet, W.L. 1979. Insect pollination: a paleontological perspective. *BioScience* **29**, 102–108.
- Futuyma, D.J. & Slatkin, M. 1983. *Co-evolution*. Sinauer, Sunderland, Massachusetts.
- Jerison, H.J. 1973. *Evolution of the brain and intelligence*. Academic Press, New York.
- Nitecki, M.H. (ed.) 1983. *Co-evolution*. University of Chicago Press, Chicago.
- Nitecki, M.H. & Kitchell, J.A. (eds) 1986. *Evolution of animal behavior: paleontological and field approaches*. Oxford University Press, New York.
- Radinsky, L. 1978. Evolution of brain size in carnivores and ungulates. *The American Naturalist* **112**, 815–831.
- Stebbins, G.L. 1981. Coevolution of grasses and herbivores. *Annals of the Missouri Botanical Garden* **68**, 75–86.
- Vermeij, G.J. 1982. Unsuccessful predation and evolution. *The American Naturalist* **120**, 701–720.

2.9 Adaptation

P. W. SKELTON

Introduction

Natural historians have long admired the ways in which the construction and activities of living organisms seem to be so well suited, or '*adapted*', to the natural circumstances in which they live. Especially striking is the extent of co-operation of their features that renders them so adapted. Illustrations readily spring to mind, such as the stiffened tail, the two backward-pointing toes on each foot, the stout beak, and the extendible tongue of the woodpecker, which together enable it to perch on tree stems and probe them for insects.

To most pre-Darwinian thinkers such adaptive traits were the essentially static attributes of fixed species, perfectly fitting them, perhaps by divine appointment, to their places in nature. Certain pre-Darwinian transformists, such as J-B. Lamarck and Erasmus Darwin, in contrast, postulated that the inheritance through many generations of habitually acquired developmental modifications (such as, say, the building up of well exercised muscle) was a means of evolutionary adaptation. But there has never been any satisfactory evidence for this.

Both the static and the Lamarckian views of adaptation were *teleological*; that is to say they appealed to final causes in placing the prospect of a function (either in God's mind, or in the 'needs' or 'strivings' of organisms) prior to the appearance of the feature adapted for it. Such explanations, though still rife in the popular imagination, are categorically denounced by most biologists today because of the inherently untestable character of the final causes.

The Darwin/Wallace theory of natural selection, on the other hand, asserted that adaptations became established in evolving populations through the preferential survival and reproduction of individuals possessing naturally occurring, heritable variations, which conferred advantage in the 'struggle for existence' arising from the excessive fecundity of the populations in relation to resources. This means that adaptation need not be perfect, as earlier naturalists had tended to opine. Rather,

adequacy in the face of competition becomes the expected rule. Persistent competition may eventually perfect some adaptations, but there are many reasons, ranging from environmental change to the inherent constraints of bodyplans, why others are not perfected.

By limiting the notion of function to the effect of any given feature on the lives of its possessors, and by placing natural selection in the creative driving seat of evolution, Darwinism avoids the teleology, and thus the unacceptable mystery of earlier explanations of adaptation.

Evolutionary thinking requires a distinction between adaptation as a *process* of gradual modification in a population, and as a *state of being* in individuals, in relation to prevailing circumstances. With natural selection, the state of being adapted, in respect of some feature or complex of features, also takes on two aspects: first, there is the element of function — the way in which the feature or complex operates — and, second, there is the selective benefit to the possessors of the feature or complex, in terms of preferential survival and/or fecundity, deriving from its operation. Varying emphasis on one or other aspect in different usages of the term 'adaptation' has led to much confusion and misunderstanding; so it is worth teasing them apart somewhat. Darwin himself still used the term in an essentially vernacular fashion, primarily stressing functional suitability: the 'best adapted' were simply those individuals possessing the most 'useful' variations for the operation of various life functions, such as feeding, locomotion, and seed dispersal. These, he repeatedly postulated, would *tend* to be favoured by natural selection in the 'struggle for existence', so fuelling the continuing process of further adaptation in populations. This Darwinian sense of adaptation, still recognizable independently of selective effects, though assumed to be both the product of, and producing them, is still widely used today, especially by palaeontologists, for reasons discussed below. Evolutionary biologists, on the other hand, have considerably refined the theory of natural selection and, in so doing, have subtly redefined the meaning of adaptation precisely in terms of selective effects.

Adaptation in evolutionary biology

The 'neo-Darwinian synthesis' of the nineteen-thirties to the nineteen-fifties attributed all evolutionary adaptation to the operation of natural selection upon the phenotypic manifestations of genes in populations. Darwin's (and others') speculations on the additional operation of processes other than selection (such as the effects of use and disuse, and other somatic influences on the germ line) were brushed aside. Statistical models of population genetics allowed the operation of natural selection to be quantified. 'Fitness', which to Darwin had been a somewhat vague expression of relative adaptedness of individuals, now became rigorously defined as the proportional survival and fecundity of a given genotype (usually simplified in scope so as to refer to all carriers of a specified pair, or pairs of alleles) relative to that genotype in the population which has the most descendants. (Unfortunately this tends to be referred to as 'Darwinian fitness'. Perhaps 'neo-Darwinian fitness' would be better, to distinguish it from Darwin's vaguer usage.) Since the effect of any truly 'useful' feature on some life function may be assumed to contribute to the fitness of its possessors, the term adaptation, too, came to be defined by neo-Darwinians in terms of the promotion or maintenance of fitness. Dobzhansky (1970), for example, cited the expression '*adaptive value*' as a synonym for '*Darwinian fitness*'. Or again, more recently, Ridley (1985) has stated 'Adaptation means good design for life. To understand how any particular property of an organism is adapted, it is necessary to think how it enhances its bearer's chances of survival and reproduction'.

This subtle shift of emphasis in the definition from mere operational suitability to selective effect means that, in order to demonstrate adaptation, it is not good enough simply to show the effectiveness of a feature in the service of some function, however impressive that may be; it has to be shown that the feature thereby confers greater fitness on its possessors relative to alternatives. Williams (1966) further qualified this selective criterion. A given feature may accidentally benefit its possessors in special circumstances without any prior adaptation for that particular effect. For example, quick reactions clearly promote the fitness of car drivers, though they obviously did not evolve by virtue of that effect. Williams considered such an '*effect*' an inadequate criterion for recognizing true adaptation; his definition of the latter also requires evidence for prior moulding of a feature by natural

selection in the service of its recognized '*function(s)*'. However, this creates an awkward grey area for the practical consideration of the origin of adaptations, when chance '*effects*' are transformed by natural selection to established '*functions*' (discussed below), and so the distinction between the two cannot always be recognized.

The neo-Darwinian formulation has the effect, worrying to some, of making Darwin's characterization of natural selection, 'the survival of the fittest', explicitly tautologous (as 'the survival of the survivors'). However, far from trivializing the theory, as might at first seem to be the case, this conclusion represents the logical outcome of purging it of teleology; selection simply acts on what organisms actually do, not what any metaphysical agent thinks they '*ought*' to do. Evolution is thus seen to be drawn in unpredictable directions by the transient effects of a myriad of immediate causes, making adaptation highly conditional. Clear illustration of this is provided by the banded snail, *Cepaea nemoralis* (Linné), common in parts of Britain and continental Europe. Most populations of this species are strikingly polymorphic, showing differences in both the colour of the shell (shades of brown, pink, or yellow) and its patterning (with, or without, a variable number of longitudinal bands). Cain and Sheppard (1954) were able to link marked differences in the relative frequencies of these variants in different habitats with preferential predation by thrushes (as estimated from broken shells around the birds' 'anvil' stones). Effective camouflage was found to be the guiding principle, with, for example, unbanded pink and brown shells dominant in the brown leaf litter of beech woods, unbanded yellow shells on shortgrass downs, and banded yellow shells in hedgerows and longer grass. So, features adaptive in one setting were found to be demonstrably maladaptive in another, in many instances only a short distance away. Nor, indeed, can one even generalize to the extent of saying that colour and pattern, although variable, are in principle adaptive for the single broad function of camouflage; subsequent work has shown that other factors, such as response to temperature change, are of overriding importance in some peripheral populations, in which different shell types acquire differing fitnesses because of their greater or lesser tendency to absorb solar heat (Jones *et al.* 1977).

The intimate linkage of adaptation with natural selection raises the issue of what, precisely, natural selection acts upon, and what is thus the focus for

adaptation. In cases like that of the *Cepaea* polymorphism cited above, particular variations can be attributed to the action of single gene loci (e.g. for shell colour, for presence or absence of bands, and so on). For a given population, simple combinations of alleles can be assigned fitness values directly derived from, say, the live and broken shell counts. From such data one may model changes in the relative frequencies of the phenotypic variants, and thus the course of adaptation, in the population. Such exercises are the stock-in-trade of population genetics (Dobzhansky 1970). But as Dobzhansky himself has pointed out, this apparent focus on genes as units of selection is illusory, since the fitness values assigned to them are statistical abstractions derived from the fate of those genes in many different total genotypes. Any given gene, no matter what its notional fitness value, will stand or fall according to the fate (survival and reproduction) of the genotype it finds itself in. Moreover, the phenotypic expression of any gene usually depends greatly upon interactions with other genes in the genotype and with the environment. So selection really only operates directly upon the phenotypic expression of whole genotypes. But then again, in sexually reproducing organisms, only the genes survive intact from one generation to the next. So we are here dealing with a hierarchical effect: selection on individuals in a population has the effect of altering gene frequencies in the gene pool of that population, and the latter changes in turn alter the genetic complexion of individuals in future generations, in adaptive ways. This means that adaptations, at no matter what organizational level within the individual, from the coadaptations of regulatory genes in chromosomes, through those of organelles in cells, and of tissues in organs, to the modifications of whole components of morphology, physiology and behaviour, must involve a net benefit to the individuals possessing them. The only case where this is not literally so is where there is 'kin selection' in favour of close relations, who are genetically similar if not identical. In such cases altruistic behaviour of individuals may promote their own eventual 'inclusive fitness' (effective genetic representation in future generations) through the reproductive efforts of the kin they assist, albeit at personal cost. So, for example, worker bees literally have an individual fitness of zero, being sterile, but are (painfully) well adapted to defend their genetically similar sisters due to become queens, who reproduce for them. The study of such adaptations has grown enormously from

seminal work done in the nineteen-sixties, particularly by W.D. Hamilton, and the theme is succinctly reviewed by Grafen (1984). Other kinds of 'group selection' arguments — involving the notion of adaptations arising 'for the good of the species' — require somewhat unrealistic circumstances to work, and so have found little favour with evolutionary biologists (Williams 1966).

Adaptive diversity

The previous section showed that in any consideration of function and selection, organism and environment are inseparable. Lewontin (1983, p. 280) further stressed that '... the environments of organisms are made by the organisms themselves as a consequence of their own life activities. How do I know that stones are part of the environment of thrushes? Because thrushes break snails on them. Those same stones are not part of the environment of juncos who will pass by them in their search for dry grass with which to make their nests. Organisms do not adapt to their environments; they construct them out of the bits and pieces of the external world.' While the claim that 'organisms do not adapt to environments' is perhaps a little over-enthusiastic, Lewontin's point about organisms defining their environment is important. The constant dynamic interplay between the *niche* that each species so defines for itself (in Lewontin's terms) and the selection imposed on the individuals of the species by the changing constraints of that niche is one major reason for the bewildering adaptive diversity of life.

Organisms of different sizes experience different environmental constraints, because of physical scaling effects, and are correspondingly diversely adapted: so, for example, the construction of an elephant has much to do with coping with gravity, while that of a pond-skater has more to do with surface tension; and bacteria (if they could think!) would probably find the notion of gravity about as abstruse as an elephant would find their experience of being jostled by molecules and ions. Then, again, there are the differences between media: streamlining is hardly an overriding design factor in the terrestrial mammalian carnivores, yet it is clearly a vital adaptation for their marine cousins, the seals. And even in the same circumstances, differences in habits create different experiences of the world: zebras see grass as food, and have the jaws and teeth to cope with it; lions see it as useful cover on the way to the zebras, for whom their jaws and

teeth are suited. Adaptation breeds diversity, as Darwin rightly emphasized.

Yet there is order in this diversity. The physical laws and functional specifications of habit mentioned above demand analogous adaptations for similar circumstances, leading to *convergence*: the streamlined shape of the seal is broadly repeated in other fast swimmers of similar size, such as penguins, porpoises and sharks, though with differences in detail, of course, reflecting inherited differences in basic architecture. Moreover, the inherited bodyplans and the modes of growth of organisms limit the adaptive possibilities open to them. For example, gas exchange in insects takes place via branching tubes reaching into the body from spiracles along their sides. The main branches of this 'tracheal' system are ventilated tidally by contractions of the body, but in the finer tubules supplying the muscle fibres, diffusion alone suffices. Muscle fibres must therefore lie close to axial tracheae, and so muscles cannot exceed a few millimetres in diameter. This in turn constrains body size, the largest known insects being some Carboniferous dragonflies with a wingspan not much greater than that of a crow. Considerations of such architectural constraints on evolution, and hence the mapping out of the adaptive potentialities that remain, are the business of *constructional morphology* (Section 4.1).

Adaptation in palaeontology

The neo-Darwinian definition of adaptation is problematical, to say the least, for palaeontologists — as the mere contemplation of trying to gauge relative survival and fecundity of genetically differing variants in fossil assemblages should suggest. However, this is not impossible.

Using an ingenious argument, which deserves wider exploitation by palaeontologists, Sambol & Finks (1977) documented natural selection in a population of the Cretaceous oyster, *Agerostrea mesenterica*. A bulk collection from an undisturbed assemblage of their markedly plicate, arcuate shells was made from a single locality in the Maastrichtian of New Jersey, U.S.A. From the annual growth increments of the shells it was possible to determine the age at death of some two and a half thousand individuals. Censuses of selective mortality could thus be carried out in relation to four morphometric parameters of the shells (Fig. 1). The censuses showed that older individuals clustered more tightly than younger ones around mean

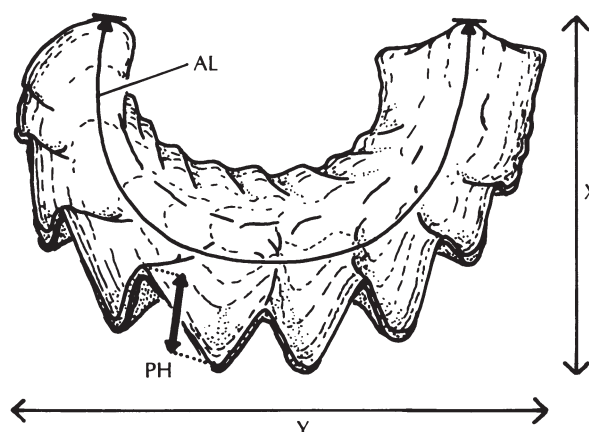


Fig. 1 *Agerostrea mesenterica*. Morphometric parameters measured by Sambol & Finks (1977): shell arc length (AL); maximum plical height (PH); number of anterior plicae (in this case, 8); and curvature index (Y/X). Exhalent flow would have issued around the concave posterior part of the shell.

values for the number of plicae and, with some unavoidable bias from ontogeny, for plical size as well as the overall arc length of the shell, indicating centripetal (stabilizing) selection on these features. The arcuate shape of the shell, in contrast, was subject to differential mortality favouring maximum curvature; i.e. directed selection had operated.

The features investigated all had well established functional linkages with gill suspension feeding, detected from comparisons with living oysters. In particular, the selection for increased curvature of the arcuate shell would have maximized the velocity of the exhalent current, thereby reducing the chances of recycling the processed water through the gills. This then is a clear 'snapshot' record of adaptation by natural selection in a fossil population. As stressed by Sambol & Finks, however, the data only show the time-averaged pattern of selection on the several generations of oysters comprising the assemblage, which probably accumulated over some 200 years. Although the selection for increased curvature is consistent with the morphological trend shown by successive species of the oyster's inferred phylogenetic lineage, only a small fragment of the history of natural selection operating in this case has been sampled.

Indeed, because of the virtually insurmountable practical difficulties attached to linking longer-term evolutionary changes in the fossil record with measurable natural selection, palaeontologists have continued to use the term 'adaptation' in the sense generally adopted by Darwin, stressing functional

suitability, rather than in direct reference to effects on neo-Darwinian fitness values. This distinction is important in that it lays some palaeontological perceptions of adaptation open to deserved criticism. The danger is that the morphology of a fossil organism can too easily become atomized in the mind of the palaeontologist to so many discrete components, to each of which a function is imaginatively assigned according to the apparent suitability of its morphology. The implicit assumption is that every feature must serve some function, or it would not be there. So, if one story is found wanting, another can be slipped into its place. This reductionist approach, branded as 'the adaptationist programme', has been criticized by Gould & Lewontin (1979) for proliferating adaptive hypotheses ('Just So Stories') where none may be warranted.

Many features are simply the geometrical consequences of the way organisms grow, and need no functional explanation *per se* (see also Section 4.1). For example, any given point on the aperture of a *Nautilus* shell traces a near perfect logarithmic spiral with growth. One could devise all manner of specious arguments for how this might be 'adaptive', and the precision with which this geometry is maintained might then be considered evidence for stabilizing selection. However, a brief consideration of the way the shell grows demolishes such arguments. If shell incrementation proceeds at fixed rates around the growing aperture of an expanding coiled tubular shell, logarithmic spiral growth is the geometrical consequence. That aspect of the *Nautilus* shell needs no adapting in order to arise, and so functional explanations for it are redundant; indeed, modification of the growth mechanism itself would be necessary to escape from such a geometry.

It is thus imperative, as stressed above, always to consider the whole organism as a developing entity, in its environmental context. Bits and pieces cannot be interpreted in isolation. Nevertheless, within that 'holistic' framework it is not only legitimate, but pragmatic to consider how particular components might have contributed to the overall conduct of an organism's life, by virtue of adaptive modification from some constructional groundplan (Mayr 1983).

Having dispensed with the erroneous reductionist demand for a function for every feature, we must now ask: how is original function to be detected at all in fossil organisms (see also Section 4.1), and by what means can adaptation for it (in the operational sense) be diagnosed? Three steps are

necessary: (1) from a consideration of what is known of the organism's affinities, construction, and autecology, a plausible function, or alternative functions, may be proposed for a given feature or set of features, where a constructional argument alone seems inadequate. Such hypotheses may be suggested by comparisons with similar living organisms, the design attributes of analogous machines, or even simply from theoretical considerations; (2) the suitability of the feature's construction (within the constraints determined by the organism's bodyplan and mode of growth) for the proposed function must be tested. This is done by comparing it with an idealized model (*paradigm*) designed for that function, to see how effective the feature would be in its service; and (3), crucially for the confirmation of adaptation, evidence that the feature has indeed been modified from some different ancestral condition, so as to approach the form of the paradigm, must be discovered, to show that the feature in question probably did perform the function attributed to it.

Testing for adaptive convergence with a paradigm really only requires an evolutionary sequence of specific modifications to be established, and this can be derived even from an outline phylogeny. There should also be reasonable evidence that the feature(s) in question consistently served the same broad function. Not all features lend themselves to such a broad-brush approach, of course, as the discussion of *Cepaea* polymorphism above illustrated. Others, however, usually concerned with such basic operational functions as feeding and locomotion, can be relied upon with greater confidence. A good example of such a test is Chamberlain's (1981) study of streamlining and static stability in ammonoids. Several lines of evidence suggest that the smooth-shelled ammonoids which he studied maintained the swimming habit. From an experimental study of accurately constructed models he was able to draw contours of drag coefficient and static stability values on a graph of possible shell shapes, and so to identify two 'adaptive peaks' where these factors were most favourable for efficient swimming. Real ammonite data show an impressive migration to the higher of these peaks with time, providing strong circumstantial evidence for adaptation.

The origin of adaptations

Ultimately, however, one is faced with the question of how an adaptation arose in the first place. Natural

selection certainly provides a mechanism for adaptation once a functional effect has become apparent, but biologists can only speculate about how a feature might have become involved in this process, from comparative studies of living forms. Here palaeontologists come into their own, for the fossil record furnishes the only concrete record of evolutionary history. But the virtual impossibility of characterizing the role of selection in much of this means that arguments about the origins of adaptations have to be cast in terms of structural change and likely functional consequences.

Two modes of origin are conceivable: either the feature newly appeared, or it was derived from some precursor. The former implicates a 'hopeful monster', presumably generated by some macro-mutation. Quite apart from the vanishingly small probability of such an extreme mutation yielding a fitter genotype, the main problem with the hopeful monster model is its untestable nature. Recognition of a homologous precursor to an adapted feature, on the other hand, allows testing for functional co-option. The point to be established is that the precursor did not originally serve the eventual function, but that with some slight modification (for whatever reason) or change in environment, it fortuitously manifested effects similar to the eventual function, for which it thus became adapted. Such a precursory feature is conventionally termed a *preadaptation* and such, for example, would have been the grasping hands of human ancestors, for tool use, when liberated by bipedalism. Some biologists instinctively recoil from the term, believing it to smell of teleology. Its literal ambiguousness (it might be construed, erroneously, as referring to some mysterious process of adaptive priming prior to the acquisition of a function) is a trivial problem, for which the remedy is simply learning the correct definition, with its crucial reference to *fortuitous* co-option. A second complaint, that it does nevertheless seem to suppose some end-directed evolution in its reference to eventual function, is a misunderstanding rooted in the essentially different working methods of many biologists and palaeontologists. Biologists investigating microevolution can directly analyse processes, but frequently stress the unpredictability of the longer term outcomes of evolution because of the myriad influences at work. Palaeontologists are in the opposite situation, knowing (some of) the outcomes of evolution, but with little direct evidence for the processes involved. The apparent 'end-directedness' of a preadaptive hypothesis is simply

the benefit of hindsight. While the danger of supposing history to have been inevitable must be avoided, it is legitimate to try to determine *in retrospect* at least some of the more prominent factors which made it take the unique course that it did follow. Any explanatory hypothesis must be tested against other historical models (in much the same way that Sherlock Holmes might have reconstructed the true nature of a crime).

In testing a preadaptive hypothesis, it is necessary to predict (strictly, retrodict) in detail the probable historical outcome of that model, beyond what has already been established, and to show how the retrodictions of alternative models significantly differ. Closer inspection of the fossil record can then point to one or other model being the more probable, perhaps on a statistical basis. Skelton (1985) adopted this approach in testing a preadaptive hypothesis for the evolution of rudist bivalves (Fig. 2). Constructional analysis had suggested that the spirogyrally coiled primitive forms were constrained by their growth geometry from exploiting the 'adaptive zones' (broad styles of adaptive morphology) occupied by their uncoiled tubular descendants. Shortening and eventual invagination of the external ligament in some spirogyrate forms had been identified as the preadaptive step which allowed the constructional changeover to uncoiled growth. The retrodiction of this model was that uncoiled taxa should have undergone an initially exponential diversification, focusing on the inception of ligamentary invagination, unmatched by their contemporaneous spirogyrate cousins. Other historical models (including a null hypothesis of random speciation and extinction) gave different retrodictions. An analysis of stratigraphical range data yielded the pattern given by the preadaptive hypothesis, at the generic level (though the species data were less enlightening, probably because of preservational bias), and this was taken to confirm the novel adaptive exploitation of the preadapted condition in the uncoiled clade.

Terminology

Gould and Vrba (1982) have expressed dissatisfaction with this terminology for discussing the origin of adaptations. Noting the literal connotation of the word 'adaptation' to imply that something has been progressively 'fitted to' (*ad* + *aptus* means 'towards a fit') the execution of some function, they followed Williams (1966) in restricting the use of that term to those features which can be shown to

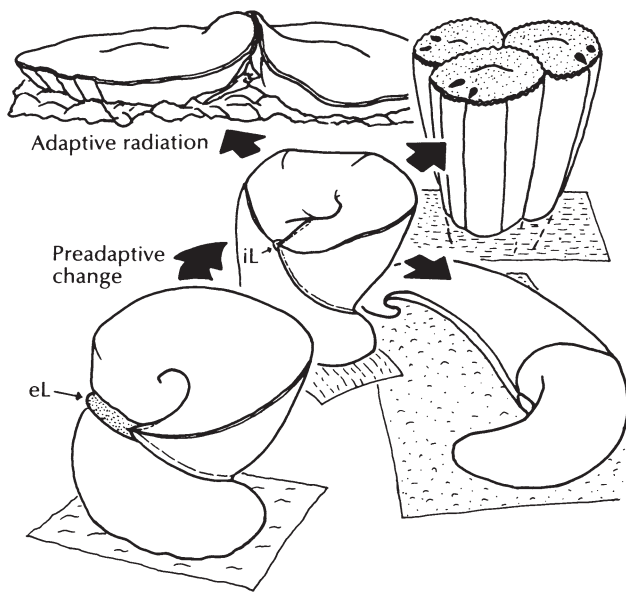


Fig. 2 Synoptic evolutionary history of uncoiling and its consequences in the rudists. (After Skelton 1985.) In primitive forms such as *Diceras* (lower left), the external ligament (eL) constrained the shell to grow spirogyrally, limiting its adaptive scope. Shortening and invagination of the ligament (iL) in *Monopleura* (centre) allowed 'uncoiled' growth. Adaptive diversification ensued (e.g. clockwise from top left, *Durania*, *Hippurites*, and *Pachytraga*) as uncoiled taxa entered new adaptive zones.

have been shaped by natural selection for their current use. Other features, which have some useful effect by virtue of their construction, but which show no clear evidence of having been produced by natural selection through the expression of that effect, they termed *exaptations* ('fit (*aptus*) by reason of (*ex*)'). However, they conceded that exaptations may undergo 'secondary adaptation', so enhancing their effectiveness. As an example, they cited the useful role of the skull sutures in young mammals in aiding parturition; these are also to be found in young birds and reptiles, where they obviously have no such role to play. Together, they designated (their) adaptations and exaptations as *aptations* — simply meaning features fitted to some function or effect.

Exaptations were seen as being co-opted ('co-optation') either from pre-existing adaptations for other functions, or from constructional elements with no previous functional effect ('nonaptations'). Both constitute forms of what has been labelled earlier, here, as 'preadaptation', though Gould and Vrba criticized this term. They argued that: (1) it fails to distinguish the two kinds of exaptation, and

so appears to have an adaptationist bias in hinting at an earlier adaptive role for the precursory feature; and (2) it is misconstrued, implying prior fitting towards some subsequent function (the teleological odour detected by some biologists).

Time will tell if these neologisms are adopted. However, their practical value has to be queried. To deal first with the more trivial aspect of etymological correctness, if the 'fitting-to' of adaptation is simply taken to mean 'suitability for', as is the implication both in common neo-Darwinian and Darwinian ('useful variations') usage, then preadaptation literally connotes no more than 'prior suitability for' (... some fortuitous functional effect). More importantly, Williams' (1966) distinction between 'effects' and 'functions' (discussed above), upon which Gould and Vrba base their 'exaptations' and 'adaptations', breaks down when origin of adaptations is considered. In so far as any adaptation is derived from a precursory feature (whether some preadaptive trait or even a mutational novelty), then the latter must have passed through the stage of being an exaptation — i.e. exhibiting fortuitous beneficial effects — to have become subject to the selection that produced the adaptation. But the very moment that such 'exaptive' benefits were expressed, fitness would have been affected and selection would have started adapting the feature. Thus, although skull sutures are indeed 'exaptive' for mammalian parturition, the extended delay in their closing up is clearly adaptive for that process. Gould and Vrba might term this a 'secondary adaptation', but surely this is no different from any 'primary adaptation' if we accept that all are founded on exaptations. The only reason we may choose to call one thing an exaptation and another an adaptation relates to the degree of modification shown. For example, it is easy to think of the skull sutures mentioned above, with the slight adaptive delay in their closure, as an exaptation, but the hooked beak of an eagle would be branded by most people as a clear adaptation for tearing flesh. Yet it is only the shape of the beak which is thus adaptive; the beak itself was, again, an exaptation for the role. In other words exaptation and adaptation are really just two aspects of the same thing, the former emphasizing derivation and the latter, destiny. To attempt to distinguish them as separate entities (which is implicit in any statement that some feature is an exaptation and not an adaptation or vice versa) seems to be as illogical as classifying the 'arrivals' and 'departures' at a railway station as two fundamentally different kinds of train.

Again, in view of the difficulties attached to detecting natural selection in fossil material, how practicable is it to attempt to distinguish between an 'exaptation' derived from a previous adaptation and one derived from a 'nonadaptation'? How could one show that some feature of a fossil organism (even if clearly a product of the mode of construction) *did not* somehow adapt the organism to its niche?

The conventional toolkit of terms — adaptation and preadaptation — appear to suffice for the nature of the material to be studied, with the proviso that they are used with well understood and precisely defined (preferably explicitly stated) meanings:

Adaptation. In neo-Darwinian usage, this is a feature or complex of features in an organism which promotes or sustains the (neo-Darwinian) fitness of its possessors; or, in a palaeontological context, which has some identifiable functional effect predicted to have been of selective benefit to its possessors (a prediction which if borne out by some means of analysis would render the feature a neo-Darwinian adaptation as well); or, in both cases, the associated historical process of modification of features in an evolving population.

Preadaptation. This is a feature or complex of features of an organism, whether already serving some functional role or merely a constructional product, which, by virtue of its fortuitous suitability for novel functional effects, becomes co-opted as a new adaptation (in the senses given above) in descendants of the organism.

It should be clear that, despite the apparently simple meaning of adaptation as a vernacular term, and its fundamental importance in evolutionary theory, it actually opens onto a terminological

minefield. Safe routes across can only be picked out by adhering to clear definitions and thinking very carefully about their practical applications.

References

- Cain, A.J. & Sheppard, P.M. 1954. Natural selection in *Cepaea*. *Genetics* **39**, 89–116.
- Chamberlain, J.A., Jr. 1981. Hydromechanical design of fossil cephalopods. In: M.R. House & J.R. Senior (eds) *The Ammonoidea*. Systematics Association Special Volume, No. 18, pp. 289–336. Academic Press, London.
- Dobzhansky, T. 1970. *Genetics of the evolutionary process*. Columbia University Press, New York.
- Gould, S.J. & Lewontin, R.C. 1979. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London*, **B205**, 581–598.
- Gould, S.J. & Vrba, E.S. 1982. Exaptation — a missing term in the science of form. *Paleobiology* **8**, 4–15.
- Grafen, A. 1984. Natural selection, kin selection and group selection. In: J.R. Krebs & N.B. Davies (eds) *Behavioural ecology: an evolutionary approach*, pp. 62–84. Blackwell Scientific Publications, Oxford.
- Jones, J.S., Leith, B.H. & Rawlings, P. 1977. Polymorphism in *Cepaea*: a problem with too many solutions? *Annual Review of Ecology and Systematics* **8**, 109–143.
- Lewontin, R.C. 1983. Gene, organism and environment. In: D.S. Bendall (ed.) *Evolution from molecules to men*, pp. 273–285. Cambridge University Press, Cambridge.
- Mayr, E. 1983. How to carry out the adaptationist program? *The American Naturalist* **121**, 324–334.
- Ridley, M. 1985. *The problems of evolution*, p. 26. Oxford University Press, Oxford.
- Sambol, M. & Finks, R.M. 1977. Natural selection in a Cretaceous oyster. *Paleobiology* **3**, 1–16.
- Skelton, P.W. 1985. Preadaptation and evolutionary innovation in rudist bivalves. In: J.C.W. Cope & P.W. Skelton (eds), *Evolutionary case histories from the fossil record*. Special Papers in Palaeontology No. 33, pp. 159–173.
- Williams, G.C. 1966. *Adaptation and natural selection — a critique of some current evolutionary thought*. Princeton University Press, Princeton.

2.10 Evolution of Large Size

M. J. BENTON

Introduction

Many plants and animals of the past and present are very large compared to the human scale. In particular, vertebrates, gymnosperms, and angiosperms achieved giant dimensions on occasion, and apparently several times independently in each group (Table 1; Fig. 1). The focus here, however, will be on truly large organisms on the human scale. The key macroevolutionary questions to be asked are:

- 1 Why do certain groups achieve giant size while others do not? Is it simply chance, or are there historical and mechanical reasons?
- 2 Why do some groups never produce giants?
- 3 Does evolution always go from small to large, or can it reverse?
- 4 How long does it take for large size to evolve in a lineage?
- 5 Are large organisms better adapted than small ones?

Giants and mechanical constraints

The bony internal skeleton of vertebrates is ideally suited to supporting great weights in terrestrial giants. The acquisition of a fully upright posture in both dinosaurs and mammals, where the limb bones are tucked immediately beneath the body, permitted giants to evolve.

The major constraints on large size in a terrestrial vertebrate are limits to the strength of bones and to the power of muscles. As animals become larger, the bones and muscles in the legs come under increasing strain, and there have to be modifications in their shape and design. Hokkanen (1986) made simple biomechanical calculations of bone and muscle strengths in order to determine the size of the largest feasible terrestrial tetrapod.

Each leg bone must be strong enough to support one-quarter of the total body weight, or more if the weight is concentrated at the back, as is often the case, and there has to be a fairly large safety factor in order to allow the animal to walk or run. The strength of a bone is proportional to its cross-sectional area (a two-dimensional measure), while

body weight is proportional to volume (a three-dimensional measure). Thus, bone cross-sectional area has to increase relatively faster than body weight, which is why elephants and dinosaurs have legs like tree trunks (Fig. 2). Under high stress, leg bones can buckle, or they can break without bending much. The strength of muscles also limits the size of an animal. A large animal has to be able to pull itself up from a lying position, and the heavier the animal is the more massive its muscles must be. So, muscle dimensions and muscle strength also limit the maximum size of a land animal.

Locomotion is yet another limiting factor. A hypothetical animal weighing 140 tonnes could stand safely enough, but if it walked its legs would break. This is because, in walking, the force of the weight of the animal is expressed at an angle through the leg bones. Even if a giant animal could stand safely with its legs positioned vertically beneath it, it might not be able to walk because the breaking force of the bone is relatively greater. Hokkanen (1986) concluded that the heaviest possible animal able to walk on four legs would have weighed no more than 100 tonnes.

The largest dinosaurs have estimated weights in the range of 80–140 tonnes, but the larger forms are poorly known. The 78 tonne weight of *Brachiosaurus* is the greatest generally accepted weight known for a terrestrial animal. The strength of bone and muscle, as described above, would have limited *Brachiosaurus* to a sedate walking pace of about 1 m/s with strides of only 2.5 m or so (quite short for an animal with 3 m legs) (Alexander 1985).

In land plants, the continuously growing supporting tissues (lignin-lined xylem cells) within a tree trunk allow vast heights and weights to be achieved. The maximum height is probably limited in part by the ability of a plant to raise sap. Water has to be 'pumped' from the ground and raised up the trunk, against the force of gravity, by means of osmosis (the sap has a higher salt content than the ground water), and the hydrostatic effect of transpiration (water loss through leaves exposed to the air).

There are also mechanical constraints imposed by

Table 1 A selection of large organisms, giving some key dimensions. Fossil forms are preceded by †, and the weights quoted for these are estimates (a question mark implies that estimates are very uncertain because complete skeletons are unknown).

Organism	Max. length (m)	Max. height (m)	Max. weight (t)
<i>Plants</i>			
<i>Algae</i>			
<i>Macrocystis</i> , Pacific giant kelp	60	—	—
<i>Gymnospermophyta</i>			
<i>Sequoiadendron</i> , Giant sequoia	—	84–112	c. 2500
<i>Pseudotsuga</i> , Douglas fir	—	126.5	—
<i>Angiospermae</i>			
<i>Eucalyptus</i> , Mountain ash	—	114.3	—
<i>Animals: Vertebrata</i>			
<i>Class Placodermi</i>			
<i>Dunkleosteus</i>	9	—	—
<i>Class Chondrichthyes</i>			
<i>Cetorhinus</i> , Basking shark	10.5	—	—
† <i>Carcharodon</i>	13	—	—
<i>Rhincodon</i> , Whale shark	12.6	—	15
<i>Class Reptilia</i>			
<i>Suborder Squamata</i>			
<i>Eunectes</i> , Anaconda snake	8.4	—	0.23
<i>Python</i>	10	—	—
† <i>Kronosaurus</i> , Pliosaur	15.2	—	—
<i>Suborder Crocodylia</i>			
† <i>Deinosuchus</i>	16	—	—
<i>Suborder Pterosauria</i>			
† <i>Quetzalcoatlus</i> wing span	11–12	—	0.09
<i>Suborder Dinosauria</i>			
† <i>Brachiosaurus</i>	23–27	12	40–78
† <i>Diplodocus</i>	27	—	18.5
† <i>Antarctosaurus</i>	30	—	80
† ‘ <i>Supersaurus</i> ’	?24–30	?15	?75–100
† ‘ <i>Ultrasaurus</i> ’	?30–35	?16–17	?100–140
† ‘ <i>Seismosaurus</i> ’	?30–36	—	80+
<i>Class Mammalia</i>			
<i>Order Perissodactyla</i>			
† <i>Indricotherium</i> (= <i>Baluchitherium</i>)	11.3	c. 6	20
<i>Order Artiodactyla</i>			
<i>Giraffa</i> , Giraffe	—	5–6	—
<i>Order Proboscidea</i>			
<i>Loxodonta</i> , African elephant	7–10	3–4.4	2–10
<i>Elephas</i> , Indian elephant	6	3	4
<i>Order Cetacea</i>			
<i>Balaenoptera</i> , Blue whale	33.5	—	190
<i>Physeter</i> , Sperm whale	20.7	—	—
† <i>Basilosaurus</i>	21.3	—	—

the vast weight of a tall tree and the possible strength of its trunk. The weight acts vertically down the trunk, but winds can cause tremendous stresses as the crown of a tree is pushed from side to side. Experiments show that winds with speeds of 60–65

km/h exert a lateral force on the tree equal to its weight (Fraser 1962). The girth of the tree then increases in proportion to the weight (i.e. relatively more rapidly than the height increases). At 100 m tall, a tree may be as much as 30 m in circumference

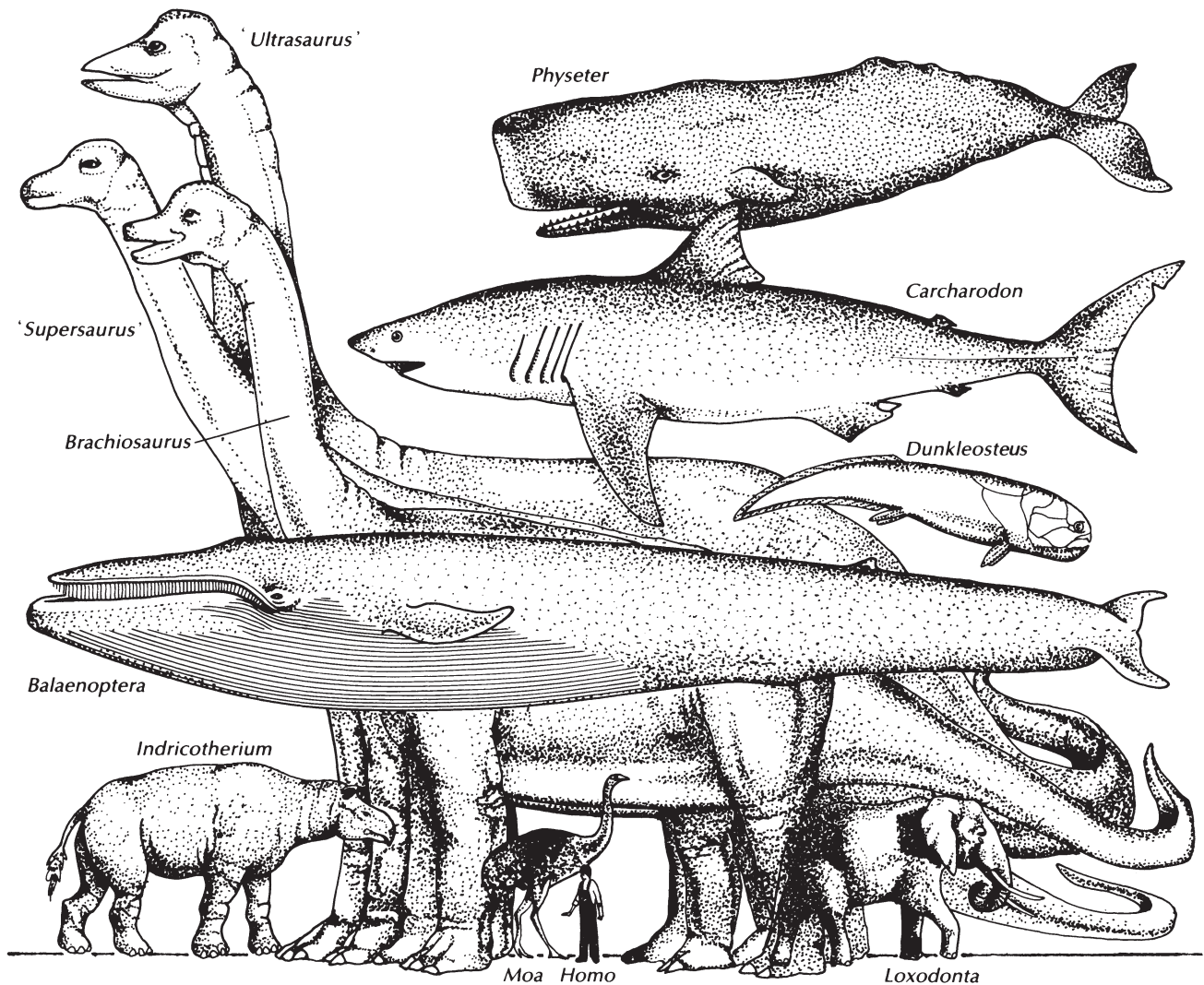


Fig. 1 A selection of large animals drawn to scale. Measurements are given in Table 1. (Drawing by Elizabeth Mulqueeny.)

(Table 1), and at much greater heights, the circumference would tend to approach the height.

Why so few giants?

Most other groups of organisms appear to be restricted from achieving large size by mechanical and physiological constraints. For example, arthropods have an external skeleton which has to be moulted frequently as the animal grows. After each moult, the animal is soft-bodied for a while, and hence vulnerable. The shed skeleton also represents a loss of body materials that have to be replaced. To achieve giant size, an arthropod would suffer the cost of moulting dozens of times. A more important constraint on large size is probably the respiratory system of tubes in the exoskeleton that allow air to

diffuse throughout the body passively. At moderate to large size, this technique would not allow all body tissues to receive an adequate supply of oxygen.

There are similar constraints on large size in most other invertebrates — e.g. the respiratory system of annelids and nematodes (simple diffusion into the body); the filter-feeding habits of brachiopods, most molluscs, coelenterates, bryozoans, graptolites, and some echinoderms; and mechanical constraints of the exoskeleton of brachiopods, most molluscs, and most echinoderms. It is assumed that filter-feeding by means of exposed cilia cannot sustain a large organism. The shells of brachiopods and molluscs can reach large sizes (e.g. the giant clam, *Tridacna*, 1 m across), but as body size increases, shell thickness has to increase in proportion to body weight to

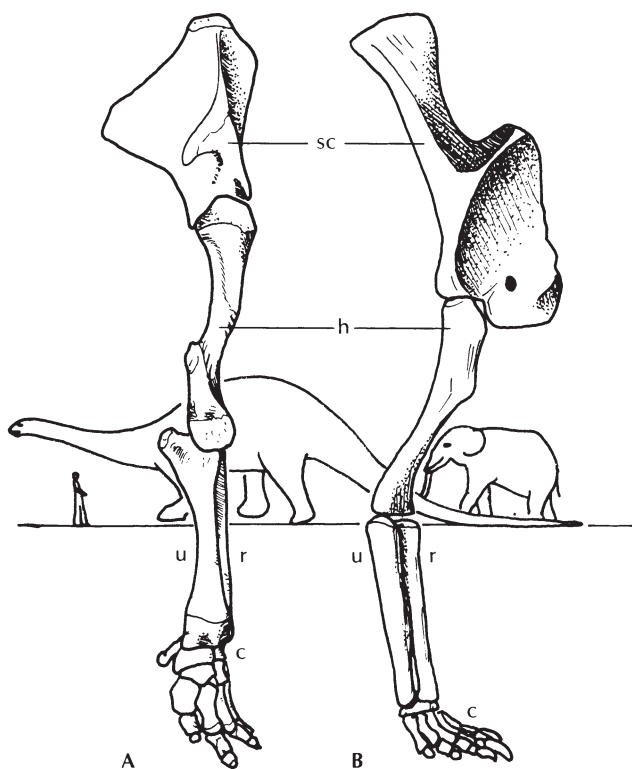


Fig. 2 The pillar-like skeleton of the forelimb of A, *Elephas*, the Indian elephant and B, *Diplodocus*, a sauropod dinosaur, showing convergent graviportal (weight-bearing) adaptations: columnar arrangement of shoulder girdle (sc= scapula) and limb bones, relatively long humerus (h), large separate radius (r) and ulna (u), block-like carpal bones (c), and relatively short finger bones spreading out over a cushioning pad.

maintain the strength of the shell. The potential weight of the shell, and the amount of particulate calcium carbonate to be extracted from the seawater, tend to prevent huge size. The same is probably true for echinoids.

Cope's Rule

In 1887, E.D. Cope presented a new principle of evolution, that organisms always tend towards large size. He could find no examples in which a lineage or clade of plants or animals evolved towards smaller size. Although Cope never explicitly defined this as a 'law' of evolution, it has since come to be known as Cope's Rule.

In considering Cope's Rule, many authors have focused on particular advantages of evolving large size (see below). However, Stanley (1973) argued that Cope's Rule had general application, not because of any particular advantages of large size, but since groups tend to arise at small body size relative

to their ecological optimum. Amongst mammals, for example, the original members of most clades in the Cretaceous and Palaeocene were small carnivores or insectivores. On the other hand, large forms are unlikely ancestors for major new lineages since they tend to be specialized to particular habitats, often by virtue of the physiological demands imposed by large size. Stanley (1973) surveyed a range of animal taxa, and found that the ancestors of a clade were generally smaller, on average, than a random sample of their descendants. Histograms of body size tended to be concentrated initially at small sizes and to be rather symmetrical. Through time, the histograms developed longer and longer tails to the right as larger body sizes arose (Fig. 3).

Size decrease also does take place in many lineages, but it is rare. For example, modern horsetails and clubmosses are midgerts in comparison with their Carboniferous tree-like ancestors. Certain vertebrate groups have also shown reductions in size since the Pleistocene, but some of the former giants (e.g. mammoth, aurochs, giant kangaroo and wombat, giant ground sloth, glyptodon, moa) may have suffered because of human influence (see also Section 2.13.8).

Evolution of large size

The evidence of the fossil record is that giant size can evolve very quickly in certain groups. For example, the first (small) dinosaurs of the late Triassic date from the Carnian. By mid-Norian times, 5–10 million years later, prosauropods such as *Plateosaurus* had reached body lengths of 5 m. The sauropodomorph line then achieved a length of 12 m with *Melanorosaurus* in the Early Jurassic, and sizes continued to increase rather slowly until the Late Jurassic when the largest known dinosaurs occurred (Table 1). This last phase of size increase towards giantism — a leap from body lengths of about 12 m and weights of 10 tonnes to maxima of 30 m and 80 tonnes or more, occurred between the Bathonian and the Kimmeridgian, a time of about 20 million years.

Mammals achieved large size just as rapidly, if not more so. From a maximum of cat size just before the end of the Cretaceous, rhinoceros-sized uimatheres and astrapotheres are known 10 million years later in the Late Palaeocene and Early Eocene. The largest land mammal of all time, the rhinoceros *Indricotherium*, was in existence by the Early Oligocene, 30 million years after the radiation began. Whales achieved large size even more

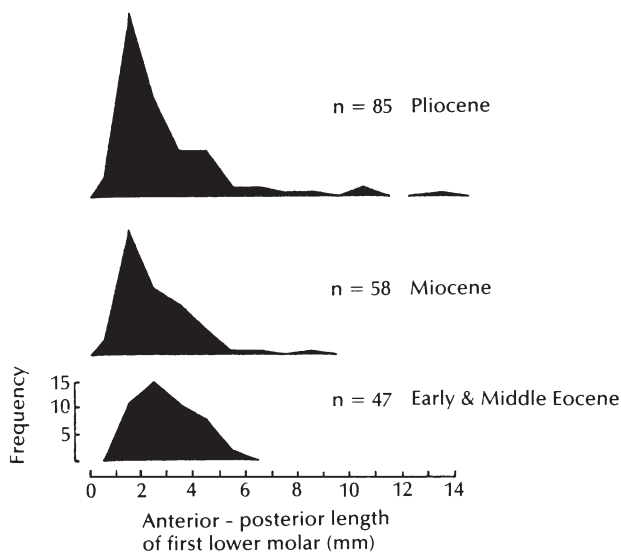


Fig. 3 The size ranges of North American rodents — an early group, and two later groups — to show the shift from small sizes to a broad range of body sizes including many large ones. The index of size is the length of the first lower molar, which varies directly with overall body size. (After Stanley 1973.)

rapidly — the Late Eocene *Basilosaurus* was 21 m long, after 15–25 million years of evolution.

Stanley (1979) noted that, in contrast, ‘large’ molluscs took much longer to evolve. The first large free-swimming clam was *Megalomoidea* which appeared after nearly 100 million years of radiation. The first large epifaunal bivalves, the inoceramid rudists of the Jurassic and Cretaceous, took nearly 400 million years to appear.

Amongst land plants, large size arose at the end of the Devonian, and especially in the Carboniferous, with the first tree-like clubmosses (*Lepidodendron*, 45 m high) and horsetails (*Calamites*, 16 m high). This had taken 50–60 million years of land plant evolution. Really giant gymnosperms (*Sequoia* and other redwoods) are known from the Jurassic, as much as 250 million years after the origin of land plants, and 150 million years after the origin of gymnosperms.

Advantages and disadvantages of large size

Numerous advantages of large size have been postulated (Stanley 1973): improved ability to capture prey or escape from predators, greater reproductive success, increased intelligence (large bodies have large brains), better stamina, expanded size range of possible food items, decreased annual

mortality, extended individual longevity, and increased heat retention per unit volume. Protection from predation would seem to be a great advantage. Adult elephants and rhinoceroses have no regular threat from carnivores today. However, thick-skinned mammals of the Oligocene to Pleistocene of the Northern Hemisphere and South America were subject to attacks by specially adapted sabre-toothed cats — the Machairodontidae in North America, Europe, Africa, and Asia, and the Borhyaenidae in South America. The sauropod dinosaurs are assumed to have been immune from attack since the largest predatory dinosaurs could only have tackled very young sauropods, or dying adults.

A disadvantage of large size may be greater proneness to extinction. This is not simply an attribute of large size, but rather an expression of specialization. Large animals are often more restricted in their niches, in their scope for adaptation, than smaller relatives. Their need for large amounts of food, or for particular environmental conditions, may make them more likely to suffer when habitats change. Also, the fact that large animals tend to have small population sizes, and hence small gene pools, makes their hold on life seem more precarious. The death of a few more individuals than normal may precipitate species extinction.

Bakker (1977) showed that terrestrial tetrapods surviving mass extinctions in the Late Palaeozoic and Mesozoic tended to be of small body size. Thus, the large dicynodonts and dinocephalians of the Late Permian died out, leaving smaller dicynodonts and cynodonts to cross the system boundary. A similar explanation has also been given for selectivity in the Cretaceous–Tertiary event on land (Section 2.13.7). In more general terms, Stanley (1979) suggested that species longevity varies with the reciprocal of body size: small species tend to survive longer than large species. This is supported by evidence from the Pliocene and Pleistocene mammalian fossil record. The only modern species that can be tracked back before 3 Ma are small mammals. All the large ones arose after that, and this is probably not an artifact of a poor fossil record since such forms are more readily fossilized than small ones.

Within any clade, lineages of large organisms may be expected to display shorter taxon durations, lower rates of speciation, and higher rates of extinction (Stanley 1979), and hence greater volatility in the face of environmental stress. These ideas have yet to be tested thoroughly. They are of added interest since they could be seen as charac-

teristics that are subject to species selection (since these are not organism-level features). They could also potentially be interpreted as examples of the 'effect hypothesis' (Vrba 1983; see also Section 2.6). This hypothesis suggests that species-level characteristics, such as species duration or broad ecological adaptation, may be incidental effects of individual characters, such as dietary or habitat preferences. Natural selection, acting on organisms, might select for large body size, which in turn might produce higher extinction rates within a lineage. These higher rates could be interpreted as an incidental *effect* of natural selection, rather than as a result of species-level selection. These ideas are still highly controversial.

References

Alexander, R.McN. 1985. Mechanics of posture and gait of

some large dinosaurs. *Zoological Journal of the Linnean Society* **83**, 1–25.

Bakker, R.T. 1977. Tetrapod mass extinctions — a model of the regulation of speciation rates and immigration by cycles of topographic diversity. In: A. Hallam (ed.) *Patterns of evolution as illustrated by the fossil record*, pp. 439–468. Elsevier, Amsterdam.

Fraser, A.I. 1962. Wind tunnel studies of the forces acting on the crowns of small trees. *Reports on Forest Research* **1962**, 178–183.

Hokkanen, J.E.I. 1986. The size of the largest land animal. *Journal of Theoretical Biology* **118**, 491–499.

Stanley, S. M. 1973. An explanation for Cope's Rule. *Evolution* **27**, 1–26.

Stanley, S.M. 1979. *Macroevolution: pattern and process*. W.H. Freeman, San Francisco.

Vrba, E.S. 1983. Macroevolutionary trends: new perspectives on the roles of adaptation and incidental effect. *Science* **221**, 387–389.

2.11 Rates of Evolution — Living Fossils

D. C. FISHER

Introduction

The study of rates of evolution encompasses a wide variety of approaches to characterization of the amount of evolutionary change within particular groups of organisms, over specified time intervals. The high level of interest that palaeontologists and evolutionary biologists have shown in this subject is not surprising, since rates are a common focus in the analysis of any process. The importance of rates, however, is often only marginally attributable to intrinsic interest in 'how rapidly' or 'how slowly' a process operates. Rather, information on rates tends to be used as a means of investigating the underlying dynamics of the process in question, or sometimes as input for analysing the dynamics of a related process. Much of the work on rates of evolution has thus been directed toward a better understanding of the dynamics of evolutionary change. Studies have been designed with the intent of comparing rates of evolution in a variety of ways — within and between particular taxonomic groups, ecological settings, and lineage geometries (e.g.,

ancestor–descendant sequences that include lineage splitting versus ones that do not). While interesting generalizations are emerging, a greater appreciation is also being gained of the difficulties of quantifying rates of evolution.

'Living fossils' is a term frequently used to denote extant representatives of groups of organisms that have survived with relatively little change over a long span of geological time. Such groups are implicitly recognized as having displayed unusually low rates of evolution. In both professional and popular literature, living fossils collectively appear to have attracted more attention than have groups displaying unusually high rates of evolution. This may be partly because, in keeping with the inherent paradox of the term 'living fossil', evolutionary history is expected to involve conspicuous change, and it is surprising when it does not. In addition, evolutionary rate statements are commonly (though not exclusively) framed in terms of putative ancestor–descendant pairs, and it is easier to recognize these when the total amount of change has been small than when it has been large. Instances of

living fossils are thus more likely to be accepted on *prima facie* grounds than are instances of higher evolutionary rates. In any event, living fossils have frequently provided a focus for discussions of evolutionary rate and have helped to clarify some of the factors that may be involved in promoting or inhibiting evolutionary change.

Three ranges of values for evolutionary rates

G.G. Simpson was one of the early major contributors to the quantitative study of evolutionary rates. He proposed that rates be classified by their absolute value as 'low', 'medium', or 'high'. Although this might be considered trivial, Simpson (1944, 1953) argued that frequency distributions of evolutionary rates for sufficiently inclusive sets of taxa typically contain three discrete modes, allowing low, medium and high categories to be recognized on non-arbitrary grounds. This empirical claim suggests some degree of disjunctness in the operation of the processes and/or constraints that interact to produce evolutionary change. Simpson coined the term 'bradytely' to refer to the phenomenon of supra-specific taxa that have shown consistently low rates of evolution. Bradytely thus encompasses the same general concept implied by 'living fossil', but without the arbitrary stipulation that a representative of the group be alive today. Simpson also suggested 'horotely' to refer to taxa comprising the middle mode in the spectrum of observed evolutionary rates and 'tachytely' to refer to supraspecific taxa showing consistently high rates of evolution. Although Simpson's (1953) demonstration of the multimodality of evolutionary rates has subsequently been shown to be flawed (Gingerich 1983; Stanley 1985), the terms denoting these rate categories (especially bradytely and tachytely) have had considerable heuristic value. They are now commonly used to refer to ranges of rate values regardless of whether multimodality has been demonstrated independently. For instance, in a study applying the terms in this latter fashion, Raup and Marshall (1980) showed that rates within several orders of mammals were significantly higher (e.g. Cetacea and Rodentia) or lower (e.g. Perissodactyla and Carnivora) than the mean for all mammalian orders. However, whether evolutionary rate distributions (at a given rank, within some more inclusive group of organisms) tend to show some 'typical' form and, if so, whether that form is multimodal, unimodal but non-normal, or unimodal and normal, are presently open questions.

Qualitative categories of evolutionary rates

Evolutionary rates may also be categorized by the aspect of evolutionary change that is measured. Three commonly discussed categories are genetic, morphological, and taxonomic rates. However, various subdivisions of each of these are also significant. For instance, genetic rates include rates of DNA nucleotide substitution and rates of gene rearrangement, among others. These two kinds of rates refer to different processes of genetic change, acting at different levels in the hierarchy of genetic structure. Each offers its own perspective on the general phenomenon of evolutionary change, and it is conceivable that each will show a different frequency distribution, even over the same large group of taxa. In the same way, morphological rates are sometimes subdivided into 'size' rates and 'shape' rates, since these two factors are commonly treated as different, though not unrelated, aspects of morphology. Finally, taxonomic rates include various approaches to measurement of the longevities and rates of origination and extinction of taxa. Terminology for categories of taxonomic rates varies somewhat among authors, and each category may be further subdivided according to the taxonomic rank treated. In each case, the meaning of such rates depends critically on the underlying taxonomic philosophy. The type of taxonomic rate that will be focused on here is the rate of origination of new taxa of specified rank, since this corresponds most closely to a 'rate of evolution' (i.e. without introducing aspects of extinction rate).

In this categorization of evolutionary rates, genetic and morphological rates refer to changes in the genotype and phenotype, respectively. An alternative convention is to distinguish between molecular and morphological rates of evolution. This retains all aspects of genotypic change within molecular evolution, but adds to it components of protein evolution that would ordinarily be considered changes in the phenotype, albeit at a molecular level.

Although molecular data are usually available only for living organisms, increasing effort is being focused on extraction of some molecular data from appropriately preserved fossil material (Section 2.1). Still, except for the success of such efforts, molecular rates can only be measured directly over relatively short timespans. Alternatively, they may be computed from the cumulative divergence of contemporaneous taxa. In this case, some parsimony assumption is used to partition change between or among the separate lineages involved. Although

this approach may seem to remove molecular rates from the domain of palaeontology, we must still relate measured divergence to the time interval over which it has developed — the time since the most recent common ancestor of these taxa. Tectonic or palaeogeographical data suffice for this in certain instances, but palaeontological data provide the most commonly applicable constraints on the time of splitting of lineages. For this reason, and because of their common focus on analysis of the pattern and process of evolution, palaeontology and studies of molecular evolution are closely related (Section 2.1).

Measurements of morphological rates may also be based on comparisons among contemporaneous taxa for which the divergence history is relatively well known. However, when morphological features can be sampled in a succession of stratigraphic intervals, we have the option of calculating rates 'directly' from the fossil record. Since any source of morphological disparity between samples will contribute to perceived evolutionary rate, it is important to be aware of, and if possible control for, non-evolutionary components of variation (e.g. differential ontogenetic representation, differential taphonomic biases, or range shifts in clinally varying populations). If it can be argued that consecutive samples represent a series of ancestors and their descendants within a species-level lineage — an ideal situation that approximates 'tracking' morphology through time — the resulting rate is referred to as a 'phyletic' rate. However, if the phylogenetic context of consecutive samples is more complex or unresolved than this, the rate is better referred to as a 'phylogenetic' rate (Raup & Stanley 1978). Phylogenetic rates imply a disclaimer recognizing that increments of change may have been measured between samples that do not bear a direct ancestor–descendant relation to one another. Depending on the history of morphological change and the pattern of phylogenetic relationships linking consecutive samples, phylogenetic rates may be either greater or less than corresponding phyletic rates (i.e. the phyletic rates that might be measured if an arguably ancestor–descendant sequence were available). Both of these types of rate represent transformation within a 'lineage' (broadly construed, possibly at a supraspecific level), but they differ in the degree of resolution with which the lineage can be traced.

Taxonomic origination rates are likewise designed to quantify change through time, but they differ fundamentally from the rates discussed thus far. To the extent that new taxa are erected to recognize some increment of morphological change

within lineages, origination rates incorporate a transformational component comparable to that assessed by molecular and morphological rates. However, origination rates also include a component representing the cladogenetic (or lineage splitting) aspect of evolutionary change. The relative contributions of these two components — lineage transformation and lineage splitting — are difficult to quantify and rarely reported. They vary from group to group depending both on taxonomic practice and on the actual evolutionary history of the group under study.

Units of measurement for evolutionary rates

Genetic or molecular rates are sometimes quantified in terms of the number of events involving a particular type of change, per time interval. Comparisons of molecular rates may be normalized for the number of entities 'at risk' for change (e.g. number of nucleotide substitutions per site, per million years), but this is not practical in all instances (e.g. computing the number of potential gene rearrangements). Molecular rates based on distance measures (e.g. DNA–DNA hybridization, immunological distance) are given in units appropriate to the distance measure utilized.

Morphological rates may be expressed as change in the value of some morphological variable (any appropriate units of measurement), per time interval. However, variables of different dimensionality (e.g. lengths versus areas) must be divided by an appropriate factor before they can be properly compared. Moreover, we are usually interested in proportional rather than absolute changes in morphology. Given the scaling relationships of most morphological variables (and their variances), a convenient solution is to measure morphological rates in terms of differences in the logarithm of the value of the variable of interest. A difference of a factor of e (base of natural logarithms, 2.718) per million years was defined by Haldane (1949) as a morphological rate of 1 darwin (d).

Rates of taxonomic origination may be measured as the number of new taxa (within a given higher taxon) per time interval. This is often expressed as a percentage increase, normalized for the length of the time interval. Rate of origination may also be calculated from the rate of change in total diversity at a given taxonomic level and the rate of extinction at that level. In interpreting origination rates, it is important to consider such possible complications as differential effects of taphonomic and mono-

graphic biases, and differential application of taxonomic practice within and between groups being compared (Raup & Marshall 1980). From an evolutionary standpoint, however, a more fundamental issue with rates of taxonomic origination is that they lump together information on lineage transformation and lineage splitting. Given the current unevenness of our detailed phylogenetic knowledge of most groups, this may be an unavoidable compromise, and indeed, it offers some benefits of convenience and succinctness in the representation of evolutionary history. However, it is to be hoped that more phylogenetically discriminating approaches to studying diversification will be developed in the future.

The effect of measurement interval on evolutionary rates

Measured rates are commonly treated as independent of the interval length over which they are measured. For processes occurring at approximately constant rates, this characterization is acceptable. However, for any variable-rate process, the measured rate is an average and may be influenced strongly by rate fluctuations during the measurement interval. Depending on the temporal structure of rate fluctuations and the range of intervals being considered, measured rates will be more or less susceptible to biasing effects from interval length.

Some molecular rates appear to behave in 'stochastically constant' fashion, at least over certain time-spans (commonly of the order of tens of millions of years). The relative constancy of these rates (with both rate and constancy varying from one molecular system to another) has led to the proposal of the 'molecular clock' hypothesis (see also Section 2.1). According to this hypothesis, molecular difference, once calibrated to reflect rate of change, can be used as a measure of time since lineage divergence (Fitch 1976). However, even for molecular clocks that are relatively 'well behaved' over a particular time interval within a given group, there is growing evidence that observable change has either accelerated or decelerated at other times during the history of that group (Goodman *et al.* 1982; Gingerich 1986). For divergence times that span periods of significant rate change, systematic biases can be anticipated.

The factors thought to influence morphological rates (see below) are known to fluctuate on a variety of time-scales. Because neither the highest nor the lowest rates are likely to be maintained over protracted periods of time, the largest range of variation

should be observed in comparing rates measured over the shortest time intervals. For the same reason, there should be a tendency toward intermediate values, which are due to averaging of rate fluctuations, when measuring over longer intervals. Since morphological rates are typically expressed in terms of net change in the value of some morphological variable, changes in the *direction* of morphological change, as well as in the rate of change *per se*, contribute to the moderation of rates measured over longer time intervals. This interaction is partly responsible for the decline in maximum observed morphological rates with increasing measurement interval (Fig. 1). However, as Gingerich (1983) pointed out, the lower, and to some extent the upper bounds of the distribution of observed rates in Fig. 1B are also influenced by factors unrelated to evolutionary process. The lower bound corresponds to a practical limit of measurement precision, beyond which earlier and later forms would not usually be recognized as different, yielding a rate of zero. The upper bound, on the other hand, represents an effective limit beyond which pairs of earlier and later forms differ so strongly that their relationship, and hence their appropriateness for a rate calculation, is likely to be questioned. The result is a tendency for longer measurement intervals to yield lower rates. Because of these biasing factors, comparison of rates measured over very different time intervals is a non-trivial problem. Many comparative studies of evolutionary rates have not adequately dealt with this issue.

Taxonomic rates are also affected by measurement interval, but not in all the ways noted above. As with morphological rates, rates of origination calculated over longer intervals are likely to be damped by averaging a range of shorter-term values. However, rates of origination are not moderated by changes in the 'direction' of evolution; 'new taxa' are new taxa, even if they show reversals in certain attributes. In addition, with rates of origination, low values do not suffer an interval-related bias based on measurement precision, nor do high values necessarily engender suspicion of lack of relationship.

The effect of stratigraphic completeness on evolutionary rates

Stratigraphic completeness (see also Section 3.12) could in principle affect the precision of palaeontologically documented divergence times, but in practice, phylogenetic uncertainties and disconti-

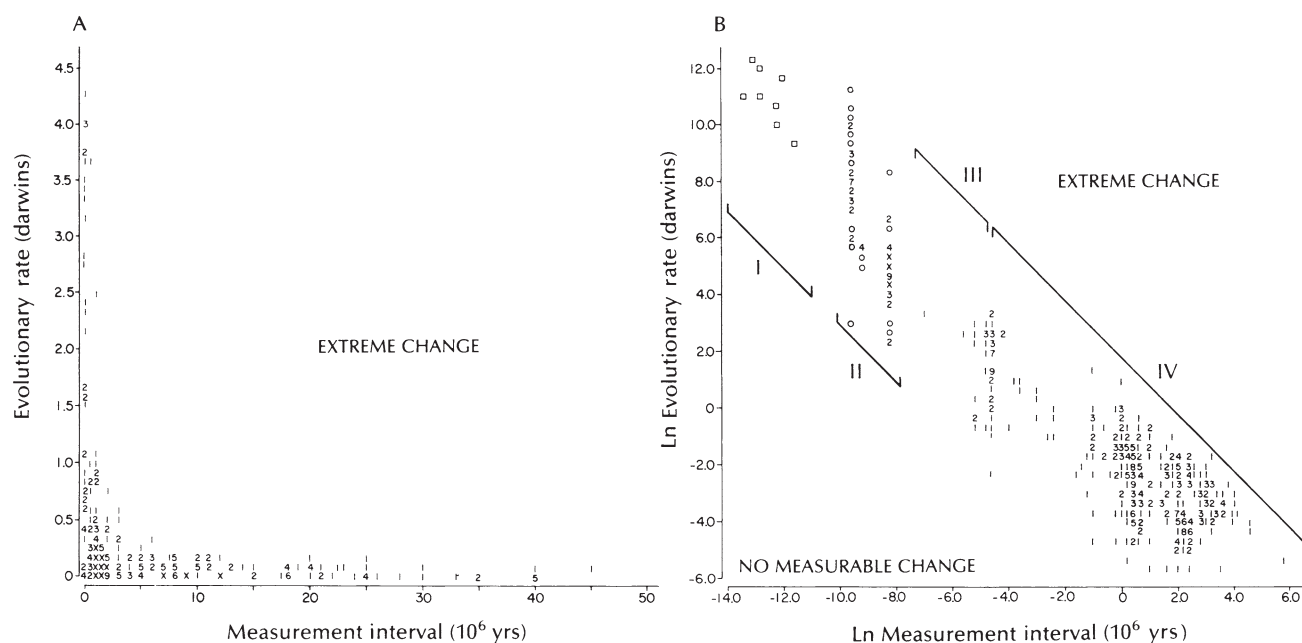


Fig. 1 Inverse relationship between morphological rates of evolution and the time intervals over which they are measured (after Gingerich 1983), illustrating some of the biasing effects discussed in the text. A, Subset of rates shown in B, plotted on linear axes. B, Logarithmic transformation of rate distribution; time intervals range from 1.5 to 350 million years. Rates plotted in domain I (open squares) represent laboratory selection experiments; domain II (open circles, and digits for multiple cases; $X > 9$) represents historical colonization events; domain III (digits) represents post-Pleistocene events; and domain IV (digits) is drawn from the pre-Holocene record of invertebrates and vertebrates.

nunities in the preserved record of taxa (even within intervals that have a sedimentary record) are more important sources of error in these estimates. Stratigraphic completeness increases in importance, however, when morphological or taxonomic rates read 'directly' from the fossil record are considered. In a relatively incomplete section, the actual age difference between two samples may be either much greater or much smaller than their estimated age difference based on linear interpolation from dated levels. This translates into substantial imprecision in rate measurements. In order to minimize this problem, Dingus & Sadler (1982) suggested that rates only be measured at levels of resolution for which stratigraphic sections can be considered complete (i.e. for which each included interval of given magnitude is likely to be represented by sediment). Following this recommendation, relatively incomplete sections limit us to longer time intervals for rate measurement and thus, through the biasing effect of interval length, tend to yield lower rates than might be seen in more complete sections. Relatively incomplete sections also tend to reduce measured rates of origination.

Factors affecting actual rates of evolution

Having explored some of the factors that tend to distort perceptions of evolutionary rates, the sources of real variation in such rates will now be discussed. Among the more conspicuous of these are controls of the rate of transformation within established, species-level lineages. These include: mutation rate; generation time; degree of resource specialization; and the nature, amount, and distribution of variability within populations. Population size may also be important but is probably overshadowed by population structure — the pattern and scale of subdivision of populations and the degree of reproductive interaction between those subdivisions. Other factors are at least partly extrinsic to the species in question: rate of environmental change; ecological factors such as the level of interspecific competition; and, in general, the intensity of selection (assuming selection and fitness are defined so that intensity of selection is not trivially equivalent to rate of evolution).

Another group of controls overlaps somewhat with the first but may be distinguished as operating at a different level in the genealogical hierarchy. It

consists of factors that determine the rate of initiation of new species-level lineages. Speciation rate assumes particular importance in a punctuated view of evolution, but its role in influencing evolutionary rate is not dependent on the predominance of a punctuated mode of evolutionary change. Intrinsic controls on speciation rate include such factors as dispersal ability (also relevant as a determinant of population structure) and degree of resource specialization. There are also extrinsic controls, such as rate or incidence of habitat fragmentation by geomorphic or tectonic processes.

Living fossils — alternative definitions

Living fossils figure in discussions of evolutionary rates as a conspicuous and yet potentially tractable case in which the relationship between a large-scale evolutionary pattern and its underlying causes may be explored (Eldredge & Stanley 1984; Schopf 1984). As noted above, the central concept in the definition of living fossils is survival over long periods of time with minimal morphological change. Auxiliary criteria have been appended by various authors and do indeed apply to certain cases traditionally recognized as living fossils. However, they are much less applicable to others. For instance, a relict geographical distribution and greatly diminished present (relative to past) diversity characterize *Sphenodon* (a rhynchocephalian) and *Nautilus* (a nautiloid cephalopod), but not *Limulus* and related genera (horseshoe crabs). Likewise, *Latimeria* (a coelacanth) and *Neopilina* (a monoplacophoran) represent clades once thought to be extinct, but *Lepisosteus* (a gar) and *Lingula* (an inarticulate brachiopod) have long been known from both fossil and Recent biotas. Living fossils are sometimes referred to as 'species' that have persisted for inordinately long periods of time, but few if any instances are actually founded on well documented species-level identity. The most generally useful definition therefore focuses on supraspecific taxa that have shown unusual morphological conservatism.

One of the most commonly cited living fossil groups is the Xiphosurida, or horseshoe crabs. Fig. 2 provides some sense of the morphological conservatism that can be seen within this group, comparing the extant species *Limulus polyphemus* with the Triassic *Limulus vicensis*. While the generic identity of these two species may be questioned (Fisher in Eldredge & Stanley 1984), their overall anatomical similarity is evident. Other species with-

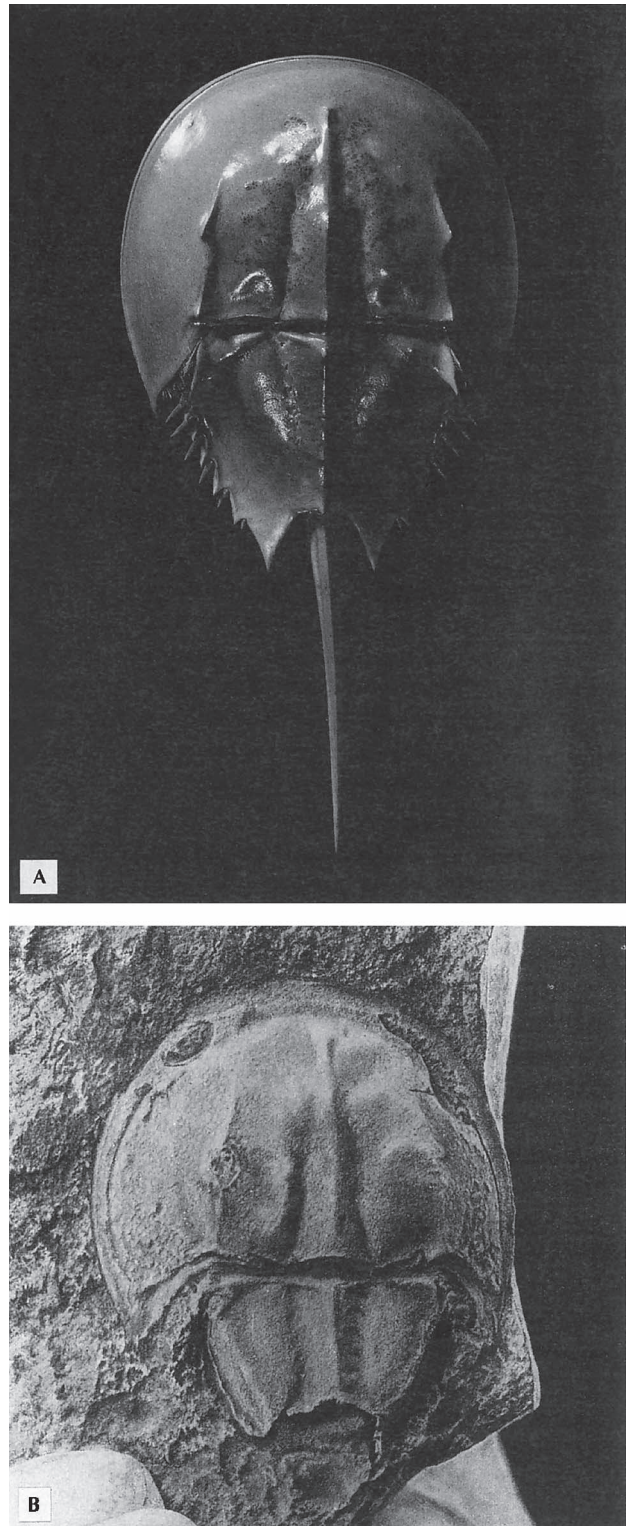


Fig. 2 Horseshoe crabs, a commonly cited living fossil group. A, Dorsal aspect of a juvenile *Limulus polyphemus*, Recent, distributed along much of the eastern coast of North America; c. one half actual size. B, Dorsal aspect of a specimen of *Limulus vicensis*, Triassic, France; c. actual size. The tail spine is not preserved on this specimen, but it was presumably present originally. (From Bleicher 1897.)

in the group show greater morphological divergence, but the reputation for bradytely has focused on comparisons such as that given here.

Bradytely — alternative explanations

The problem posed by living fossils is to explain the general phenomenon of bradytely. Simpson's (1944, 1953) interpretation was that the low rates of long-term morphological evolution shown by bradytelic lineages are a consequence of unusually low rates of intraspecific phyletic transformation (Fig. 3A). This appears to be a testable proposition, but it has thus far received little direct, empirical evaluation (perhaps because few bradytelic groups have a sufficiently continuous fossil record). However, some of the factors that have been suggested as responsible for low rates of phyletic transformation (e.g. unusually low levels of morphological or genetic variability) have been assessed within bradytelic groups and found not to differ significantly from values typical of nonbradytelic taxa (e.g. Selander *et al.* 1970). Other factors that could in principle be responsible (e.g. extreme habitat stability, or strongly canalized development) are difficult to test. Some factors do seem to hold for a wide range of bradytelic groups and have been thought to contribute directly to low rates of intraspecific change (e.g. ecological generalization and broad physiological tolerance; Simpson 1953). Nevertheless, consideration of alternative explanations is clearly warranted.

Another approach to interpreting bradytely steps up a level in the hierarchy of evolutionary processes — from intraspecific interactions to the circumstances surrounding speciation events (cladogenesis). It depends, furthermore, on the proposition (associated with the concept of punctuated equilibrium) that most morphological change is accomplished during and driven by cladogenesis, and that the subsequent history of species tends to be dominated by morphological stasis. Under this characterization of evolution, a low rate of intraspecific transformation would be the norm and would not be seen as a sufficient cause of bradytely. However, bradytely might be due to unusually low rates of speciation within bradytelic lineages (Fig. 3B); according to this interpretation, low speciation rate would allow few opportunities for morphological change and would thus restrict a lineage to a relatively low rate of change averaged over the long term (Eldredge 1979). As long as speciation is understood as a process that is not itself dependent

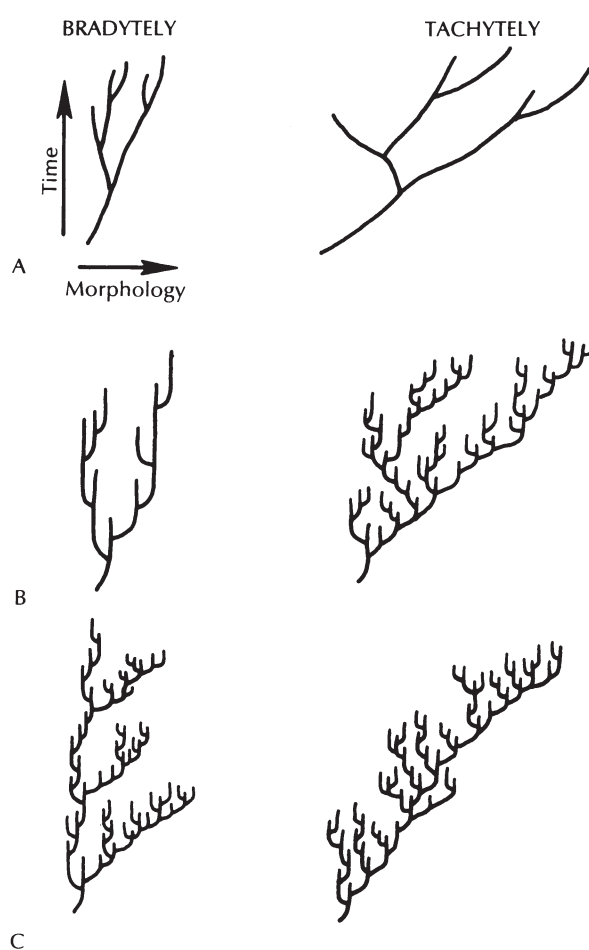


Fig. 3 Schematic representation of three explanations of controls on long-term rate of morphological evolution. (After Fisher in Eldredge & Stanley 1984.) A, The contrast between bradytely and tachytely may be due to differences in the rate of intraspecific morphological transformation. B, The same contrast may be due to differences in rate of speciation. C, Bradytely and tachytely may also reflect higher-order patterns of differential survival and cladogenesis.

on morphological change, this interpretation represents a novel perspective on bradytely. Yet there is still a question as to why certain taxa, retrospectively recognized as bradytelic, show such a low rate of speciation. One possible answer has been suggested by the observation that a number of bradytelic taxa also show tendencies toward eurytopy — i.e. they have, at least in many respects, relatively broad, generalized ecological requirements. In this, they contrast with stenotopic taxa, which have relatively narrow, specialized requirements. It has been suggested that, relative to stenotopic taxa, eurytopic taxa are less subject to directional selection, often have broader geographical

ranges, and tend to have populations that are less susceptible to range disruption and consequent reproductive isolation (Jackson 1974; Eldredge 1979; Vrba in Eldredge & Stanley 1984). This may result in a lower rate of speciation and a lower likelihood of morphological divergence during speciation. The case studies of bradytely in Eldredge & Stanley's (1984) compendium offer qualified support for the association of bradytely, low speciation rate, and eurytopy, but rigorous evaluation of this pattern is difficult because of the lack of quantitative indices of morphological conservatism or eurytopy. In addition, measurements of speciation rate are subject to significant sampling problems, such that even an evaluation of the relationship between speciation rate and subjective assessments of bradytely and eurytopy would be complicated.

A third interpretation of bradytely is that it arises at an even higher level in the genealogical hierarchy, as a result of differential survival of relatively primitive and relatively derived lineages within a clade (Fig. 3C). Treated simply as a phylogenetic pattern, bradytely may or may not have any single lower-level cause, but whether it does or not, it could be independent of any systematic difference in intraspecific rates of transformation or rates of speciation (Fisher in Eldredge & Stanley 1984).

While any of these three explanations of bradytely might operate in isolation from the others, they are not mutually incompatible. Nor can the possibility be ruled out that different instances of bradytely are traceable to different mixes of factors operating at a variety of levels. Although there are thus no simple answers, the investigation of bradytely has led to an expanded appreciation of the possible controls of long-term evolutionary rates.

References

- Bleicher, M. 1897. Sur la découverte d'une nouvelle espèce de limule dans les marnes irisées de Lorraine. *Bulletin des Séances de la Société des Sciences de Nancy* **14**, 116–126.
- Dingus, L. & Sadler, P. M. 1982. The effects of stratigraphic completeness on estimates of evolutionary rates. *Systematic Zoology* **31**, 400–412.
- Eldredge, N. 1979. Alternative approaches to evolutionary theory. *Bulletin of the Carnegie Museum of Natural History* **13**, 7–19.
- Eldredge, N. & Stanley, S.M. (eds) 1984. *Living fossils*. Springer-Verlag, New York.
- Fitch, W.M. 1976. Molecular evolutionary clocks. In: F.J. Ayala (ed.) *Molecular evolution*. Sinauer, Sunderland, Mass.
- Gingerich, P.D. 1983. Rates of evolution: effects of time and temporal scaling. *Science* **222**, 159–161.
- Gingerich, P.D. 1986. Temporal scaling of molecular evolution in primates and other mammals. *Molecular Biology and Evolution* **3**, 205–221.
- Goodman, M., Weiss, M.L. & Czelusniak, J. 1982. Molecular evolution above the species level: branching pattern, rates, and mechanisms. *Systematic Zoology* **31**, 376–399.
- Haldane, J.B.S. 1949. Suggestions as to quantitative measurement of rates of evolution. *Evolution* **3**, 51–56.
- Jackson, J.B.C. 1974. Biogeographic consequences of eurytopy and stenotopy among marine bivalves and their evolutionary significance. *American Naturalist* **108**, 541–560.
- Raup, D.M. & Marshall, L.G. 1980. Variation between groups in evolutionary rates: a statistical test of significance. *Paleobiology* **6**, 9–23.
- Raup, D.M. & Stanley, S.M. 1978. *Principles of paleontology*, 2nd edn. Freeman, San Francisco.
- Schopf, T.J.M. 1984. Rates of evolution and the notion of living fossils. *Annual Review of Earth and Planetary Sciences* **12**, 245–292.
- Selander, R.K., Yang, S.Y., Lewontin, R.C. & Johnson, W.E. 1970. Genetic variation in the horseshoe crab (*Limulus polyphemus*), a phylogenetic "relic". *Evolution* **24**, 402–414.
- Simpson, G.G. 1944. *Tempo and mode in evolution*. Columbia University Press, New York.
- Simpson, G.G. 1953. *The major features of evolution*. Columbia University Press, New York.
- Stanley, S.M. 1985. Rates of evolution. *Paleobiology* **11**, 13–26.

2.12 Mass Extinction: Processes

2.12.1 Earth-bound Causes

A. HALLAM

Introduction

The idea that mass extinctions could be caused by strictly Earth-bound phenomena is an old one, dating back to the so-called heroic age of geology in the early part of the nineteenth century. Following the pioneering extinctions research of his compatriot G. Cuvier, the French geologist Elie de Beaumont proposed that catastrophic, virtually instantaneous upheavals of mountain ranges at infrequent intervals through geological history caused drastic environmental changes leading to the destruction of a high proportion of the Earth's biota. The correlation between episodes of diastrophism and times of major organic turnover was also noted by the American geologist T.C. Chamberlin at the beginning of this century, and by European geologists such as E. Suess and J.F. Umbgrove (Hallam 1981a). Modern research on tectonic activity suggests, however, that it is too localized geographically and insufficiently 'catastrophic' in time to account satisfactorily for mass extinction events. Attention must be confined to phenomena global in scale that can give rise to drastic changes in the physical environment. The only plausible contenders are changes in sea level and climate, and episodes of increased volcanicity.

Sea-level

The American palaeontologist Newell (1967) was the first person to make an explicit correlation between mass extinction episodes among Phanerozoic marine invertebrates and eustatic falls in sea-level, attributing the extinctions to increased environmental stress consequent upon substantial reduction of habitat area of shallow epicontinental seas. He distinguished six such episodes: end-Cambrian, end-Ordovician, Late Devonian, end-Permian, end-Triassic and end-Cretaceous. The first two are especially well marked by trilobite extinctions and the last three by ammonite extinctions. Extensive communities of reef-dwelling organisms

were destroyed in the Late Devonian (Section 2.13.3) and end-Triassic (Section 2.13.5) episodes and the calcareous plankton (foraminifera and coccolithophorids) drastically reduced at the end of the Cretaceous (Section 2.13.6). The biggest event of all was at the end of the Permian (Section 2.13.4), when many important Palaeozoic groups went completely extinct, including fusulinid foraminifera, camerate and inadunate crinoids, trepostome and cryptostome bryozoans, rugose corals, and productid brachiopods.

All but the first of these extinction episodes have subsequently been accepted by palaeontologists as the most significant extinction events in Phanerozoic history (Raup & Jablonski 1986). The correlation between major sea-level falls and Newell's mass extinction events is indeed striking (Fig. 1; Jablonski 1986). On a smaller scale, there is an equally striking correlation between the extinction of environmentally sensitive groups such as ammonoids and other episodes of widespread regression, probably correlating with sea-level fall, in both the Palaeozoic and Mesozoic (e.g. Hallam 1987a). Following ecological research on island biogeography, it is clear that smaller habitat areas can accommodate fewer taxa, so reduction in area must lead to lower diversity as the extinction rate increases. Whether the extinction is due to reduced habitat diversity, increased competition, crowding effects, or whatever, the basic empirical relationship appears to be well established.

Critics have pointed out that inferred episodes of significant marine regression do not always correlate with notable mass extinctions of marine organisms. This is most obviously true for eustatic falls of sea-level in the Quaternary and Middle Oligocene, the latter being probably the largest in the Tertiary (Haq *et al.* 1987). At least two explanations can be put forward, both of which take into account the phenomenon of biological adaptation. Quaternary regressions were followed by equally rapid transgressions after geologically short time intervals, limiting the effect of reduced habitat area and permitting a sufficient number of organisms to survive and expand their populations during the succeeding transgressions. Quaternary faunas are likely to have been relatively eurytopic, or environmentally tolerant, because they represent survivors of environmentally stressful Late Cenozoic times. The same

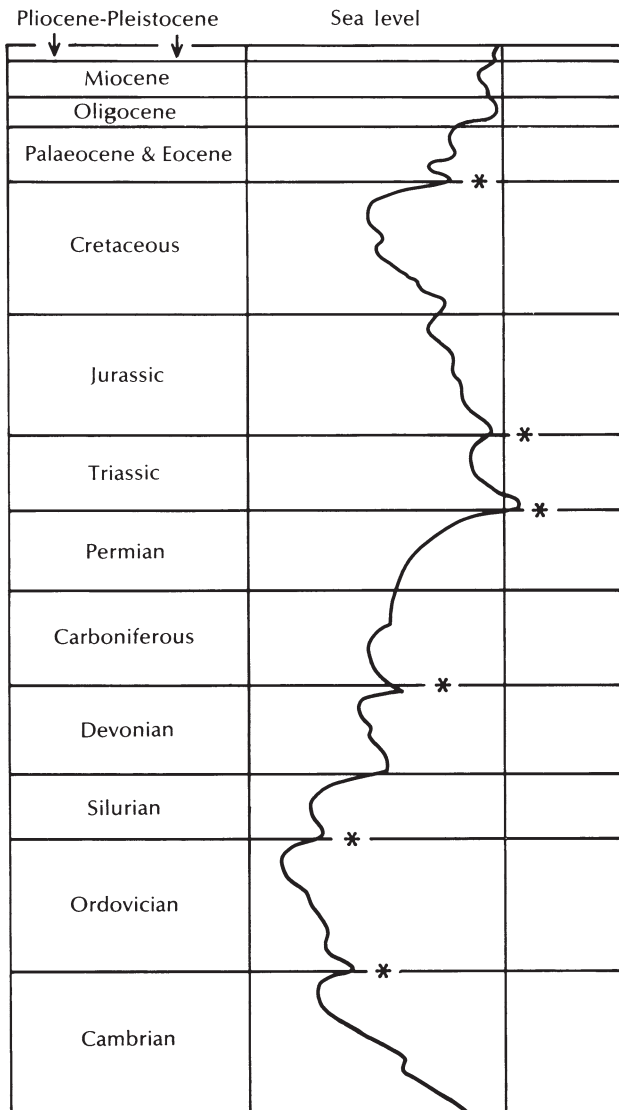


Fig. 1 Phanerozoic sea-level curve with Newell's (1967) six extinction events shown as asterisks. High sea-level to left. (After Hallam 1984.)

consideration may apply also to the Middle Oligocene regression, which followed closely on a significant increase in marine extinction rates across the Eocene–Oligocene boundary. It is likely that for long periods of Phanerozoic time most organisms became so well adapted to conditions of relative environmental stability, including equable climate, that even modest changes of sea-level could have had a striking effect on so-called 'perched faunas' in extensive and extremely shallow epicontinental seas. Such palaeogeographic phenomena cannot be closely matched at the present day, which marks an unusually regressive episode in Earth's history.

The latest Exxon sea-level curve from the Triassic

to the present (Haq *et al.* 1987) does not show unusually large falls at the times of the two greatest marine extinction episodes during this interval, the end of the Triassic and the end of the Cretaceous. The Exxon curve is based largely, however, on seismic stratigraphy, and should not be treated as more than a tentative model to be subjected to testing by other evidence. There is indeed considerable evidence of a major end-Triassic regression (Hallam 1981b), and some strong indications that the extent of the end-Cretaceous regression has been underestimated by Haq *et al.* (Hallam 1987b).

For several events, namely the end-Permian, end-Triassic and end-Cretaceous, mass extinctions in the marine realm appear to correlate closely with mass extinction of some terrestrial vertebrates, notably those large in size, which, because of their relatively low population numbers and reproductive rates, would be more vulnerable to environmental disturbance than smaller organisms (see also Section 2.10). Obviously such extinctions cannot be accounted for by reduction in land area, and a more likely explanation is bound up with the increased continental seasonal temperature contrasts induced by regression of epicontinental seas.

While much attention has been paid to regression as a promoter of extinctions it should be noted that there is a strong association between inferred sea-level rises that follow directly after falls and the spread of anoxic water in epicontinental seas, as recorded for instance by widespread laminated black shales. Habitable areas can be as severely reduced by this means as by regression, with a mass extinction event ensuing. For many extinction events, both major and minor, a clear correlation exists with extensive deposits of black shales. Among the major events the best examples are the basal Silurian and basal Famennian (Devonian), effectively equivalent to the end-Ordovician (Section 2.13.2) and end-Frasnian (Section 2.13.3) extinction events. Among minor events the clearest examples are the Cenomanian–Turonian boundary and Early Toarcian (Hallam 1987a). The spread of anoxic bottom waters may possibly also be implicated as a contributory factor in the end-Permian (Section 2.13.4) and end-Triassic (Section 2.13.5) events.

For much of Phanerozoic history the ocean might have been poorly stratified, in marked contrast to the present-day situation (Wilde & Berry 1984). In consequence the deeper ocean would be more or less anoxic and could not have served as a refuge for

shallow-water organisms at times of regression, or if they were outcompeted by other organisms. It is more than possible that the great bulk of the modern deep-sea fauna, which contains representatives of most phyla, is no older than Tertiary. Since the Late Eocene there has evidently been a system of strong currents induced by Antarctic glaciation, which have served to aerate bottom water in the deep ocean (Hallam 1981a). Lack of a deep water anoxic zone could help to explain why there is no significant extinction recorded for the major Middle Oligocene regression.

The cause of sea-level changes is bound up either with the melting and freezing of polar icecaps or with tectonics, such as the uplift and subsidence of ocean ridges and the splitting or collision of continents. The end-Ordovician event might well have had a glacioeustatic cause, associated with growth and disappearance of the Saharan ice sheet, but for the other major extinction events the most likely cause is tectono-eustatic. This poses a problem, because the rates of sea-level rise and fall produced by plate tectonics are approximately three orders of magnitude lower than for glacioeustasy, thereby allowing more time for organisms to adjust to a changed environment and hence avoid extinction. Unfortunately there are as yet insufficient data from the stratigraphic record, on amount and rate of sea-level change, to resolve this problem satisfactorily.

There remains another possibility, that rapid regressions and transgressions on a regional rather than a global scale could be produced as a result either of changes in the pattern of lateral stresses in the crust (Cloetingh *et al.* 1985) or by the rise of mantle plumes to cause epeirogenic uplift, with related volcanism associated with subsidence (Loper & McCartney 1986). The fact that such changes would not strictly come under the category of eustatic is irrelevant as far as the organisms are concerned, provided that the changes in question are both geographically extensive and rapid, thereby leading to drastic changes in the environment.

Climate

Changes of sea-level could have, as a by-product, some climatic consequences, but climate could of course fluctuate with time independent of eustasy. Stanley (1984, 1987) has been the strongest advocate of the view that temperature changes in the marine realm have been the dominant causal factor in Phanerozoic mass extinctions. This interpretation involves a gross extrapolation from his detailed

studies of Plio-Pleistocene molluscan extinctions off the Atlantic and Gulf coasts of the U.S.A. Whereas there is a high rate of species extinctions in this region, there is negligible evidence of contemporary extinctions around the Pacific margins, or the Mediterranean. Stanley maintained that, because the extinctions are regional not global in extent, eustatic changes cannot be invoked. Instead he argued for a more pronounced lowering of temperature on the American east coast than elsewhere, as a result of palaeogeographical factors.

Extending back through time, the next major marine extinction event for which temperature decline can plausibly be invoked is across the Eocene–Oligocene boundary. This 'event' is decidedly not sudden in geological terms and is marked more by a pronounced increase in extinction rate rather than a drastic change over a narrow time interval. There is good independent evidence from oxygen isotopes of a fall in both surface and bottom water temperatures, but no indication from the curve of Haq *et al.* (1987) of sea-level changes significantly larger than at other times in the Tertiary. For pre-Tertiary times, however, the evidence implicating temperature as a causal factor is weak to non-existent, forcing Stanley to resort to some special pleading (though it could be argued that the end-Ordovician event (Section 2.13.2) had an ultimate climatic causation, if the glacioeustatic interpretation is accepted). For example, the largest extinction event of all, at the end of the Permian (Section 2.13.4), took place during a period of climatic amelioration, marked by the Middle Permian disappearance of the Gondwana ice sheet. It is conceivable, of course, that the end-Permian event was induced by an episode of temperature rise, but no plausible case has been made for this.

One of the points that Stanley cited in favour of his temperature control hypothesis is that the most extinction-vulnerable organisms, such as reef dwellers, were tropical in distribution throughout Phanerozoic history. While this may be true, it does not necessarily establish temperature as the key control, because tropical organisms tend to be generally stenotopic, as they are relatively sensitive to a variety of environmental factors.

A really extensive overturn of deep anoxic water at the beginning of episodes of climatic change has been suggested as a possible contributing factor to mass extinction events in the oceans (Wilde & Berry 1984). As discussed above, the rise and spread onto continental shelves of anoxic water is often associated with marine transgressions, so that it may be

unnecessary to invoke climatic change as well.

As regards changes in air temperature, the only satisfactory record comes from Late Cretaceous to Recent terrestrial plants. No striking extinction event has been recorded among these organisms for the Cenozoic, but at the end of the Cretaceous there were significant extinctions in the North Temperate Realm of western North America and Eastern Asia. Whereas the palaeobotanical consensus has related such extinctions to gradual temperature decline through the Late Cretaceous, the most recent research in the North American Western Interior suggests a temperature rise in the Maastrichtian and no significant change across the Cretaceous–Tertiary boundary (Wolfe & Upchurch 1987). Further back in time the evidence from terrestrial plants is more obscure, and has so far not been adequate to establish a convincing picture of climatic change.

Volcanism

The end-Cretaceous extinction event is the one that has received by far the most attention (see also Sections 2.13.6, 2.13.7). Notwithstanding the claims made for extra-terrestrial impact, there is strong evidence for marine regression at this time, suggesting that this phenomenon is involved in the extinctions. Sea-level change cannot account, however, for the drastic extinctions at the Cretaceous–Tertiary boundary of calcareous plankton, nor for such physico-chemical evidence as an anomalous enrichment on a global scale of iridium, and the presence locally of quartz grains with shock-metamorphic laminae, in Cretaceous–Tertiary boundary layers (see also Section 2.12.2). Evidence of this sort has been claimed as conclusive for bolide impact, but in fact a case of at least equal plausibility can be made for terrestrial volcanism on a massive scale (Hallam 1987b). It is known that aerosols enormously enriched in iridium compared with crustal rocks can be expelled from the mantle during flood basalt eruptions. Eruptions of this kind on a sufficient scale over several 100 000 years could produce the observed global enrichment of the element. The Deccan Traps of India, erupted during the magnetic zone that embraces the Cretaceous–Tertiary boundary, are the most obvious candidate. There is good evidence of contemporary explosive volcanism in other parts of the world, and reasonable grounds for believing that such volcanism can generate the pressures required to produce shock-metamorphic laminae in mineral grains.

Massive volcanism over an extended period would have deleterious environmental consequences. It is known that flood basalt fissure eruptions that produce individual lava flows with volumes greater than 100 km³ at very high mass eruption rates are capable of injecting large quantities of sulphate aerosols into the lower stratosphere, with potentially devastating atmospheric consequences. Such volatile emissions on a large enough scale would lead to the production of immense amounts of acid rain, reduction in alkalinity and pH of the surface ocean, global atmospheric cooling, and ozone layer depletion. Atmospheric cooling would be reinforced by ash expelled into the atmosphere by contemporary explosive volcanicity.

Thus for the end-Cretaceous extinctions a compound scenario seems to be required, involving both sea-level fall and volcanicity on an exceptionally intense scale, with associated climatic changes (there is as yet, however, no evidence to support the notion that volcanicity was a direct causal factor for other mass extinction events). Loper and McCartney (1986) noted that increased end-Cretaceous volcanism correlates with a significant change in the geomagnetic field, with a long Cretaceous reversal-free period coming to an abrupt end in the Maastrichtian. They proposed a model involving periodic instability of the thermal boundary layer at the base of the mantle. This layer accepts heat from the core and transmits it upward by way of mantle plumes. As it thickens by thermal diffusion it becomes dynamically unstable and hot material erupts from it. Heat is extracted from the core at a greater rate, increasing the energy supply and hence the magnetic reversal frequency of the dynamo in the fluid outer core. Hot material rises through mantle plumes to the surface to give rise to volcanic activity. Both non-explosive and explosive volcanism can be produced, depending on the condition of the lithosphere, which varies regionally. Increased mantle plume activity has the potential for causing uplift of extensive sectors of continents and hence regression of epicontinental seas. Present-day hotspots are associated with regional topographic bulges, so it is reasonable to infer that most epeirogenic uplifts reflect hot, low density regions in the asthenosphere, derived from plume convection. Epeirogenic subsidence on the continents and marine transgression might be expected to follow episodes of substantial volcanic eruptions.

Fischer (1984) put forward a general hypothesis that relates changes of sea-level, climate, and volcan-

icity to produce two supercycles during Phanerozoic time. Times of high rates of ocean floor spreading and oceanic volcanicity correlate with buoyant ocean ridges and consequently high sea-level stands. Less carbon dioxide is removed from the atmosphere by terrestrial weathering because of reduced continental area, and the volcanicity brings more of the gas to the Earth's surface. Thus the carbon dioxide content of the atmosphere is high, and because of the greenhouse effect the climate is equable, with no polar ice caps. The converse tectonic situation gives rise to low sea-level stands, low atmospheric carbon dioxide, and stronger climatic differentiation between the tropics and the poles. The rates of change involved in such processes appear, however, to be too low to account for mass extinction events. The most promising line of approach in generating terrestrial models is probably a closer investigation of the relationship between sea-level change, continental uplift, volcanism, and mantle plume activity, as has been proposed for events across the Cretaceous–Tertiary boundary (Sections 2.13.6, 2.13.7). The end-Permian extinction episode (Section 2.13.4) is an especially promising candidate for this type of investigation.

References

- Cloetingh, S., McQueen, H. & Lambeck, K. 1985. On a tectonic mechanism for regional sea level variation. *Earth and Planetary Science Letters* **75**, 157–166.
- Fischer, A.G. 1984. The two Phanerozoic supercycles. In: W.A. Berggren & J.A. van Couvering (eds) *Catastrophes and Earth history*, pp. 129–150. Princeton University Press, Princeton.
- Hallam, A. 1981a. *Facies interpretation and the stratigraphic record*. W.H. Freeman, Oxford.
- Hallam, A. 1981b. The end-Triassic bivalve extinction event. *Palaeogeography, Palaeoclimatology, Palaeoecology* **35**, 1–44.
- Hallam, A. 1984. Pre-Quaternary sea-level changes. *Annual Review of Earth and Planetary Sciences* **12**, 205–243.
- Hallam, A. 1987a. Radiations and extinction in relation to environmental change in the marine Lower Jurassic of northwest Europe. *Paleobiology* **13**, 152–168.
- Hallam, A. 1987b. End-Cretaceous mass extinction event: argument for terrestrial causation. *Science* **238**, 1237–1242.
- Haq, B.U., Hardenbol, J. & Vail, P.R. 1987. Chronology of fluctuating sea levels since the Triassic. *Science* **235**, 1158–1167.
- Jablonski, D. 1986. Causes and consequences of mass extinctions. In: D.K. Elliott (ed.) *Dynamics of extinction*, pp. 183–229. Wiley, New York.
- Loper, D.E. & McCartney, K. 1986. Mantle plumes and the periodicity of magnetic field reversals. *Geophysical Research Letters* **13**, 1525–1528.
- Newell, N.D. 1967. Revolutions in the history of life. *Special Papers of the Geological Society of America* **89**, 63–91.
- Raup, D.M. & Jablonski, D. (eds) 1986. *Patterns and processes in the history of life*. Report of Dahlem Workshop, 1985. Springer-Verlag, Berlin, Heidelberg.
- Stanley, S.M. 1984. Marine mass extinction: a dominant role for temperature. In: M.H. Nitecki (ed.) *Extinctions*, pp. 69–117. University of Chicago Press, Chicago.
- Stanley, S.M. 1987. *Extinction*. Scientific American Books, New York.
- Wilde, P. & Berry, W.B.N. 1984. Destabilisation of the oceanic density structure and its significance to marine extinction events. *Palaeogeography, Palaeoclimatology, Palaeoecology* **48**, 143–162.
- Wolfe, J.A. & Upchurch, G.R. 1987. North American non-marine climates and vegetation during the late Cretaceous. *Palaeogeography, Palaeoclimatology, Palaeoecology* **61**, 33–78.

2.12.2 Extra-terrestrial Causes

D. JABLONSKI

Introduction

Extra-terrestrial causes have long been invoked for mass extinctions, but only in the past decade has the general scientific community taken the idea seriously. Geochemical, sedimentary, and other signals in the stratigraphic record are sufficient to suggest that it is impossible to ignore extra-terrestrial impacts as potential explanations for the biotic crises that punctuate the fossil record. The case is not fully proven for any single mass extinction, although it is strongest for the end-Cretaceous event (W. Alvarez 1986; L.W. Alvarez 1987; see Hallam 1987 and Officer *et al.* 1987 for different views; see also Sections 2.12.1, 2.13.6, 2.13.7). In any event, the initial discovery of iridium and other geochemical anomalies at the Cretaceous–Tertiary boundary has sparked an immense amount of interdisciplinary research on the problem of mass extinctions and potential extra-terrestrial forcing agents.

Potential mechanisms

Proposed extra-terrestrial causes for mass extinctions have included variation in solar heat output, massive solar flares, sudden influx of cosmic rays owing to a nearby supernova or the Solar System's crossing of the Galactic plane, and collisions with comets, asteroids, or other extra-terrestrial objects

(collectively termed bolides). Until recently such factors were at best subject to only the weakest verification based on approximate correlations in timing, and at worst simply reflections of desperation in the face of seemingly inexplicable biotic upheavals. New lines of evidence for possible bolide impacts at one, and perhaps as many as five, extinction events have shifted these speculations into the realm of testability.

Earth-crossing asteroids (asteroids whose orbits cross that of the Earth or could cross as a result of long-range gravitational perturbations) are sufficiently common that significant bolide impacts must have occurred in the geological past. The Earth should suffer impacts by *c.* six 1 km asteroids per million years, and by *c.* two asteroids of 10 km or more per 100 million years, i.e. about a dozen large impacts since the beginning of the Phanerozoic (Shoemaker 1984). Effects of 1 km objects are uncertain but, as discussed below, most workers believe that impact by a 10 km bolide would have severe, global consequences.

The average collision rate for comets is almost certainly lower than that for asteroids. Cometary impact rates could occasionally be raised, however, by perturbing the Oort cloud of comets that surrounds the Solar System far beyond the outermost planets (inner edge about 10^4 Astronomical Units (AU) from the Sun, where 1 AU is the distance from the Sun to the Earth). Passage through the higher stellar densities in the spiral arms of the Galaxy might raise collision rates by about 10% (Shoemaker 1984). This low-frequency modulation of cometary impacts would be punctuated approximately once per 100 million years by short-lived bursts (1–3 million years) triggered by close passage of individual stars (Hut *et al.* 1987).

Evidence for periodic extinctions, still hotly debated, suggests (but does not prove) a more regular and frequent perturbation of the Oort cloud. Hypothesized mechanisms include: oscillations around the Galactic plane, where encounters with stars and molecular clouds would be most probable; a tenth planet in a highly eccentric orbit beyond Pluto (at *c.* 100 AU); and a dim solar companion star, christened Nemesis in advance of discovery (at distances variously estimated in the order of 10^4 – 10^5 AU). Debates on the astronomical plausibility of these mechanisms, with Nemesis maintaining a slight edge, are reviewed by Shoemaker & Wolfe (1986) and Hut *et al.* (1987) (see also Section 2.12.3).

The magnitude and geographical scale of an impact's effects depend on bolide size and velocity but

thresholds have not been determined. An asteroid 10 km in diameter was estimated for the end-Cretaceous event on the basis of global iridium levels, and although potential effects are still poorly understood they would probably have been severe. W. Alvarez (1986), L.W. Alvarez (1987) and Prinn & Fegley (1987) emphasize the following possibilities:

1 *Darkness* caused by the global cloud of fine dust particles generated by the impact. For 2–11 months, this darkness may have been sufficiently profound to halt photosynthesis, thereby causing the collapse of marine and terrestrial food chains.

2 *Cold* would accompany the darkness, with temperatures dropping below freezing in continental interiors. Maritime climates would be less severely perturbed, owing to the thermal inertia of oceanic waters.

3 *Greenhouse effects* and global warming could follow the cold-temperature excursion if the bolide(s) struck in the ocean. After dust grains coagulated and settled from the atmosphere, the remaining burden of water vapour could trap infrared energy reflected from the Earth and raise global temperatures by as much as 10°C. The duration of this greenhouse episode is uncertain, with estimates ranging from months or years to much longer spans than the immediate cold, dark aftermath — perhaps as long as 1000 years (Prinn & Fegley 1987).

4 *Nitric acid rain* might result from shock heating of the Earth's atmosphere during impact (see Prinn & Fegley 1987, whose calculations are followed here). Energy from atmospheric entry and, especially, the supersonic plume ejected upon impact would produce very large amounts of nitric oxides. These compounds would undergo a series of reactions and ultimately rain out as nitric and nitrous acid. On land this would severely damage foliage (and, presumably, animals) both directly and through mobilization of trace metals. In the ocean, within a decade or less, the acid rain could lower the pH of the mixed layer (especially the upper 30 m) to 7.5–7.8, sufficient to dissolve calcite and thus severely stress calcareous organisms. Further, injection of so much strong acid into the atmosphere would elicit a significant exhalation of oceanic CO₂, which, combined with the accumulation of CO₂ in the atmosphere owing to depressed activity of marine phytoplankton, would yield greenhouse warming over thousands of years.

This impressive menu of impact-driven perturbations could be expected to cause mass extinctions of the observed magnitudes. Indeed, a

number of palaeontologists have argued that the hypothesized perturbations are too severe for the observed extinctions, even at the Cretaceous–Tertiary boundary (e.g. Hallam 1987). However, the impact-effect models are very poorly constrained and require extrapolation far beyond hard observational data; a new generation of more realistic and sophisticated models may provide an improved basis for critically comparing hypothesized causes with observed extinction patterns.

Biological evidence

The initial impetus for seeking extra-terrestrial impacts was of course the biological pattern of extinction in the fossil record, whether perceived as peaks in global extinction rates or as disappearances of taxa or biomass in local sections. Unfortunately, the biological consequences of impacts, massive volcanism, and other alternatives are not sufficiently understood or sufficiently unique to provide critical tests. Complex biological upheavals enacted on scales of months, years or decades, as postulated by impact scenarios, are extremely difficult, often impossible, to resolve in single stratigraphic sections, and challenge the limits of global correlation. Short-term events are superimposed on more protracted patterns in the expansion and contraction of taxa, owing to Earth-bound physical and biotic factors, so that the effect of a given boundary event on a particular taxon (particularly a waning one) is debatable. At present, the strongest constraints that palaeontological data can provide involve consistency between a given mechanism and the biological pattern observed in an imperfect fossil record.

Onset and aftermath. For extra-terrestrial impacts, biological responses include abrupt onset of extinction with an extremely brief crisis period, and a relatively short-lived reorganization and rebound during return to pre-impact conditions. In end-Cretaceous impact models, for example, most environmental perturbations would last only 1–10 years, an interval impossible to correlate among distant localities, and within which events are virtually unresolvable in the geological record. Geologically abrupt onset of mass extinction is a requirement but not a unique prediction of impact hypotheses: even such gradual processes as marine regression or transgression could in principle carry threshold effects that would produce sudden extinction pulses on stratigraphically-resolvable time-scales.

Hypothesized greenhouse warming, and possibly other palaeoceanographic anomalies, would persist for some thousands of years beyond the impact itself. Some palaeontological (and geochemical) evidence supports a geologically brief — but ecologically protracted — recovery period, particularly in terrestrial plants (reviewed by Wolfe 1987) and marine plankton (reviewed by Zachos & Arthur 1986), although, again, these would not be unique to extra-terrestrial events.

Extinction patterns observed at critical boundaries cannot be taken at face value. Seemingly abrupt extinction can result from erosion or non-deposition of sediments during the critical time interval, so that biological events are compressed into single beds. At the same time, artificially gradational extinction patterns result when sampling deteriorates, or is simply uneven, in the interval approaching the boundary (a phenomenon termed backwards-smearing, or the Signor–Lipps effect — see Jablonski 1986a; Raup 1987).

Stepwise patterns of extinction, with pulses of extinction arrayed around a mass extinction boundary, have been claimed to reconcile the requirements of abrupt extinction with observations that seemed to suggest gradual loss of taxa. Such stepwise patterns — with up to 12 discrete extinction events claimed near the Cretaceous–Tertiary boundary — are also taken as the geologically rapid succession of extinction events expected during cometary bombardment. These stepwise patterns are distinct from prolonged patterns of decline such as suggested for Late Cretaceous ammonites, and are recorded near the Cenomanian–Turonian, Cretaceous–Tertiary, and Eocene–Oligocene boundaries (Hut *et al.* 1987). Unfortunately, such patterns cannot yet be taken at face value, because they can also be generated by sampling effects, local ecological changes, and/or minor breaks in sedimentation imposed on *either* abrupt or gradational extinction.

Lazarus taxa (which seem to suffer extinction but then reappear later in the stratigraphic record; Jablonski 1986a; Raup 1987) provide one means of partially controlling for unevenness in sampling and preservation: the proportion of Lazarus taxa, i.e. of observed last appearances that represent artificial extinction, permits a rough quantitative assessment of the reliability of extinction data within and around critical time intervals. Most stepwise extinction sequences contain some Lazarus taxa, suggesting that sampling effects are indeed a factor. More rigorous and comprehensive approaches are

required to place confidence limits on bed-by-bed extinction patterns.

Detailed studies of critical time intervals are urgently needed, but the plea for more centimetre by centimetre sampling near extinction events is somewhat misguided. At that scale, local ecological effects, the vagaries of sampling, and even bioturbation are likely to overwhelm the fine structure of global events. Careful sampling of relatively long geological sequences that encompass extinction events would be especially valuable, so that absences as well as presences could be recorded throughout, to provide some statistical control. Consistency of extinction patterns among widely separated localities also should be sought in a critical fashion; caution is necessary, particularly for apparent stepwise patterns, because different taxa — say, ammonites and benthic gastropods — have different sampling characteristics, even on broad geographical and temporal scales (see Jablonski 1986a on the biology of Lazarus taxa).

Selectivity has been claimed for most mass extinctions: large-bodied taxa, reef-dwellers or tropical organisms in general, and endemic taxa all appear to suffer preferential extinction (Jablonski 1986a, b). Critics (and some supporters!) of impact hypotheses have claimed that impact-driven extinction would be random rather than selective, so that any observed taxonomic or ecological selectivity would be contrary evidence. This claim seems inappropriate, however: taxa differ in their vulnerability to environmental change, so that any given perturbation, regardless of scale, should affect some groups more severely than others. Survivorship of widespread taxa, non-tropical taxa, small-bodied taxa, members of detrital food chains, freshwater taxa, deciduous plants, and plankton whose life cycles include resting cysts, has been claimed for the end-Cretaceous extinction (Jablonski 1986a, b; Hallam 1987). All are consistent with, but not exclusive to, impact hypotheses. Similarly, the possibility that mass extinctions are qualitatively different from background extinctions in their victims (e.g. see Jablonski 1986a, b) does not require impact events — any perturbation of sufficient magnitude could, for example, cross a threshold of extinction effects so that broad geographical range could determine survivorship but species richness was no longer important.

Periodicity. The apparent periodicity of post-Palaeozoic extinction events has sparked much re-

search and speculation on extra-terrestrial forcing factors (see also Section 2.12.3). The periodicity itself, however, is not an adequate test for extra-terrestrial causes, although few alternatives have been advanced (Hallam 1987 reviewed a hypothesis of endogenous periodicity in mantle plumes; see also Section 2.12.1). Clearly, the critical role for palaeontological data in testing for extra-terrestrial causes of mass extinctions lies in the degree of correspondence between biological events and independent physico-chemical evidence for impacts or other extra-terrestrial forcing mechanisms. As discussed below, however, assembling such evidence is not as straightforward as was once hoped.

Physical evidence

Several physico-chemical phenomena have been proposed as independent evidence for extra-terrestrial impact. Although each has its critics, and some may not be strictly diagnostic, taken together the data make a strong case for the end-Cretaceous and Late Eocene extinctions, with weaker but suggestive evidence for several other post-Palaeozoic events (Raup 1987). The strongest Earth-bound alternative at this time appears to be volcanism (Hallam 1987; Officer *et al.* 1987; see also Section 2.12.1).

Geochemical. The anomalously high concentrations in Cretaceous–Tertiary boundary sediments of iridium, and other elements scarce in the Earth's crust but abundant in asteroids, launched the Alvarez hypothesis that an end-Cretaceous impact caused the mass extinction. Since 1979 this anomaly has been found at over 75 localities world-wide (Fig. 1) in deep-sea, shallow-marine, and continental palaeoenvironments, usually in a distinctive clay layer that coincides (within stratigraphic uncertainty limits) with the extinction event (W. Alvarez 1986; L.W. Alvarez 1987). Excursions in oxygen and carbon isotopes near the boundary also suggest a low-productivity episode that may have lasted 1.0 million years or more, accompanied by detectable but unexceptional temperature oscillations (Zachos *et al.* 1989). The direction of the stable isotopic fluctuations is appropriate to impact hypotheses, but the duration seems too long and the temperature changes too mild (but see above discussions on uncertainties in impact models and limits in stratigraphic resolution).

None of the other four major mass extinctions of the Phanerozoic has such strong geochemical

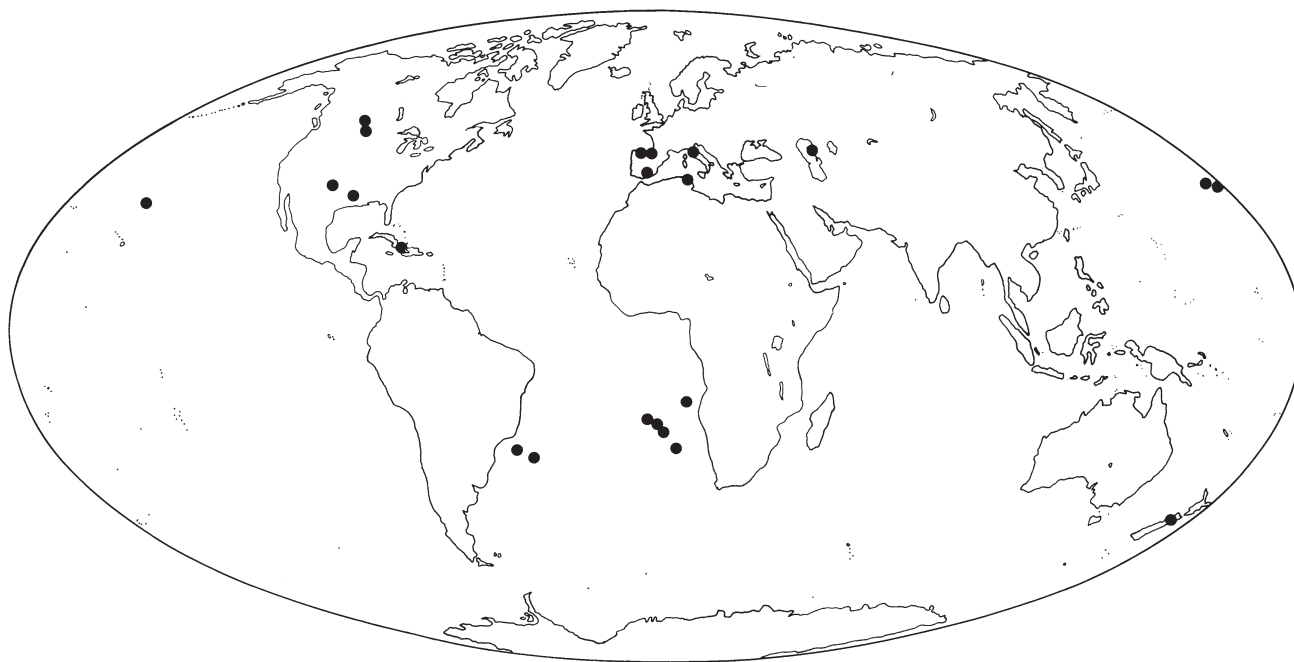


Fig. 1 Global distribution of iridium anomalies in Cretaceous–Tertiary sediments. (After L.W. Alvarez 1987.)

anomalies known from so many localities, although far less effort has been devoted to the search (Jablonski 1986a; Donovan 1987a; Raup 1987). Slight end-Ordovician iridium enrichments seem to be terrestrial in origin; the end-Triassic results are negative so far; the reported end-Permian anomaly, at the largest mass extinction of them all, has not been repeated by other laboratories, and the boundary clays seem volcanic in origin; the Late Devonian (Frasnian–Famennian) anomaly occurs in an unusual stromatolitic deposit and has not been replicated in other boundary sections.

Among lesser extinction events, iridium anomalies are geographically widespread near the Eocene–Oligocene extinction boundary, along with a series of microtektite horizons whose impact origin is virtually uncontested (Hut *et al.* 1987). An iridium anomaly was recently discovered (L.W. Alvarez 1987) for the small Middle Miocene extinction that forms the most recent peak in periodicity analyses, although the global extent of the iridium is as yet unknown. The Cenomanian–Turonian boundary has excess iridium, but other impact signatures are lacking and a terrestrial origin may be involved. An anomaly at the Middle–Upper Jurassic boundary — where no extinction event occurs but is predicted by periodicity models — occurs (like the Frasnian–Famennian example) in stromatolitic sediments, raising the spectre of biological or dia-

genetic concentration. Age uncertainty of an iridium anomaly in a 2–3 mm iron-rich crust at an unconformity in the Southern Alps overlaps with another weak or ‘missing’ (i.e. predicted by periodicity models) extinction peak in the Bajocian (Rocchia *et al.* 1986). An iridium anomaly, with other cosmic debris, is recorded from Late Pliocene sediments in the Southern Ocean, coinciding in time but not in space with a regional extinction event in the North Atlantic. The situation is further complicated by an anomaly near the base of the Cambrian, at a level lacking mass extinction and well after the beginning of the Cambrian radiation of skeletonized organisms (Donovan 1987b).

The degree to which all of these iridium anomalies denote impacts is still debated (Hallam 1987; Officer *et al.* 1987; Section 2.12.1). Iridium enrichments may extend for metres around the Cretaceous–Tertiary boundary in some key sections; the significance of these new observations is unclear, with interpretations ranging from diagenetic mobilization from an impact-fallout layer to prolonged deposition from volcanic aerosols. An aerosol from the Hawaiian volcano Kilauea was highly enriched in iridium, apparently derived from the deep mantle; however, other elements in the aerosol do not mimic the extra-terrestrial abundances in end-Cretaceous boundary sequences (W. Alvarez 1986) so that, again, the significance of these data is uncer-

tain. Boundary clay compositions do not always correspond to extra-terrestrial elemental abundances and isotope ratios, and yield conflicting evidence regarding the nature of the hypothesized bolide. It is not clear whether post-impact diagenetic overprint or multiple impacts by bolides of different compositions (expected in cometary bombardment?) can account for such inconsistencies. New analytical techniques (L.W. Alvarez 1987) will permit much more extensive stratigraphic coverage, both at extinction boundaries and at quiet times in between, and thus greatly improve understanding of the global iridium flux and potential nonextra-terrestrial enrichment mechanisms.

Mineralogical. Potential independent evidence for impact comes from shock-metamorphosed quartz and other sedimentary particles. Like iridium, quartz grains with at least two and up to nine intersecting sets of shock lamellae have been found in Cretaceous–Tertiary boundary sequences throughout the world, in both marine and continental settings (Fig. 2) (Bohor *et al.* 1987a; Izett 1987). Such multiple lamellae are known only in particles from nuclear testing sites and impact craters. Shock-metamorphosed minerals do form near certain explosive volcanic eruptions (Hallam 1987), but the multiple lamellae and the world-wide distribution of the relatively large grains (0.1–0.2 mm in North Pacific and New Zealand sediments, up to 0.6 mm in North America) are difficult to reconcile with volcanic activity (W. Alvarez 1986; Bohor *et al.* 1987a). The search for shock-metamorphosed minerals at other extinction events has been negative so far, except for an intriguing preliminary report near the Triassic–Jurassic boundary in Austria (Badjukov *et al.* 1987).

Sedimentological. Microtektites (glassy droplets formed by bolide impacts) are almost undoubtedly present at three horizons near the Eocene–Oligocene boundary (Hut *et al.* 1987). A similar origin has been suggested for spherules of disordered potassium-feldspar (sanidine), glauconite, goethite, and magnetite found world-wide in Cretaceous–Tertiary sequences (W. Alvarez 1986), but recent evidence suggests an authigenic, non-impact origin for at least some spherules (Hallam 1987; Izett 1987). Microspherules of varying composition occur in Permo-Triassic boundary sediments in Sichuan, China (Gao *et al.* 1987); their significance is uncertain in the light of the seemingly volcanic origin of the boundary clays in China.

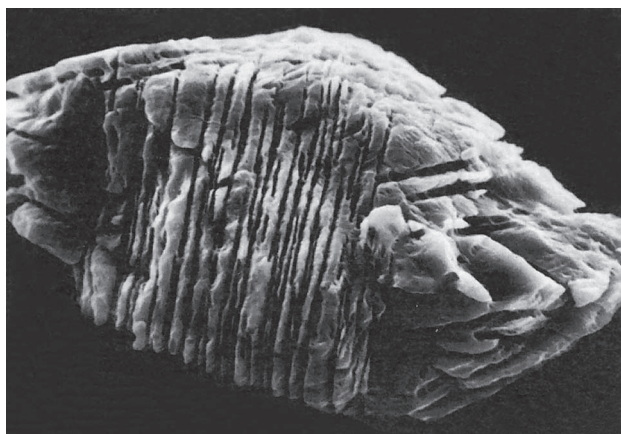


Fig. 2 Shocked quartz grain from Cretaceous–Tertiary boundary clay in a non-marine section at Brownie Butte, Garfield County, Montana. Scanning electron micrograph, width of field 0.14 mm. (Courtesy of B.F. Bohor.)

More work is needed in separating spherules of different origins before interpretations are possible (Bohor *et al.* 1987b).

Soot particles are abundant in Danish and New Zealand Cretaceous–Tertiary boundary clays (W. Alvarez 1986; L.W. Alvarez 1987). If these clays represent only one year of deposition, as postulated by most impact models, the carbon flux would have been 10^3 – 10^4 above background levels, suggesting extensive wildfires triggered by the heat of impact or propagated among the remains of forests killed by the hypothesized post-impact cold interval. However, the uniqueness of such soot occurrences is uncertain, and the high flux depends on the duration of clay layer deposition, which is still debated (Hallam 1987).

Cratering. Major impacts should leave craters at least an order of magnitude larger than the bolide itself. Age uncertainties are troublesome and the data are extremely sparse, but the association between extinction events over the past 250 million years and the 26 well dated craters of 5 km or more in diameter may be statistically significant (reviewed by Shoemaker and Wolfe 1986, who are sceptical). Simulations by Trefil & Raup (1987) suggest that this cratering record comprises about one-third periodic impacts (presumably comet showers) and two-thirds random collisions with asteroids. Shoemaker & Wolfe (1986) reach a similar conclusion by different means.

Questions emerge about the best-studied extinction event, however. The only well dated craters of appropriate size near the Cretaceous–Tertiary

boundary are in the U.S.S.R. (Shoemaker & Wolfe 1986), but the size and density of shocked quartz grains suggests an impact in North America (Bohor *et al.* 1987a; Izett 1987). Further, the shocked quartz suggests an impact in sedimentary rocks, i.e. a continental or shallow-water setting, whereas magnetite and other spherules suggest altered basalt, and thus an oceanic impact (although impact derivation of the spherules is now questioned, as noted above). These contradictions are perhaps resolvable with an end-Cretaceous comet shower and the consequent multiple impact, but the problem of impact site(s) remains (Hallam 1987). A volcanic interpretation is no more satisfactory in this regard.

Conclusion and prospects

Although no one indicator is definitive, at present the diverse physical and chemical evidence at the Cretaceous–Tertiary boundary is most readily explained by a bolide impact. Volcanism is the chief rival, but as W. Alvarez (1986) argued, evidently only quiet basaltic eruptions yield iridium aerosols and melt microspherules, whereas violent siliceous eruptions are needed to produce shocked minerals. Neither kind of eruption will produce all of the observed impact signatures, nor can either account for the world-wide distribution of shocked quartz, iridium and other geochemical anomalies. The periodic mantle plume hypothesis might yield both explosive and non-explosive volcanism on a global scale (Hallam 1987; Section 2.12.1), but this model awaits evaluation. The palaeontological data are generally consistent with, but provide little conclusive support for, impact-driven extinction mechanisms. As many authors have noted, marine regression at this and other extinction events obscures biological and physico-chemical signals and may even play a role in extinction (Section 2.12.1).

The most definitive evidence for or against extra-terrestrial factors in mass extinctions (apart from the discovery of the hypothesized solar companion, Nemesis) will come with an assessment of the strength of temporal association between Phanerozoic mass extinctions and physico-chemical signatures of bolide impacts. This work is under way, and it is impressive that the three or four most recent extinction peaks recognized in global data sets and/or local stratigraphic sections (Middle Miocene, Eocene–Oligocene, Cretaceous–Tertiary, and Cenomanian–Turonian) bear at least some impact indicators. The weak but significant clustering of crater ages at extinction events over the past

250 million years should prompt analyses around other boundaries, with ongoing refinement of hypotheses. Assessment of negative evidence remains a problem, however, so that impact hypotheses can be remarkably elastic and difficult to falsify: absence of craters, shocked quartz and even iridium anomalies are consistent with impact on now-subducted ocean, basaltic impact site, and cometary rather than meteorite impact, respectively. Additionally, not all major craters, microtektite horizons, or iridium anomalies coincide with extinction events. Better understanding of the potential effects of impacts, and of the distribution of potential impact signatures through the stratigraphic record, should lead to the framing of more refined hypotheses regarding the role of extra-terrestrial factors in the evolution of life on Earth.

References

- Alvarez, L.W. 1987. Mass extinctions caused by large bolide impacts. *Physics Today* **40**, 24–33.
- Alvarez, W., 1986. Toward a theory of impact crises. *Eos* **67**, 649, 653–655, 658.
- Badjukov, D.D., Lobitzer, H. & Nazarov, M.A. 1987. Quartz grains with planar features in the Triassic–Jurassic boundary sediments from Northern Limestone Alps, Austria. *Lunar and Planetary Science* **18**, 38–39.
- Bohor, B.F., Modreski, P.J. & Foord, E.E. 1987a. Shocked quartz in the Cretaceous–Tertiary boundary clays: evidence for a global distribution. *Science* **236**, 705–709 (see also 666–668).
- Bohor, B.F., Triplehorn, D.M., Nichols, D.J. & Millard, H.T., Jr. 1987b. Dinosaurs, spherules, and the ‘magic’ layer: a new K–T boundary clay site in Wyoming. *Geology* **15**, 896–899.
- Donovan, S.K. 1987a. Iridium anomalous no longer? *Nature* **326**, 331–332.
- Donovan, S.K. 1987b. Confusion at the boundary. *Nature* **329**, 288.
- Gao Zhengang, Xu Daoyi, Zhang Qinwen & Sun Yiyin 1987. Discovery and study of microspherules at the Permian–Triassic boundary of the Shangsi section, Guangyuan, Sichuan. *Geological Review* **33**, 203–211 (in Chinese with English abstract).
- Hallam, A. 1987. End-Cretaceous mass extinction event: argument for terrestrial causation. *Science* **238**, 1237–1242.
- Hut, P., Alvarez, W., Elder, W.P., Hansen, T., Kauffman, E.G., Keller, G., Shoemaker, E.M. & Weissman, P.R. 1987. Comet showers as a cause of mass extinctions. *Nature* **329**, 118–126.
- Izett, G.A. 1987. Authigenic ‘spherules’ in K–T boundary sediments at Caravaca, Spain and Raton Basin, Colorado and New Mexico, may not be impact derived. *Bulletin of the Geological Society of America* **99**, 78–86.
- Jablonski, D. 1986a. Causes and consequences of mass extinctions: a comparative approach. In: D.K. Elliott (ed.)

- Dynamics of extinction*, pp. 183–229. Wiley, New York.
- Jablonski, D. 1986b. Evolutionary consequences of mass extinctions. In: D.M. Raup and D. Jablonski (eds) *Patterns and processes in the history of life*, pp. 313–329. Springer-Verlag, Berlin.
- Officer, C.B., Hallam, A., Drake, C.L. & Devine, J.D. 1987. Late Cretaceous and paroxysmal Cretaceous/Tertiary extinctions. *Nature* **326**, 143–149.
- Prinn, R.G. & Fegley, B., Jr. 1987. Bolide impacts, acid rain, and biospheric traumas at the Cretaceous–Tertiary boundary. *Earth and Planetary Science Letters* **83**, 1–15.
- Raup, D.M. 1987. Mass extinction: a commentary. *Palaeontology* **30**, 1–13.
- Rocchia, R., Boclet, D., Bonte, P., Castellarin, A. & Jehanno, C. 1986. An iridium anomaly in the Middle–Lower Jurassic of the Venetian region, northern Italy. *Journal of Geophysical Research* **91**, E259–E262.
- Shoemaker, E.M. 1984. Large body impacts through geologic time. In: H.D. Holland & A.F. Trendall (eds) *Patterns of change in Earth evolution*, pp. 15–40. Springer-Verlag, Berlin.
- Shoemaker, E.M. & Wolfe, R. 1986. Mass extinctions, crater ages, and comet showers. In: R.S. Smoluchowski, J.N. Bahcall & M.S. Matthews (eds) *The galaxy and the solar system*, pp. 338–386. University of Arizona Press, Tucson.
- Trefil, J.S. & Raup, D.M. 1987. Numerical simulations and the problem of periodicity in the cratering record. *Earth and Planetary Science Letters* **82**, 159–164.
- Wolfe, J.A. 1987. Late Cretaceous–Cenozoic history of deciduousness and the terminal Cretaceous event. *Paleobiology* **13**, 215–226.
- Zachos, J.C. & Arthur, M.A. 1986. Paleocyanography of the Cretaceous/Tertiary boundary event: inferences from stable isotopic and other data. *Paleocyanography* **1**, 5–26.
- Zachos, J.C., Arthur, M.A. & Dean, W.E. 1989. Geochemical evidence for suppression of pelagic productivity at the Cretaceous/Tertiary boundary. *Nature* **337**, 61–64.

2.12.3 Periodicity

J. J. SEPKOSKI, Jr

Introduction

Periodicity of extinction is a hypothesis that extinction events (both mass extinctions and their less severe analogues) have occurred at regularly spaced intervals through geological time. It is an empirical claim based upon statistical analyses of the fossil record which indicate that maxima in extinction intensity, recognized in both biostratigraphic studies and taxonomic data compilations, are decidedly non-random with respect to time and seem to fit a regular, periodic time series. This hypothesis was introduced by Fischer & Arthur (1977) for

patterns of diversity in open-ocean pelagic communities and later supported by Raup & Sepkoski (1984), who claimed a 26 million year periodicity in extinction of global marine families. The hypothesis has since proved very controversial largely as a result of association with suggestions of catastrophic, extra-terrestrial forcing agents.

Meaning of periodicity

A perfectly periodic time series has regularly spaced events separated by invariant waiting times (Fig. 1). In most of the debate about periodicity, this pattern has been contrasted with a 'random', or Poisson, time series. A Poisson series can arise when events are independent of one another and determined by a large number of unrelated factors. A classic example is coin flipping, in which the outcome (heads or tails) of each trial results from a multitude of independent forces. The lower time series in Fig. 1 was generated by flipping a pair of coins and recording when both came up heads. The frequency of events (one in four trials) is the same as in the upper, periodic series, but the appearance is very different. The lower series is composed of loose clusters of events with irregular gaps in between; waiting times approach an exponential distribution with the median waiting time shorter than the average frequency.

The relevance of these considerations to the study of extinction events is that traditionally each event has been analysed in isolation from others and independent causal hypotheses have been formulated. Implicit in this is the assumption that extinction events must be randomly spaced in time. Observation of regular spacing, however, implies some organizing principle to extinction events, either some set of factors that governs waiting times so that they appear invariant, or some single ultimate forcing agent that has clock-like properties. Periodicity can also imply that the proximate agent of any one extinction event is the same for all, although this is not a necessary implication if the chain of causation is complex.

The association of periodicity with catastrophism comes from these last considerations. In particular, it has been suggested that: (1) the claimed 26 million year periodicity of extinction events is too long to have been produced by any known terrestrial process with periodic behaviour, leaving some astronomical clock as the likely forcing agent; and (2) the association of the Cretaceous–Tertiary mass extinction with evidence of a large extra-terrestrial



Fig. 1 The contrast between a periodic and a random time series. Both series have the same frequency of 'events', but the random series is characterized by irregular clusters of events with variable gaps between them. (After Sepkoski 1986.)

impact (Section 2.12.2) suggests (as a hypothesis to be tested) that other events in the periodic series might have been similarly caused. Note that these arguments suggest only a possible association, and other, terrestrial mechanisms are still conceivable (Section 2.12.1).

Evidence for periodicity

The hypotheses of periodicity put forward by Fischer & Arthur (1977) and Raup & Sepkoski (1984) were based upon compilations of diversity data and extinction times for taxa in the marine fossil record. Fischer & Arthur were concerned with recurrent fluctuations in the diversity of globigerinid species, ammonoid genera, and large pelagic predators through the Mesozoic and Cenozoic. They argued, without rigorous statistical testing, that these fluctuations were cyclic with a 32 million year waiting time. Using family-level data for the entire marine ecosystem, updated time-scales, and a variety of statistical tests, Raup & Sepkoski corroborated the Fischer–Arthur hypothesis but concluded that the period length was closer to 26 million years. Their statistical tests (which included parametric Fourier and autocorrelation analyses and non-parametric randomization analysis) all indicated a significant non-randomness in the distribution of extinction events and a good, but not perfect, fit to a periodic series.

Raup & Sepkoski's (1984) treatment and testing of familial extinctions were somewhat complex and have led to some confusion. Their analysis was limited to families in the Late Permian through Neogene, where stratigraphic stages are shorter and more accurately dated than in the preceding Palaeozoic. To enhance resolution, only families with extinctions known to the stage level were used

and taxa of soft-bodied and lightly sclerotized animals, or of very uncertain taxonomic position, were rejected. These manipulations left a data set of 567 extinct families ranging over 39 stratigraphic stages. Extinction intensity was measured by percent extinction, the number of extinctions in a stage divided by diversity. This metric (statistical measure) scales extinction to the number of families at risk in any stage but does not incorporate estimates of stage duration, which have limited accuracy. Percent extinction for families exhibits very low values over the Cenozoic, leaving peaks of extinction difficult to discern; Raup & Sepkoski therefore used only the diversity of families extinct before the Recent in the denominator of the metric, inflating its values in the Cenozoic.

The time series constructed by this treatment (Fig. 2) contains 'peaks', or local maxima, that vary considerably in height. Raup & Sepkoski recognized that some of these (e.g. the Guadalupian, Rhaetian, and Maastrichtian) correspond to well documented mass extinctions, but that some of the lower peaks might be spurious. Nevertheless, they chose to analyse all peaks rather than a selected subset, in order to avoid possible subjective bias. Unfortunately, they referred to all 12 peaks as 'mass extinctions'.

A randomization test for periodicity was favoured by Raup & Sepkoski (1984) because it permitted fitting a wide band of period lengths and was not sensitive to unequal spacing of data (imposed by the stratigraphic time-scale) or to variation in magnitudes of extinction peaks (which were presumed to fluctuate freely). The test (which is akin to bootstrap procedures) involved fitting periodicities to the observed extinction peaks and then comparing the goodness of fits to randomized (i.e. shuffled) versions of the data. The peaks were treated as if they all fell at the ends of stages; this, however, was merely a formalization, and equivalent results would have obtained if the peaks were consistently placed at the middles or beginnings. The shuffling procedure converted the data into what was essentially a random walk with the only constraint being that peaks must be separated by at least two stages.

The randomization test showed that periodicity fits the observed data better than 99.99% of random walks at 26 million years, even though the fit to the peaks (especially the smaller peaks) was not perfect (Fig. 2). On this basis, Raup & Sepkoski concluded that there was a 26 million year periodicity to 'mass extinctions' through the Mesozoic and Cenozoic Eras. No periodicity was found in the Palaeozoic, however. Rampino & Stothers (1984) corroborated

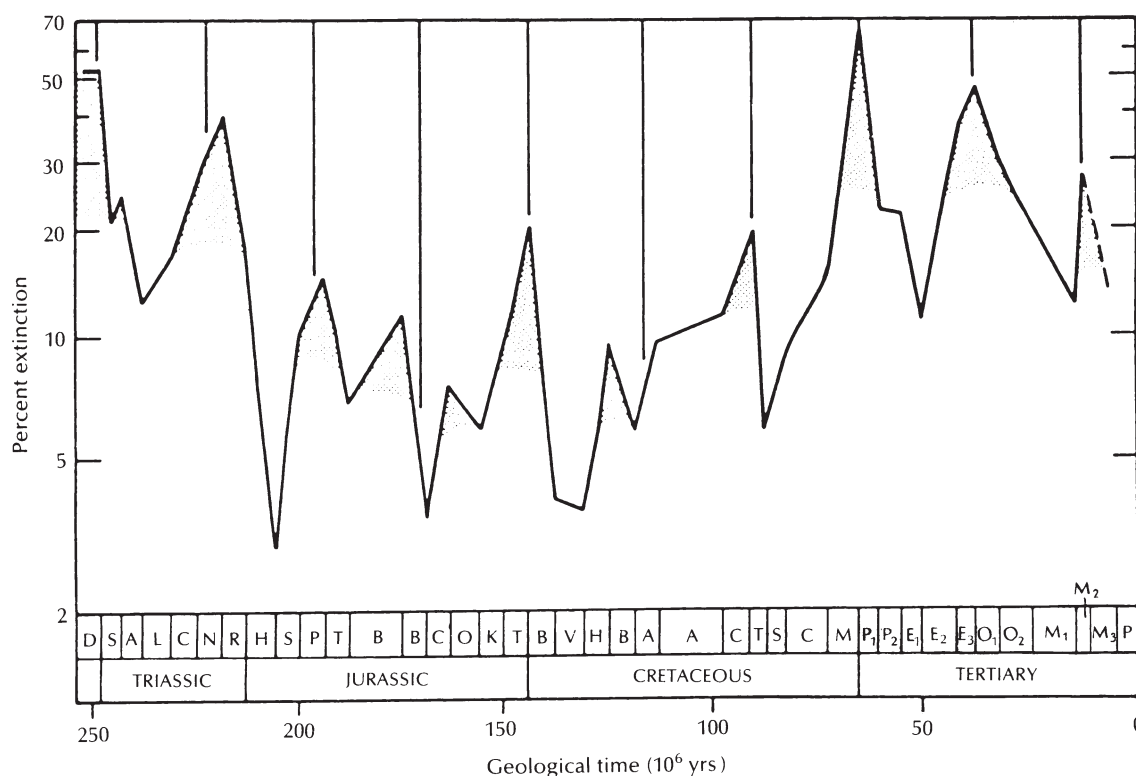


Fig. 2 Raup & Sepkoski's (1984) time series for familial extinction from the Late Permian to Neogene, computed from their highly culled sample of marine animal families. The 12 'peaks' of extinction are highlighted by stippling, and the fit to the 26 million year periodicity is indicated by the vertical lines (note the lack of fit in the Middle Jurassic and Early Cretaceous). Geological systems and stages (small boxes with initial letters of names) are indicated along the bottom of the graph (note the variable stage lengths, especially through the Cretaceous and Tertiary). The y axis is logarithmic. Abbreviations along the time axis: D = Dzhulfian; S = Scythian, A = Anisian, L = Ladinian, C = Carnian, N = Norian, R = Rhaetian; H = Hettangian, S = Sinemurian, P = Pliensbachian, T = Toarcian, B = Bajocian + Aalenian, B = Bathonian, C = Callovian, O = Oxfordian, K = Kimmeridgian, T = Tithonian; B = Berriasian, V = Valangian, H = Hauterivian, B = Barremian, A = Aptian, A = Albian, C = Cenomanian, T = Turonian + Coniacian, S = Santonian, C = Campanian, M = Massstrichtian, P₁ = Danian, P₂ = Upper Palaeocene, E₁ = Lower Eocene, E₂ = Middle Eocene, E₃ = Upper Eocene, O₁ = Lower Oligocene, O₂ = Upper Oligocene, M₁ = Lower Miocene, M₂ = Middle Miocene, M₃ = Upper Miocene, P = Pliocene.

this result, even after eliminating the three smallest peaks in the time series. A similar period-fitting technique applied to the nine remaining peaks gave a 26 million year period. However, a regression-based technique resulted in a 30 million year period, which they favoured on other grounds.

Subsequent analyses performed by Raup & Sepkoski were designed to counter criticisms of their data manipulation and statistical procedures, and to explore the correspondence between global taxonomic data and information from biostratigraphic studies. Sepkoski & Raup (1986) reanalysed the familial data using all extinctions (other than those of soft-bodied animals tied to Konservat-Lagerstätten; Section 3.11) and employing total diversity in the metrics. Fig. 3 illustrates the time

series for percent extinction in this analysis. Three other metrics of extinction were also computed and an attempt was made to assess which extinction peaks could be considered statistically significant. Sepkoski & Raup determined only eight of their previous 12 peaks to be significant and found that the heights of these peaks were generally lower than in the highly culled data set. They argued, however, that seven of the peaks corresponded to extinction events recognized by palaeontologists working at the species level with material collected from outcrops and cores. This indicated to the authors that global familial data could be trusted to reflect important extinction patterns among species in the fossil record.

Sepkoski & Raup (1986) found that the random-

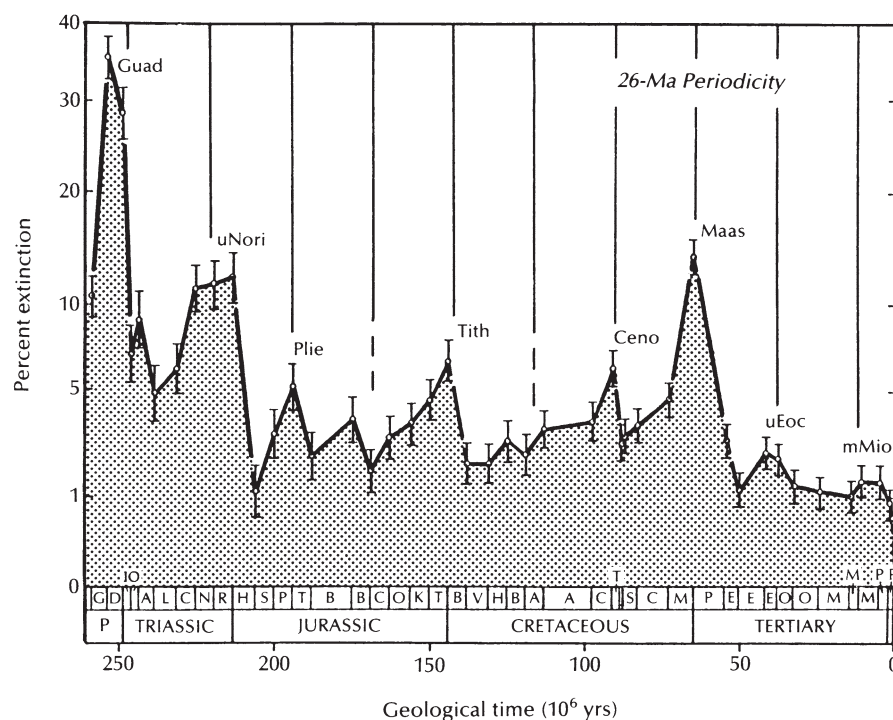


Fig. 3 Percent extinction for all marine animal families based on a data set with 970 extinctions. Peaks that are 'significantly' higher than background are labelled. (Guad = Guadalupian, uNori = Upper Norian or Rhaetian, Plie = Pliensbachian, Tith = Tithonian, Ceno = Cenomanian, Maas = Maastrichtian, uEoc = Upper Eocene, mMio = Middle Miocene, other abbreviations as in Fig. 2). Error bars indicate one estimated standard error on either side of the data points. The y axis is a square root scale. (After Sepkoski & Raup 1986, by permission of the AAAS.)

ization test applied to the eight extinction peaks still indicated a significant periodicity at 26 million years, with the standard error judged to be about ± 1 million years when imprecisions in the time-scale were accommodated. In a companion paper, Raup & Sepkoski (1986) showed that the level of statistical significance of the randomization test varied somewhat if different ages were assigned to the less precisely dated events (end-Permian and end-Triassic). Still, they concluded that most fits of the 26 million year periodicity were significant at or above the 95% level, even after adjustment for the problem of multiple tests (i.e. testing many frequencies in the 12–60 million year band).

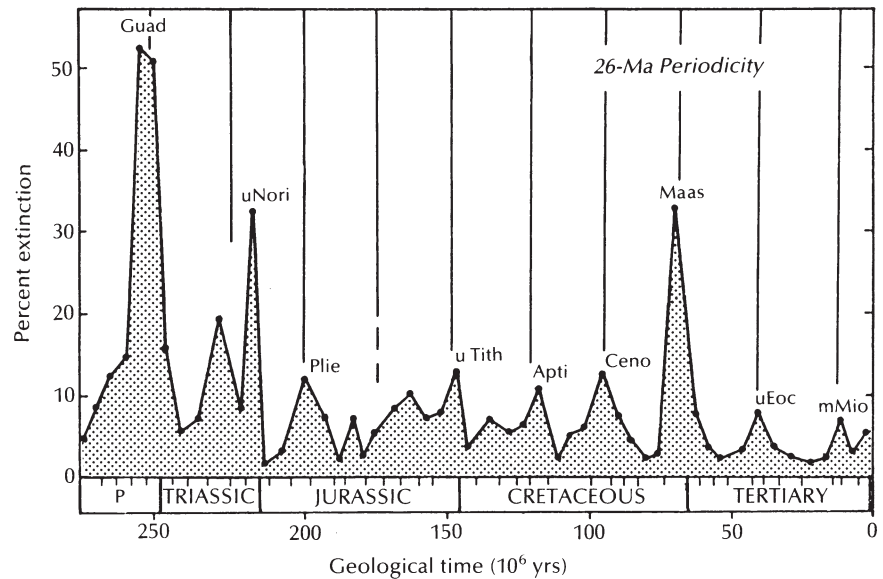
Raup & Sepkoski (1986) and Sepkoski (1986) also conducted analyses at the generic level, using a new compilation for global marine animals. This was done to increase sample size and to obtain a better approximation of species patterns. Higher taxa tend to damp the signal of species extinction since all species within a polytypic taxon must disappear for the taxon to register an extinction event. The new data set contained nearly 10 000 genera in the interval from Upper Permian to Recent. It also incorporated a refined stratigraphic time-scale with 51 intervals (in contrast to the previous 39–43 stages).

Fig. 4 illustrates one of four time series for generic extinction. As expected, the eight peaks of extinction

are more prominent than in the familial data. The peak in the Middle Miocene seems to be confirmed and an extinction event is suggested in the Aptian, which previously appeared as a gap in the periodic sequence (Fig. 3). A gap still exists in the Middle Jurassic despite two questionable peaks (Lower Bajocian and Oxfordian). These two peaks, as well as the Carnian peak (to the left of the Upper Norian peak in Fig. 4), fluctuate erratically with different metrics of extinction, suggesting that they are not robust features of the data.

Raup & Sepkoski (1986) performed the randomization test on these data and concluded that they contained the 26 million year periodicity of extinction. Sepkoski (1986) also performed autocorrelation analyses (i.e. correlating a time series with itself at a given time lag, which assesses amplitude as well as wavelength) and obtained statistically significant results consistent with a 26 million year periodicity. Finally, Fox (1987) performed an elaborate series of Fourier analyses on the generic data and also found a significant 26 million year periodicity. This was true even when he split the time series into two parts: both halves displayed a periodicity with the same wavelength and, very importantly, nearly the same phase. None of these analyses of the generic data showed decisive evidence for a periodicity prior to the Permian, however, although Sepkoski

Fig. 4 Percent extinction for marine genera. (After Sepkoski 1986.) The data illustrated comprise 9773 genera, of which 5594 are extinct. Note that peaks of extinction (labelled as in Fig. 3) are better defined than in the total familial data, and a seemingly periodic Aptian ('Apti') peak appears in the Early Cretaceous. No clearly periodic peak is present in the Middle Jurassic, although low, possibly spurious peaks appear over the Lower Bajocian and Oxfordian; a peak also appears over the Carnian within the Late Triassic.



(1986) did suggest a possible longer (greater than 30 million years) periodicity in the late Palaeozoic.

Critiques of periodicity

The hypothesis of periodicity in extinction engendered immediate attention from scientists as well as the popular press. Not surprisingly, this led to intense scrutiny of both the data and the statistical analyses. The result has been a complex series of critical discussions with various responses by Raup & Sepkoski (see Sepkoski 1989), which can only be briefly summarized here.

Data. The validity of compilations of taxonomic data has been questioned by several authors. Hoffman (1985) argued that familial data are very noisy and that different treatments, including application of alternative time-scales, results in different, seemingly random patterns of extinction peaks. This claim was countered by Sepkoski & Raup's (1986) demonstration of consistency of eight extinction peaks under four different metrics and by Sepkoski's (1986) argument that even Hoffman's composite data display strong periodicity. The presence of the same periodic extinction peaks in the much larger generic data would also seem to indicate signal rather than noise.

Stigler & Wagner (1987), however, argued that periodicity even in the generic data could be an artifact of imperfect sampling of the fossil record. Failure to sample taxa in their last stage of existence will smear the record of extinction backward in time. This will tend to swamp some minor extinction

peaks between major maxima and cause the time series to appear more regular than expected for a Poisson distribution. The counterargument to this claim (Sepkoski & Raup 1986) is simply that detailed biostratigraphic investigations corroborate most of the extinction peaks evident in the generic data, and do not indicate many smaller extinction events in other stages (although some major extinction events may be composites of tightly clustered steps).

Patterson & Smith (1987) questioned the accuracy of any taxonomic compilation that contains paraphyletic taxa (see Section 5.3). They claimed that three-quarters of the families of echinoderms and vertebrates used by Raup & Sepkoski were paraphyletic, monotypic, and/or misdated. When a corrected monophyletic component (equivalent to 10% of Raup & Sepkoski's total data set) was examined, no periodicity was evident. Sepkoski (1987) responded that paraphyly in itself should not be a problem since family extinctions simply represent a sample of species extinctions. He further noted that the monophyletic taxa in Patterson & Smith's analysis failed to show some well documented extinction events (e.g. the Maastrichtian mass extinction) and suggested that this might be due to small sample size, idiosyncracies in the echinoderm and vertebrate records, or biases inherent in the cladistic culling.

Inaccuracies in the estimated ages of stratigraphic intervals used in the data sets pose numerous problems. As noted above, Hoffman (1985) argued that use of different time-scales causes differences in extinction peaks. Shoemaker & Wolfe (*in* Smoluchowski *et al.* 1986) assessed the estimated

ages of Raup & Sepkoski's (1984) 12 extinction peaks and concluded that only three (the Cenomanian, Maastrichtian, and Upper Eocene) were reliable; this was too small a sample to support periodicity. Raup & Sepkoski (1986), however, showed that their randomization test did give significant results for the last four, best-dated extinction events (including the Middle Miocene event which Shoemaker & Wolfe rejected on the basis of familial data, but which Raup and Sepkoski accepted on the evidence of the generic data).

Stigler & Wagner (1987) questioned the strength of this test, arguing that the 26 million year periodicity might be embedded in the time-scale. This is not surprising, however, since some stratigraphic boundaries are placed at points of major turnover (e.g. the Palaeozoic–Mesozoic and Mesozoic–Cenozoic boundaries). Potential coupling of the stratigraphical and biological records was recognized by Raup & Sepkoski (1984, 1986), who shuffled the time-scale in their tests in order to avoid this problem. It should be noted that the 51-interval time-scale used in the generic data, with longer stages subdivided and shorter stages amalgamated, does not display any embedded 26 million year periodicity.

Statistical analyses. Many technical aspects of the statistical tests conducted by Raup & Sepkoski have been questioned. Hoffman & Ghiold (1985) claimed that the analyses did not properly test for a random walk. They argued that the familial data displayed a mean frequency of one peak in every four stages, which is indistinguishable from the expectation of a random walk. But these authors (and likewise Noma & Glass 1987) failed to recognize that Raup & Sepkoski's randomization procedure in essence converted the extinction data into random walks (although perhaps with less variance than proper, as pointed out by Quinn (1987)). Also, Sepkoski's (1986) autocorrelation analysis with the refined time-scale showed a peak every fifth interval in the generic data, which is not consistent with a random walk.

Noma & Glass (1987) used turning points in the familial data to argue that the hypothesis of randomness could not be rejected. However, their test was very sensitive to variance in stage durations (which range from 1 million years for the Coniacian to 15.5 million years for the Albian), and it is unclear whether Noma & Glass demonstrated anything more than this variance. They also argued that there were flaws in the selection of 'significant' extinction

peaks by Raup & Sepkoski (1986) (as well as Sepkoski & Raup 1986). This argument is valid, and at best Sepkoski & Raup merely eliminated demonstrably insignificant peaks from their familial analyses. However, other evidence presented by Sepkoski & Raup suggests that the remaining eight peaks were not insignificant since: (1) the same peaks appeared even more prominently in the generic time series (Fig. 4); and (2) most of the peaks correspond to independently identified events in biostratigraphic analyses.

Kitchell & Pena (1984) reanalysed the familial data assuming equal durations of stages and applying a series of autoregressive models (i.e. regression equations in which values in each time interval are predicted from values in preceding intervals). They rejected a simple model with periodic impulses but found adequate fits with a model incorporating five-stage memory, which they concluded demonstrated only pseudoperiodicity in the data. However, the rejected simple periodic model imposed a regular amplitude as well as wavelength, and required equal numbers of stages between extinction peaks. (The number of stages between Raup & Sepkoski's periodic peaks varied from two to six; see Fig. 2). Again, Sepkoski's (1986) autocorrelation analysis of the generic data suggested that a simple periodic impulse model could provide a statistically significant fit when the stratigraphic intervals were adjusted to be more equal in length.

Quinn (1987) criticized Raup & Sepkoski's (1984) randomization test for ignoring the autocorrelation in the data (although Stigler & Wagner (1987) did not consider this to be a problem). Quinn failed to note that Raup and Sepkoski had recognized this problem and used only randomizations that had the same number of peaks as observed in the data. Quinn offered an alternative test that compared waiting times between peaks to the expectation of random events (a broken-stick distribution). This test, he claimed, failed to demonstrate any evidence of periodicity in either the familial or the generic data. Unfortunately, he used an arbitrary definition of 'mass extinction' (either all stages with extinction intensities in the upper quartile of the data, or all peaks exceeding the mean intensity after log–linear adjustment for temporal trends). His test appears to be sensitive to the number of points selected and could reject a moderately noisy sine curve if the number of points exceeded the number of cycles.

Running Quinn's test for different numbers of cycles or points would have presented difficulty in assessing the significance level for multiple tests.

Quinn (1987) complained that Raup & Sepkoski (1984) did not calculate the joint significance level for the 49 independent tests that were conducted in assessing all periodicities between 12 and 60 million years (although Raup & Sepkoski did attempt to tackle this, albeit incorrectly). Quinn claimed the joint significance level was only 39%, given a significance of 99% for the fit of the 26 million year period. Tremaine (*in* Smoluchowski *et al.* 1986) calculated the joint significance level to be 95.4%, using a recomputed significance of 99.74% for the 26 million year period. Tremaine went on to argue, however, that random simulations run over the 12–60 million year band indicated a joint significance level of less than 90% for the 12 peaks of Raup & Sepkoski (1984) and less than 50% for the eight peaks of Sepkoski & Raup (1986). But these results may have been sensitive to his assumption that variance in fit was directly proportional to period length in his tests. Raup & Sepkoski (1986) used Tremaine's procedure without this assumption and obtained joint significance levels greater than 95%.

All of these tests and arguments have used a Poisson model of randomness as a basis of comparison. Lutz (1987) argued that this is not the only alternative in testing for periodicity. He tested Raup & Sepkoski's (1984) familial time series against models for Poisson distributions, 'noisy' periodicities, and constrained episodicities (i.e. γ distributions in which the standard deviation in waiting times is less than the mean waiting time). He found that the Poisson model could be rejected at the 95% significance level, but he could not distinguish between fits of noisy periodicities and of episodicities with variances less than 30% of mean waiting time (although it is not clear how sensitive these results are to selection of events and to errors in the time-scale).

Lutz (1987) concluded that an exogenous forcing agent with clock-like behaviour was not necessary to explain the data. Stanley (1987) proffered a similar argument on qualitative grounds. He suggested that extinction events eliminate particularly vulnerable taxa and that there is a lag time after each event during which few vulnerable taxa are available for extinction. Thus, palaeontologically recognizable perturbations should be spaced more widely than expected from a Poisson distribution. The counter to this argument is that recovery times observed for most extinction events in the Mesozoic and Cenozoic are only one or two stages, which is within the lag time built into Raup & Sepkoski's randomization procedure.

It cannot be claimed that any of these arguments and counterarguments is decisive, and it is doubtful whether new, more accurate data could settle the matter (although more precise data would certainly promote better understanding of extinction in the fossil record). A definitive settlement will be reached only if a clear agent of periodic extinction is discovered.

Possible causes of periodicity

Both terrestrial and extra-terrestrial mechanisms have been suggested as ultimate causes of periodicity in extinction. The terrestrial mechanisms involve hypothetical quasiperiodic processes in the deep Earth that lead to episodes of intense volcanism. The extra-terrestrial mechanisms involve a variety of observed and hypothesized astronomical clocks that might induce periodic cometary bombardments of the Earth.

Evidence that extra-terrestrial impacts might be important in periodic extinction come from two sets of observations (see also Section 2.12.2):

- 1 Materials presumed to be of impact origin (excess iridium, microtektites, and/or shocked mineral grains) are associated with several periodic extinction events, including the Cenomanian, Maastrichtian, Upper Eocene, and Middle Miocene.
- 2 Ages of terrestrial craters seem to exhibit a weak periodicity, involving 25–50% of impacts, that has a phase and period length (variously estimated at 27–32 million years) that are roughly congruent with the extinction periodicity (see Shoemaker & Wolfe *in* Smoluchowski *et al.* 1986).

The periodic impactors are presumed to be comets derived from the Oort Cloud at the outer fringes of the Solar System. It has been hypothesized that a gravitational perturbation from a body as small as four times Jupiter's mass could induce a comet shower that would bring up to 10^9 comets into the inner Solar System; about 25 of these on average would strike the Earth over a 1 million year interval. Four mechanisms, all of which are flawed, have been suggested to produce such comet showers periodically (reviewed by Sepkoski & Raup 1986; Shoemaker & Wolfe *in* Smoluchowski *et al.* 1986):

- 1 A dim binary companion to the Sun, dubbed 'Nemesis'. This small star is hypothesized to have a highly eccentric orbit with a mean revolution time of 26–28 million years. At aphelion, it would pass through the Oort Cloud and induce a comet shower. However, a distant companion has never been ob-

served, and simulations indicate it would be unstable and easily stripped from its orbit by passing field stars and molecular clouds.

2 An unobserved tenth planet, usually called 'Planet X'. If it had a slightly eccentric orbit inclined to the plane of the Solar System, orbital precession could bring the perihelion into the solar plane twice every 52–56 million years, at which time the planet would scatter comets from the inner edge of the Oort Cloud. However, a tenth planet has never been observed, and it is not clear whether it would have sufficient mass to scatter enough comets to leave a recognizable periodic signature on Earth.

3 Oscillation of the Solar System perpendicular to the Galactic plane. This well known behaviour moves the Solar System every 31–33 million years through the dense plane of the Galaxy, where gravitational encounters with molecular clouds might perturb the Oort Cloud. However, the oscillation is out of phase with the extinction periodicity, and it has been argued that the mass of the Galaxy is not sufficiently concentrated in the plane to affect any distinct periodicity over a 270 million year interval.

4 Quasiperiodic transit of the Solar System through the spiral arms of the Galaxy. During its galactic orbit, the Solar System passes through either two or four arms, where concentrated mass may perturb the Oort Cloud. However, the intervals between transits are either about 60 or 125 million years, which is much longer than the observed periodicity of extinction.

Alternative hypotheses that deep-Earth processes could induce periodic extinction are based on two lines of evidence (see also Section 2.12.1): (1) there is an arguable periodicity of around 30 million years in the frequency of reversals of the Earth's magnetic field, suggesting some kind of regularity in deep-Earth dynamics (Loper *et al.* 1988); and (2) several periodic extinction events are associated with immense volcanic deposits (e.g. the Siberian traps, Deccan traps, and Columbia River basalts), which were produced during major episodes of basaltic volcanism. Such episodes could release large quantities of particulates, sulphates, and carbon dioxide into the atmosphere, perturbing climate and inducing extinction.

Loper *et al.* (1988) argued that major volcanic episodes would be quasiperiodic if they were caused by variation in the thickness of the thermal layer at the base of the mantle. Thickening of this layer through time could lead to dynamical insta-

bilities that would spawn mantle plumes and cause widespread basaltic volcanism. Release of the plumes would draw material from the thermal layer, re-establishing stability and thus limiting the duration of the volcanic episode.

This hypothesis of terrestrial forcing challenges, but does not negate, a role for extra-terrestrial impacts in producing the observed distribution of extinction events: coincidental impact during a volcanic episode could greatly amplify a biotic crisis. Both sets of hypotheses are consistent with the implication from periodicity that most Mesozoic–Cenozoic extinction events share a common ultimate cause. But, as Lutz (1987) noted, the deep-Earth mechanism is not strictly clocklike but would operate by constraining waiting times between events to generate the non-random distribution that is seen in the fossil record of extinction.

References

- Fischer, A.G. & Arthur, M.A. 1977. Secular variations in the pelagic realm. In: H.E. Cook & P. Enos (eds) *Deep-water carbonate environments*. Special Publication of the Society of Economic Paleontologists and Mineralogists No. 25, pp. 19–50.
- Fox, W.T. 1987. Harmonic analysis of periodic extinctions. *Paleobiology* **13**, 257–271.
- Hoffman, A. 1985. Patterns of family extinction: dependence on definition and geologic time scale. *Nature* **315**, 659–662.
- Hoffman, A. & Ghiold, J. 1985. Randomness in the pattern of 'mass extinctions' and 'waves of originations'. *Geological Magazine* **122**, 1–4.
- Kitchell, J.A. & Pena, D. 1984. Periodicity of extinctions in the geologic past: deterministic versus stochastic explanations. *Science* **226**, 689–692.
- Loper, D.E., McCartney, K. & Buzyna, G. 1988. A model of correlated episodicity in magnetic-field reversals, climate, and mass extinctions. *Journal of Geology* **96**, 1–16.
- Lutz, T.M. 1987. Limitations to the statistical analysis of episodic and periodic models of geologic time series. *Geology* **15**, 1115–1117.
- Noma, E. & Glass, A.L. 1987. Mass extinction pattern: result of chance. *Geological Magazine* **124**, 319–322.
- Patterson, C. & Smith, A.B. 1987. Is the periodicity of extinctions a taxonomic artefact? *Nature* **330**, 248–251.
- Quinn, J.F. 1987. On the statistical detection of cycles in extinctions in the marine fossil record. *Paleobiology* **13**, 465–478.
- Rampino, M.R. & Stothers, R.B. 1984. Terrestrial mass extinctions, cometary impacts and the Sun's motion perpendicular to the galactic plane. *Nature* **308**, 709–712.
- Raup, D.M. & Sepkoski, J.J., Jr. 1984. Periodicity of extinctions in the geologic past. *Proceedings of the National Academy of Sciences, U.S.A.* **81**, 801–805.
- Raup, D.M. & Sepkoski, J.J., Jr. 1986. Periodic extinction of

- families and genera. *Science* **231**, 833–836.
- Sepkoski, J.J., Jr. 1986. Global bioevents and the question of periodicity. In: O. Walliser (ed.) *Global bio-events*, pp. 47–61. Springer-Verlag, Berlin.
- Sepkoski, J.J., Jr. 1987. Is the periodicity of extinction a taxonomic artefact? Response. *Nature* **330**, 251–252.
- Sepkoski, J.J., Jr. 1989. Periodicity in extinction and the problem of catastrophism in the history of life. *Journal of the Geological Society of London* **146**, 7–19.
- Sepkoski, J.J., Jr. & Raup, D.M. 1986. Periodicity in marine extinction events. In: D.K. Elliott (ed.) *Dynamics of extinction*, pp. 3–36. Wiley, New York.
- Smoluchowski, R.S., Bahcall, J.N. & Matthews, M.S. (eds) 1986. *The galaxy and the solar system*. University of Arizona Press, Tucson.
- Stanley, S.M. 1987. *Extinction*. Scientific American Books, New York.
- Stigler, S.M. & Wagner, M.J. 1987. A substantial bias in non-parametric tests for periodicity in geophysical data. *Science* **238**, 940–945.

2.13 Mass Extinction: Events

2.13.1 Vendian

M. A. S. McMENAMIN

Introduction

The earliest known, reasonably well documented mass extinction is of Vendian age, and seems to have occurred in the middle part of the Vendian, about 650 Ma. The severity and timing of this extinction is somewhat obscured by the difficulty of obtaining precise dates for Vendian sediments. Also, some losses of Vendian diversity appear to be the continuation of declines that began before the beginning of the Vendian, such as the loss of many different types of stromatolites.

Micro-organisms

Stromatolites reached a peak in diversity (nearly 100 recognized taxa) in the Late Riphean (c. 850 Ma). Following this acme, stromatolites underwent a precipitous decline (see also Section 1.5) starting in the second half of the Late Riphean and continuing through the Vendian. Stromatolite diversity bottomed out at less than 30 taxa by the beginning of the Cambrian. Although this decline does not necessarily represent the extinction of any of the individual microbial species that participated in the formation of stromatolites, it does indicate that the conditions became much less favourable for many formerly successful types of benthic microbial communities. For example, well formed specimens of the conical Proterozoic stromatolite *Conophyton*

are unknown after the Vendian. The advent of burrowing and grazing metazoans, and disturbance to microbial mats as a result of their activities, has been hypothesized as the factor responsible for the decline of stromatolites.

Individual taxa of benthic microbial organisms (Section 1.2), represented by delicate unicells and filamentous chains of cells preserved in chert, seem to have been largely unaffected by extinction during the Vendian, although it is difficult to recognize taxonomic turnover in floras consisting primarily of morphologically simple coccoidal and filamentous microbes. This problem is further compounded by the fact that fossilized benthic microbiotas are rare after the beginning of the Cambrian; apparently the conditions necessary for fossilization of microbes in chert became much less common after the end of the Vendian.

A different situation exists with acritarchs, a heterogeneous group of organic-walled microfossils recovered from sediment by acid maceration. By comparison with modern dinoflagellate cysts, most acritarchs are thought to represent the resting stages of planktic, eukaryotic marine algae (Section 1.7.2). Both within-flora and total taxonomic diversity of these planktic microfossils underwent a severe decline during the Middle to Late Vendian, which Vidal & Knoll (1982) regarded as indicative of major extinctions in the eukaryotic phytoplankton. Diagnostic acritarch taxa such as *Trachysphaeridium laufeldi* and the distinctively striate *Kildinella lophostriata* (Vidal & Knoll 1982) disappeared by the Middle Vendian.

These distinctive Late Riphean–Early Vendian acritarchs were succeeded by a depauperate flora typified by *Bavlinella faveolata* (an acritarch that

resembles the existing colonies of spherical cyanobacteria called chroococcaleans) and the ribbon-shaped vendotaenid algae. The sediments containing this depauperate flora also have curiously large amounts of organic matter (sapropel) derived from the burial of acritarchs and other organic-walled objects. The re-radiation of the plankton from this low-diversity interlude was slow. Acritarch diversity in most stratigraphic sections did not recover to Early Vendian levels until well into the Lower Cambrian, when very spiny forms such as *Skiagia* became abundant (but see Zang & Walter 1989).

Metazoans

The soft-bodied fossils of the Ediacaran fauna are generally thought to be metazoans (Section 1.3). Frondose or leaf-like Ediacaran forms such as *Charnia* and *Charniodiscus* are known throughout the world in sediments of Vendian age. Some of these organisms attained sizes of up to one metre in length. The second half of the Vendian (the Kotlin Horizon) is marked by local extinction on the Russian Platform of many of these large, distinctive soft-bodied creatures. Possibly coincident with the decline in phytoplankton diversity, Late Vendian metazoan faunas of the Russian Platform were reduced to rare problematic forms of medusoids and small trace fossils (Fedonkin 1987; Sections 1.3, 1.5).

The Ediacaran fauna seems to have died off by the end of the Vendian (the top of the Rovno Horizon of the Siberian Platform), although a few of these soft-bodied forms may have survived into the Early Cambrian. Seilacher (1984) argued that the end of the Vendian witnessed a mass extinction of the soft-bodied Ediacaran forms, and that these extinctions were real and were not an artifact of preservation. It must be noted, however, that the intensity of burrowing increased greatly in the terminal Vendian. The trace fossils at this time became more complicated, deeper and larger, indicating an increase in the dimensions of infaunal animals. This development may have reduced the potential for preservation of soft-bodied animals.

The Late Vendian increase in burrowing intensity was accompanied by an explosion in the diversity of trace fossils. Numerous new ichnotaxa appeared that have ranges continuing through most or all of Phanerozoic time. Of the dozens of new ichnogenera that first appeared in the Vendian, only six became extinct by the end. Of these, *Neonoxites*, and *Palaeopascichnus* were horizontal grazing or very shallow

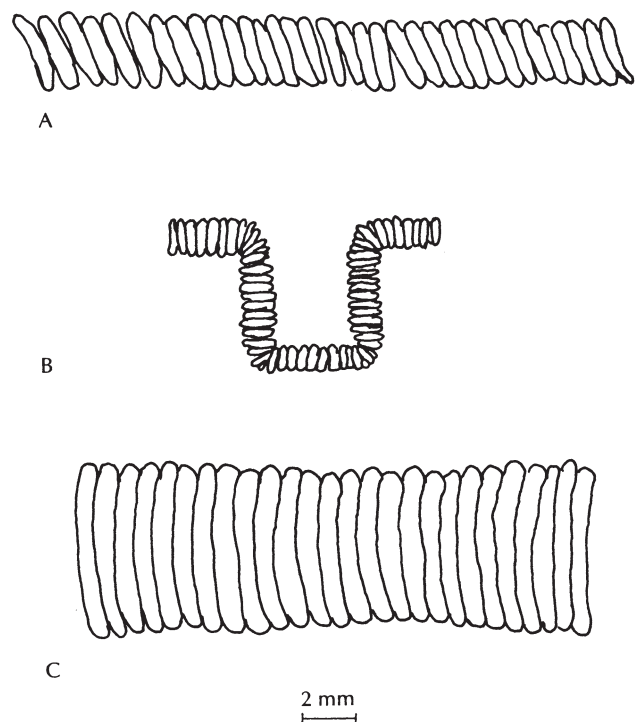


Fig. 1 Three ichnogenera known from the late Vendian. A, *Harlaniella*. B, *Neonoxites*. C, *Palaeopascichnus*. All three are shallow deposit feeding or grazing, horizontal bedding plane trace fossils.

deposit feeding traces (Fig. 1). If there was indeed a mass extinction at the end of the Vendian, it was overshadowed by the metazoan diversification occurring at this time.

The study of Vendian extinctions is hampered by a paucity of well preserved macrofossils. Nevertheless, the disappearance of acritarchs suggest that the Middle Vendian was marked by a mass extinction event that rivalled in magnitude the better known mass extinctions occurring later in the Phanerozoic. This Middle Vendian acritarch extinction event was linked to the Varangian glaciation by Vidal & Knoll (1982), who invoked climatic cooling as a causal mechanism. More evidence is needed to clarify the timing, severity, and possible climatic control of these extinction events. Of particular interest is the unresolved question of whether global metazoan mass extinctions occurred in the Vendian, and whether or not they were coincident with the phytoplankton extinctions.

References

- Fedonkin, M.A. 1987. *The non-skeletal fauna of the Vendian and its place in the evolution of metazoans*. Nauka, Moscow (in Russian).

- Seilacher, A. 1984. Late Precambrian and Early Cambrian Metazoa: preservational or real extinctions? In: H.D. Holland & A.F. Trendall (eds) *Patterns of change in Earth evolution*, pp. 159–168. Springer-Verlag, Berlin.
- Vidal, G. & Knoll, A.H. 1982. Radiations and extinctions of plankton in the late Proterozoic and early Cambrian. *Nature* **297**, 57–60.
- Zang, W.L. & Walter, M.R. 1989. Late Proterozoic plankton from the Amadeus Basin in central Australia. *Nature* **337**, 642–645.

2.13.2 End-Ordovician

P. J. BRENCHLEY

Introduction

About 22% of all families became extinct in the Late Ordovician, which makes this one of the largest episodes of mass extinction (Raup & Sepkoski 1982). Although there were some extinctions throughout the Ashgill, the main phase of extinction was in the Late Ashgill. The Late Ordovician extinctions cannot be related to a single stratigraphic level, but occurred in at least two steps. One phase coincided with the start of a major regression at the end of the Rawtheyan (the penultimate stage in the Ashgill) and a second phase coincided with a transgression at the end of the Hirnantian (the last Ashgill stage), about 1–2 million years later (Brenchley 1984). There may in addition have been some extinctions throughout Hirnantian times. The two major phases of extinction have been best documented from clastic sequences in Europe. Upper Ordovician extinctions of comparable magnitude are known from carbonate sequences in North America but have not been clearly differentiated into two phases.

Extinction patterns

The Late Ordovician extinctions involved significant changes to a number of groups:

Trilobites. Only about 14 of the 38 families of Middle Ashgillian trilobites survived into the Silurian. The main decline in diversity occurred at the end of the Rawtheyan when about 15 families disappeared. Generic diversity which had been fairly constant throughout the Ashgill declined by 45–75% at the top of the Rawtheyan (Fig. 1) and species diversity was possibly reduced even more.

Brachiopods. Thirteen families of brachiopods became extinct at or near the Ordovician–Silurian boundary. Of the 27 families which crossed the boundary, nine showed a marked decline in abundance (Sheehan 1982). Amongst the rich brachiopod faunas of the Ashgill of northwest Europe, 25% of genera disappeared at the top of the Rawtheyan and another 40% at the top of the Hirnantian (Fig. 1).

Graptolites. The diversity of graptolite species decreased from a high point in the Late Caradoc to a nadir in the *Climacograptus extraordinarius* and *Glyptograptus persculptus* zones, when the total world graptolite fauna consisted of only a few genera.

Primitive echinoderms. The diversity of cystoid, edriasteroid, and cyclocystoid families declined sharply in the Late Ashgill. The sharpest drop in numbers of cystoid genera in the families Diploporita and Dichoporita was at the Rawtheyan–Hirnantian boundary, when the rich and varied Rawtheyan fauna with 26 genera was reduced to a small but distinctive Hirnantian fauna with only eight. Most of the latter fauna apparently disappeared at the end of the Hirnantian.

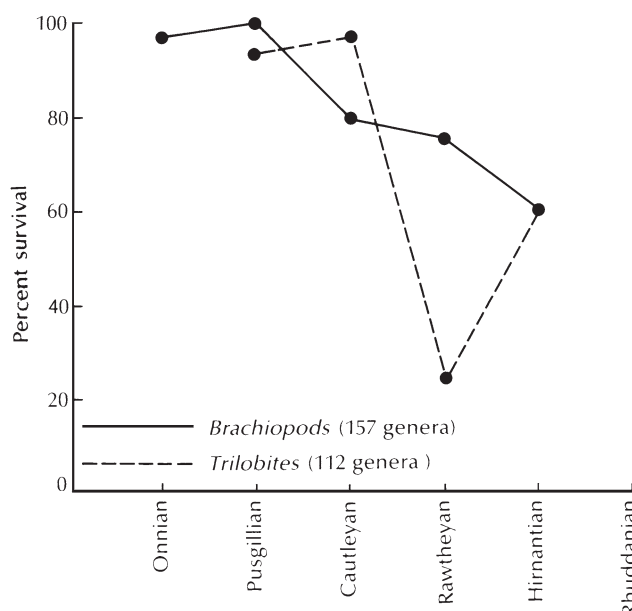


Fig. 1 Percentage survival of genera from stage to stage in the Upper Ordovician, e.g. 98% of Cautleyan trilobite genera survived into the Rawtheyan and 25% of Rawtheyan genera survived into the Hirnantian. The graphs are based on genera found in sequences with good stratigraphical control mainly in Europe. (After Brenchley 1984, by permission of John Wiley & Sons Ltd.)

Conodonts. The mainly clastic Hirnantian sequences of Europe have yielded very few species of conodonts, even where collecting has concentrated on the more promising limestone horizons. In the carbonate sequences of North America diverse conodont faunas declined a little in the Gamachian and disappeared almost completely at the Ordovician–Silurian boundary.

Chitinozoa, acritarchs, and ostracodes. All three groups show major decreases in diversity and changes in taxonomic composition at or near the Ordovician–Silurian boundary.

Corals. The best data for the Late Ordovician show that c. 50 of the 70 tabulate and heliolitoid genera became extinct in Late Ordovician times. It is not clear whether this was an end-Rawtheyan or Hirnantian extinction.

Following the first wave of extinction at the end of the Rawtheyan there was a residual fauna, dominated by brachiopods, which is usually referred to as the *Hirnantia* fauna. This fauna is unusually cosmopolitan and appears to have ranged from circumpolar to sub-tropical latitudes, though it was not well developed in the carbonate environments of tropical regions. The *Hirnantia* fauna is commonly considered to have been a relatively cool water fauna.

The second wave of extinction at the top of the Hirnantian (top Gamachian in Canada) was relatively modest in the clastic sequences of Europe. Several elements of the *Hirnantia* fauna disappeared at this level, and coral and ostracode faunas may have been heavily depleted. Coral-stromatoporoid reefs which occur at the top of the Hirnantian are rare or absent in the lower levels of the succeeding Silurian.

In North America the diversity of brachiopods, trilobites, conodonts, acritarchs, and ostracodes greatly diminished at the end of the Ordovician (Lespérance 1985), but because the detailed stratigraphy is uncertain the extinctions could be Early or Late Hirnantian.

Environmental changes

In most shelf sequences there is a change of facies at the Rawtheyan–Hirnantian boundary, reflecting the start of the regression which reached its maximum in the Middle or Upper Hirnantian. The regression partially drained many clastic shelves leaving a variety of shallow-marine sandy deposits.

A major part of the world's carbonate platforms became exposed with widespread development of karst surfaces and disconformities.

At the top of the Hirnantian there is generally a sharp change in facies indicating a rapid transgression. In many clastic sequences the shallow-marine rocks of the Upper Hirnantian are overlain by black graptolitic shales. In carbonate regions there is a progressive return to more offshore carbonate facies.

It has been estimated that the regression involved a fall in sea-level of 50–100 m (Fig. 2). In several Hirnantian sequences there is some evidence of fluctuations of sea-level (two to four regressions) but the pattern is not clear on a global scale.

Causes

The cause or causes of the extinctions are debatable. The stepped nature of the extinctions makes an extra-terrestrial cause, such as meteorite impact, unlikely. Furthermore no iridium anomaly was discovered in detailed investigations of the Ordovician–Silurian stratotype at Dob's Linn or in the carbonate sequence of Anticosti Island. The very precise correlation between the disappearance of faunas in many sections and the first evidence of regression makes it likely that the extinctions were related to contemporaneous environmental changes such as the following:

1 Sea-level changes. The fall in sea-level during the Hirnantian would have drastically reduced the size of continental shelves and platforms and hence the habitable area for shelf benthos. Many very extensive platforms (N. America, Baltica and the Russian Platform) were covered by shallow seas during most of the Ordovician so a sea-level fall of tens of metres would have had a profound effect.

The main argument against a major role for sea-level change in causing extinctions is that the faunal changes were concentrated at the Rawtheyan–Hirnantian boundary while the regression appears to have continued throughout the early part of the Hirnantian. The second phase of extinction at the top of the Hirnantian coincides with a rise in sea-level, and consequently a potential increase in habitable area. However, following the transgression, black shales were deposited on many clastic shelves, indicating widespread anoxic or dysaerobic conditions hostile to benthic faunas (see also Section 2.12.1).

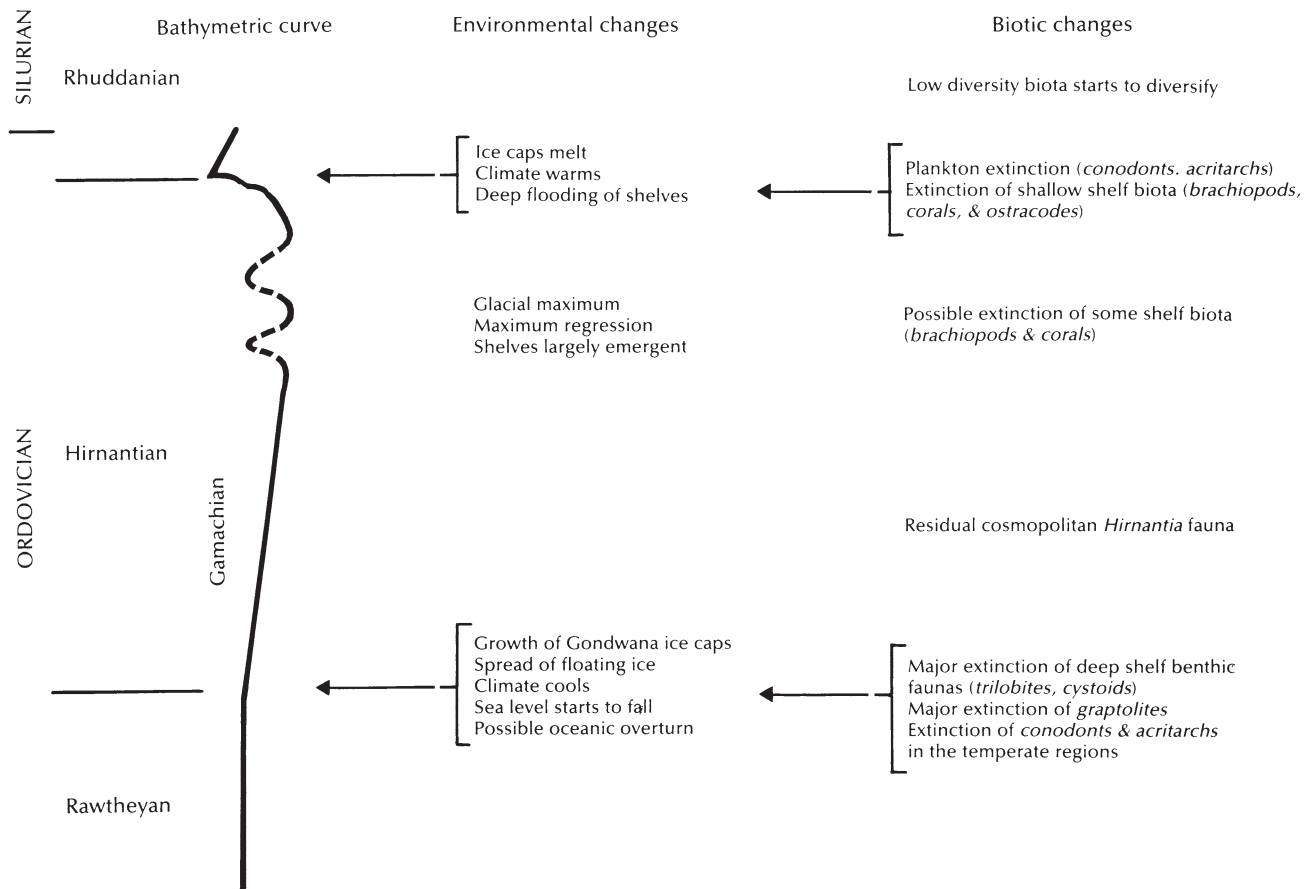


Fig. 2 Summary diagram showing the Late Ordovician–Early Silurian bathymetric curve and associated environmental and biotic changes.

Although changes in sea-level may have had some effect on the shelf benthos they do not satisfactorily account for the major extinctions amongst the plankton.

2 Temperature. There is evidence from the wide extent of upper Ordovician continental glaciation on Gondwanaland, and the occurrence of marine tilloids peripherally, for a very substantial extension of arctic and subarctic conditions in the Hirnantian. On the other hand, tropical carbonate environments survived at the same time in equatorial latitudes. There might well have been a substantial reduction in the area of temperate seas and some restriction of tropical regions which could account for the reduction in diversity of microplankton and graptolites. The widespread distribution of the *Hirnantia* fauna might be related to the spread of cool water, which could even have impinged on formerly tropical areas during major phases of glaciation.

The second phase of extinction at the top of the

Hirnantian should have been at a time of rising temperature so it is unlikely that this phase of extinction was the result of changing climate.

3 Oceanic overturn. It has been suggested that during periods of climatic change the stability of weakly stratified ocean waters might be disturbed, and that they could overturn bringing 'unconditioned' biologically-toxic bottom water to the surface (Wilde & Berry 1984). Such overturns might occur either during a period of climatic deterioration, when cold water from high latitudes intruded below weakly stratified ocean waters, or during times of climatic amelioration.

The model of oceanic overturn has several attractions, as it can account for the coincidence of the two phases of extinction with times of maximum climatic change, and for extinction of both plankton and shelf benthos. Furthermore, the effects of overturn would have been rapid, which accords with the disappearance of faunas at precise levels in

many sections. Unfortunately no evidence of oceanic overturn has yet been detected in the sedimentary record.

In summary, changes in both temperature and sea-level, for which there is good evidence, might have played a significant role in the faunal extinctions. A related oceanic overturn could have been important, but remains hypothetical.

References

- Brenchley, P.J. 1984. Late Ordovician extinctions and their relationship to the Gondwana glaciation. In : P.J. Brenchley (ed.) *Fossils and climate*, pp. 291–315. John Wiley & Sons, Chichester.
- Lespérance, P.J. 1985. Faunal distributions across the Ordovician–Silurian boundary, Anticosti Island and Percé, Quebec, Canada. *Canadian Journal of Earth Sciences* **22**, 838–849.
- Raup, D.M. & Sepkoski, J.J., Jr 1982. Mass extinctions in the marine fossil record. *Science* **215**, 1501–1503.
- Sheehan, P.M. 1982. Brachiopod macroevolution at the Ordovician–Silurian boundary. *Proceedings of the Third North American Paleontological Convention* **2**, 477–481.
- Wilde, P. & Berry, W.B.N. 1984. Destabilisation of the oceanic density structure and its significance to marine 'extinction' events. *Palaeogeography, Palaeoclimatology, Palaeoecology* **48**, 143–162.

2.13.3 Frasnian–Famennian

G. R. McGHEE, Jr

Extinction patterns

The massive deterioration in ecosystems which occurred throughout the world during the Frasnian–Famennian event can be described correctly as catastrophic in effect (McLaren 1982). Frasnian ecosystems were ecologically very diverse and equitable in structure. Early Famennian ecosystems, in contrast, were impoverished in ecological diversity and in overall species richness. The effect of the biotic crisis can easily be seen in the 'bottleneck' constriction of ecological complexity which occurred in the Appalachian region of eastern North America. There the diverse Frasnian ecosystem is replaced by an ecologically depauperate Famennian ecosystem proportionately overdominated by reduced species numbers of brachiopods, bivalves, and glass sponges (McGhee 1982).

The analysis of the fossil remains of organisms around the globe which perished during the extinction event, as well as those which survived, reveals the following ecological patterns:

Latitudinal effect. Tropical reefal and perireefal marine ecosystems were particularly hard hit. The low-latitude, geographically widespread and massive stromatoporoid-tabulate reefal ecosystems vanished, and perireefal rugose coral-tabular stromatoporoid bioherms were decimated. The stromatoporoids suffered a severe reduction in biomass, but they did not become extinct nor did they totally lose their reef-building potential (Stearn 1987). Post-Frasnian stromatoporoid structures are of small dimensions, and are generally found in the warm water equatorial region of the Palaeotethys. Famennian stromatoporoids found outside this area are generally labechiids, which are believed to have been better adapted to cool water than the majority of Frasnian species, which were tropical and low-latitude in distribution.

Differential survival of high-latitude, cool-water adapted species is also exhibited by the brachiopods, which were the dominant form of shelly animal in Frasnian benthic ecosystems (Copper 1986). Of the total brachiopod fauna, approximately 86% of Frasnian genera did not survive into the Famennian. However, 91% of brachiopod families whose species were generally confined to low-latitude, tropical regions perished in the extinction event, in contrast to a loss of 27% of those families with species which ranged into high-latitude, cool-water regions.

Other elements of the marine benthos which exhibit latitudinal patterns of survival include the foraminifera. They suffered major losses in species diversity with the substantial reduction which occurred in the area of the global belt of carbonate sedimentation. Species of the high-latitude regions differentially survived the event; species of the cool-water Siberian realm expanded their geographical ranges into low-latitude regions with the latitudinal contraction in the range of Palaeotethys species.

Bathymetric effect. In general, shallow-water marine ecosystems were much more severely affected during the Frasnian–Famennian interval than deeper-water systems. The bathymetric selectivity in extinction is seen most dramatically within the rugose corals, a group which suffered a massive loss in biomass. Only 4% of the shallow-water species

survived. Deeper-water species suffered a 60% extinction in their numbers, and while this reduction was severe it pales in comparison with the 96% loss of species in the shallow waters. The decimation of the shallow-water corals was actually more severe than that of the stromatoporoids (Stearn 1987).

A particularly intriguing bathymetric pattern of selective extinction and diversification occurs across the Frasnian–Famennian boundary in the Appalachian marine ecosystems of eastern North America. Simultaneously with the extinction of many shallow-water benthic species, the hyalosponges (glass sponges) migrated from deeper water into the shallows and underwent a burst of diversification in species numbers. Modern glass sponges are generally found in water depths in excess of 200 m, and are considered to be better adapted to colder waters than most other invertebrate species. Blooms in other siliceous organisms, most notably the radiolarians, are also reported during the Frasnian–Famennian interval.

Habitat effect. A marked habitat effect in selective survival can be observed in the Devonian fish groups which included both marine and freshwater species. Only 35% of marine placoderm species survived, in contrast to 77% of those which lived in freshwater. A similar pattern occurs in the acanthodian fishes: only 12% of marine species survived, in contrast to 70% of freshwater species.

A key environmental parameter which differentiates the two habitat regions (other than salinity) is temperature. In general, freshwater species are adapted to seasonal and diurnal fluctuations in temperature, in contrast to those species in temperature-buffered shallow-water marine regions. The differential survival of freshwater fish may reflect their greater tolerance to temperature changes. Other elements of the terrestrial ecosystem appear to have been unaffected by the event. Floras exhibit no major disruptions, and plant biomass productivity appears to have been unchanged, or even perhaps enhanced, during the Frasnian–Famennian interval.

Within the shallow-water marine benthos, epifaunal filter-feeding organisms appear to have been most affected by the extinction event; infauna and detritus feeders were relatively unaffected.

In common with other extinction events, the upper oceanic water habitat of the marine plankton was massively disrupted. Approximately 90% of the preservable phytoplankton was affected, and

massive biomass reductions also occurred among the zooplankton.

Summary. The ecological signature preserved in the fossil record of the Frasnian–Famennian extinction event appears to indicate a significant drop in global temperatures during the crisis interval. The decimation of low-latitude tropical reef ecosystems and of warm-water shallow marine faunas, combined with the relatively higher survival of high-latitude faunas, deep-water faunas, and terrestrial fauna and flora, seems most compatible with lethal temperature decline at a global level. At the local and regional level the extinction event doubtless records additional local environmental factors.

Evolutionary dynamics

The precise timing of the Frasnian–Famennian extinction event is still uncertain. Present evidence, however, indicates that extinction rates were elevated above average during a geologically significant span of time during the latter half of the Frasnian, for a period of perhaps 3–4 million years (Fig. 1). There appears to have been no single synchronous extinction peak shared by all species, but a series of stepwise extinctions of different species groups. It has been consistently observed, however, that a marked drop in standing species diversity occurred at the very end of the Frasnian. The fact that extinction rates were elevated above average for a significant period of time before the terminal Frasnian suggests that the drop in species diversity at the Frasnian–Famennian boundary was not a simple function of extinction rate magnitudes (McGhee 1982).

In the analysis of ecosystem evolution it is often misleading to consider the pattern and timing of extinction rates alone. Species diversity is a function of the relationship of two evolutionary variables: the rate at which species were lost from the system (extinction rate), and the rate at which new species were added (origination rate). While either rate alone is of considerable interest, the evolutionary behaviour of the total ecosystem can best be characterized by the sign and magnitude of species diversity changes (the turnover rate, i.e. origination rate minus the extinction rate). If origination and extinction rates were of equal magnitude, the ecosystem was in a state of dynamic equilibrium with no diversity change. Where origination exceeded extinction (positive turnover rates), the system was

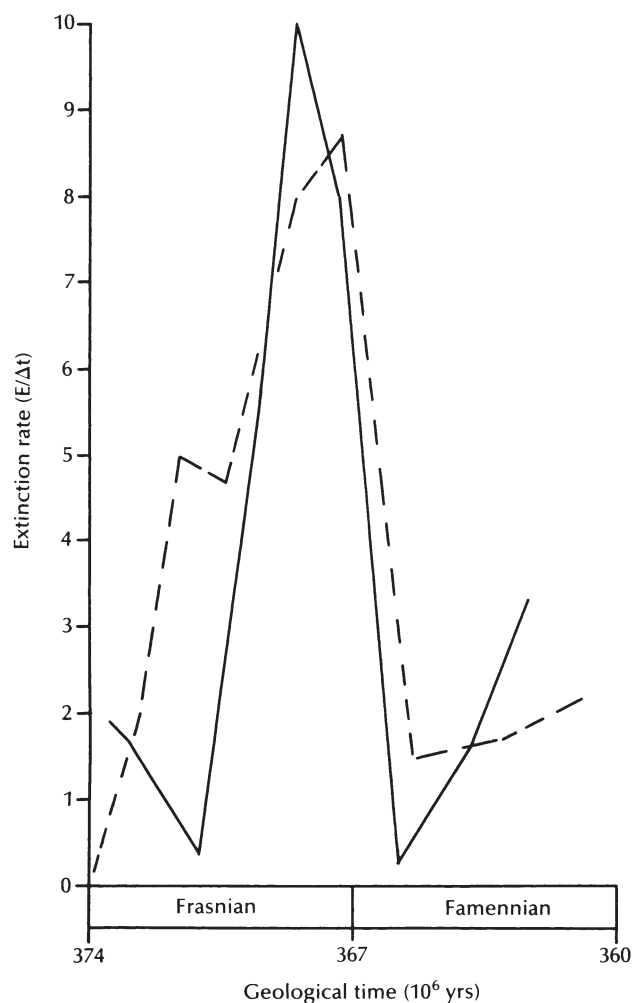


Fig. 1 Elevated extinction rates during the Middle and Late Frasnian exhibited by brachiopods from the Appalachians, U.S.A. (solid line) contrasted with the Urals, U.S.S.R. (dashed line). Extinction rate metric is the number of species extinctions (E) per million years (Δt).

diversifying, whereas if extinction exceeded origination (negative turnover rates) the system was losing species diversity.

Marine ecosystems appear to have been flourishing (in terms of standing species diversity) during the interval of time characterized by some of the highest extinction rates which occurred in the latter half of the Frasnian. This phenomenon is due to the fact that species origination rates were even higher per time interval than the corresponding extinction rates.

This pattern of relative origination–extinction rate magnitudes reversed abruptly during the latest Frasnian, precipitating a rapid loss of species diversity (Fig. 2). Extinction rates in many cases remained the same, or actually declined in some, but species

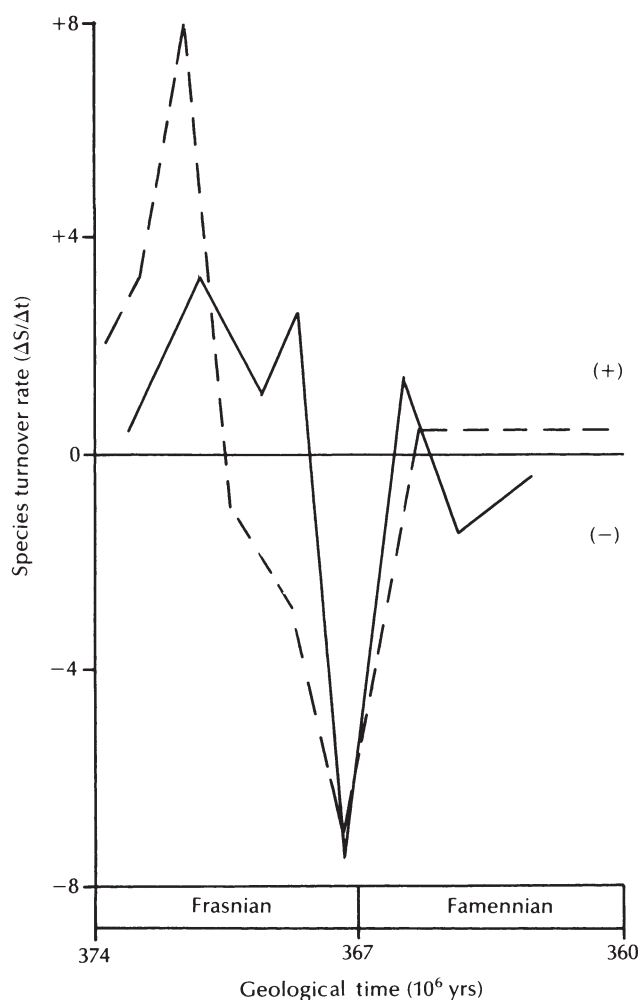


Fig. 2 Species turnover rates for brachiopods from the Appalachians, U.S.A. (solid line) and the Urals, U.S.S.R. (dashed line) during the Frasnian–Famennian interval. A sharp negative pulse in turnover rates occurs in both regions at the very end of the Frasnian, signalling a severe and rapid loss of species diversity in this interval of time. Turnover rate metric is the change in the number of species (ΔS) per million years (Δt).

turnover rates became sharply negative (Fig. 2). Whether extinction rates were rising or falling, it was the decline in species originations which drove species turnover rates sharply negative at the very end of the Frasnian.

Thus, while there was no single synchronous extinction rate pulse during the Frasnian, the ecosystem did exhibit a rather abrupt and massive drop in species diversity in the terminal Frasnian. In understanding the ultimate cause of the extinction event the most important question may not be what triggered the elevated extinction rates, but what was the inhibiting factor that caused the cessation of new species originations.

References

- Copper, P. 1986. Frasnian/Famennian mass extinction and cold-water oceans. *Geology* **14**, 835–839.
- McGhee, G.R., Jr. 1982. The Frasnian–Famennian extinction event: a preliminary analysis of Appalachian marine ecosystems. *Special Paper of the Geological Society of America* **190**, 491–500.
- McLaren, D.J. 1982. Frasnian-Famennian extinctions. *Special Paper of the Geological Society of America* **190**, 447–484.
- Stearn, C.W. 1987. Effect of the Frasnian–Famennian extinction event on the stromatoporoids. *Geology* **15**, 677–679.

2.13.4 End-Permian

D. H. ERWIN

Introduction

During the latest Permian 54% of all marine families became extinct (Table 1), as did 83% of all marine genera (Sepkoski 1986). Several authors have estimated that as many as 90–96% of all durably-skeletonized marine invertebrate species became extinct (e.g. Sepkoski 1986). This extinction was the most severe of the Phanerozoic and eliminated twice as many families as the second largest, the end-Ordovician mass extinction (Section 2.13.2). Many major taxa were eradicated or declined drastically in diversity, eliminating the shallow-water, sessile, epifaunal, brachiopod–bryozoan–pelmatazoan echinoderm communities which dominated the Palaeozoic. This permitted the expansion of the mobile, infaunal, molluscan-dominated communities which dominated the post-Palaeozoic. The effects of the extinction on land are less clear, but extinctions and changes in faunal dominance occurred in both terrestrial vertebrates and plants throughout the Permian.

Despite its magnitude and significance, analysis of the patterns and causes of the extinction has been hampered by the restricted number of marine sections of latest Permian age. The number of well-studied sections has increased recently, particularly in South China and elsewhere in the Tethyan region. However, facies changes at the boundary indicate that no continuous Late Permian–Early Triassic sections have been discovered.

The assembly of the supercontinent Pangaea was largely completed with the collision of the

Table 1 Extinction percentages for 17 major groups of marine families during each series of the Permian. Families not resolved to series were not used in the analysis. A = Asselian, S = Sakmarian, L = Leonardian, G = Guadalupian, D = Dzulian. (Data from Sepkoski 1982.)

Marine family	Percentage extinction				
	A	S	L	G	D
Foraminifera	0	0	3	6	38
Porifera	0	0	18	24	10
Tabulata	0	14	15	42	100
Rugosa	0	6	38	62	100
Gastropoda	0	0	15	25	11
Bivalvia	0	3	2	12	11
Cephalopoda	17	0	20	43	47
Other Mollusca	0	17	40	33	0
Other Arthropoda	0	0	21	33	25
Ostracodes	4	8	8	35	29
Bryozoa	10	4	4	23	65
Brachiopoda	0	3	12	34	71
Crinoidea	0	16	5	93	0
Other Echinodermata	5	5	5	37	8
Conodonta	0	20	0	20	25
Other taxa	0	0	0	9	3
Marine vertebrates	0	0	0	39	0

Kazakhstan, Tarim, and Siberian blocks in the Late Carboniferous and the accretion of this unit to the Russian platform by the end of the Artinskian Stage (Fig. 1). The North China block collided with Kazakhstan in the latest Permian. The South China block closely approached the North China block in the latest Permian but rotation and accretion of the two blocks was not completed until the Late Triassic or Early Jurassic. (Considerable movement and rotation occurred between tectonic blocks during the Permian. Consequently, palaeocontinental reconstructions are poorly constrained until the Late Triassic.)

The Late Carboniferous–Early Permian glaciation in Gondwanaland ended during the Asselian–Sakmarian as the South Pole moved off the continent and the formation of Pangaea led to increased temperatures and seasonality (see Fig. 2 for time-scale). Continuing climatic oscillations into the Late Permian are suggested by sea-level fluctuations on a 2–2.5 million year cycle. Scattered reports of Late Middle to Late Permian glaciation, however, involve only restricted mountain glaciations. Global warming continued into the Triassic and there is no evidence for widespread cooling or glaciation during the Late Permian.

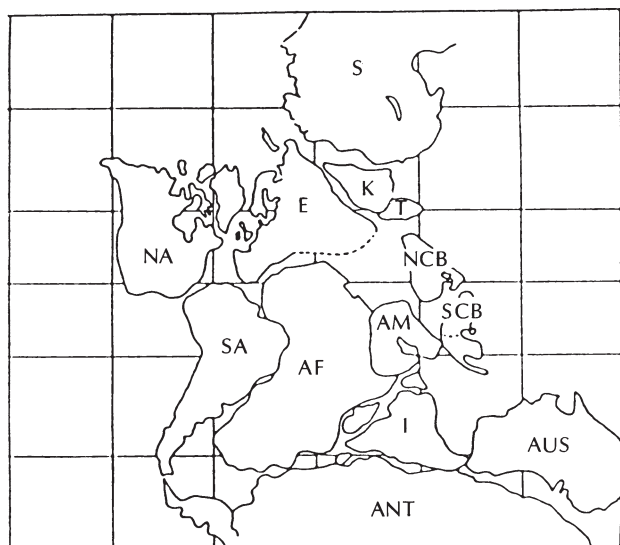


Fig. 1 Palaeocontinental reconstruction for the Late Permian. AF = Africa, AM = Asia Minor, ANT = Antarctica, AUS = Australia, E = Europe, I = India, K = Kazakhstan, NA = North America, NCB = North China block, S = Siberia, SCB = South China block, T = Tarim. (From Lin *et al.* 1985. Reprinted by permission from *Nature* vol. 313 pp. 444–449. Copyright © 1985 Macmillan Magazines Ltd.

The increased temperatures and seasonality associated with the formation of Pangaea are indicated by evaporites and red beds. As continents become more exposed during a regression, the ameliorating effects of the ocean (due to the high heat capacity of water) decline, climates become more severe, and seasonality increases, a condition described as increased continentality or inequability (Valentine & Moores in Logan & Hills 1973; Jablonski 1986). The concentration of land area in one unit exacerbated the trend, leading to high seasonality in continental interiors. The effects were not limited to continental interiors. Storm activity, particularly monsoons in the Tethyan realm, and consequent disturbances in shallow-marine ecosystems, increased. The largest Permian evaporite deposits are of Kungurian age, coinciding with the initial formation of Pangaea, although these are dwarfed by later Triassic deposits. Finally, a sharp marine regression occurred at the end of the Permian.

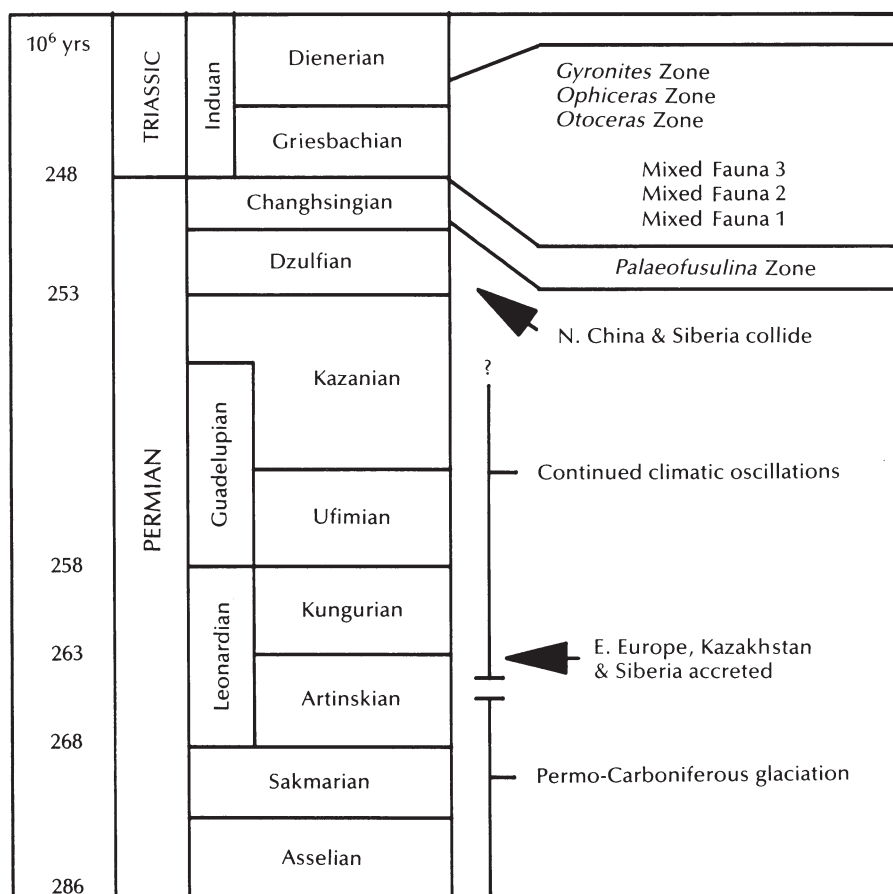


Fig. 2 Permian and lowest Triassic geological stages, Permo-Triassic boundary biostratigraphic zones, and major tectonic and climatic events. Most workers consider the *Palaeofusulina* Zone the final zone of the Changhsingian Stage, and the mixed faunas the lower-most units of the Triassic. The *Gyronites* Zone marks the Griesbachian–Dienerian boundary. (Geological dates from Harland *et al.* 1982.)

Extinction patterns

Taxa which became extinct include tabulate and rugose corals, conularids, eurypterids, leperditiid ostracodes, several gastropod groups, goniatitic ammonites, orthid and productid brachiopods, blastoids, inadunate, flexible, and camerate crinoids, and the few remaining trilobites (Table 1; Fig. 3). A number of other groups suffered sharp drops in diversity, including the cryptostomate and trepostomate bryozoans, foraminifera, ammonoids and fish. Reefs were eliminated and tropical ecosystems in general were severely affected. Jablonski (1986) analysed survival patterns of articulate brachiopods and noted that 75% of the families confined to the tropics became extinct, while only 56% of extra-tropical families died out. All fusulinid foraminifera and 54% of all foraminiferan families became extinct, including both planktic and benthic taxa. In general the zooplankton, sessile filter feeders, and the high-level carnivores (ammonoids and fish) were the most strongly influenced trophic groups.

The diversity history of marine vertebrates parallels that of invertebrates, with the decline beginning in the Guadalupian and accelerating in the Dzulian. Elasmobranchs, Holocephali and marine Chondrostei and Holostei follow this pattern. Freshwater and euryhaline fish and amphibians, however, reach a diversity low in the Leonardian and appear to be diversifying across the boundary.

Sepkoski identified three distinct assemblages of taxa during the Phanerozoic, each with characteristic diversity maxima (Section 1.6). His Palaeozoic evolutionary fauna includes the groups which dominated the Palaeozoic: articulate brachiopods, crinoids and other pelmatozoan echinoderms, and bryozoans. These taxa suffered disproportionate extinction during the end-Permian, with a 79% familial extinction, while bivalves, gastropods, some arthropod taxa, and others which constitute the Mesozoic–Cenozoic evolutionary fauna declined 27%. This differential extinction pattern contributed to the development of burrowing, infaunal, molluscan-dominated communities in the post-Palaeozoic.

The Permian extinction produced a large number of 'Lazarus taxa': taxa which disappear from the record during the Late Permian, only to reappear in the Triassic (Jablonski 1986). As Batten (*in* Logan & Hills 1973) noted, 'Palaeozoic'-aspect gastropods are better represented in the Triassic than in the latest Permian. The number of Lazarus taxa indicates that the record across the boundary is too fragmentary

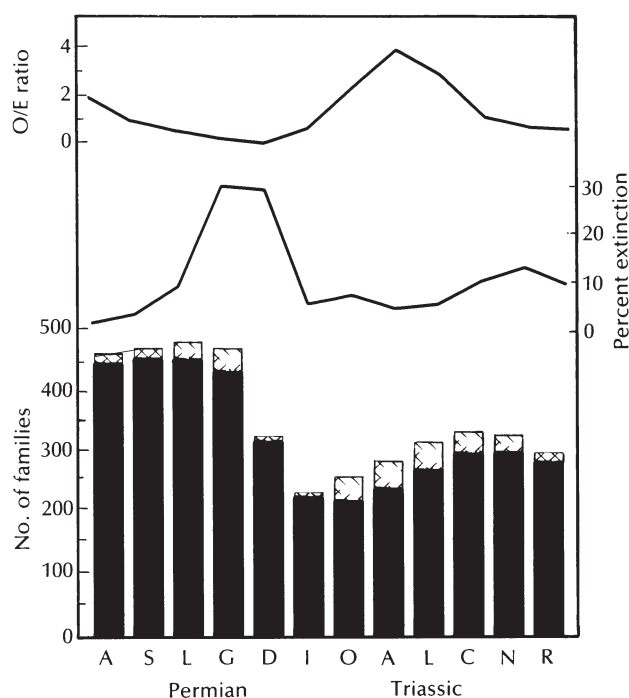


Fig. 3 Permian and Triassic data for marine families. The histogram shows the total number of families in each series, with the new families indicated by the cross-hatched symbol. The percentage extinction [calculated as percentage (total extinctions)/(initial diversity + total originations)] is shown above. Families whose originations or extinctions could not be resolved to series are not included. The graph at the top shows the ratio of originations to extinctions (O/E). The ratio falls below 1.0 during the Leonardian, begins to rise in the Induan, but does not climb above 1.0 until the Olenekian. A = Asselian, S = Sakmarian, L = Leonardian, G = Guadalupian, D = Dzulian, I = Induan, O = Olenekian, A = Anisian, L = Ladinian, C = Carnian, N = Norian, R = Rhaetian. (Data from Sepkoski 1982.)

and sparse to accurately reflect the rate and duration of the extinction. Furthermore, many groups which first appear in the Triassic, including zygopleurid gastropods and scleractinian corals, are clearly descended from Palaeozoic ancestors but with a significant gap between ancestor and descendant. The restricted number of sections, the paucity of the fossil record, and the obvious sampling problems make it difficult to determine whether the extinction began in the Guadalupian or was restricted to the Dzulian–Changhsingian (see also Jablonski 1986). Preliminary generic diversity data (Fig. 4) show the extinction peak in the Guadalupian, although this may be a preservational artifact (Sepkoski 1986).

An indirect method of determining the duration of an extinction event is to analyse the replacement ratio, or the ratio of originations to extinctions (O/E

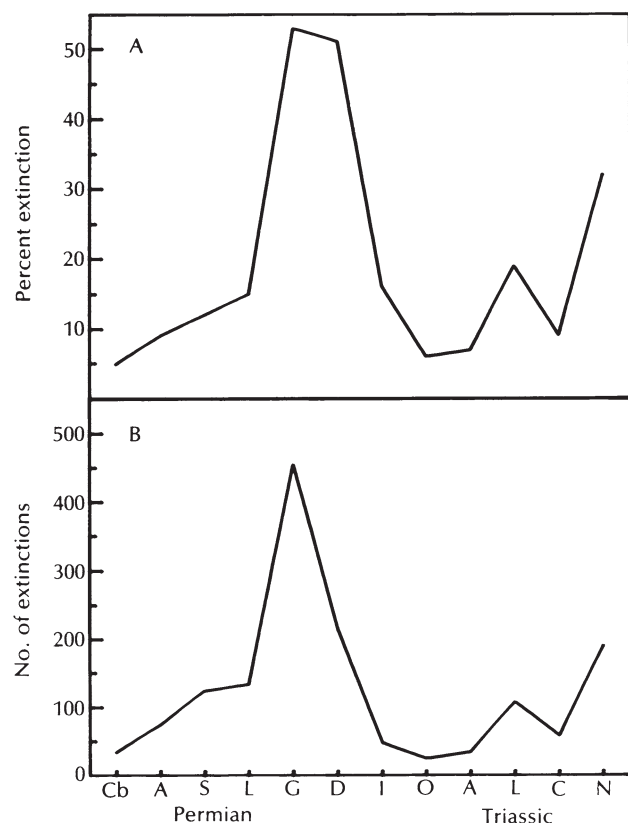


Fig. 4 Permian and Triassic extinctions in marine genera. A, Percentage extinction (calculated as in Fig. 3). B, Total number of generic extinctions. Cb = Carboniferous. Other abbreviations as in Fig. 3. (Data from Sepkoski 1986.)

Fig. 3). This ratio is 2.0 in the Asselian, but falls below 1.0 in the Leonardian and reaches a minimum of 0.1 in the Dzulfian. Extinctions exceed originations from the Leonardian onward. Note that originations continue, with 38 new families in the Guadalupian or a 9% origination rate for the stage; however, extinctions total 152 (32%), and 100 (31%) in the Dzulfian. Several authors have even argued that an 'extinction' never occurred but origination rates were severely depressed. While originations did decline, this cannot explain a drop in familial diversity of 54%.

Terrestrial vertebrates and plants experienced a series of extinctions and ecological replacements from the Early Permian through the Triassic (Fig. 5). Eighty-one per cent of amphibian families became extinct near the end of the Leonardian; interestingly, most of these families first appeared during the preceding glaciation. A second extinction pulse at the close of the Permian primarily affected therapsids, and was followed closely by a third extinction episode at the close of the Lower Triassic.

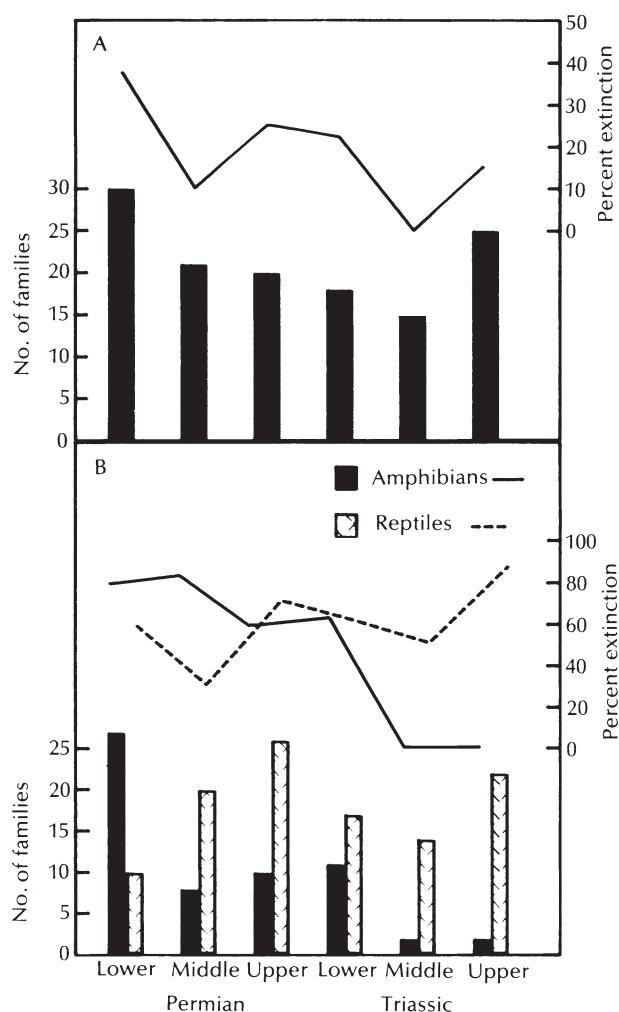


Fig. 5 Number of families and percentage extinction (calculated as in Fig. 3) per stage of the Permian and Triassic. A, Terrestrial plants. (Data from Knoll 1984.) B, Terrestrial amphibians (lepospondyli and labyrinthodonts) and reptiles (captorhinomorphs, pelycosaurs, therapsids, and archosaurs). (Data from Olson *in* Silver & Schultz 1982.)

The fourth occurred at the end of the Triassic (Section 2.13.5) and removed archosaurian reptiles (Olson *in* Silver & Schultz 1982). R. Sloan distinguished a more complex pattern of vertebrate extinction episodes: two in the Sakmarian–Leonardian, two in the Guadalupian, and six in the Dzulfian. His analysis showed that the survivors of each extinction had a higher percentage of mammalian characters and smaller body size than those which became extinct. Sloan suggested that the regular cycle of extinction and replacement was caused by post-glacial climatic oscillations. Preservation quirks are apparent as well, for the magnitude of the Late Permian–Early Triassic extinction is accentuated by an unusual upland fauna from the

Beaufort Series (Dzulfian) of South Africa with a large number of endemic families. If mass extinctions are characterized as short-term extinction episodes significantly above background rates (Jablonski 1986), the vertebrate extinctions at the close of the Permian are merely one of a series of extinctions and replacements. There is no compelling evidence of a mass extinction.

Permian floras included a mesic, equatorial assemblage of broad-leaved pteridosperms, cordaites, and pectopterid ferns, the Glossopterid flora dominated by pteridosperms, and an endemic Asiatic (Angaran) flora of cordaites. Vascular plant diversity dropped by 50% from the Early Permian to the Middle Triassic (Fig. 5A; Knoll *in* Nitecki 1984). The Palaeophytic floras were replaced briefly by a mixed flora of Palaeophytic and Mesophytic affinities before development of a truly Mesophytic flora of conifers, cycads, ginkgoes, and new groups of pteridophytes and pteridosperms. Local transitions took 5 million years or less but the global change occurred diachronously over 25–30 million years. Knoll (*in* Nitecki 1984) attributed these changes to the disappearance of warm, equable climates in low latitudes as the glaciation ended, and to the spread of drier climates during the coalescence of Pangaea. The drier climates allowed conifers to invade the lowlands and replace archaic pteridosperms and pteridophytes. Despite the transition between floral types, there is no evidence of a major extinction event at the close of the Permian. Olson (*in* Silver & Schultz 1982) discounted a link between the end of the Permo-Carboniferous glaciation and vertebrate diversity patterns, but such a link now seems likely. Vertebrate diversity may have responded directly to climatic change or have been indirectly affected through restructuring of plant communities.

The Permo-Triassic boundary

The total number of marine Permo-Triassic boundary sections remains few, but good exposures are present in south China, the Salt and Surghar Ranges of Pakistan, Malaysia, Kashmir, northern Iran and southern Soviet Armenia, and east Greenland. Unfortunately the completeness of several of these sections is questionable and a widespread disconformity may be present. A distinctive fauna combining Permian-type brachiopods and Triassic-type bivalves and ammonoids is widespread in South China and occurs in Kashmir, east Greenland, Dzulfia, and Pakistan (Sheng *et al.* 1984). The Permian elements are similar to a siliceous facies

assemblage lower in the section, but individual species are frequently dwarfed. This fauna is deposited in units 0.2–2.0 m thick with three distinct assemblages (Fig. 2). In south China 'mixed fauna bed 1' is a 10 cm thick yellowish or greenish-brown shale containing numerous Permian-type brachiopods and Triassic-type cephalopods. Bed 2 is a 20–30 cm marly dolomite with Permian brachiopods and a few foraminifera. Mixed-fauna beds 1 and 2 appear to be equivalent to the *Otoceras* Zone, the lowest zone of the Triassic. Bed 3 is a greenish shale of variable thickness containing the Triassic bivalve *Claria* and cephalopod *Ophiceras*. Few Permian taxa are present (Sheng *et al.* 1984).

Earlier work suggested that the beds were formed by post-depositional mixing, but recent analysis of their sedimentology, preservation, and areal extent suggests that the mixed faunas are real. However, the abrupt facies shifts present in the three beds signifies a disconformable rather than strictly conformable boundary. The best evidence for a rapid mass extinction at the close of the Permian is the disappearance of seven genera and 22 species of Permian brachiopods in beds 1 and 2, which actually occurs in the lowest Triassic!

Causes

Suggested causes for the end-Permian extinction have included trace element poisoning, regression-induced species–area effects, cosmic radiation, high temperatures, global cooling, salinity changes, trophic-resource fluctuations, tectonically-induced changes in marine faunal provinces, and extra-terrestrial impact. The range of explanations reflects the scarcity of information on latest Permian–earliest Triassic faunas, and the variety of geological events during the Late Permian.

Several suggestions can be disposed of quickly. Cosmic radiation, long a favourite of O. Schindewolf, does not penetrate the surficial ocean layer and cannot affect benthic organisms. Salinity changes would have been greater in the Middle Permian and Middle Triassic, when the major evaporite deposits were formed, than in the Late Permian. The possibility of an extra-terrestrial impact, and thus a connection with the Cretaceous–Tertiary extinction, was fuelled by reports of an iridium anomaly of 2.0 ppb in Chinese boundary sections (Xu *et al.* 1985) but attempts to reproduce the results have failed (Clark *et al.* 1986). At present there is no evidence for an impact event at this boundary.

Thermal overheating of marine faunas is negated as an explanation by the lack of terrestrial extinction (since vertebrates would be influenced before marine taxa), the prevalence of evaporites during the Kungurian and Middle Triassic (when no mass extinctions occurred), and the ability of most taxa to migrate to cooler (more polar) regions (Stanley *in* Nitecki 1984). Stanley (*in* Nitecki 1984) suggested a global cooling, rather than glaciation-associated polar cooling at the close of the Permian, as a proximal cause of the extinction. He also argued that the effects of global cooling would be especially severe when preceded by a period of mild climate and equability — yet the Permian was a time of climatic instability. Despite the climatic oscillations there is no evidence for widespread Late Permian glaciation or a global cooling event, and maximum glaciation occurred in the Asselian–Sakmarian without a mass extinction.

Marine productivity is highly dependent on nutrients derived from land, particularly those delivered through estuarine ecosystems. Tappan (*in* Silver & Schultz 1982) suggested that the gradual decline in marine phytoplankton, zooplankton, and many filter feeders (corals, articulate brachiopods, stalked echinoderms, bryozoans) resulted from sequestering of nutrients on land (as new plant communities developed) and in coal beds. Yet the groups most at risk diversified in the Lower Permian when the floral transition began, and coal beds (and maximum nutrient sequestering) predominated during the Carboniferous and Lower Permian. Declining marine productivity may well have played a role in the extinction, but if so the cause was unrelated to nutrient sequestering by land plants.

The marine regression at the close of the Permian has been the most frequently cited ultimate cause of the extinction, but a causal connection to the extinctions is unclear. Sea-level declined steadily throughout the Permian, with a drop of about 210 m in the Dzulfian–Changhsingian to a point only 40 m below present day sea-level (Hallam 1984). This left about 13% of the continental shelves covered. Short-term regressions of equal or greater magnitude occurred in the Upper Miocene, the Middle Oligocene, Lower Cretaceous, and Middle Triassic without associated mass extinctions, but began from higher sea stands, leaving more continental shelf area exposed than in the Permian. The cause of the Permian regression is unclear, but the most probable explanation is a cessation in mid-ocean ridge spreading as Pangaea formed, subsequent sinking of the ridges, and expansion of the volume of the ocean basins.

Logarithmic plots of species number versus area are frequently linear, giving the relationship: $S = kA^z$, where S is the number of species, A is total area, z is the slope of the regression of the log–log plot, and k is a constant. This empirically derived relationship, or species–area effect, implies that a regression will increase competition and thus extinctions. The species–area effect has been criticized for a variety of ecological and palaeontological reasons but only a few are important here. Stanley (*in* Nitecki 1984) noted that since an island approximates to a cone in shape, the amount of habitable area increases during a regression. Jablonski (1986) calculated that 87% of the 276 families of molluscs, echinoderms and coelenterates that he analysed are present on one or more of the 22 oceanic islands studied. Consequently most marine families will be protected from the effects of a regression by occurrence on islands which expand in size. Perhaps of greatest importance, diversity does not seem to correlate well with available shelf area. For instance the Hawaiian Islands contain approximately 1000 molluscan species while the tropical Pacific–Panamic province, without the diverse coral reefs, contains about 3000 such species. Both regions have very small shelf area, and correspond well to the narrow, linear provinces which existed during the latest Permian. This suggests that high diversity faunas can exist in small regions and the species–area effect may be spurious (Stanley *in* Nitecki 1984).

Marine provinces result from oceanic current patterns and differing temperature optima among taxa. At least 12 Middle Permian marine faunal provinces can be confidently identified and additional work in Asia and South America should increase this number; many of these provinces, particularly in the Tethyan region, had very high species diversity. However earliest Triassic faunas are composed of cosmopolitan ammonoids, bivalves, inarticulate brachiopods, and chondrosteian fish with low species diversity but high abundance. Yin (1985) identified six lowest Triassic bivalve provinces, while others have suggested a maximum of three provinces. Valentine & Moores (*in* Logan & Hills 1973; see also Schopf 1979) argued that the formation of Pangaea eliminated most marine provinces and played an important role in the extinction.

The low Early Triassic provinciality and the cosmopolitan faunas may be a consequence of the extinction rather than a cause, however. The major continental suturing occurred in the Lower and Middle Permian, yet provinciality apparently re-

mained high until the end of the period. The declining provinciality is unlikely to be a consequence of the regression, since regressions restrict the areal extent of marine provinces but only reduce the total number if extensive epeiric seas are present, which was not true in the Late Permian. The large number of 'Lazarus-taxa' demonstrates that many taxa persisted in as yet unidentified refugia.

Marine regressions have important collateral side-effects which may be the effective cause of extinction, including a loss of habitats, climatic changes and resulting resource instability, reduction or loss of facies-controlled community assemblages, and destruction of estuaries (a primary source of nutrients). Carbonate biomes are heavily affected at the end-Permian, and it may not be fortuitous that the Permian brachiopods found in the Tethyan mixed faunas are derived from underlying siliceous-facies assemblages.

Marine regression also changes the Earth's albedo, increasing continentality. As noted above, increasing continentality produces severe storms, high seasonality and exaggerated climatic fluctuations (Valentine & Moores *in* Logan & Hills 1973; Jablonski 1986). While many of these effects are terrestrial, increased seasonality will increase habitat disturbance, particularly in tropical regions. During such periods generalists, which utilize a broad range of resources, and disturbance-tolerant, eurytopic taxa will be favoured. Communities with large numbers of trophic specialists, typical of tropical areas, are highly efficient during periods of trophic stability but may be eliminated during unstable periods. Non-planktotrophic developers increase reproductive success by producing a few well adapted offspring and predominate in high latitudes and seasonal environments. Valentine & Jablonski (see Valentine 1986) recorded the preferential loss of planktotrophic taxa in crinoids, articulate brachiopods, and archaeogastropods. This may have been due to the more speciose nature of non-planktotrophic clades, or the increased trophic instability may have selected for non-planktotrophic developers. The high-latitude faunas where non-planktotrophs predominate are dominated by trophic generalists.

The most likely cause of the end-Permian mass extinction was tectonically-induced climatic instability and marine regression which brought about increased trophic instability (see Erwin 1990). Independently none of these would have produced an extinction of this magnitude, but their simultaneous occurrence produced a synergistic reaction,

magnifying the result (this pattern corresponds well with those seen in non-linear dynamics).

The extinction of many marine plankton, sessile filter feeders, and high-level carnivores, and the relative success of disturbance-tolerant taxa, particularly in high latitudes, emphasizes the role of trophic instability. This occurred as seasonality and habitat destruction increased. Highly specialized communities were most affected, leaving broadly dispersed trophic generalists to populate the earliest Triassic.

The end-Permian extinction also begins the 26.4 million year cycle of mass extinctions (Section 2.12.3) postulated by Raup & Sepkoski (the cycle may extend into the Palaeozoic, but the data are not sufficiently resolved at present). This implies a causal connection between the end-Permian and end-Cretaceous extinctions (the two largest of the series) which should produce similar extinction patterns. In fact, the two differ strongly. No iridium anomaly is present and the extinction patterns contrast sharply between the two events. The end-Permian event shows no marked vertebrate extinctions, unlike the extensive vertebrate extinctions of the Cretaceous (Section 2.13.7), and present evidence suggests that the end-Permian extinction was of significant duration. The selective removal of marine invertebrates with a planktotrophic developmental stage (crinoids, articulate brachiopods, and perhaps archaeogastropods) contrasts with the end-Cretaceous when larval developmental mode had no effect on survival. The extensive evidence for tectonically-induced climatic change and trophic instability in the Late Permian casts doubt on any causal link to the extinctions at the end of the Cretaceous.

References

- Clark, D.J., Wang, C-Y., Orth, C.J. & Gilmore, J.S. 1986. Conodont survival and low iridium abundances across the Permian-Triassic boundary in south China. *Science* **233**, 984-986.
- Erwin, D.H. 1990. The end-Permian mass extinction. *Annual Review of Ecology and Systematics* **21**, 69-91.
- Hallam, A. 1984. Pre-Quaternary sea-level changes. *Annual Review of Earth and Planetary Sciences* **12**, 205-243.
- Harland, W.B., Cox, A.V., Llewellyn, P.G., Pickton, C.A.G., Smith, A.G. & Walters, R. 1982. *A geologic time scale*. Cambridge University Press, Cambridge.
- Jablonski, D. 1986. Causes and consequences of mass extinctions: a comparative approach. In: D.K. Elliot (ed.) *Dynamics of extinction*, pp. 183-229. John Wiley & Sons, New York.
- Knoll, A.H. 1984. Patterns in the fossil record of gasvascular plants. In: M.H. Itaki (ed.) *Extinctions*, pp. 21-68.
- Lin, J-L., Fuller, M. & Zhang, W-Y. 1985. Preliminary Pha-

- nerozoic polar wander paths for the north and south China blocks. *Nature* **313**, 444–449.
- Logan, A. & Hills, L.V. (eds) 1973. *The Permian and Triassic systems and their mutual boundary*. Memoir of the Canadian Society of Petroleum Geologists, No. 2.
- Nitecki, M.H. (ed.) 1984. *Extinctions*. University of Chicago Press, Chicago.
- Schopf, T.J.M. 1979. The role of biogeographic provinces in regulating marine faunal diversity through geologic time. In: J. Gray & A.J. Boucot (eds) *Historical biogeography, plate tectonics and the changing environment*, pp. 449–457. Oregon State University Press, Corvallis.
- Sepkoski, J.J., Jr. 1982. *A compendium of marine families*. Milwaukee Public Museum Contributions to Biology and Geology, No. 51.
- Sepkoski, J.J., Jr. 1986. Global bioevents and the question of periodicity. In: O. Walliser (ed.) *Global bio-events*. Springer-Verlag, Berlin.
- Sheng, J.Z., Chen, C.Z., Wang, Y.G., Rui, L., Liao, Z.T., Bando, Y., Ishii, K., Nakazawa, K. & Nakamura, K. 1984. Permian–Triassic boundary in Middle and Eastern Tethys. *Journal of the Faculty of Sciences, Hokkaido University, Series IV* **21**, 133–181.
- Silver, L.T. & Schultz, P.H. (eds) 1982. *Geological implications of impacts of large asteroids and comets on the Earth*. Special Paper of the Geological Society of America No. 190.
- Valentine, J.W. 1986. The Permian–Triassic extinction event and invertebrate developmental modes. *Bulletin of Marine Science* **39**, 607–615.
- Xu, D-Y., Ma, S-L., Chai, Z-F., Mao, X-Y., Sun, Y-Y., Zhang, Q-W. & Yang, Z-Z. 1985. Abundance variation of iridium and trace elements at the Permian–Triassic boundary at Shangsi in China. *Nature* **314**, 154–156.
- Yin, H.F. 1985. Bivalves near the Permian–Triassic boundary in south China. *Journal of Paleontology* **59**, 572–600.

2.13.5 End-Triassic

M. J. BENTON

Introduction

A mass extinction event in the Late Triassic has been recognized for some time. The decline and virtual disappearance of the ammonoids at the Triassic–Jurassic boundary has long been clear to cephalopod workers, while in the nineteen-forties Edwin Colbert described the major extinctions of terrestrial tetrapods at that time.

Recent surveys of mass extinction events in the sea (e.g. Raup & Sepkoski 1982) have identified

the end-Triassic event as one of the five major Phanerozoic extinctions, equal in magnitude overall to the end-Ordovician (Section 2.13.2), Late Devonian (Section 2.13.3), and end-Cretaceous (Sections 2.13.6, 2.13.7) events, with a loss of over 20% of approximately 300 families of marine invertebrates and vertebrates. It is also one of the key extinctions in considerations of periodicity (Raup & Sepkoski 1984; Section 2.12.3), occurring as it does about 26–30 million years after the end-Permian event.

Extinction patterns

In the sea, several major lineages of invertebrates and vertebrates went extinct. The main groups to be affected were the cephalopods (58 families became extinct), the gastropods (13), various marine reptiles (13), the brachiopods (12), the bivalves (8), and the sponges (8). The effects on the ceratitid cephalopods, which were abundant and widespread in the Triassic, were dramatic — all 46 Late Triassic families died out — and the ammonoids as a whole barely survived into the Jurassic. When genera are considered, the Ceratitida reached a peak of *c.* 150 genera in the Carnian, which fell to *c.* 100 in the Norian, and to single figures in the latest Norian, finally disappearing at the Triassic–Jurassic boundary. This corresponds to an extinction rate of 100% at all levels. Only the Phylloceratina passed, at very low diversity, from the Triassic into the Jurassic. The family extinction rate for bivalves was not so marked, but the generic extinction rate was 42%, and the species extinction rate, in Europe at least, was 92% (Hallam 1981). This suggests that both the cephalopods and the bivalves barely scraped through the end-Triassic event into the Jurassic to establish new radiations. The last strophomenid brachiopods, conodonts, conulariids, nothosaurs, and placodonts also disappeared in the Late Triassic.

On land, major extinctions occurred amongst the insects (35 families), the freshwater bony fishes (8), and the thecodontians (8). There was a major faunal turnover amongst non-marine tetrapods in the Late Triassic, during which the formerly dominant labyrinthodonts, mammal-like reptiles, thecodontians, procolophonids, prolacertiforms, and rhynchosaurs died out, or were greatly depleted, and new groups, such as the dinosaurs, crocodiles, pterosaurs, turtles, lepidosaurs (lizards and their relatives), lissamphibians (frogs and salamanders), and mammals came on the scene (Benton 1986; Benton; Olson & Sues *in* Padian 1986).

Terrestrial events: competition or mass extinction?

There was a global faunal turnover amongst vertebrates on land in the Late Triassic. In several early papers, E. Colbert drew attention to the loss of a whole range of groups, as noted above. The nature of this massive replacement has been controversial. Initially, Colbert argued that the new groups radiated into effectively empty ecospace after a series of extinctions. Thus, the new lizard-like animals occupied the niches that procolophonids and prolacertiforms had held before, crocodiles filled the niches of the recently-extinct phytosaurs, and so on. However, views changed during the nineteen-sixties and seventies to focus more on competition-based models for the faunal replacement. The idea was that the dinosaurs outcompeted the formerly dominant mammal-like reptiles, rhynchosaurs, and thecodontians. The dinosaurs were said to have advantages in their style of locomotion (upright, instead of sprawling) and/or their thermal physiology. Thus, some authors argued that the dinosaurs must have been fully warm-blooded (endothermic) in order to compete successfully, while others stressed the advantages of cold-bloodedness (the need for less food and water).

However, recent detailed analyses of the fossil data (reviewed in Benton; Olsen & Sues *in* Padian 1986) suggest that the long-term competitive models are not likely. The relative abundances of the major terrestrial tetrapod groups throughout the Triassic show that there was no gradual long-term decline of the earlier groups, and matching radiation of the replacing groups. The dinosaurs were in existence near the beginning of the Late Triassic (in the Carnian), but they were very rare elements in their faunas (1% or less of all individuals) (Fig. 1). Many ecologically important groups then disappeared at the end of the Carnian, as far as the data indicate (rhynchosaurs, various mammal-like reptiles, and thecodontians), and the dinosaurs radiated thereafter in the Early Norian. Further groups of thecodontians and mammal-like reptiles disappeared at the end of the Triassic, and the dinosaurs apparently radiated again.

Other taxonomic studies on the diversity of tetrapod families in the Triassic (Benton 1986; Benton; Olson & Sues *in* Padian 1986) confirm the importance of mass extinction as the triggering factor for the remarkable faunal replacements in the Late Triassic. Indeed, there seems to have been more than one extinction event in the Late Triassic.

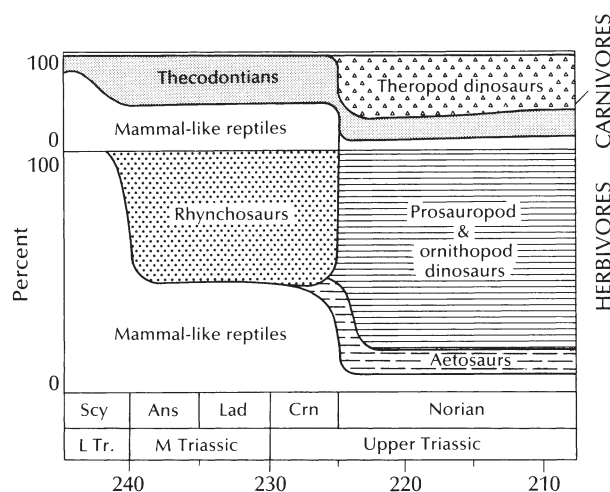


Fig. 1 The pattern of evolution of Triassic non-marine tetrapods, showing the extinction of major groups at the end of the Carnian stage. The dinosaurs, and other replacing groups, appear to have radiated only *after* the extinction event, thus suggesting an opportunistic replacement rather than one involving long-term competition. For abbreviations see the legend to Fig. 2. (After Benton 1988, by permission of Macmillan Magazines Ltd.)

Timing of the extinction events

Was there one mass extinction event in the Late Triassic or several? Many studies (e.g. Raup & Sepkoski 1982, 1984) identify a single event, but a great deal of evidence now appears to disagree with that view. Several authors had already noted that the extinctions in the Late Triassic were either not synchronous in the sea and on land (e.g. Hallam 1981), or that the extinction lasted for much of the Late Triassic, through the Carnian and Norian (the latter including the 'Rhaetian'), a time-span of 18–25 million years (depending on the time-scale used).

The timing of the Late Triassic marine extinction event has not been determined precisely for all groups. The bivalves declined in diversity from a Carnian–Early Norian peak, and were affected by a major extinction event at the end of the Norian (Hallam 1981). Similarly, the ceratite ammonoids reached their peak of diversity in the Carnian, and declined thereafter. The last genera disappeared at the end of the Norian ('Rhaetian'). The mass extinctions of brachiopods and conodonts appear to have occurred at the end of the Norian, while Benton (*in* Padian 1986); Olson & Sues (*in* Padian 1986) identified two extinction events for non-marine tetrapods, one at the end of the Carnian, and another at the end of the Norian.

The timing of the extinction of the marine vertebrates has been disputed. It has generally been assumed (e.g. Raup & Sepkoski 1982) to coincide with the end-Triassic invertebrate extinctions. However, most of the Late Triassic marine reptile families died out in the Carnian (five families), the timing of one is uncertain, and only one died out at the Triassic–Jurassic boundary:

Benton (1986) presented three separate analyses that indicate at least two mass extinction events in the Late Triassic, one probably at the end of the Carnian, and the other 12–17 million years later, at the Triassic–Jurassic boundary:

1 A detailed analysis of ammonoid families suggests that there were several declines in family diversity, the largest two in the Carnian and Late Norian (Fig. 2A). Total extinction rates for ammonoid families vary considerably during the Triassic and Early Jurassic (Fig. 2B), showing high peaks in the Late Scythian, the Late Ladinian, and the Early Carnian, and smaller peaks in the Anisian and in the Middle and Late Norian. Per-taxon extinction rates (Fig. 2C) show high peaks in the Late Scythian, the Late Anisian, the Early Carnian, and the Late Norian. The Late Triassic ‘mass extinction’ of ammonoid families was not a single event, but at least two — one in the Carnian, and a larger one at the end of the Norian (the ‘Rhaetian’).

2 Triassic and Early Jurassic families of non-marine tetrapods (Fig. 3A) show declines in diversity in the Early and Late Scythian, at the end of the Carnian, and at the end of the Norian. These declines are matched by peaks in total extinction rate (Fig. 3B) and in per-taxon extinction rate (Fig. 3C). The end-Carnian mass extinction of non-marine tetrapods was apparently the larger of the two Late Triassic events, according to Benton (1986), while the analysis by Olsen & Sues (*in* Padian 1986) suggested that the end-Norian event was larger than the end-Carnian one. These differences arise from the use of slightly different data sets.

3 Diversity and extinction rate data for all marine and non-marine animals in the Late Triassic (Benton 1986) indicate average marine diversity of about 340 families, and non-marine diversity of about 190 families. The numbers of family extinctions in marine taxa per time unit ranged from 3 to 54 (mean: 23.0), and in non-marine taxa from 4 to 45 (mean: 18.1). The plots of marine and non-marine family diversity (Fig. 4A) showed declines at the end of the Carnian, and smaller ones at the Triassic–Jurassic boundary. In all cases, both the total extinction rates (Fig. 4B), and the per-taxon extinction rates

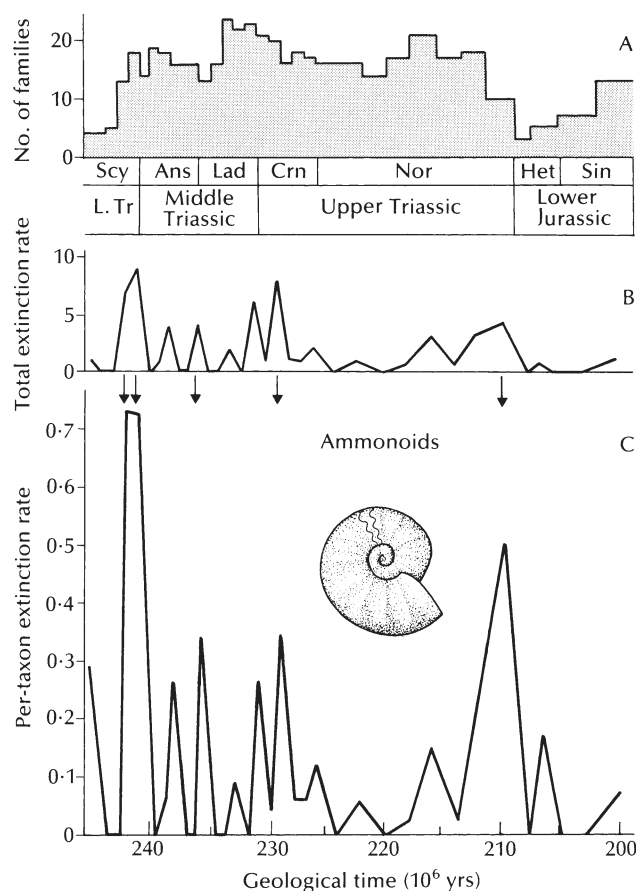


Fig. 2 Triassic ammonoid families. A, Total diversity. B, Total extinction rates. C, Per-taxon extinction rates. The total extinction rates were calculated as the numbers of families dying out per million years, and the per-taxon extinction rates as the total extinction rate divided by the number of taxa at risk. The five highest per-taxon extinction rates (possible mass extinctions) are indicated with arrows. Diversity and extinction rate data were calculated by Triassic and earliest Jurassic substages (Scythian, Hettangian, Sinemurian), and by zones (Anisian–Norian). The ‘Rhaetian’ stage is included in the Late Norian. Abbreviations: Ans = Anisian, Crn = Carnian, Het = Hettangian, Lad = Ladinian, L. Tr = Lower Triassic, Nor = Norian, Scy = Scythian, Sin = Sinemurian. The fossil is *Ceratites*. (After Benton 1986.)

(Fig. 4C) showed two similar peaks, with a slightly higher one in the Late Norian, and a lower one in the Carnian.

Present evidence points to as many as three Triassic extinction events: at the end of the Scythian, in the Middle to Late Carnian, and at the Triassic–Jurassic boundary; none of these is clearly larger than either of the other two. The temporal spacings of the three events are as follows:

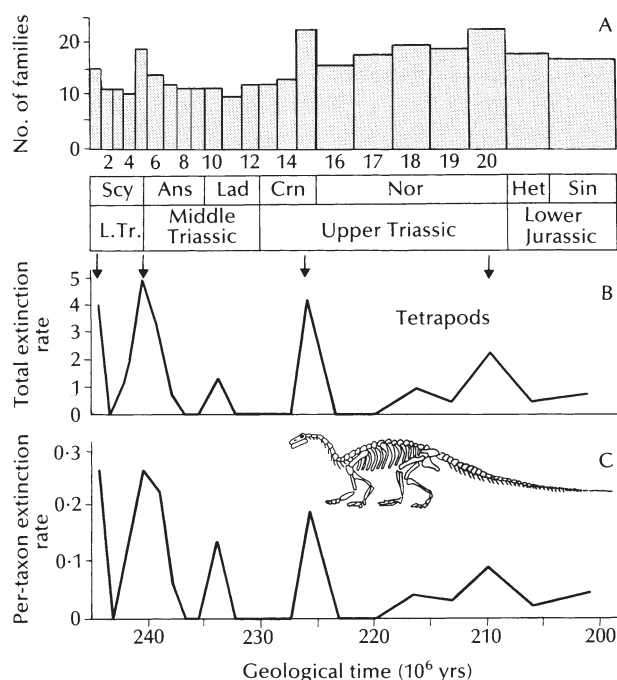


Fig. 3 Triassic and earliest Jurassic non-marine tetrapod families. A, Total diversity. B, Total extinction rates. C, Per-taxon extinction rates. The four highest total extinction rates (possible mass extinctions) are indicated with arrows. The time-scale is subdivided into informal 'substages', numbered 1–20 in the Triassic. For abbreviations see the legend to Fig. 2. The fossil is *Thecodontosaurus*. (After Benton 1986 by permission from Macmillan Magazines Ltd.)

end-Permian to end-Scythian: 5–6 Myr

end-Scythian to end-Carnian: 15–19 Myr

end-Carnian to end-Norian: 12–17 Myr

If the periodicity theory of mass extinctions (Raup & Sepkoski 1984) is to be valid (Section 2.12.3), there should have been a single event 26–30 million years after the end-Permian event, thus 219–222 Ma, in the Early Norian, according to current time-scales. Present evidence on the Late Triassic record of marine and non-marine animals strongly contradicts this prediction.

Causes

The Late Triassic tetrapod extinctions have been linked to an increasing aridity observed in reptile-bearing beds in various parts of the world (Benton in Padian 1986). Associated with these climatic changes were abrupt floral replacements in the Norian. The *Dicroidium* flora of Gondwanaland was replaced by a world-wide conifer-bennettitalean flora at the end of the Norian ('Rhaetian') and in the Early Jurassic. It has been suggested that these

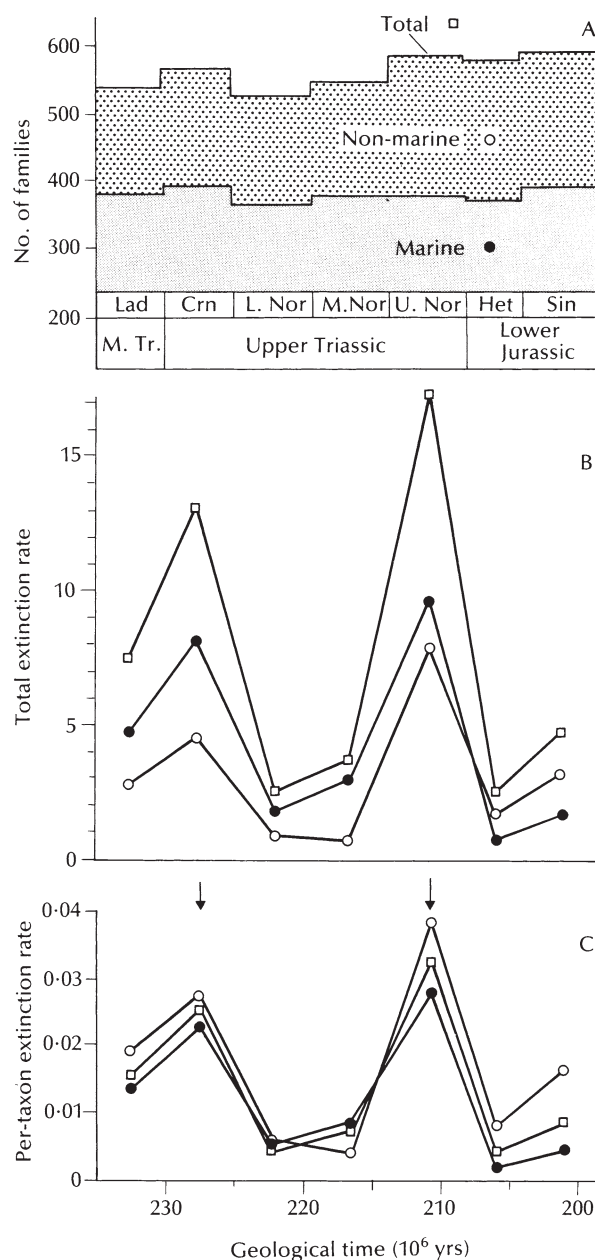


Fig. 4 Late Triassic and earliest Jurassic plant and animal families. A, Total diversity. B, Total extinction rates. C, Per-taxon extinction rates. The data are plotted separately for marine (closed circles), non-marine (open circles), and total (open squares) families. The two mass extinctions (end-Carnian; end-Norian) are indicated with arrows. The data were calculated by stratigraphic stage, except for the Norian which was subdivided into Lower, Middle, and Upper substages. Each Norian substage was assumed to have the same duration. For abbreviations see the legend to Fig. 2. (After Benton 1986, by permission from Macmillan Magazines Ltd.)

climatic and floral changes could have led to the extinction of various tetrapod groups. Another view links the extinctions with marine regressions and

reduced orogenic activity. These would have resulted in lower habitat diversity on land as new lowlands appeared, removal of reproductive barriers, and lower speciation rates. Indeed, much of the decline in diversity of tetrapods at this time is linked to depressed origination rates.

Several kinds of explanations have also been given for extinctions in the marine realm: widespread marine regression followed by an anoxic event (Hallam 1981; see also Section 2.12.1), temperature changes, or extraterrestrial impact (Section 2.12.2). Indeed, the last proposal is supported by the Manicougan crater in Canada (70 km in diameter) which is dated at 206–213 Ma. However, to date, no iridium anomaly or shocked quartz occurrence has been reported that coincides with either of the possible Late Triassic extinction events.

References

- Benton, M.J. 1986. More than one event in the late Triassic mass extinction. *Nature* **321**, 857–861.
- Benton, M.J. 1988. The origins of the dinosaurs. *Modern Geology* **13**, 14–56.
- Hallam, A. 1981. The end-Triassic bivalve extinction event. *Palaeogeography, Palaeoclimatology, Palaeoecology* **35**, 1–44.
- Padian, K. (ed.) 1986. *The beginning of the age of dinosaurs*. Cambridge University Press, New York.
- Raup, D.M. & Sepkoski, J.J., Jr. 1982. Mass extinctions in the marine fossil record. *Science* **215**, 1501–1503.
- Raup, D.M. & Sepkoski, J.J., Jr. 1984. Periodicity of extinctions in the geologic past. *Proceedings of the National Academy of Sciences* **81**, 801–805.

2.13.6 Cretaceous–Tertiary (Marine)

F. SURLYK

Introduction

The alleged mass extinction at the end of the Mesozoic Era has been one of the most intensively debated subjects within geology and palaeontology (see also sections 2.12.1, 2.12.2). In the forefront of the discussion has been the meteorite impact hypothesis (see Alvarez *et al.* 1984). The Mesozoic–Cenozoic Era boundary is placed at the boundary

between the Maastrichtian and Danian Stages. The Danian was, however, originally considered the end-Cretaceous Stage.

Discussions about the stratigraphic position of the Danian first started with Bramlette & Martini's (1964) observation that a major turnover of calcareous marine plankton took place at the Maastrichtian–Danian boundary. The broad pattern of marine extinctions is now known for both microfauna and flora, and macrofauna (see reviews by Kauffman 1984; Stanley 1987). The detailed extinction pattern, however, is known only for planktic foraminifera and calcareous nannoplankton, although new data are continually appearing for other microfossil groups. High-resolution stratigraphic information on the extinction across the boundary is virtually non-existent for marine invertebrates. Only a few groups have been studied and lend themselves to investigation on the basis of closely spaced sample series.

While the broad extinction patterns at the stage level can be relatively easily assessed on the basis of the literature, all discussions on the detailed patterns across the boundary come down to the question of sampling. It is well known that sampling effects can modify diversity patterns on both regional and local scales. Signor & Lipps (1982) particularly emphasized the effect of reduced sample size and artificial range truncation in the top Maastrichtian. Although this may seem trivial, it is overlooked in most reviews of the end-Maastrichtian extinction. Oceanic micro- and nannoplankton have a more complete record because the sediment itself is commonly composed of their skeletal remains. This is particularly true of biogenic sediments deposited above the carbonate compensation depth. Data based on randomly collected macroinvertebrate fossils, however, are poorly suited to illuminate the detailed nature of the extinction, and the existing data-base for invertebrates is totally inadequate to illustrate the short term nature and rate of extinction and diversification across the boundary.

A standard boundary sequence

A major breakthrough in high-resolution stratigraphy across the boundary was reached with the recognition that the same lithological and faunal succession could be traced over all major ocean basins (Smit & Romein 1985). This 'standard K–T boundary event sequence' contains a succession of five lithological units (1–5) which reflect the sequence of events across the boundary. The sequence

(or parts of it) was recognized in almost every boundary section world-wide. Unit 1 represents the uppermost Cretaceous and usually consists of pelagic calcareous oozes, limestones or marls. Unit 5 is in almost all respects comparable to unit 1 except for the completely new planktic biota of the Palaeocene. The thickness of the transitional interval between the Cretaceous and the Tertiary periods represented by units 2–4 amounts to no more than a few tens of centimetres in most sections. An exception is El Kef in Tunisia which shows the most complete evolutionary development of planktic foraminifera at the base of the Palaeocene. Here the transitional interval is almost 2 m thick. The analysis of Smit & Romein (1985) showed that the shift from dominantly Cretaceous to dominantly Palaeocene forms is consistently later for the nannoplankton than for the planktic foraminifera. The latter disappear within centimetres above the boundary, the former 10–50 cm higher in the section. This raises a very important question concerning the precise definition of the boundary. Should it be drawn at the mass extinction level, the iridium level (these two levels may coincide), the level of extinction of the last Cretaceous species, or at the first appearance of true Palaeocene taxa? This question still remains to be settled and it requires further detailed work on selected sections. Smit & Romein (1985) suggested that the standard K–T sequence could serve as a reference for all K–T boundary sections. It was interpreted as caused by a major impact, followed by a series of longer lasting biotic stresses. This interpretation can probably not be fully upheld, as will be shown below.

Microplankton

Work on the planktic foraminifera of the El Kef section by Keller (1988) demonstrated that the species extinctions prior to the assumed impact event cannot be explained by such a single impact, but suggest that multiple causes may be responsible. Keller's detailed study of the boundary at El Kef (Fig. 1A, B) revealed that the mass extinction of planktic and benthic foraminifera occurred over an extended period beginning well before and ending after the boundary. In contrast, geochemical data indicate a geologically instantaneous event at the boundary. The expanded sediment record at El Kef shows that the boundary extinctions of planktic foraminifera extend over an interval from 25 cm below the geochemical boundary (iridium anomaly) to 7 cm above.

Species extinctions appear sequential with complex, large, ornate forms disappearing first and smaller, less ornate forms surviving longer. Cretaceous species survivorship is greater than previously thought. Twenty-two percent (ten species) survive into the third subzone (*Globigerina eugubina*). The survivors are small primitive forms which are generally smaller than their ancestors. Species evolution after the boundary event occurred in two pulses. The first new Palaeocene species are found in the basal black clay immediately after the major Cretaceous extinctions. These evolving species, which define the first two subzones, are small and primitive, similar to the survivor species. The second pulse in species evolution occurred in the lower part of the third subzone and is characterized by larger, more diverse species. The first major increase in carbonate productivity occurred at this time, reflecting the recovery of the ecosystem nearly 300 000 years after the boundary event. The boundary extinctions largely affected deeper-dwelling planktic species, and surface dwellers survived longest. Keller concluded that the species extinctions prior to the generally assumed impact event (implied by the iridium anomaly), and the long recovery period of the ecosystem thereafter, cannot be explained by a simple impact. Multiple causes may be responsible, such as climatic changes, a sea-level drop, production of warm saline bottom water, and the chemical consequences associated with increased salinity.

Benthic foraminifera were, on the other hand, much less affected by the boundary crisis. However, Keller recorded a 50% reduction in species diversity at the boundary at El Kef, and the diversity remained 37% lower during deposition of the first 3 m of Danian sediments. The surviving species were generally tolerant of low O₂ conditions and productivity was low.

Keller worked with very densely spaced sample series and, even more importantly, also analysed smaller sample size fractions (63–150 µm) than the normal > 150 µm. This procedure fundamentally changed earlier concepts of the detailed nature of the microplankton boundary extinctions. All species in the earliest Tertiary assemblages are smaller than 150 µm, leaving Cretaceous survivors, as well as the rapidly evolving earliest Tertiary forms, unrecorded in the normally studied fraction.

Nannoplankton

Most other pelagic biota have been studied in less

detail across the boundary. The Coccolithophorida were once thought to go virtually extinct at the boundary. Cretaceous species are common in the basal Tertiary strata but were until recently always described as reworked forms. By ingenious isotope work, however, Perch-Nielsen *et al.* (1982) succeeded in demonstrating that the Cretaceous nannofossils in the lowest Tertiary sediments actually survived the boundary. These relic species became extinct some tens of thousands of years after the actual boundary, probably as a consequence of the environmental stress following the boundary events. The latest Cretaceous and the earliest Tertiary oceans contained significantly different isotopic signals, which were incorporated into the tests of the calcareous nannofossils. However, the Cretaceous species which dominate (to nearly 100%) the very lowermost Tertiary sediments have isotopic values different from those below the boundary. Perch-Nielsen *et al.* (1982) accordingly suggested that, in order to carry an Early Tertiary isotope signal, the nannofossils must have lived in the Early Tertiary oceans. That is, most of the typically Cretaceous species actually survived the mass extinction associated with the boundary. Survivors are most common in the basal Danian of high- and mid-latitude sites, whereas they are rare and only occur sporadically in low latitudes. The earliest Danian assemblages are of low diversity and short-term blooms of so-called disaster forms *Thoracosphaera* and *Braarudosphaera* are characteristic.

Other groups of marine microplankton have not received the same detailed study. Dinoflagellates were evidently much less affected by the boundary events, but an unusually high rate of species turnover across the boundary has been recorded in Danish localities by J.M. Hansen. A number of characteristic Maastrichtian species disappear at the top of the Maastrichtian while a succession of new Tertiary species rapidly appear in the lowermost Danian. The concentration of dinoflagellates in the boundary clay is extremely high, possibly representing blooms. Diversity is also high and there is a dearth of opportunistic species; this indicates non-stress conditions for this microplankton group.

Invertebrates

The problem of sampling effects becomes much more important for the record of marine invertebrates. Few facies are sufficiently fossiliferous to allow dense sampling. Even more limiting is the

rarity of richly fossiliferous shallow-marine Upper Maastrichtian sections, not to speak of relatively complete boundary sections. The sections most commonly mentioned in the literature are those at Stevns Klint and Nye Kløv in Denmark, and at Braggs and Brazos River in the U.S.A. The frequently cited Spanish localities around Zumaya show deep-water sequences which are extremely poor in benthic invertebrate fossils, while ammonites are common and diverse in the Lower Maastrichtian part of the sections.

Characteristic Cretaceous bivalve groups such as rudists and inoceramids experienced severe attrition before the end of the Cretaceous. Only about four inoceramid genera made their way into the Maastrichtian and no true inoceramid species are known from Upper Maastrichtian strata. The reef-forming rudists flourished in great diversity earlier in the Cretaceous, reaching a climax in the Late Campanian to Early Maastrichtian. They were decimated at the beginning of Late Maastrichtian time and were virtually extinct by the end of this stage.

The problems of sampling and preservation are illustrated by Heinberg's (1979) study of the bivalves across the boundary at Stevns Klint. The topmost 3.5 m of the Maastrichtian is developed here as low mounds of bryozoan chalk (see Alvarez *et al.* 1984, fig. 1). The mound summits are incorporated in a complex early (but not basal) Danian hardground. Early lithification predated aragonite dissolution and ensured the fortuitous mould preservation of minute aragonite shelled bivalves otherwise unknown from the Maastrichtian chalk. The fossil record of most chalks is thus highly distorted by the artificial dominance of calcite shelled taxa; Heinberg's work gives a tantalizing glimpse of the complete bivalve fauna once present. His data (on genera only) show that 13 typically Tertiary genera first appear in the lithified top of the Maastrichtian. He therefore concluded that the terminal Cretaceous faunal turnover in bivalves was under way before the boundary, indicating a graded faunal transition. Twelve of these genera are dominantly aragonitic and the ranges of ten were substantially extended back in time when they were recognized in the hardground. They almost certainly represent older Cretaceous groups that survived the extinction event, rather than the sudden advanced appearance of Tertiary forms before the mass extinction. Of the ten genera last appearing in this hardground, four had partially calcitic shells and their apparent extinction at this level is probably real. It is very likely

that the extinction at species level was more dramatic, such as that seen in the brachiopods and to some extent also bryozoans.

The fate of the chalk brachiopods has been studied in most Danish sections and a particularly detailed sample sequence has been described from the locality Nye Kløv (Fig. 1D). The uppermost Maastrichtian at Nye Kløv contains 27 species. Thirteen of these also occur in the 3 cm thick boundary clay; they have never been found higher in the Danian sequence, are broken, worn and non-transparent, and were probably contained in smears of Maastrichtian chalk which are common in the clay. The possibility that they represent very short term survivors cannot, however, be excluded. Six species are common to the Maastrichtian and the Danian Stages and represent true survivors. They are long-ranging, morphologically unspecialized forms which were probably environmentally tolerant generalist species. Finally, 23 species appear for the first time in the Danian 4–5 m above the boundary.

The same picture is known, albeit in less detail, from other Danish boundary localities. The end-Maastrichtian extinction seems to coincide with the base of the boundary clay, although several species last appear a few centimetres below the boundary. This may be taken to reflect their true extinction but is more likely a trivial sample effect (cf. Signor & Lipps 1982). The species diversity curve shows no systematic decrease at the end of the Maastrichtian. Several workers have focused on the barren interval above the boundary clay and have suggested the possibility that some of the Maastrichtian species survived for a time, possibly in the same refuge as the six survivors. This is certainly very likely but, with only a few exceptions, species found at a later date in the 'barren' interval in Nye Kløv and in correlative deposits at other Danish localities belong to the six surviving generalist species.

The extinction pattern noted for the brachiopods (Fig. 1C, D) was interpreted on purely ecological grounds. The fauna was specialized to the extensive macrohabitat of the chalk sea bottom and is virtually unknown outside this environment. Chalk productivity essentially stopped at the end of the Maastrichtian as a result of the calcareous microplankton crisis and extinction. The combined effects of the cessation of chalk production, the onset of clay deposition, hardground formation, and possibly low oxygenation at the sea floor caused the geologically instantaneous destruction of a unique macrohabitat of great longevity. The immediate

result was the mass extinction of faunal groups specialized to the chalk substrate, such as the brachiopods described above. When chalk production was eventually resumed in the Early Danian, adaptive radiation within surviving groups led to rapid restoration of the chalk macrohabitat and its fauna.

Much attention has been focused on the ammonites because they are a very characteristic Mesozoic faunal element which forms the basis for Mesozoic stratigraphy, and because they became extinct at the boundary. It is widely accepted that reduction in the diversity of ammonites had been in progress for several million years before the boundary. Genera and families fall into two groups. The first includes long-lived taxa, which neither increased nor decreased dramatically in diversity through time. The second group, composed of short-lived taxa, accounts for most of the variations in abundance and diversity. Nine species representing seven genera and six families occur in the topmost Maastrichtian of Denmark. They belong to a wide range of highly different morphotypes, probably reflecting different modes of life. The extinction of these taxa was the critical and unpredictable event. Without it the ammonites would not have disappeared. The Zumaya locality contains a diverse Lower Maastrichtian ammonite fauna, but the diversity rapidly decreases and the topmost Maastrichtian contains only few ammonites. This has been taken possibly to reflect an earlier extinction in a tropical, Tethyan environment, as compared to the Boreal chalk sea, but this may be putting just too much weight on an extremely poorly fossiliferous interval in a section. For comparison, ammonites are extremely rare in the main part of the Lower Maastrichtian of Denmark while they occur in abundance in the uppermost Lower and Upper Maastrichtian.

Our knowledge of the fate of marine vertebrates is very limited compared to that of the microplankton and the main invertebrate groups. The mosasaur and plesiosaur reptiles became extinct some time during the Maastrichtian together with the largest marine turtles, but nothing is known concerning the timing of the extinction.

The end-Cretaceous extinction is thus of a complex nature and the scenario has been variously described as invoking a 'multiplicity of interacting factors', as 'multicausal' or as a 'compound crisis'. Taken alone such phrases are merely clichés which are difficult to test, and they should not

provide a let-out from undertaking a more precise interpretation of the available data.

References

- Alvarez, W., Kauffman, E.G., Surlyk, F., Alvarez, L.W., Asaro, F. & Michel, M.V. 1984. Impact theory of mass extinction and the invertebrate fossil record. *Science* **223**, 1135–1141.
- Bramlette, M.N. & Martini, E. 1964. The great change in calcareous nannoplankton fossils between the Maastrichtian and the Danian. *Micropaleontology* **10**, 291–322.
- Heinberg, C. 1979. Bivalves from the latest Maastrichtian of Stevns Klint and their stratigraphic affinities. In: T. Birkelund & R.G. Bromley (eds) *Cretaceous-Tertiary Boundary Events Symposium. I. The Maastrichtian and Danian of Denmark*, pp. 58–64. University of Copenhagen, Copenhagen.
- Johansen, M.B. 1987. Brachiopods from the Maastrichtian–Danian boundary sequence at Nye Kløv, Jylland, Denmark. *Fossils and Strata* **20**, 1–99.
- Johansen, M.B. 1988. Brachiopod extinctions in the Upper Cretaceous to lowermost Tertiary chalk of northwest Europe. In: *Paleontology and evolution*, pp. 41–56. Revista Española de Paleontología, no. Extraordinario 1988.
- Kauffman, E.G. 1984. The fabric of Cretaceous marine extinctions. In: W.S. Berggren & J.A. van Couvering (eds) *Catastrophes and Earth history: the new uniformitarianism*, pp. 151–246. Princeton University Press, Princeton.
- Keller, G. 1988. Extinction, survivorship and evolution of planktonic foraminifera across the Cretaceous/Tertiary boundary at El Kef, Tunisia. *Marine Micropaleontology* **13**, 239–263.
- Perch-Nielsen, K., McKenzie, J. & He, Q. 1982. Biostratigraphy and isotope stratigraphy and the ‘catastrophic’ extinction of calcareous nannoplankton at the Cretaceous/Tertiary boundary. *Special Paper of the Geological Society of America* **190**, 353–371.
- Signor, P.W., III & Lipps, J.H. 1982. Sampling bias, gradual extinction patterns and catastrophes in the fossil record. *Special Paper of the Geological Society of America* **190**, 291–296.
- Smit, J. & Romein, A.J.T. 1985. A Sequence of events across the Cretaceous–Tertiary boundary. *Earth and Planetary Science Letters* **74**, 155–170.
- Stanley, S.M. 1987. *Extinction*. Scientific American Books, New York.
- Surlyk, F. & Johansen, M.B. 1984. End-Cretaceous brachiopod extinctions in the chalk of Denmark. *Science* **223**, 1174–1177.

2.13.7 Cretaceous–Tertiary (Terrestrial)

L. B. HALSTEAD

Introduction

The extinction of the dinosaurs is generally considered to be one of the most dramatic events in the Earth’s history. The potential significance of the event was enormously enhanced by the suggestion by Alvarez *et al.* (1980) that the demise of dinosaurs could be attributed to the impact of a large Earth-crossing asteroid (Section 2.12.2), indicated by iridium enrichment at the Cretaceous–Tertiary boundary. This publication sparked off a considerable debate which emphasized the differing methodologies of physicists and palaeontologists (see analysis by Clemens 1986). An alternative view (Section 2.12.1) is that it is not necessary to postulate the impact of an extra-terrestrial object to account for the iridium enrichment, but that volcanic activity provides a more reasonable explanation. This is supported by the accumulated evidence that the end-Cretaceous extinctions of different animal and plant groups on both land and in the seas were not synchronous (Section 2.13.6). When attempting any discussion of the Cretaceous–Tertiary extinction on land, it is important to recognize that an unbroken sedimentary transition of terrestrial sediments is only known in an area in Western Canada and the U.S.A. extending from Alberta in the north through Montana, Wyoming, and Colorado to New Mexico in the south.

Plants

A study of the record of plants by Wolfe & Upchurch (1986) revealed a dramatic change at the Cretaceous–Tertiary boundary with a sharp peak of distribution of fern spores. This widespread fern spike is usually taken as a sign of wildfires. Once fire has swept through an area, the first plants to recolonize are usually ferns and their allies, and this event is picked up in the fossil record as a fern spike. Wildfires can be attributed to either volcanic activity or asteroid impact. It is perhaps significant that a similar fern spike was identified at the Cretaceous–Tertiary boundary in marine sediments from Hokkaido island, Japan, by Saito, *et al.* (1986),

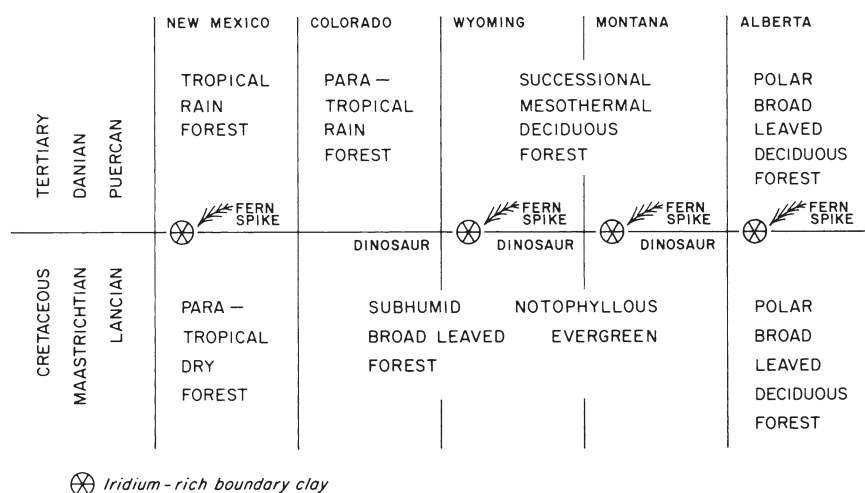


Fig. 1 Changes in vegetation at the Cretaceous-Tertiary boundary. (After Wolfe & Upchurch 1986.)

again indicating a major widespread event that triggered off wildfires.

The cause of the wildfires is not the important issue, but whether or not there was a major change in plant life at this juncture (Fig. 1). The evidence from terrestrial plants points to a sudden event at the Cretaceous-Tertiary boundary with widespread forest fires. But the pattern of change across the boundary does not seem to involve much in the way of extinction; rather it seems to indicate a climatic change with an increase in seasonality and the concomitant change in the pattern of distribution of vegetation zones. It is striking that in Alberta broad-leaved deciduous forests both above and below the boundary indicate a comparable climatic regime, whereas in New Mexico the subtropical dry forest of the end-Cretaceous is replaced by tropical rain forest in the Tertiary (Fig. 1).

Mammals

The record of terrestrial vertebrates has been carefully documented by Archibald & Clemens (1982; 1984). Microvertebrates (mainly mammal teeth) have been carefully collected across the Cretaceous-Tertiary boundary. A clear pattern emerges of gradual change (Fig. 2). The multituberculates, with their chisel-like incisors, must have occupied the gnawing-nibbling niche, and continued into the Tertiary with relatively little change. The small herbivorous condylarthrans began a modest increase in both numbers and diversity, and by Early Tertiary times had become the major herbivores. By far the most significant change in the mammalian fauna was the drastic reduction in the number of species of marsupials, with basically

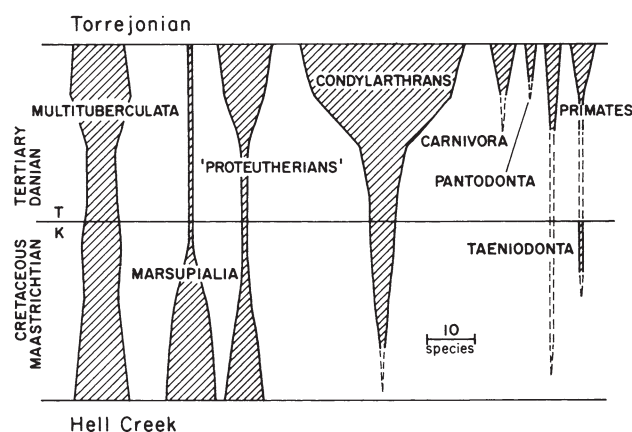


Fig. 2 Changes in mammalian faunas across the Cretaceous-Tertiary boundary. (After Archibald & Clemens 1984, by permission from Macmillan Magazines Ltd.)

only the opossum continuing through to the Tertiary. The primitive proteutherian insectivorous mammals seem to have declined in the Late Cretaceous and then recovered in the Tertiary, achieving a modest increase in diversity compared with the preceding Cretaceous.

The pattern of mammalian distribution through time demonstrates a gradual replacement of more typical Cretaceous forms by Tertiary types. This gradual turnover began well before the Cretaceous-Tertiary boundary. There is clear evidence that significant faunal changes occurred around the time of the Cretaceous-Tertiary boundary, but there was no sign of any dramatic event taking place at the boundary itself.

Reptiles

The general consensus is that the dinosaurs all died out during the Cretaceous. It has been claimed, however, that seven dinosaur genera survived into the basal Tertiary. Dinosaur bones were discovered in channel deposits containing typical Palaeocene mammals (Sloan *et al.* 1986). The claim has been disputed with some vigour (see Retallack & Leahy 1986 and discussion).

Whether or not the existence of Palaeocene dinosaurs is confirmed, what is clearly established is that from the Campanian (as seen in the Dinosaur Provincial Park, Alberta) through the Horseshoe Canyon Formation to the Hell Creek and Lance Formations the dinosaurs were beginning to go into a serious decline and that during the last 300 000 years of the Cretaceous this decline was greatly accelerated (Fig. 3). Both numbers of specimens and diversity were drastically reduced. Van Valen & Sloan (1977) assembled evidence which suggested a gradual southwards migration of dinosaurs in the wake of an invasion by large numbers of herbivorous condylarthran mammals, and by more temperate plant life which replaced the humid broad-leaved evergreen forests.

The distinct impression is left, however, that the end for the few remaining dinosaurs was sudden. It is therefore instructive to examine the record of reptiles across the Cretaceous–Tertiary boundary (Sullivan 1987) (Fig. 4).

Among the chelonians (turtles and tortoises) only the family Protostegidae died out at the end of the Campanian. Six other families crossed the boundary well into the Tertiary and ranged beyond the Palaeocene. The crocodile-like eosuchians (the Champsosauridae) ranged through the latter part of the Campanian, the Maastrichtian, and the Palaeocene. Three families of crocodile ranged through into the Tertiary. Several families of lizard similarly survived into the Tertiary with no evidence of any change: the agamids, anguinids (slow-worms), the iguanas, teids, and varanids. A number of groups known from the Tertiary, however, are not certainly known from the uppermost Cretaceous, such as the Gila monster, the helodermatids, the necrosaur, skinks, and the xenosaurids. Of the snakes only the boids (boas) spanned the entire Campanian through to the present. Thus the Cretaceous–Tertiary boundary was no barrier to chelonians, eosuchians, crocodiles, lizards, and snakes.

When the record of the dinosaurs is examined in

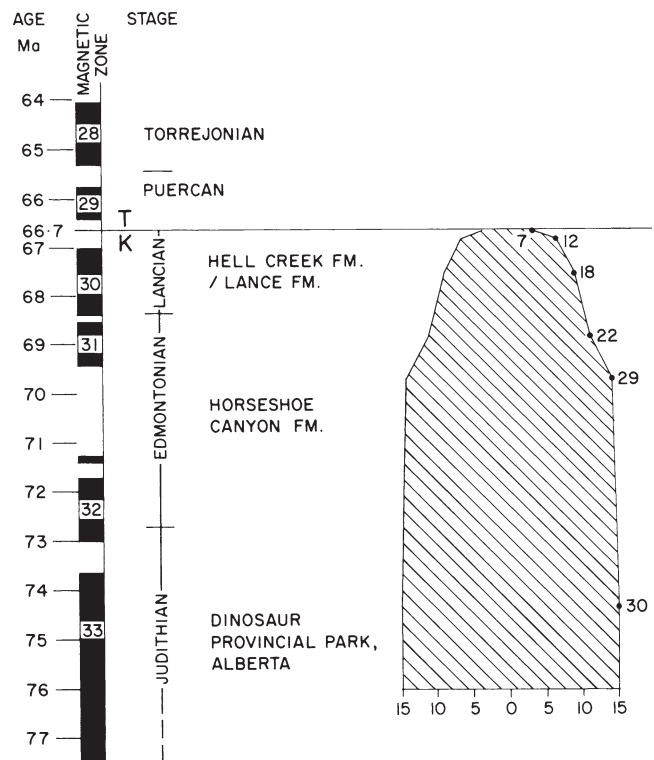


Fig. 3 Diversity of dinosaur genera in the Upper Cretaceous showing the decline below the Cretaceous–Tertiary boundary. (After Sloan *et al.* 1986.)

comparable detail it is strikingly evident that most genera became extinct before the Cretaceous–Tertiary boundary (Fig. 4). The giant scavenging carnivores *Tyrannosaurus* and *Albertosaurus* reached the boundary; the lightly built clawed carnivorous dromaeosaurs are known from a few fragments at the boundary. The ornithomimids (ostrich dinosaurs) died out just prior to the boundary, and the last surviving sauropod *Alamosaurus* did not reach it.

The primitive small ornithopod *Thescelosaurus*, a surviving hypsilophodont, reached the end of the Cretaceous. The primitive non-crested hadrosaur *Edmontosaurus* reached the boundary. One of the armoured dinosaurs *Euplococephalus* made it, whereas another genus *Panoplosaurus* became extinct at the base of the Lancia. Among the bone-heads, the well known genera *Stegoceros* and *Pachycephalosaurus* had died out before the boundary but one genus *Stygimoloch* managed to reach it. The only North American protoceratopsid *Leptoceratops* died out in the Edmontonian and only two ceratopsians continued to the end: *Triceratops* reached the boundary, while *Torosaurus* did not quite make it.

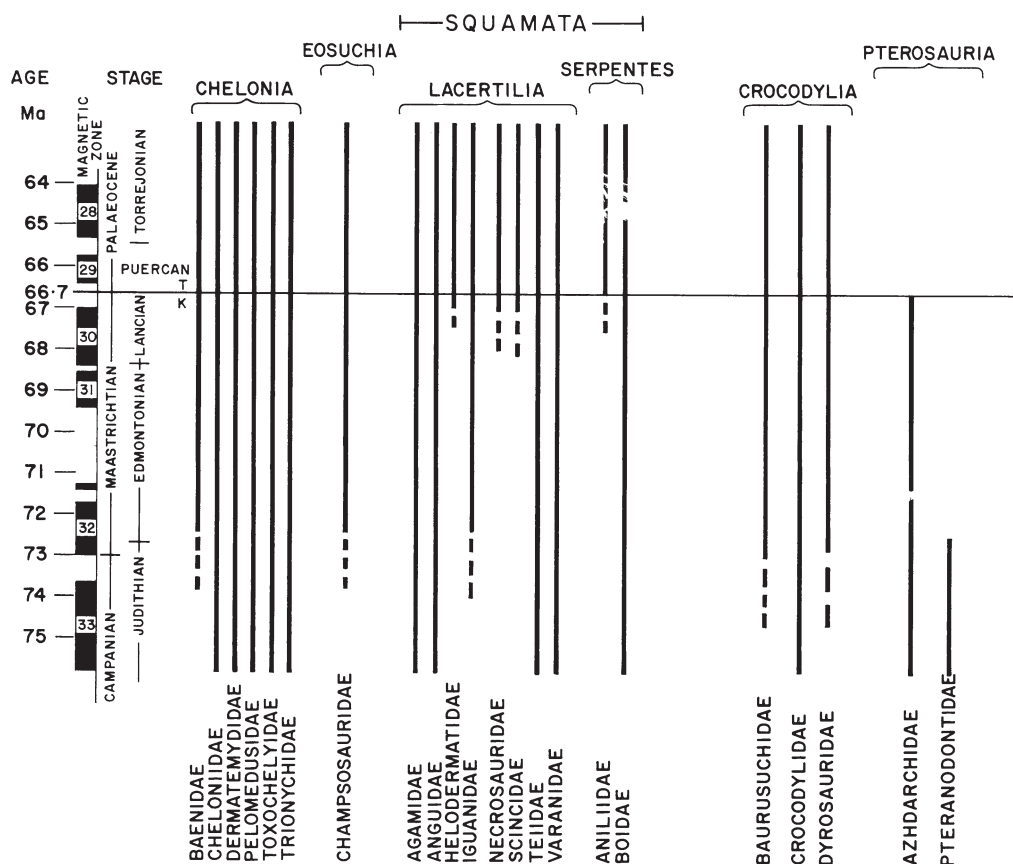
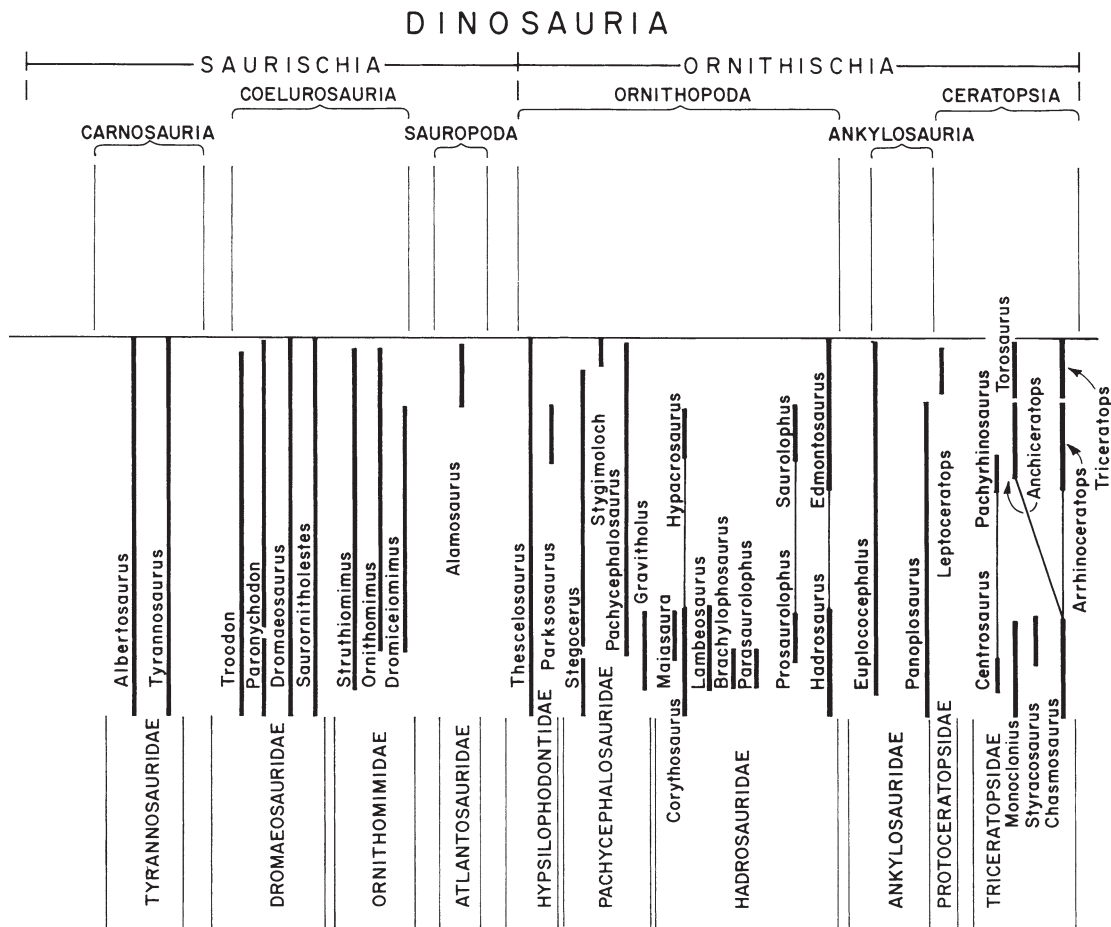


Fig. 4 Changes in reptilian faunas across the Cretaceous-Tertiary boundary. (After Sullivan 1987; Sloan *et al.* 1986. [© copyright 1986 by the AAAS].)

It appears that only 12 species of dinosaur reached the Cretaceous-Tertiary boundary. It becomes difficult to regard the end-Cretaceous event as a mass extinction of the terrestrial biota. The fossil record certainly documents important biological changes, but they indicate a serious decline in part of the fauna during the Late Cretaceous, with a significant acceleration of the process towards the boundary itself.

References

- Alvarez, L.W., Alvarez, W., Asaro, F. & Michel, H.V. 1980. Extraterrestrial cause for the Cretaceous-Tertiary extinction. *Science* **208**, 1095-1108.
- Archibald, J.D. & Clemens, W.A. 1982. Late Cretaceous extinctions. *American Scientist* **70**, 377-385.
- Archibald, J.D. & Clemens, W.A. 1984. Mammal evolution near the Cretaceous-Tertiary boundary. In: W.S. Berggren & J.A. van Couvering (eds) *Catastrophes and Earth history: the new uniformitarianism*, pp. 339-371. Princeton University Press, Princeton.
- Clemens, E.S. 1986. Of asteroids and dinosaurs: the role of the press in the shaping of scientific debate. *Social Studies of Science* **16**, 421-456.
- Retallack, G., Leahy, G.D., & Sheehan, P.M. 1986. Cretaceous-Tertiary dinosaur extinction. *Science* **234**, 1170-1171.
- Saito, T., Yamanoi, T. & Kaiho, K. 1986. End-Cretaceous devastation of terrestrial flora in the boreal Far East. *Nature* **323**, 253-255.
- Sloan, R.E., Rigby, J.K., Van Valen, L.M. & Gabriel, D. 1986. Gradual dinosaur extinction and simultaneous ungulate radiation in the Hell Creek Formation. *Science* **232**, 629-633.
- Sullivan, R.M. 1987. A reassessment of reptilian diversity across the Cretaceous-Tertiary boundary. *Contributions in Science* **391**, 1-26.
- Van Valen, L.M. & Sloan, R.E. 1977. Ecology and extinction of the dinosaurs. *Evolutionary Theory* **2**, 37-64.
- Wolfe, J.A. & Upchurch, G.R. 1986. Vegetation, climatic and floral changes at the Cretaceous-Tertiary boundary. *Nature* **324**, 148-154.



2.13.8 Pleistocene

E. L. LUNDELIUS, Jr

Introduction

One of the major world-wide biological events of the last million years was the extinction of numerous taxa of mammals and a few lower vertebrates. This extinction event differed from those at the end of the Permian (Section 2.13.4) and the Cretaceous (Sections 2.13.6, 2.13.7) in that it primarily affected large (i.e. more than 44 kg) terrestrial mammals, rather than the much greater variety of vertebrate and invertebrate animals from both marine and terrestrial habitats that characterized earlier extinctions.

Extinction patterns

In North America, 33 genera of large and four genera of small (i.e. less than 44 kg) mammals disappeared. South America lost up to 46 genera of large mammals. In Europe, of 13 genera that disappeared, only three became totally extinct while others survived elsewhere (Martin *in* Martin & Klein 1984) and there was no significant loss of small animals. Data from Africa and Asia are too sparse to give a comprehensive picture, but it appears that Southern Africa lost at least six species and Northern Asia lost about three. Fifteen genera (48 species) of mammals, three large reptiles and one large bird became extinct in Australia.

Many of the North American species that became extinct, such as several ground sloths (*Glossotherium*, *Eremotherium*, *Nothrotheriops*, and *Megalonyx*), saber-toothed cats (*Smilodon*, *Homotherium*), giant short-faced bears (*Arctodus*), large peccaries (*Platygonus*, *Mylohyus*), glyptodonts (*Glyptotherium*), and giant

armadillos (*Holmesina*) were the last representatives of formerly much more diverse genera and families. The extinction of the mastodon (*Mammuthus*) and the last two species of mammoths (*Mammuthus jeffersoni* and *M. primigenius*) terminated a whole order of mammals, the Proboscidea, in North America. The extinction of all these taxa resulted in the disappearance of a large number of adaptive types.

In South America, the extinct taxa included the last representatives of the indigenous orders Notoungulata and Liptopterna, which contained one and two genera respectively. In the order Edentata four families, the Glyptodontidae (nine genera), and three families of sloths — the Megalonychidae (three genera), the Megatheriidae (two genera), and the Mylodontidae (five genera) — disappeared. Two other groups of indigenous South American mammals, the armadillos, and the caviomorph rodents, suffered losses. Several groups of North American origin, such as the proboscideans and horses, disappeared completely, while others, such as the camels, felids, and cervids, underwent a loss of diversity.

Eurasia and Africa had fewer losses. Most of the forms that disappeared in Europe survived in other continents, primarily Africa, and only one taxon, the saber-tooth cat *Homotherium*, represented a totally extinct adaptive type. There were few extinctions in Africa during the Late Pleistocene and these were at the specific or generic level. In Australia, three families of marsupials — Palorchestidae, Diprotodontidae, and the Thylacoleonidae — became extinct (Murray in

Martin & Klein 1984). The number of genera and species in other families, such as the Macropodidae and Vombatidae, were reduced.

The exact time and duration of the extinction events were not the same on all continents. In North America an extensive list of radiocarbon dates indicate that most of the extinctions occurred between 13 000 and 9 000 years BP. A critical review of the youngest radiocarbon dates for 20 genera (Mead & Meltzer in Martin & Klein 1984) suggests that most of the taxa disappeared in a narrow time interval between 12 000 and 10 000 BP (Fig. 1). In Europe the disappearance of the large mammals spanned a longer period of time, perhaps beginning as early as 20 000 BP and ending about 11 000 BP. In southern Africa the major extinction event took place between 12 000 and 9 500 BP (Klein in Martin & Klein 1984). In Australia the extinctions began about 25 000 BP and continued until 11 000 BP. The major effect of these extinctions was a great reduction in diversity of the large mammal component of the faunas of these continents.

An examination of the pattern of extinction in North America in the last ten million years indicates that the terminal Pleistocene extinction was only the last in a series of extinctions that occurred at 9, 6, 5, 1.9, and 0.5 Ma. For reasons that are not understood, the earlier extinctions exterminated a larger proportion of small animals than did the last one. Tentative correlations with the deep sea glacial/interglacial chronology indicate that the extinctions coincided with deglaciation events, particularly with those that ended longer and/or

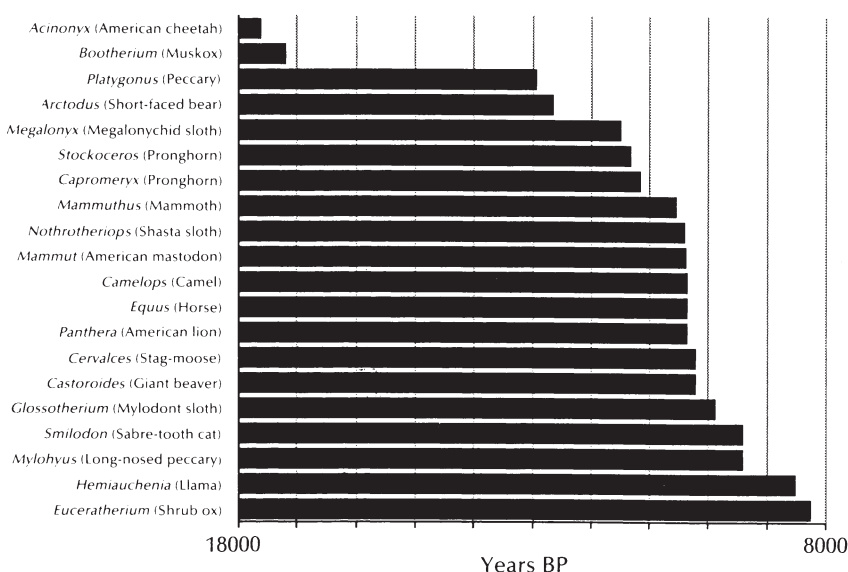


Fig. 1 Terminal dates of some North American mammals based on reliable Carbon 14 dates. (Data from Mead & Meltzer in Martin & Klein 1984.)

more severe glacial stages (Webb *in* Martin & Klein 1984).

Causes

The two most commonly proposed hypotheses as to the causes of the terminal Pleistocene extinctions are human overpredation and the destabilization of habitats caused by climatic change. The model for human overpredation relies on the apparent coincidence in time between the arrival of humans in North America and the disappearance of the large mammals (Martin *in* Martin & Klein 1984). On other continents, where these migrations did not occur, it has been proposed that new, and presumably more sophisticated, hunting techniques were developed and contributed to the extinction of prey species (Klein *in* Martin & Klein 1984).

The case for the climatic hypothesis depends on a coincidence in time between the extinction event and the extensive environmental changes that

occurred at the end of the last glacial stage. A shift from an equable to a more seasonal climatic regime is indicated by the presence, within the same fossil deposits, of sympatric associations of extant species whose ranges have become markedly allopatric since 11 000 BP. Palynological data show a loss of similar associations of plant species at this time with a consequent depletion of niches resulting from a reduction in local plant diversity. The loss of these associations closely coincides with the extinctions of the large mammals in all parts of the world for which there are adequate data (Lundelius 1983; Graham & Lundelius *in* Martin & Klein 1984).

References

- Lundelius, E.L., Jr. 1983. Climate implications of late Pleistocene and Holocene faunal associations in Australia. *Alcheringa* 7, 125–149.
- Martin, P.S. & Klein, R.G. (eds) 1984. *Quaternary extinctions: a prehistoric revolution*. University of Arizona Press, Tucson.

