

The quality of the fossil record of the vertebrates

MICHAEL J. BENTON

Department of Earth Sciences, University of Bristol, Bristol, BS8 1RJ, U.K.

[Cite as: Benton, M. J. 1998. The quality of the fossil record of vertebrates. Pp. 269-303, in Donovan, S. K. and Paul, C. R. C. (eds), *The adequacy of the fossil record*. Wiley, New York, 312 pp.]

Fossil vertebrates include a great diversity of animals of all sizes and shapes, ranging in age back to the Cambrian. The history of the vertebrates has been recounted many times (for example, Romer, 1966; Carroll, 1987; Benton, 1990a, 1997a) and the outlines of the story are well known. These broad outlines were worked out during the nineteenth century, and the sequence includes the armoured ostracoderms and placoderms of the Devonian, Carboniferous amphibians, Permian mammal-like reptiles, Mesozoic dinosaurs, ichthyosaurs, plesiosaurs, and pterosaurs, birds, Tertiary mammals, and Plio-Pleistocene hominids. This succession is usually recalled as a one-way progression from essentially toothed worm-like creatures of the early Palaeozoic to humans, even though such a vision is merely a didactic device, and does not properly depict the branching bushy pattern of vertebrate evolution.

How good is this record of the diversification of vertebrates? There are two intuitive answers. One is to suggest that the record is good because nothing much has changed in our understanding of the timing of events in the past 100 years of research. The other is to say that the record is terrible because many vertebrates, particularly tetrapods, live on land, and they are much less likely to be preserved than marine shelf invertebrates. This criticism cannot, of course, apply to fishes.

The purpose of this paper is to summarize current knowledge of the history of backboned animals, but only in the simplest of outline forms, and to present some recent work on the quality of the fossil record of vertebrates.

A GOOD FOSSIL RECORD?

It is likely that the numbers of fossils, localities, and vertebrate palaeontologists have multiplied by several orders of magnitude during the twentieth century, and yet there have been no surprises in the accepted broad-scale pattern of vertebrate evolution. Of course, the impressive efforts of collectors have pushed the origins of various groups backwards in time, but these stratigraphic range extensions have all been predictable. This century, the origin of agnathan fishes has been pushed back from the Silurian to the Cambrian, the origin of amphibians from the early Carboniferous to the latest Devonian, the origin of reptiles from the early Permian to the mid Carboniferous, and the origin of mammals back from the mid Jurassic to the late Triassic. Arguably, the origin of birds has remained unchanged in the latest Jurassic, since the work of Owen and Huxley in the 1860s and 1870s. However, if *Protoavis* is a bird (Chatterjee, 1995), then the point of origin of the group moves back to the late Triassic, and that would distort many parts of the phylogeny, not only of birds, but also of Dinosauria in general.

A critic of the quality of the vertebrate fossil record might have expected more surprises. Recall that Charles Lyell, a supporter of the idea that time proceeded in cycles, and an opponent of the idea of progression, or unidirectional change through time, quite expected in the 1830s that human fossils might be found in the Silurian. He campaigned hard in the 1850s to convince colleagues that new discoveries of Silurian arthropod tracks from North America had actually been made by land vertebrates. He also lent his strong support, in the early 1850s, to the view that the aeolian yellow sandstones round Elgin in north-east Scotland, which had just yielded supposed turtle tracks and the skeleton of an apparently lizard-like animal, were actually Devonian in age, rather than Triassic (Benton, 1983). Had he been right, then the generally accepted pattern of vertebrate evolution would have looked very different. So far, in 150 years of searching, palaeontologists have not found human remains in the Silurian, nor have they found modern-style reptiles in the Silurian or Devonian. It can be asserted that, the longer such out-of-place fossils do not turn up, the greater the likelihood that our knowledge of vertebrate evolution approximates the truth.

The notion of a good fossil record of vertebrates was confirmed in a quantitative analysis by Maxwell and Benton (1990). These authors compared several stages in the development of knowledge about the history of tetrapods over the past 100 years. They used a number of publications, dated 1900, 1933, 1945, 1966, and 1987, as snapshots of then-current knowledge of the former diversities and distributions in time of families of fossil tetrapods. There had certainly been huge changes in palaeontological knowledge from 1900 to 1987, not least a doubling of the known diversities of all groups, presumably as a result largely of intensive collecting efforts. In addition, the snapshots of palaeontological understanding included revisions of stratigraphy and taxonomy. However, the results of all these changes appeared to be randomly distributed with respect to time. Global diversities essentially doubled throughout the whole fossil record of tetrapods, from the late Devonian to the present-day, but without any biases becoming evident. The overall pattern of diversification, and the timing and magnitudes of major extinction events, were unchanged. Maxwell and Benton (1990) concluded that all the changes in understanding of the tetrapod fossil record in the past 100 years had not altered the broad-scale macroevolutionary patterns derived from it. These findings have been confirmed for marine animals in an analogous study by Sepkoski (1993).

A POOR FOSSIL RECORD?

Perhaps a commoner intuitive view of the vertebrate fossil record is that it is poor or very poor, especially when compared to the fossil records of marine shelf skeletonized invertebrates (for example, Valentine, 1969; Raup, 1979; Benton, 1985; Cowen, 1990; Flessa, 1990; Jablonski, 1991). This assumption has been made by scaling up from field observations. Typically, limestones and clastic rocks laid down on the shallow continental shelf yield abundant fossils of skeletonized invertebrates, such as brachiopods, molluscs, corals, arthropods, bryozoans and echinoderms (Kidwell, 1986; Fürsich, 1990). Continental sediments, on the other hand, generally yield much less abundant fossil faunas of freshwater fishes and molluscs, and terrestrial insects and vertebrates (Behrensmeyer and Hill, 1980; Retallack, 1984).

This differentiation may be largely an effect of the nature of the sediments: sedimentation in river systems and lakes is highly episodic compared to the more continuous deposition on marine shelves, and particularly in abyssal areas of oceans (Sadler, 1981). In addition, there may be biological factors. Many groups of skeletonized marine shelf invertebrates include forms with relatively short life spans, forms that live in huge abundances, and some which

moult, and hence produce several potential body fossils during a lifetime. Many vertebrates, and tetrapods in particular, often have lifespans lasting several years, and populations are often not counted in the thousands or millions.

Many of these observations are qualitative, but recent taphonomic and palaeoecological studies (for example, papers in Briggs and Crowther, 1990; Allison and Briggs, 1991; Donovan, 1991) show huge differences in the abundance and closeness of spacing between fossil species of marine invertebrates and planktonic forms on the one hand, and continental vertebrates on the other.

In conclusion, there are two intuitive views of the quality of the fossil record of vertebrates, and each is supported by observational evidence and by quantitative studies. How can they be reconciled?

THE PATTERN OF THE EVOLUTION OF VERTEBRATES

The evolution of any group can be represented diagrammatically in various ways. One useful kind of graphic presentation is a 'spindle diagram', in which the evolution of a group is represented by symmetrical spindle shapes which indicate the waxing and waning of a group through time. The y-axis is proportional to time, and the x-axis to species numbers, or some similar measure. Each group originates as a narrow point, and then typically expands into a wider spindle as it radiates. Any individual spindle may remain narrow (low diversity) through time, or it may expand, or vary in width through time, depending on the relative fortunes of the group.

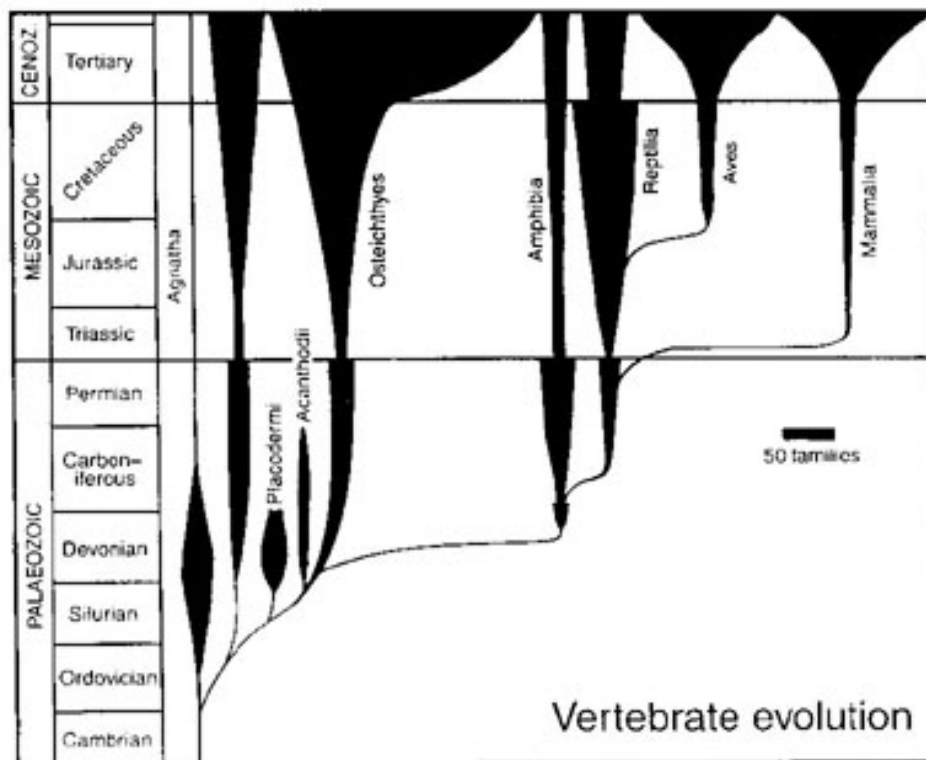


Figure 1. The pattern of evolution of the vertebrates, showing the relative importance of the major groups through time. This is a 'spindle diagram', in which the vertical axis represents geological time, and the horizontal axis represents the diversity of each group. In this case, the horizontal

dimension is proportional to the number of families in each group, based on data compiled by various authors in Benton (1993). The groups include some clades (that is, monophyletic groups), such as Chondrichthyes, Placodermi, Acanthodii, Aves, and Mammalia, but the others are paraphyletic groups (that is, a group that includes the ancestor, but not all of the descendants of that ancestor). All groups are treated in their traditional sense. Mass extinctions show up in the Late Devonian, late Permian, and end-Cretaceous, indicated by relatively rapid contractions in the diversities of several clades.

For vertebrates, the latest compilation of data (chapters in Benton, 1993) gives a diagram (Fig. 1) that has not changed much since earlier comparable attempts (for example, Romer, 1966; Carroll, 1987). The relative widths of different spindles have changed somewhat, as also the currently oldest indications of some of the groups. Certain mass extinctions, particularly those in the late Devonian, late Permian, and at the end of the Cretaceous, are highlighted by relatively rapid contractions in the widths of several of the spindles. This implies high rates of extinction of several groups at the same time.

These data on diversity may be rendered as more precise plots of actual counts of families through time (Fig. 2). In this case, the same database (Benton, 1993) was trawled for data on the diversities of families of the major groups. The data are plotted separately for fishes (Fig. 2a) and tetrapods (Fig. 2b), and, in each case, the presentation is cumulative, with each labelled curve adding on top of the curves below. The upper curve represents the total sum of diversity for all fishes or all tetrapods at any time in the past 500 Myr.

In detail, the pattern of diversification of the fishes (Fig. 2a) shows a rapid rise in the early Palaeozoic, to a broad peak, at about 50-70 families in the Devonian. This initial rise in diversity may be more apparent than real; Sansom *et al.* (1996) have argued that the pre-Devonian record of fishes is unnaturally impoverished, and that true diversity was much higher than is currently believed. Diversity declines at the end of the Devonian, and diminishes further through the late Permian, before slowly climbing during the Mesozoic. The diversity of fishes reaches Devonian levels of 50-70 families again in the Cretaceous, and then diversification is apparently explosive through the Tertiary, although apparently slowing down in the last 40 Myr.

The high levels of fish diversity in the Devonian are dominated by ostracoderm agnathans and placoderms, both of which groups were hard hit by the late Devonian mass extinction. Bear in mind, however, that 'fishes' is a paraphyletic group, and the drop in post-Devonian diversity was matched by the diversification of Palaeozoic tetrapods, just another branch of fish diversity that happened to move partly or fully on to land, through the late Devonian and Carboniferous. The subsequent diversification of fishes is dominated by the chondrichthyans (sharks and rays) and actinopterygians (bony fishes), and the dramatic early Tertiary burst in diversification is driven largely by the radiation of two clades, the Neoselachii (modern sharks) and the Teleostei (the majority of modern bony fishes).

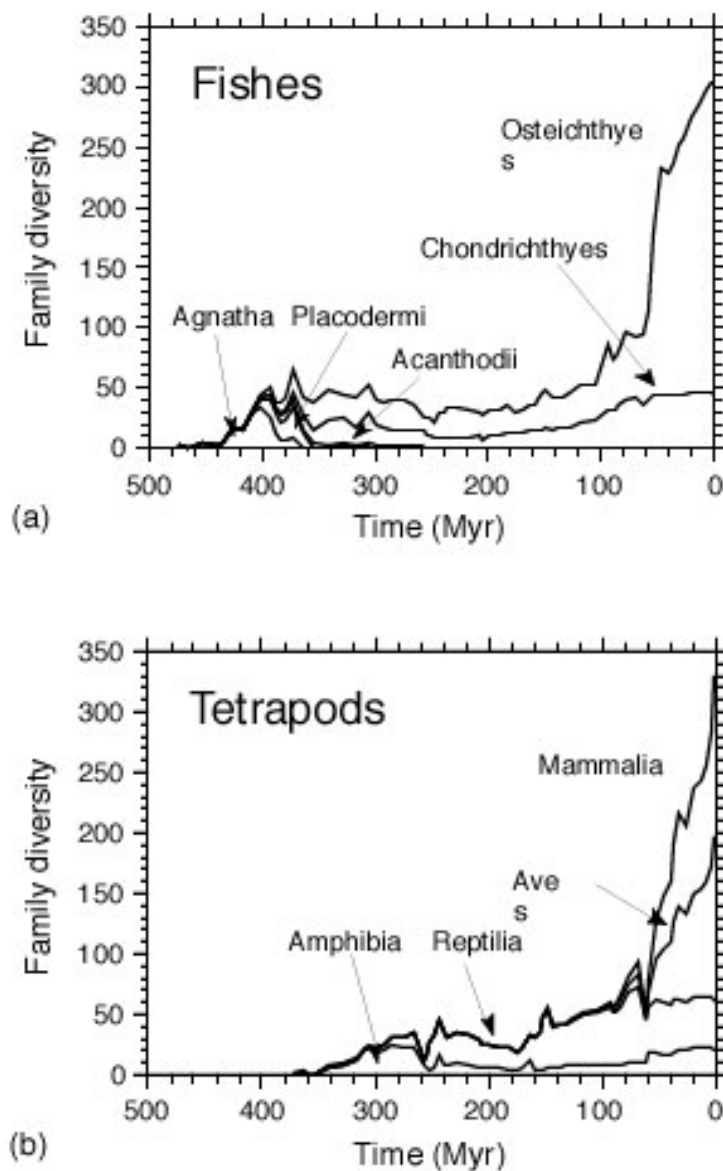


Figure 2. Patterns of the diversification of fishes (a) and tetrapods (b) through time, based on counts of numbers of families present during each geological stage (data from chapters in Benton, 1993). Major groups, some monophyletic, some paraphyletic, are shown, and the upper curve in each case is the sum of the family diversities of these groups. The effects of mass extinctions may be detected in the late Devonian (390 Ma), late Permian (250 Ma), late Triassic (225, 205 Ma), and end-Cretaceous (65 Ma). Abbreviations: Carb, Carboniferous; Cen, Cenozoic; Cret, Cretaceous; Dev, Devonian; Jur, Jurassic; Ord, Ordovician; P, Permian; S, Silurian; Tert, Tertiary; Tr, Triassic.

The diversification of tetrapods (Fig. 2b) does not show such a rapid early rate of increase, and there is a steady rise to a total of about 40 families by the end of the Palaeozoic. Diversity levels remain roughly constant at that level throughout most of the Triassic and Jurassic, but there is a steady rise from 50 to 70 families during the early Cretaceous, followed by a rapid increase during the late Cretaceous to 100 families, and a further accelerating rate of increase during the Cenozoic.

The Palaeozoic diversity record is dominated by 'amphibians' and these basal tetrapods declined in diversity dramatically during the Mesozoic, and finally disappeared in the Cretaceous. The post-Palaeozoic record of amphibians represents the clade Lissamphibia, and their diversity has risen slowly from Mesozoic levels of about 10 families to about twice that at the end of the Cenozoic. Note that the diversification pattern of 'Amphibia' focuses on a paraphyletic group in the Palaeozoic (that is, all tetrapods except amniotes) and a descendant clade, the Lissamphibia, after the Palaeozoic, so there is no need to seek special reasons for the double-peaked pattern of diversification of the 'Amphibia', as Carroll (1977) attempted. The pattern is an artefact of the conventional classification of the group and need not represent an unusual gap in the fossil record.

The diversification pattern for reptiles is similarly unrealistic, but the groups are represented in this way to indicate the common understanding of these terms. Through most of the Mesozoic at least, the patterns are meaningful, although the descendant clades, the birds and mammals, rise to prominence especially in the late Cretaceous. Reptile diversity increased marginally through the Mesozoic from 20-30 families in the Triassic to 50-60 in the late Cretaceous, mainly dinosaurs. Birds and mammals existed through much of the Mesozoic, but at low diversities, and both groups showed dramatically accelerating rates of diversification through the Cenozoic.

Changes in diversity may be tracked also by documenting origination and extinction rates. Here, the percentage rates are presented (Figs 3, 4), in other words the numbers of families arising or becoming extinct in a geological stage as a proportion of the numbers extant at the time. This measure of origination and extinction gives a measure of risk, but it is not normalized to geological time. This has not been done since the stratigraphic stages are largely of comparable length, typically 5-10 Myr, and the precise durations of many of these stages are not known with confidence.

Origination rates for fishes (Fig. 3) were particularly high in the late Cambrian (radiation of agnathans, including conodonts), during the Silurian, Devonian and Carboniferous, in the early Triassic (after the late Permian extinction), in the mid Cretaceous and Eocene (after the KT event). Tetrapods also show high rates of origination in the late Devonian and early Carboniferous, just after the late Permian extinction, and just after the KT event. The late Jurassic peaks may be real, or may reflect some sites of exceptionally good preservation.

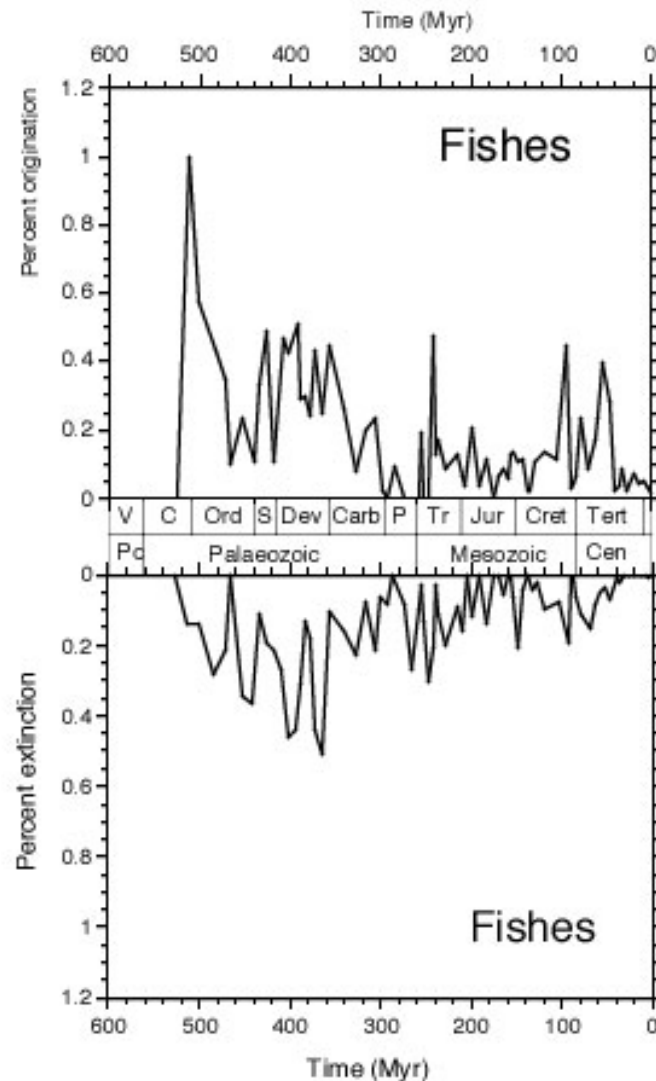


Figure 3. Percentage origination and extinction rates for fishes, calculated as numbers of originations/ extinctions per stage in proportion to total diversity at the time. Abbreviations: as for Fig. 2, and C, Cambrian; Pc, Precambrian; V, Vendian.

The fossil records of both fishes and tetrapods show the influences of mass extinctions (Figs 2-4). The late Devonian mass extinction affected fishes severely, with ostracoderm agnathans and placoderms virtually wiped out. The diversity of fishes, essentially chondrichthyans and sarcopterygians, fell roughly to half during the whole of the Permian, and the end-Permian mass extinction did not particularly affect them. The effects of the end-Permian event were much greater on tetrapods, with a reduction in amphibian diversity from about 20 families to five, and reptile diversity from 10 families to about five. Fish diversity was apparently little affected by Mesozoic events, showing some small reductions, but nothing profound, and seemingly very little at the KT boundary, 65 Myr ago. Tetrapods, on the other hand, were affected by a number of the Mesozoic extinction events, notably the two late Triassic events, and possibly an event at the end of the Jurassic, 150 Myr ago (although that drop is preceded by a sudden rise, possibly reflecting the exceptional preservation of the Solnhofen beds of southern Germany). The KT event was profound, marked by a drop in the diversity of families of tetrapods from over 90 to 60.

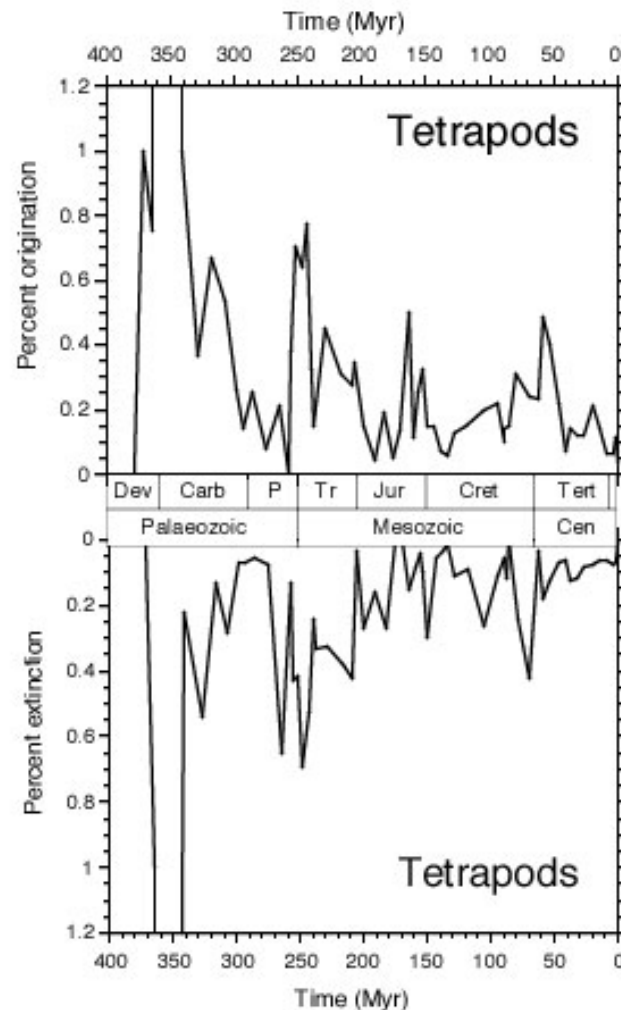


Figure 4. Percentage origination and extinction rates for tetrapods, calculated as numbers of originations/ extinctions per stage in proportion to total diversity at the time. Abbreviations: as for Figs 2.

There is some evidence for a coupling of patterns of origination and extinction (Figs 3, 4), as noted before for vertebrates (Benton, 1989), hence suggesting a possible 'Lagerstätten effect'. [If a fossil record is affected excessively by the influence of specific localities of exceptional preservation, the patterns of origination and extinction are often tightly coupled.] For fishes, apparent high rates of turnover (high origination and high extinction rates) occur in the mid Devonian and early Late Cretaceous (Fig. 3). However, the evidence for coupling is not strong, and more often high extinction rates are followed by high origination rates. The same appears to be largely true for tetrapods (Fig. 4), although there were coupled high rates of origination and extinction in the early Triassic and late Jurassic. The coupling seems to be less than it was in earlier analyses (Benton, 1985, 1989): perhaps our knowledge of the background fossil record is improving to the extent that Lagerstätten are no longer distorting the patterns unduly.

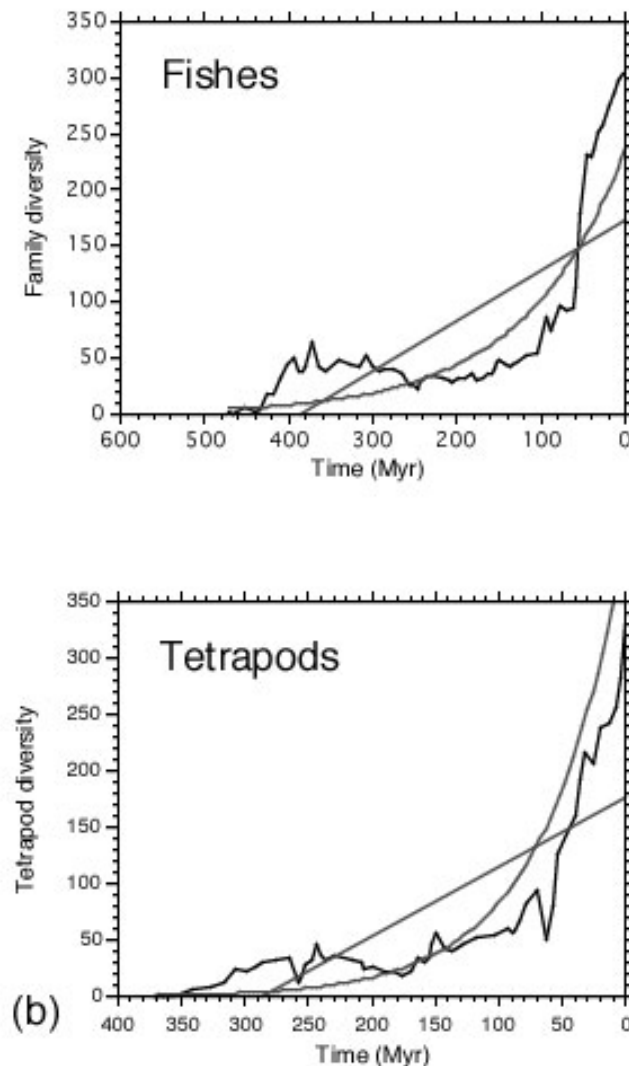


Figure 5. Patterns of the diversification of fishes (a) and tetrapods (b) through time, with best-fitting linear and exponential models. In both cases the exponential model fits best: for fishes, the linear model ($y = 173.7 - 0.45x$, $r = 0.719$) is a poorer match to the data than the exponential model ($y = 236.9 - 0.0085x$, $r = 0.925$); and for tetrapods, the linear model ($y = 177.0 - 0.62x$, $r = 0.779$) is also a poorer match to the data than the exponential model ($y = 409.5 - 0.0159x$, $r = 0.979$).

The overall diversification curves both appear to follow exponential patterns of increase, from relatively low numbers in the Palaeozoic, through slightly increasing diversities in the Jurassic and Cretaceous to dramatically accelerating diversifications from the Late Cretaceous onwards. It would be hard to interpret these curves as logistic, as Sepkoski (1984) has done for the diversification of families of marine animals. The diversity curves for both fishes and tetrapods show a good fit to an exponential curve (Fig. 5). The fit is markedly better than to a straight line (linear model), and the distribution of points prevents the calculation of any kind of meaningful logistic fit. Benton (1995a) found the same result for all continental organisms, including essentially tetrapods, insects, and land plants.

EXPONENTIAL OR EQUILIBRIAL SYSTEMS

The implications of exponential, rather than logistic, curves of increase in diversity are profound (Benton, 1997b, c; Fig. 6 herein). An exponential curve, allowing for the temporary reverses caused by extinction events, implies that there is no limit to the global diversity that can be achieved, or at least that vertebrates (or life on land, or all life) have yet to approach that maximum global carrying capacity for life. A logistic curve, or series of logistic curves, implies that there are global-scale caps to diversity, that the world reaches a stage where all ecospace is full, and this limiting level, or steady state, can be breached only by some major revolution (a mass extinction, the origin of a substantial new adaptive complex, a dramatic environmental shift). Which is the true interpretation of the history of life, or of vertebrates at least?

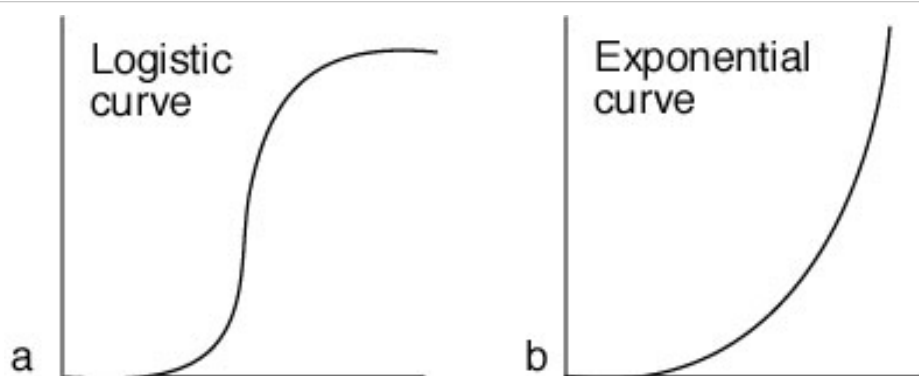


Figure 6. The basic shape of a logistic (a) and an exponential (b) curve.

Logistic models

The logistic model for the expansion of the diversity of life has developed from an influential body of ecological theory. MacArthur and Wilson (1967) presented their 'theory of island biogeography' as a simple means of estimating the rate of filling of an island, or other defined patch of living space. They showed how the rate of arrival of organisms is initially high, but that, as the space fills up, the rate of successful colonization diminishes, and the rate of local extinction increases. At some point, the two rates, colonization and extinction, stabilize, and this dynamic equilibrium represents the carrying capacity, or ideal diversity of life on the island. The carrying capacity is not fixed, but can be modified by major changes: for example, movements in the relative position of the island with respect to the nearest sources of species, changes in topography, climatic shifts, and the like.

Rosenzweig (1975, 1995), and others, have extended this island-scale model to regional and global scales, allowing essentially for intraclade competition, so that MacArthur and Wilson's island becomes the world, and their local rates of colonization and extinction become global rates of origination and extinction of species, or higher taxa. Time scales move from tens or hundreds of years to millions.

Sepkoski (1978, 1979, 1984, 1996) has developed an equivalent system of modelling, but in which he focuses on more complex intraspecific competition rather than intraclade competition. The end-result is similar, however, and he finds evidence for equilibrium levels in the Cambrian and in the post-Cambrian Palaeozoic, and these are interpreted as representing real steady states. The post-Palaeozoic record of diversification of marine animals, according to his analysis, shows no further equilibrium level, and hence the past 250

million years is interpreted as the rising part of a logistic curve. Courtillot and Gaudemer (1996), on the other hand, in a re-analysis of Benton's (1995a) data, confirmed a logistic model for the diversification of life, but they found evidence for only one equilibrium level in the Palaeozoic, and they believed they had identified the beginning of a slow-down in diversification rates in the late Cenozoic, and hence a hint that life is approaching a new equilibrium level.

The logistic model, and the implications of global steady states or equilibria, have dominated discussion in this area. Sepkoski's interpretation has been reproduced in many textbooks (for example, Allen and Briggs, 1989; Clarkson 1993; Skelton, 1993; Benton and Harper, 1997). However, some of the key assumptions behind such a view have been criticized (for example, Hoffman, 1985).

The critical issue with assumptions of equilibrium levels, or even less stable steady states, is the notion of a carrying capacity, or ideal number of species that can be accommodated on the Earth. Rieppel (1984), for example, showed that the idea of a global steady state for species is equivalent to the old 'principle of plenitude'. This was a pre-Darwinian idea, founded on the key assumption of natural theology, that God had created all organisms perfect and that he had fitted some plant or animal to each available task (we would say niche) in the economy of nature. There was no space left.

Darwin borrowed the principle of plenitude and made it explicitly evolutionary in his analogy of a barrel of apples. In his unpublished *Natural Selection* (see Stauffer, 1975, p. 208), Darwin compared the present-day diversity of species to a number of apples floating on the surface of a barrel filled with water. The surface is covered by exactly the right number of apples, and it is impossible to add a new apple without displacing one that is already there. Similarly, he argued, each species had been honed by evolution to fit its niche, and if a new species arises, it has to displace a pre-existing species before it can become established. Walker and Valentine (1984) questioned Darwin's assumption, that all niches are full. This has also typically been an assumption of regional- and global-scale equilibrium models, but it is unnecessary. These authors estimated that the mean proportion of empty niches ranged from 12 to 54% for eight marine invertebrate groups, and that species turnover could occur readily and rapidly, but without the need to assume that there is constant evolutionary pressure from competitively superior species.

Should we believe mathematical models? There is no question that a logistic model is a much better fit to the pattern of diversification of families of Palaeozoic marine animals than a straight-line, exponential or power-law curve (Sepkoski, 1979, 1984, 1996; Courtillot and Gaudemer, 1996). However, if the basic ecological and evolutionary assumptions behind such a logistic model are contradicted, then perhaps a fresh consideration is justified.

Exponential models

The question of mathematics vs. basic observations does not arise for vertebrates, since a logistic curve, or series of logistic curves, would be hard to fit to the known patterns of diversification of fishes or tetrapods (Figs 2, 5). The patterns as a whole suggest expansion, and especially rapid expansion over the past 50-100 Myr, the part of the fossil record that is probably better known and better dated than earlier segments.

Intuitive observations of the fossil record of vertebrates confirm the idea of continuing expansion based on evolutionary innovation. The bursts of radiation among fishes, for

example (Fig. 2a), may be related to specific new adaptations, such as extensive armour in ostracoderm agnathans and placoderms in the late Silurian and Devonian, jaws in placoderms and acanthodians in the Devonian, increased swimming speeds and efficiency of jaws in Jurassic bony fishes (and especially the teleosts), expansion of trophic levels in the sea (especially the modern sharks, Neoselachii), adaptations to feeding on specific foods (such as plankton, corals, buried bivalves and echinoderms), and parasitic modes.

A similar sequence of dramatic expansions of ecospace characterize the evolution of tetrapods; terrestrial adaptations among early tetrapods in the late Devonian and Carboniferous, insect-eating in Carboniferous tetrapods, the cleidoic (enclosed) egg in late Carboniferous amniotes, herbivory in certain early Permian amniotes, large size in some late Permian herbivorous mammal-like reptiles, fully upright posture in several amniote groups in the late Triassic, true flapping flight in late Triassic pterosaurs and late Jurassic birds, endothermy in Mesozoic mammals and birds, very large size in Jurassic and Cretaceous dinosaurs, new marine top-predator roles in the Mesozoic (ichthyosaurs, plesiosaurs, mosasaurs), burrowing and tree-climbing among some Mesozoic and Cenozoic amniote groups, further expansion of niches to include polar regions and nocturnal habits among Cenozoic mammals, and so on.

These additions of dietary modes and habitats have been a key feature of the evolution of tetrapods (Benton, 1990b), and presumably also of fishes, and indeed many other expanding clades. The tetrapods shifted from being essentially 100% fish-eaters and 100% the inhabitants of freshwaters and adjacent land areas in the late Devonian and early Carboniferous, to a much wider array of niches. Indeed, fish-eating and amphibious freshwater niches fell to a steady 10% of all families from the late Mesozoic onwards (Benton, 1990b). With a fine magnifying glass, perhaps one could argue that there were 'steady state' conditions worldwide between each of these adaptive bursts of radiation (Rosenzweig, 1995). However, the sum total of new adaptations that can be identified is larger than the lists just given (for example, colonial nesting in mole rats, feeding on garbage by various urbanizing mammals, ant-eating,...). How small do the 'steady states' have to become before they evaporate altogether into a picture of opportunistic expansion in diversification?

Can the expansion of diversity go on forever? Of course there is an ultimate limit to the numbers of families, or other taxa, that can inhabit the Earth at any time; such a limit would be caused not least by the amount of standing room on the Ark. Presumably, though, if a limit of living space were approached, ever smaller organisms might perhaps be favoured by evolution. Equally, as has happened so many times during evolution, organisms would take unexpected measures to survive by, for example, occupying the air, burrowing into sediments and, in the case of some bacteria, living deep within the Earth's crust. With size reduction, the ultimate limit to the diversification of life might then become the availability of the chemical components of life, principally carbon. Of course, other partial escape mechanisms from such a limiting factor is to speed up the rate of cycling of carbon through biogeochemical cycles, to retain such chemicals in the organic realm for longer, and to reduce the amount of time they remain buried.

Clade replacements

An expectation of the logistic model for diversification is that there is a phase of rapid increase in diversity, followed by a levelling-off phase as the gradient of the curve diminishes (Fig. 6a). This occurs as diversity approaches the equilibrium level, and the rate of increase diminishes progressively the closer the diversity approaches that level. Just as the rate of

diversification declines, so the phase of rapid expansion (filling of ecospace) switches to a prolonged phase of dynamic equilibrium; new taxa may arise, but they will tend to displace pre-existing taxa (Darwin's barrel of apples analogy). In a real case of the evolution of a major clade, one might expect to find a change in the nature of taxon originations, from expansionist to equilibrial, from taxa moving into unoccupied ecospace, to more and more competitive displacement.

Is there any evidence for such a switch from expansionist to competitive originations among tetrapods, say in the later Palaeozoic, or perhaps in the late Cenozoic, if Courtillot and Gaudemer (1996) were correct in suggesting that the diversification of 'all life' is entering the levelling-off phase of a logistic curve? Benton (1996a, b) has carried out a comprehensive census of all 840 families of tetrapods that have a fossil record, and which include more than a single species. He found that 13% of familial origins *could* be (but need not be) explained by competitive interaction with a pre-existing family on a reasonable estimate. Even using a maximal estimate of possible competitive interactions, where it was assumed that the fossil record was extraordinarily incomplete, the proportion of familial originations that were candidate competitive replacements (CCRs) rose to 26%. In a plot of CCRs (Fig. 7), the distributions of peaks appears to be largely random with respect to time. There is no evidence for a rise in CCRs in the late Palaeozoic or the late Cenozoic, and hence no suggestion that tetrapod niches were filling up at either of those times. This is evidence against a logistic model of diversity increase.

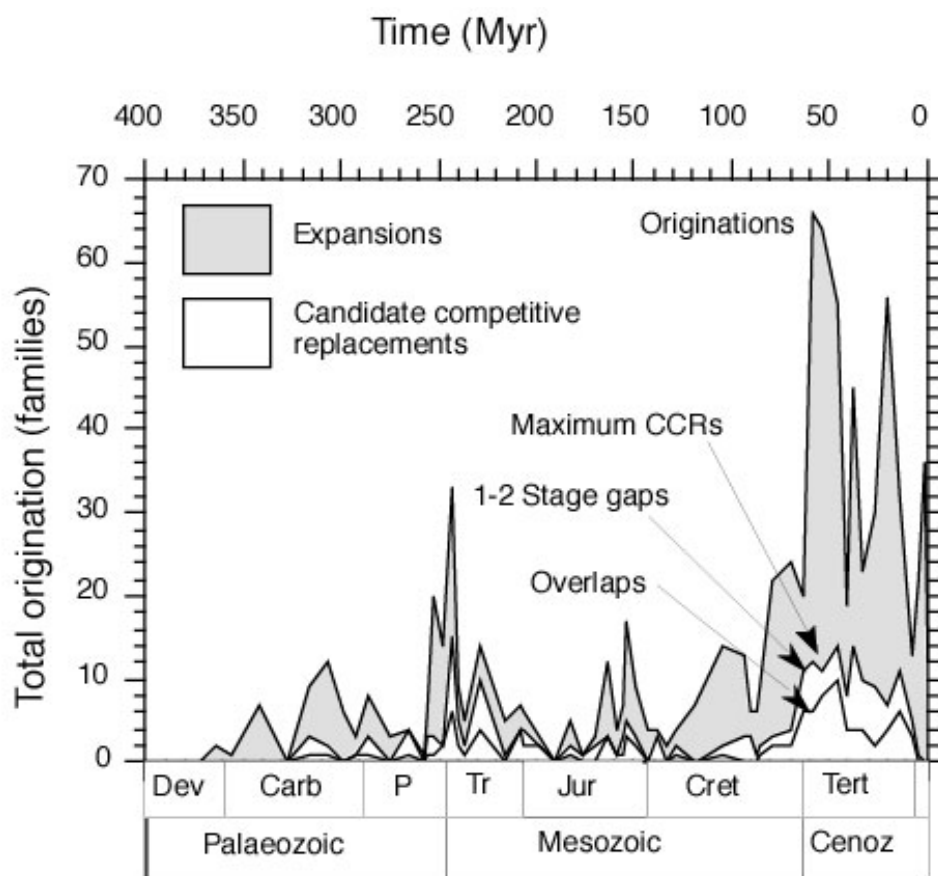


Figure 7. Distribution in time of originations of tetrapods of all habitats, showing also the occurrences of overlaps and of maximum candidate competitive replacements (CCRs). CCRs were identified by comparison of pairs of families. First, stratigraphic range charts were plotted for each combination of body size, diet, and habitat, and maximum geographic ranges of each family were

noted. Then, the point of origin of each of the 840 families was scrutinised to determine whether it was a CCR or an expansion. CCR cases were subdivided into overlaps, where the stratigraphic range of the family overlapped another family, and situations where the family apparently originated at the precise time of extinction of another (gap 0), or after a gap of one or two (gap 1, 2) stratigraphic stages. Overlaps give evidence that the two families could have encountered each other, while the gap 0, 1, and 2 cases allow for possible incompleteness of the fossil record. CCRs are plotted as overlaps and maximum CCRs (the sum of all overlaps and gap 0, 1, and 2 cases). Abbreviations: Carb, Carboniferous; Cenoz, Cenozoic; Cret, Cretaceous; Dev, Devonian; Jur, Jurassic; P, Permian; Tert, Tertiary; Tr, Triassic. Based on data in Benton (1996a, b).

QUALITY OF THE FOSSIL RECORD OF VERTEBRATES

The quality of the fossil record, or of some segment of it, may be assessed in a qualitative way, and assertions may be made based upon field experience, or upon overall surveys of a group. These intuitive approaches have been discussed above, especially with regard to the fossil record of vertebrates. These observations, however, are often hard to justify, especially to non-palaeontologists, and they may give misleading evidence: cases were set out above that the fossil record of vertebrates is either good (stability in our understanding; no real surprises) or poor (incompleteness of sedimentary record; rarity of fossils). Ultimately, of course, a decisive test on the quality of any fossil record cannot be applied, since no mortal can know what really happened in the past. Without a yardstick of the truth, it is clearly impossible to test assertions about the fossil record once and for all. However, three quantitative approaches may shed some light on completeness.

Taphonomic tests of the completeness of the fossil record

The first quantitative approach to assessing the completeness of the fossil record is to compare like with like through time. For example, it is a fair assumption that deposits containing exceptionally preserved fossils (Lagerstätten) from the Cambrian are not necessarily worse than equivalent sedimentary settings of Jurassic or Eocene age. Lagerstätten of specific types may then be treated as comparable snapshots of the true diversity of life at particular times and in specific environments/ regions, since they include soft parts and entirely soft-bodied organisms that are otherwise missed. They may then act as standards against which other, more typical, deposits (those that do not preserve soft parts and soft-bodied organisms) may be compared. This may be a fruitful approach (Allison and Briggs, 1993; Chapter 3 herein) and it can provide a semi-quantitative assessment of how the quality of 'normal' fossil deposits has varied through time.

Comparing like with like need not stop at Lagerstätten. Certain other kinds of fossil accumulations may be treated as equivalent and unaffected by time-related destructive phenomena (see also Chapters 1, 5, 11 herein). For example, coquinas, or winnowed accumulations of fossil shells, may survive in equally unmetamorphosed and uncrushed condition from the early Palaeozoic and the late Cenozoic. Kidwell and Brenchley (1996) found no diminution of the quality of preservation backwards in time when they compared large samples of Ordovician-Silurian, Jurassic, and Neogene coquinas. They did find other time-related trends, some of them actually making the Mesozoic fossil record poorer than the Palaeozoic, for example a dramatic increase in the diversity and effectiveness of predatory organisms that crush shells, and increases in the diversity of organisms that burrow and bore through sea-bed sediments. Further, in line with the evolution of new and ever-more fiendish

groups of shell-crushers and shell-borers, most shells of the potential prey became thicker, and hence the younger coquinas are themselves thicker (more and thicker shells are introduced into the shell beds), but they are also probably more time-averaged (that is, representing a longer time period of accumulation).

Historical tests of the completeness of the fossil record

Historical approaches may also be helpful. The 'collector curve' (Fig. 8) is a useful approach that is often used in ecology, and it may be used in various palaeontological studies. This technique was devised as a time-saving approach to help field ecologists decide when to stop collecting specimens. A primary requirement of an ecological survey is to produce a list of all the species in an area. At first, the collecting goes well, and most specimens that are picked up represent additions to the list. In time, the 'hit' rate declines, and it becomes harder and harder to find a new species that has not already been identified. The collector curve simply quantifies the collecting effort (assessed as number of specimens picked up, or time) against the identifications of new species. When the rate declines markedly, it is assumed that the collector is approaching the true total diversity, and he then decides to stop collecting at the 90% or 95% level: the effort required to find the very last species in an area might well equal the effort expended in finding all the others.

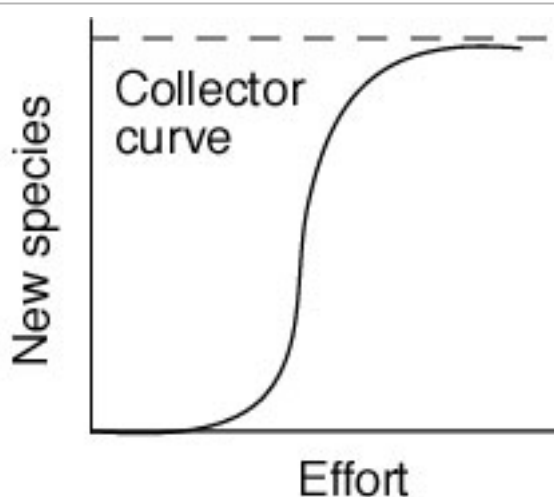


Figure 8. The basic collector curve is a way of estimating when to stop collecting. At first, the 'effort' devoted to collecting is richly rewarded with discoveries of new species, but as the collector approaches the maximum possible total (dashed line), the effort expended to find a new species increases dramatically.

Palaeontologists can readily use such an approach when collecting fossils from a new locality. This also provides a technique for standardizing collecting effort when a palaeontologist is attempting to draw up range charts and assessments of the relative abundance of particular groups through a section. The collector curve may also provide an approach to assessing the quality of the fossil record.

Maxwell and Benton (1990) suggested that, given enough time, all the fossils that are out there in the rocks will be collected. This sample of fossils in the rocks does not, of course, represent all of the life of the past, since multitudes of species must have come and gone, and yet never have been fossilized. Nonetheless, those organisms that were fossilized are potentially knowable. Of course, like the ecologist setting out to compile a faunal list on a new tropical island, the palaeontologist has to exert ever more effort to find those last rare specimens (see Chapter 8).

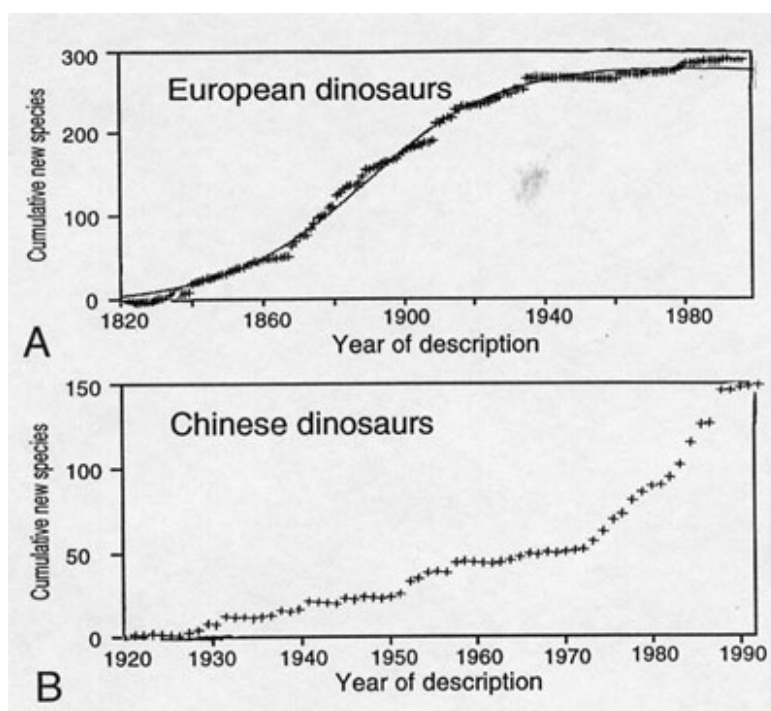


Figure 9. Collector curves and discovery of dinosaurs in Europe (a) and China (b). In Europe, the rate of recovery of new dinosaurs has been fairly static since the 1920s, and perhaps all the species that are in the rocks to be found have been found. In China, on the other hand, collecting began 100 years later, in the 1920s, and the rate of determination of new species increased in the 1970s, with an equivocal flattening in the 1990s.

Large-scale palaeontological collector curves may be compiled in various ways. One simple approach is to use years of study as the measure of effort (since it would be hard to count accurately the numbers of specimens palaeontologists have inspected, and perhaps discarded), and then to count the rate of accretion of new species. One caveat is that the rate of accretion of new species can be distorted horribly by an enthusiastic splitter (a taxonomist who names species based on small differences).

One example is a preliminary study of the rate of discovery of new dinosaurs (Fig. 9). Here, the effort scale runs from 1824, the date of publication of the first formal dinosaur name, *Megalosaurus*, by William Buckland, up to 1990. The 'new species' axis represents 'net new species', that is, actual new species named, minus those new species that were later synonymised. If the literal number of new species had been plotted, the final total diversity would have been much higher, but any interpretations would then have been based partly on fantasy. Admittedly, there is no guarantee that later synonymies were always justified, nor that all currently accepted dinosaur names are actually valid. In addition, the synonymy rate falls off for newer dinosaur names since these have not all gone through the normal processes of reassessment by other workers.

Dinosaurs from two broad regions are shown, those from Europe (Fig. 9a), and those from China (Fig. 9b), approximately equivalent areas of the Earth's surface, and thus potentially roughly equally likely to yield dinosaurs of different kinds, and potentially likely to yield comparable total numbers of dinosaur species. Intuitively, palaeontologists would expect Europe to be a 'mature' region in terms of the discovery of new species of dinosaurs, since the first finds were made there, and there is a record of nearly 200 years of collection in generally

populous areas that are actively combed by large numbers of collectors and academics. China, on the other hand, is a mere juvenile, since the first dinosaurs were identified there only in the 1920s, much of the area is remote, and there have been far fewer collectors and palaeontologists in China than in Europe. Indeed, the intuitive expectation is borne out: the European collector curve shows a fully developed logistic pattern, and the rate of determination of new species of dinosaurs has been levelling off since the 1920s. In China, on the other hand, the long slow phase of accretion, from the 1920s to the 1960s, has been followed by the accelerating phase since the 1970s, and there is only an equivocal hint of a levelling-off in the 1990s. These figures are also rather more immature than the European totals, since more revision and synonymy of the Chinese material may occur in the future. Up to 1990, about 220 species of dinosaurs had been identified from Europe, and 160 species from China. Can Chinese palaeontologists expect to push their totals to 220, or even higher? Until the logistic curve definitely begins to bend over, no-one can tell!

The historical approach was used, in a slightly different form, in Maxwell and Benton's (1990) assessment of the fossil record of tetrapods, and in Sepkoski's (1993) assessment of the fossil record of marine animals. In the former case, the historical time span under investigation ran from 1900 to 1987, and especially 1966 to 1987. In the latter case, the time span was 1982 to 1992. The premise in both cases was the same, that many changes had occurred in our knowledge of the fossil record (new specimens, revised stratigraphy, revised taxonomy), but had these had a significant effect on palaeontological knowledge as a whole? In both studies, it was shown that the changes were randomly distributed with respect to time and major clades: in other words, many changes, but the overall pattern stays the same. Specific changes were that overall diversity increased, and that extinction events became sharper (as new fossils were found that filled gaps). The fossil record is, then, good enough to read empirically, as a valid indicator of the true history of life. But how much change has there been?

Maxwell and Benton (1990) hoped that palaeontological knowledge was improving. They, and Sepkoski (1993), were only able to show, however, that palaeontological knowledge was changing. Perhaps all the new fossils, new stratigraphies, and new taxonomies were actually making things worse. How could it be shown that the changes in knowledge were tending in the right direction, towards a full knowledge of the fossil record? This required a different approach, the use of some external yardstick, and luckily such a yardstick exists.

Phylogenetic tests of the completeness of the fossil record

The fossil record may be tested against cladistic and molecular phylogenies, since phylogenies are independent of stratigraphy. Cladograms are constructed by the search for patterns in the distribution of characters among fossil or living organisms (Platnick, 1979), and there is no test of geological age in the assessment of a character or a taxon. Likewise, molecular phylogenies are based on comparisons of sequence or general similarity data, and stratigraphy is not involved. This means that the order and distribution of fossils in the rocks may be compared with morphological cladograms or molecular phylogenies, and these independent sources of data on the true phylogeny of life may be assessed for congruence (Benton, 1994, 1995b; Benton and Hitchin, 1996, 1997).

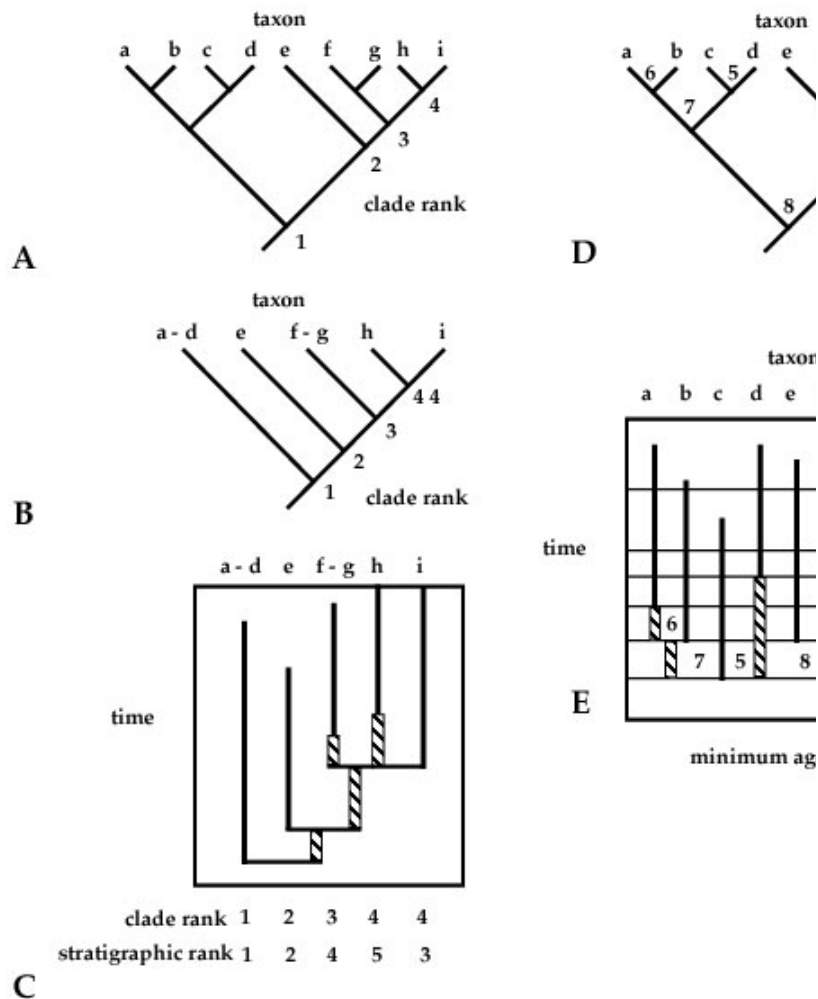


Figure 10. Techniques for assessing the quality of the fossil record. Comparisons are made between branching order in cladograms and stratigraphic data (A-E), and between the relative amount of gap and the known record (E). The example is a cladogram with nine terminal branches (A-I). For comparisons of clade order and age order, cladistic rank is determined by counting the sequence of primary nodes in a cladogram (A): nodes are numbered from one (basal node) upwards to the ultimate node. In cases of non-pectinate cladograms (A), the cladogram is reduced to pectinate form (B), and groups of taxa that meet the main axis at the same point are combined and treated as a single unit. The stratigraphic sequence of clade appearance is assessed from the earliest known fossil representative of sister groups, and clade rank and stratigraphic rank may then be compared (C). Matching of clade rank and stratigraphic rank may be tested by Spearman rank correlation (SRC). SRC coefficients may range from 1.0 (perfect correlation) through 0 (no correlation) to -1.0 (perfect negative correlation). For assessing the proportion of ghost range, or minimum implied gap (MIG), and known stratigraphic range, the whole cladogram is used (E). MIG (diagonal rule) is the difference between the age of the first representative of a lineage and that of its sister, as oldest known fossils of sister groups are rarely of the same age. The proportion of MIG to known range is assessed using the relative completeness index (RCI), according to the formula:

$$\Sigma (MIG)$$

$$RCI = (1 -) \times 100\%.$$

$$\Sigma (SRL)$$

RCI values may range from 100% (no ghost range) through 0 (ghost range = known range) to high negative values (ghost range >> known range). Stratigraphic consistency is assessed (D, E) as a comparison of the ratio of nodes that are younger than, or of equal age to, the node

immediately below (consistent), compared to those that are apparently older (inconsistent). The stratigraphic consistency index (SCI) is assessed on the full cladogram (D, E). SCI values range from 1.0 (all nodes stratigraphically consistent) to 0 (no nodes stratigraphically consistent). Based on data in Benton and Hitchin (1997).

The assessment metrics. There are a variety of metrics for comparing phylogenies and fossil records (Fig. 10); Spearman rank correlation (SRC), the relative completeness index (RCI), and the stratigraphic consistency index (SCI). SRC is an established nonparametric statistical test, and it has been used in comparing the order of fossils in the rocks with the implied order of appearance of groups based on the sequence of nodes (branching points) in a cladogram. The first applications of the SRC test for this purpose were by Gauthier *et al.* (1988) and Norell and Novacek (1992a, b).

The RCI was proposed (Benton, 1994; Benton and Storrs, 1994) as an additional metric that took account of the actual time spans between branching points, and of implied gaps before the oldest-known fossils of lineages. Sister groups, by definition, originated from an immediate common ancestor, and diverged from that ancestor. Thus, both sister groups should have fossil records that start at essentially the same time. In reality, usually the oldest fossil of one lineage will be older than the oldest fossil of its sister lineage. The time gap between these two oldest fossils is the 'ghost range' or minimal cladistically-implied gap. The RCI assesses the ratio of ghost range to known range, and high values imply that ghost ranges are short, and hence that the fossil record is good.

[It has been suggested (Paul, 1992; Wagner, 1995) that ghost ranges may be an artefact of the cladistic technique, which assumes that sister taxa generally originate at the same time (by dichotomous branching). If one sister includes ancestors of the other, however, and branching occurs after the origin of one of the sister taxa, the ghost range might disappear, and hence the RCI technique would be invalid. It is not clear whether the ancestor model applies to a majority of cases or not. Certainly, the criticism of ghost ranges is not valid for groups analysed at low taxonomic levels where the fossil record is patchy, since the taxa available for analysis are necessarily only a sample of all those that ever existed, and the chances of hitting on a true ancestor are small. This is probably the case for the vertebrates, echinoderms, arthropods, and other groups for which cladograms are abundant, and on which we have based our tests. On the other hand, many fossil molluscs and foraminifera may be ancestors of other known forms, and the RCI technique would not perhaps work for them. If most sister taxa split at a single point of origin, as asserted by cladists, then the technique works.]

The SCI was proposed by Huelsenbeck (1994) to test how well the nodes in cladograms corresponded to the known fossil record. Nodes are dated by the oldest known fossils of either sister group subtended from the node. Each node is compared with the node immediately below it. If the upper node is younger than, or equal in age to, the node below, the node is said to be stratigraphically consistent. If the node below is younger, the upper node is stratigraphically inconsistent. The SCI for a cladogram compares the ratio of the sums of stratigraphically consistent to inconsistent nodes. SCI values can indicate cladograms whose nodes are all in line with stratigraphic expectations through to cladograms that imply a sequence of events that is entirely opposite to the known fossil record.

These metrics may be applied to individual phylogenetic problems - that is, which of these ten cladograms of marsupial relationships provides the best fit to current stratigraphic evidence? - or they may be used to assess large samples of cladograms. In the cases to be presented here,

the latter approach is used. The assumption here is that variations in cladistic or molecular techniques, variations in taxic level, and stratigraphic variations are subsumed in the overall variation within a large sample. So far, the sample of cladograms assessed in these ways amounts to 384, composed of 174 cladograms of tetrapods, 147 cladograms of fishes, and 63 cladograms of echinoderms (Benton and Hitchin, 1996, 1997; see also <http://palaeo.gly.bris.ac.uk/palaeo/cladestrat.html>).

Testing fossil record quality: branching order. The key question to be answered by this regime of testing is whether the fossil record is good enough. First results were encouraging: Norell and Novacek (1992a) found that 18 out of 24 test cases of family-level and generic-level cladograms of vertebrates (75%) gave statistically significant ($P < 0.05$) correlations of clade and age data, using the SRC test. In larger samples, Norell and Novacek (1992b) found significant correlation in 24 of 33 test cases (73%), while Benton and Storrs (1994) found significant correlation in 41 of 74 test cases (55%). In other words, for tetrapods, there was apparently good agreement between stratigraphic and cladistic evidence in most cases.

Subsequent assessments, however, based on larger samples of cladograms, provided more disappointing results. For echinoderms, Benton and Hitchin (1996) found that only 24 out of 63 cladograms (36%) show statistically significant ($P < 0.05$) matching of clade rank and age rank data. For fishes, the figure is 37 out of 147 cladograms (25%), and for tetrapods, 87 out of 174 cladograms (50%). The results for all cladograms in the test sample is that 148 out of 384 showed significant SRC values (38%). Does this mean that only a minority, something from one-quarter to one-half, of cladograms are congruent with fossil records, that the fossil record is equally likely to give the wrong order of fossils as it is to give the correct order?

Two points may be made. The comparisons by Benton and Hitchin (1996) were made on large samples of cladograms, and these included some categories of cladogram that performed badly in the SRC test (Benton and Hitchin, 1997):

- (1) Cladograms based on species or genera. Taxonomically low-level cladograms generally gave very poor matches to stratigraphy, largely because the fossil records were not adequate. Often only one or two out of ten or fifteen genera had any fossil representatives.
- (2) Rejected cladograms. Our sample of cladograms was comprehensive, and we included a number of cladograms that had been published as possible, but not preferred, solutions. Many of these performed badly.
- (3) Cladograms with small numbers of terminal taxa. When there are only four or five terminal taxa, the SCI in particular has only two or three nodes to compare with the basal node. The sequence has to be perfect in order to achieve a good result, and one mismatch gives a very low value. For the SRC test also, cladograms with fewer than five or six terminal taxa must have a perfect match of the sequences of clade and age order before they produce a statistically significant coefficient. For both the SCI and SRC metrics, acceptable scores may be achieved by larger cladograms even if there are some mismatches.

Testing fossil record quality: RCI and SCI metrics. The tests of the quality of the fossil records of echinoderms, fishes and tetrapods using the SRC statistic were disappointing. However, much better results were obtained with the RCI and the SCI metrics (Benton and Hitchin, 1996, 1997). It may seem unusual that many cladograms that apparently failed the SRC test of matching between age and clade data should pass with another metric. The reason is that the RCI and the SCI measure different aspects of cladogram and fossil record quality, and perhaps

the SRC test is too indiscriminate for many purposes. The SRC simply compares raw orders of fossils and branching points. It takes no account of the overall amounts of time involved, nor especially of the seriousness of a mismatch. We found that many cladograms failed the SRC test because they had many nodes packed within a narrow time band. Some of the nodes were out of sequence by only 1-2 Myr or less, which is an insignificant amount of time in most analyses.

For all three groups assessed, most cladograms have RCI values equal to, or greater than, 0.5, than values less than 0.5 (Fig. 11). The pass rates are 49 out of 63 cladograms (78%) for echinoderms, 124 out of 147 cladograms (84%) for fishes, and 128 out of 174 cladograms (74%) for tetrapods (Fig. 12). The pass rate for all cladograms was 78%. In other words, 301 of the 384 cladograms tested have more than twice as much of their ranges represented by fossils than represented by ghost range. The differences in mean values of RCI for echinoderms (mean, 62.3%) and fishes (mean, 69.4%) are modest, but continental tetrapods have a much lower value (mean, 49.8%).

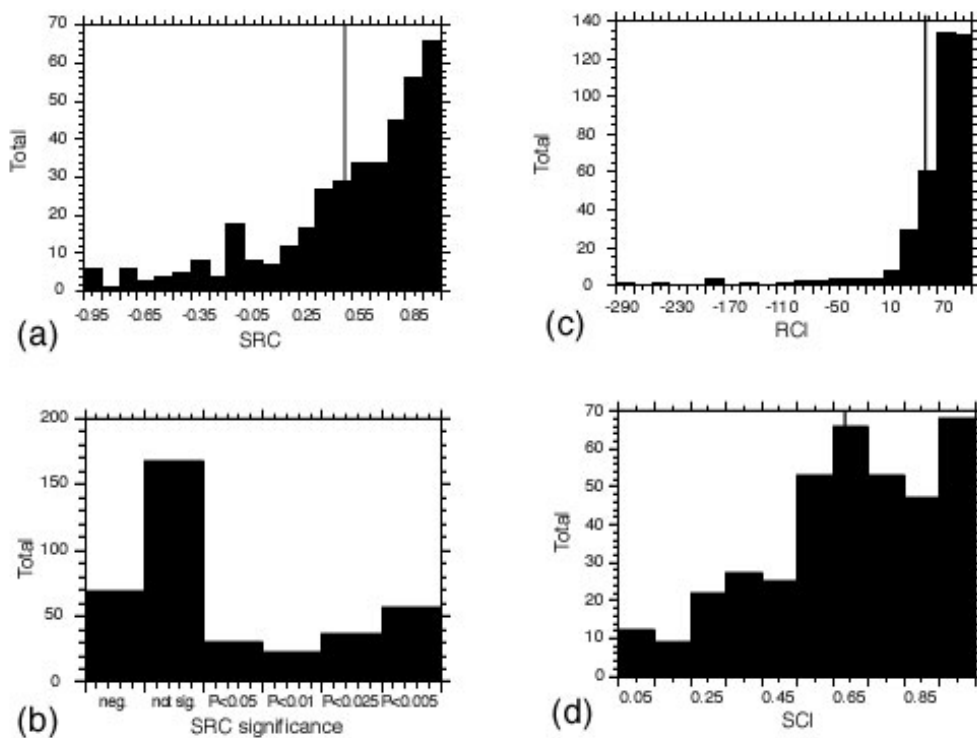


Figure 11. Assessments of congruence between stratigraphic and cladistic data show highly skewed distributions. Values for three metrics calculated on a sample of 384 cladograms of echinoderms, fishes, and tetrapods: Spearman rank correlation (SRC) coefficients (a), measures of the significance of those SRC coefficients, which take account of cladogram size (b), relative completeness index (RCI) values (c), and stratigraphic consistency index (SCI) values (d). Mean values for each sample are indicated by dotted lines. Based on data in Benton and Hitchin (1997).

The pass rates are similarly favourable for the SCI measure (Fig. 11). In these cases, all three sets of cladograms have significantly more than half their nodes showing stratigraphic consistency than inconsistency. The pass rates are 60 out of 63 cladograms (95%) for echinoderms, 102 out of 147 cladograms (69%) for fishes, and 152 out of 174 cladograms (87%) for tetrapods (Fig. 12). The pass rate for all cladograms is 82%, based on 314 of the 384 cladograms (the SCI metric could not be calculated for 70 small cladograms in the full sample). The reasons for significantly higher SCI values for echinoderms (mean, 0.78) than fishes (mean, 0.55) and tetrapods (mean, 0.66) are not immediately evident.

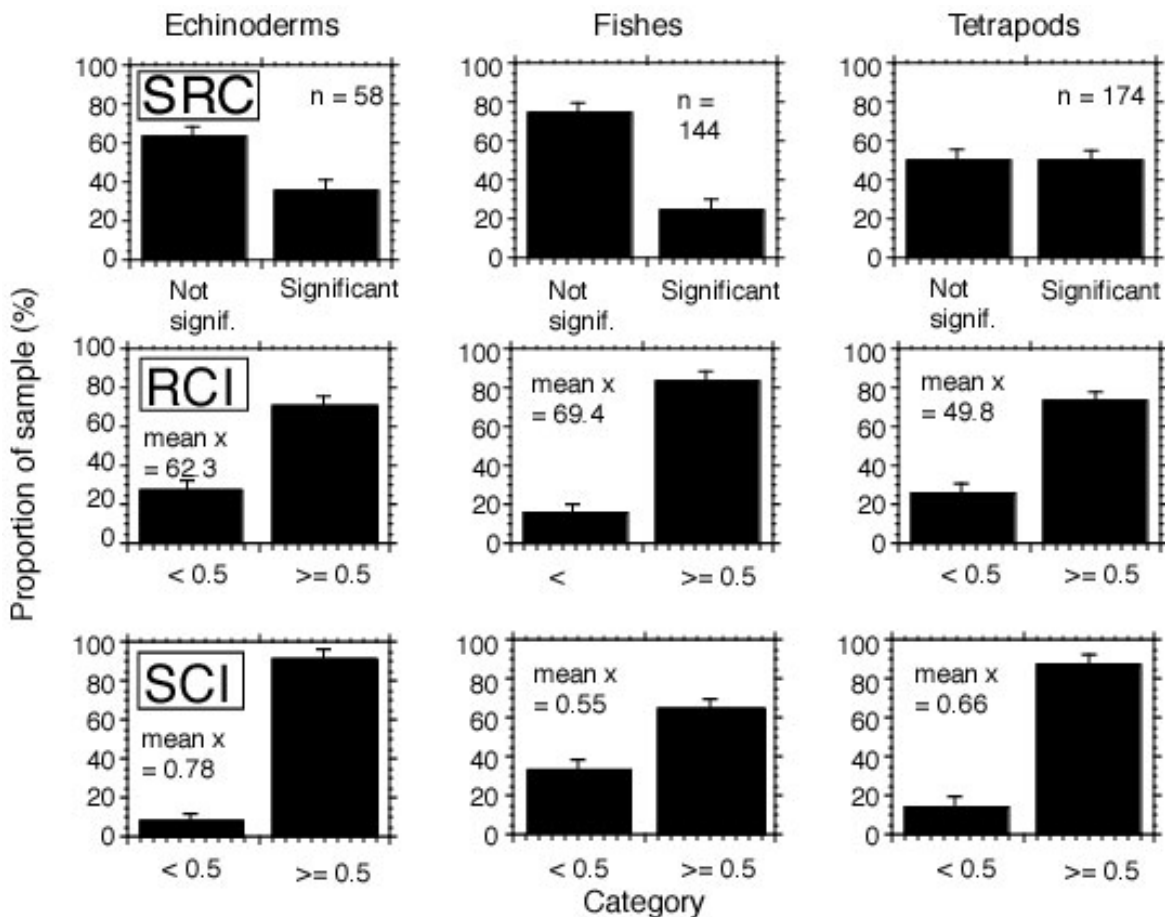


Figure 12. Summary of the metrics for comparison of cladogram data and stratigraphic age data. Metrics indicated are Spearman Rank Correlation (SRC) of age and clade data, the Relative Completeness Index (RCI), based on comparisons of known and implied stratigraphic ranges, and the Stratigraphic Consistency Index (SCI) of nodes in cladograms. The metrics have been applied to large samples of cladograms (n, number of cladograms in sample) for echinoderms, fishes, and tetrapods. Comparisons are between significant and non-significant SRC coefficients, and between frequencies of values of the RCI above and below 50%, and frequencies of values of the SCI above and below 0.5. The differences in values among the three groups are significant, based on comparison of the binomial error bars. Based on data in Benton and Hitchin (1996).

The results of the RCI and SCI metrics show that fossil records are on the whole good for echinoderms, fishes and tetrapods. Comparisons among the three groups shows that none of them consistently has a better fossil record, or better cladistic resolution, than the others. Each of the animal groups under study could be said to have the best fossil record since each

is supported by one of the three tests: tetrapods by SRC, fishes by RCI and echinoderms by SCI (Fig. 12). Only one group comes out worst on two of the tests: fishes have the poorest showing according to the SRC and SCI metrics. Tetrapods have the worst fossil records according to the RCI metric, while echinoderms are not worst of the three animal groups according to any of the metrics.

Comparing continental and marine habitats. Benton and Simms (1995) obtained some counter-intuitive results when they showed that continental tetrapods have a fossil record that is as good as, or better than, that of echinoderms, based on comparisons of results obtained with the SRC and RCI metrics. This shocking result could not have been predicted from observations of the field occurrence of both groups: tetrapods are found in sporadic and unpredictable sedimentary settings, while echinoderm remains are hugely abundant in many marine shelf deposits.

A more detailed comparison of SRC, RCI and SCI metrics for all marine cladograms and all continental cladograms yielded mixed results (Benton and Hitchin, 1996; Fig. 13). The SRC test showed that 87 out of 174 continental cladograms (50%) had significant matching of age and clade order, while the value for marine groups was only 61 out of 210 cladograms (29%), much worse than the results for echinoderms alone reported by Benton and Simms (1995). The pass rate for RCI values was much more comparable, with 173 of the 210 marine cladograms yielding values higher than 50% (82%), compared to 128 of 174 continental cladograms (74%). Mean values confirmed that marine cladograms show a lower proportion of implied gaps (mean RCI, 67.3%) than do continental cladograms (mean RCI, 49.8%). The pass rate for SCI values, on the other hand, favoured the continental cladograms, where 152 of 174 cladograms had values equal to, or better than, 0.500 (87%), compared to 162 of the 210 marine cladograms (77%). Mean values suggested that continental cladograms (mean SCI, 0.66) perform slightly better than marine cladograms (mean SCI, 0.62) in the SCI test.

The finding that continental vertebrates have a fossil record of similar quality to marine echinoderms and fishes suggests two observations. (1) The relative abundance of specimens at individual fossil localities is no indicator of the completeness of their fossil record on a large scale: this depends on the number of stratigraphic horizons that have yielded fossils, and on the packing of those horizons in time. (2) The fossil record of continental tetrapods has probably been more intensively studied than has that of echinoderms, and indeed fishes. Hence, current knowledge of the tetrapod fossil record is now higher on the collector curve (numbers of taxa vs. effort), and may be assumed to approach closer to the level of complete sampling and full knowledge of all fossil taxa that exist in the rocks.

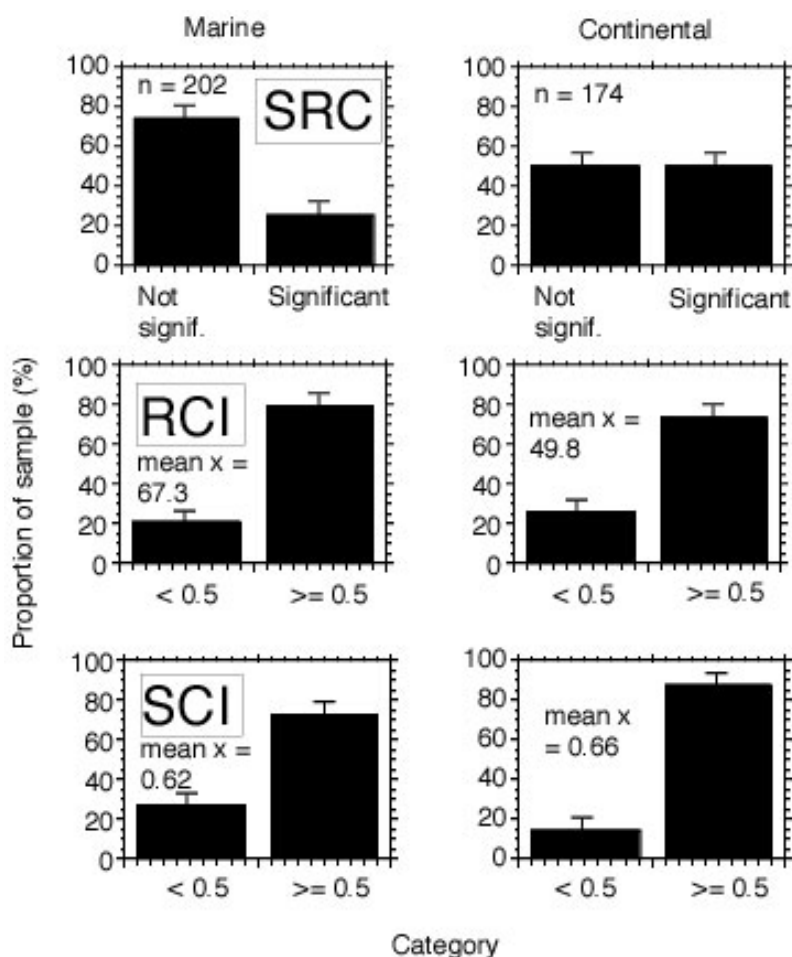


Figure 13. Summary of the metrics for comparison of cladogram data and stratigraphic age data. Comparisons of cladograms of marine (echinoderm + fish) and continental (tetrapod) cladograms according to the SRC, RCI and SCI metrics. The differences in values among the three groups are significant, based on comparison of the binomial error bars. Abbreviations as in Fig. 12. Based on data in Benton and Hitchin (1996).

Comparing changes in knowledge of the fossil record. The SRC and RCI metrics have been used to compare historical aspects of the understanding of the fossil record. It might be expected that the addition of new fossil finds and reanalysis of older ones would improve the fit of age data to a fixed sample of cladograms, by the filling of gaps, and corrections of former taxonomic assignments. However, in a comparison of a 1967 data set (Harland *et al.*, 1967) and one from 1993 (Benton, 1993), Benton and Storrs (1994, 1996) found no change at all in the proportions of cladograms that showed statistically significant ($P < 0.05$ and $P < 0.01$) matching of clade and age order, although there had been a change in the status of 28 of the 71 cladograms compared (39%; Fig. 14 herein). In other words, as a result of 26 years of work, new discoveries and reassignments had improved the fit in 20% of cases, but had caused mismatches of clade and age data in a further 20% of cases. Sometimes, a new fossil does not fill a gap, but creates additional gaps on other branches of a cladogram.

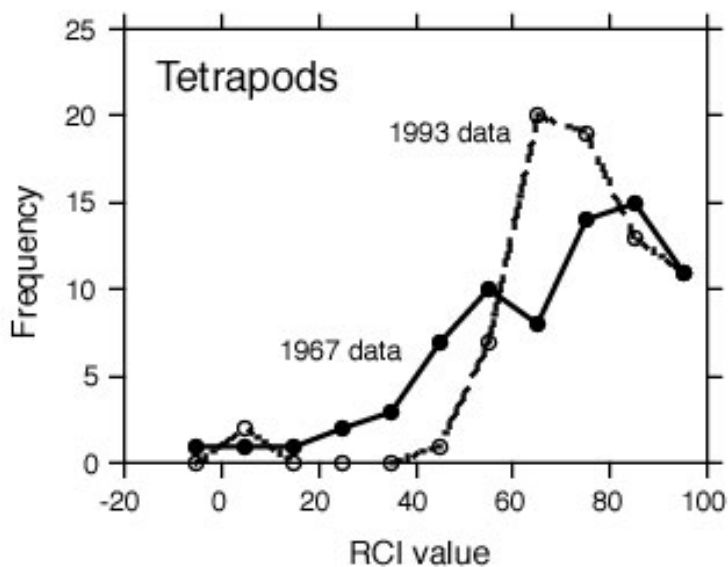


Figure 14. Relative improvement in fossil record quality from 1967 (Harland *et al.*, 1967) to 1993 (Benton, 1993). During these 26 years, gaps in the record were filled, and there is a clear shift in the distribution of RCI (relative completeness index) values to the right from 1967 to 1993, indicating improvement in palaeontological knowledge (significant shift at $P < 0.05$; t-test and non-parametric signs and Wilcoxon signed ranks tests). Based on data in Benton and Storrs (1994).

This discovery of a lack of improvement in the congruence of clade vs. age rank order is important, since it highlights the fact that mismatches may arise from subtle changes in knowledge. Non-correlation may result from minor variations in fossil dating, and may not imply wildly different evidence about the history of life from cladograms and from fossil occurrences.

The RCI metric, however, detected a significant improvement in knowledge of the tetrapod fossil record from 1967 to 1993. In their study of cladograms of vertebrates, Benton and Storrs (1994, 1996) found that the mean RCI value shifted from 67.9% to 72.3%, a statistically significant difference, according to a Wilcoxon signed ranks test ($P = 0.026$). In other words, comparisons of the relative completeness of cladograms shows a significant *improvement*, by about 5%, in knowledge of the fossil record over the past 26 years of research. Hence, new fossil discoveries, and reassignments of older ones, do positively affect the amount of ghost range, although such changes in knowledge did not apparently affect the match of clade and age rank order, as assessed by the SRC test.

CONCLUSIONS

The adequacy of the fossil record is hard to assess, as the various contributions in this book indicate. In a strict sense, it will never be possible to assess the adequacy of any segment of the fossil record, since the true picture will be forever unknown. Perhaps there have been whole phyla, or even kingdoms, of extraordinary organisms that lived at different times in the past, but which have left no fossil indications. One could imagine whole tribes of giant purple worms with bodies 100 metres long, squirming around on Carboniferous forest floors, or an entirely unknown kingdom of photosynthesising organisms that lived in Cambrian seas and moved by means of floppy wheels made from protoplasm. Such organisms are not impossible, but they are unlikely. The unlikelihood increases day-by-day as ever more palaeontological effort fails to turn up any hint of such unknown major groups of macroscopic organisms.

There are a variety of powerful new techniques for assessing the quality of the fossil record. One approach is to adopt a uniformitarian approach to specific kinds of fossil-preservation sites, and to compare like with like across vast spans of geological time. The kind of intuitive

argument presented in the previous paragraph represents a second approach, where the pattern of discovery over research time is investigated. It is statistically valid to quantify effort against discovery rate, and to assert that the longer some unknown organism remains undiscovered the less likelihood there is of its former existence. (This statement assumes that there is a premium attached to finding such an unknown organism, and this is certainly very much the case.)

The third new approach, comparing phylogenetic and stratigraphic evidence to assessing the adequacy of the fossil record, has been applied with considerable success to vertebrates. This approach is based on the observation that phylogenies (cladograms founded on morphological characters and molecular phylogenies) are constructed independently of geological evidence. The order and timing of splitting events in phylogenies may then be cross-compared with stratigraphic evidence on the order and timing of the appearance of groups in the fossil record in order to assess the degree of congruence. Good matching of the data sets implies that both the fossil record and the phylogeny are probably good, while a mismatch implies either a misleading fossil record or an inaccurate phylogenetic hypothesis.

The phylogenetic congruence assessments have indicated that there is no evidence that vertebrates have a fossil record that is either any better, or indeed any worse, than that of any other major group of animals. In addition, there is no evidence that the record of continental (that is, terrestrial and freshwater) tetrapods is worse than that of marine echinoderms or fishes. These kinds of assessments have proved highly fruitful, and they provide a sound quantified answer to the old cry of 'the fossil record is pretty incomplete and uninformative'.

ACKNOWLEDGEMENTS

I thank Gilles Cuny, Steve Donovan, Becky Hitchin, Chris Paul and David Unwin for comments on this MS, and the Leverhulme Trust (Grant F182/AK) for funding.

REFERENCES

- Allen, K. C. and Briggs, D. E. G. 1989. *Evolution and the fossil record*, Belhaven Press, London, 265 pp.
- Allison, P. A. and Briggs, D. E. G. 1991. *Taphonomy; Releasing the data locked in the fossil record*, Plenum, New York, 560 pp.
- Allison, P. A. and Briggs, D. E. G. 1993. Exceptional fossil record: distribution of soft-tissue preservation through the Phanerozoic, *Geology*, 21, 527-530.
- Behrensmeyer, A. K. and Hill, A. P. 1980. *Fossils in the making: Vertebrate taphonomy and paleoecology*, University of Chicago Press, Chicago, 338 pp.
- Benton, M. J. 1983. Progressionism in the 1850s: Lyell, Owen, Mantell and the Elgin fossil reptile *Leptopleuron* (*Telerpeton*), *Archives of Natural History*, 11, 123-136.
- Benton, M. J. 1985. Mass extinction among non-marine tetrapods, *Nature*, 316, 811-814.
- Benton, M. J. 1989. Patterns of evolution and extinction in vertebrates, in *Evolution and the Fossil Record*, (Ed. K. C. Allen and D. E. G. Briggs), Belhaven, London, pp. 218-241.

- Benton, M. J. 1990a. The causes of the diversification of life, in *Major Evolutionary Radiations*, (Ed. P. D. Taylor and G. P. Larwood), Systematics Association Special Volume, 42, Clarendon Press, Oxford, pp. 409-430.
- Benton, M. J. 1990b. *Vertebrate palaeontology*, Unwin Hyman, London, 377 pp.
- Benton, M. J. 1993. *The Fossil Record 2*, Chapman & Hall, London, 839 pp.
- Benton, M. J. 1994. Palaeontological data, and identifying mass extinctions, *Trends in Ecology and Evolution*, 9, 181-185.
- Benton, M. J. 1995a. Diversification and extinction in the history of life, *Science*, 268, 52-58.
- Benton, M. J. 1995b. Testing the time axis of phylogenies, *Philosophical Transactions of the Royal Society, Series B*, 348, 5-10.
- Benton, M. J. 1996a. On the nonprevalence of competitive replacement in the evolution of tetrapods, in *Evolutionary paleobiology*, (Ed. D. Jablonski, D. H. Erwin, and J. H. Lipps), University of Chicago Press, Chicago, pp. 185-210.
- Benton, M. J. 1996b. Testing the roles of competition and expansion in tetrapod evolution, *Proceedings of the Royal Society of London, Series B*, 263, 641-646.
- Benton, M. J. 1997a. *Vertebrate palaeontology* (2nd edition), Chapman & Hall, London, 452 pp.
- Benton, M. J. 1997b. The history of life: large data bases in palaeontology, in *Statistical methods in palaeobiology*, (Ed. D. A. T. Harper), Wiley, London, in press.
- Benton, M. J. 1997c. Models for the diversification of life. *Trends in Ecology and Evolution*, in press.
- Benton, M. J. and Harper, D. A. T. 1997. *Basic palaeontology*, Addison-Wesley, London, 325 pp.
- Benton, M. J. and Hitchin, R. 1996. Testing the quality of the fossil record by groups and by major habitats, *Historical Biology*, 12, 111-157.
- Benton, M. J. and Hitchin, R. 1997. Congruence between phylogenetic and stratigraphic data on the history of life, *Proceedings of the Royal Society, London, Series B*, in press,
- Benton, M. J. and Simms, M. J. 1995. Testing the marine and continental fossil records, *Geology*, 23, 601-604.
- Benton, M. J. and Storrs, G. W. 1994. Testing the quality of the fossil record: paleontological knowledge is improving, *Geology*, 22, 111-114.
- Benton, M. J. and Storrs, G. W. 1996. Diversity in the past: comparing cladistic phylogenies and stratigraphy, in *Aspects of the genesis and maintenance of biological diversity*, (Ed. M. E. Hochberg, J. Clobert, and R. Barbault), Oxford University Press, Oxford, pp. 19-40.
- Briggs, D. E. G. and Crowther, P. R. 1990. *Palaeobiology: a synthesis*, Blackwell Scientific, Oxford, 583 pp.

- Carroll, R. L. 1977. Patterns of amphibian evolution: an extended example of the incompleteness of the fossil record, in *Patterns of evolution as illustrated by the fossil record*, (Ed. A. Hallam), Elsevier, Amsterdam, pp. 405-437.
- Carroll, R. L. 1987. *Vertebrate paleontology and evolution*, W. H. Freeman, San Francisco, 698 pp.
- Chatterjee, S. 1995. The Triassic bird *Protoavis. Archaeopteryx*, 13, 15-31.
- Clarkson, E. N. K. 1993. *Invertebrate palaeontology and evolution* (3rd edition), Chapman & Hall, London, 434 pp.
- Courtillot, V. and Gaudemer, Y. 1996. Effects of mass extinctions on biodiversity, *Nature*, 381, 146-148.
- Cowen, R. 1990. *History of life*, Blackwell Scientific, Boston, 470 pp.
- Donovan, S. K. 1991. *The processes of fossilization*, Belhaven Press, London, 303 pp.
- Flessa, K. W. 1990. The "facts" of mass extinctions, *Geological Society of America, Special Paper*, 247, 1-7.
- Fürsich, F. T. 1990. Fossil concentrations and life and death assemblages, in *Palaeobiology; a synthesis*, (Ed. D. E. G. Briggs and P. R. Crowther), Blackwell Scientific, Oxford, pp. 235-239.
- Gauthier, J., Kluge, A. G., and Rowe, T. 1988. Amniote phylogeny and the importance of fossils, *Cladistics*, 4, 105-209.
- Harland, W. B., Holland, C. H., House, M. R., Hughes, N. F., Reynolds, A. B., Rudwick, M. J. S., Satterthwaite, G. E., Tarlo, L. B. H., and Willey, E. C. 1967. *The fossil record; a symposium with documentation*, Geological Society of London, London, 827 pp.
- Hoffman, A. 1985. Island biogeography and palaeobiology: in search for evolutionary equilibria, *Biological Reviews*, 60, 455-471.
- Huelsenbeck, J. P. 1994. Comparing the stratigraphic record to estimates of phylogeny, *Paleobiology*, 20, 470-483.
- Jablonski, D. 1991. Extinctions: a paleontological perspective, *Science*, 253, 754-757.
- Kidwell, S. M. 1986. Models of fossil concentrations: paleobiologic implications, *Paleobiology*, 12, 6-24.
- Kidwell, S. M. and Brenchley, P. J. 1996. Evolution of the fossil record: thickness trends in marine skeletal accumulations and their implications, in *Evolutionary palaeobiology*, (Ed. D. Jablonski, D. H. Erwin, and J. H. Lipps), University of Chicago Press, Chicago, pp. 290-336.
- MacArthur, R. H. and Wilson, E. O. 1967. *The theory of island biogeography*, Princeton University Press, Princeton, 203 pp.

- Maxwell, W. D. and Benton, M. J. 1990. Historical tests of the absolute completeness of the fossil record of tetrapods, *Paleobiology*, 16, 322-335.
- Norell, M. A. and Novacek, M. J. 1992a. The fossil record and evolution: comparing cladistic and paleontologic evidence for vertebrate history, *Science*, 255, 1690-1693.
- Norell, M. A. and Novacek, M. J. 1992b. Congruence between superpositional and phylogenetic patterns: comparing cladistic patterns with fossil records, *Cladistics*, 8, 319-337.
- Paul, C. R. C. 1992. The recognition of ancestors. *Historical Biology*, 6, 239-250.
- Platnick, N. I. 1979. Philosophy and the transformation of cladistics, *Systematic Zoology*, 28, 537-546.
- Raup, D. M. 1979. Biases in the fossil record of species and genera, *Bulletin of the Carnegie Museum of Natural History*, 13, 85-91.
- Retallack, G. 1984. Completeness of the rock and fossil record: some estimates using fossil soils, *Paleobiology*, 10, 59-78.
- Rieppel, O. 1984. The problem of extinction, *Zeitschrift für Zoologischer Systematik und Evolutionsforschung*, 22, 81-85.
- Romer, A. S. 1966. *Vertebrate paleontology* (3rd edition), University of Chicago Press, Chicago, 468 pp.
- Rosenzweig, M. L. 1975. On continental steady states of species diversity, in *The ecology of species communities*, (Ed. M. L. Cody and J. M. Diamond), Belknap Press of Harvard University Press, Cambridge, Mass., pp. 121-140.
- Rosenzweig, M. L. 1995. *Species diversity in space and time*, Cambridge University Press, Cambridge, 436 pp.
- Sadler, P. M. 1981. Sediment accumulation rates and the completeness of stratigraphic sections, *Journal of Geology*, 89, 569-584.
- Sansom, I. J., Smith, M. M., and Smith, M. P. 1996. Scales of thelodont and shark-like fishes from the Ordovician of Colorado, *Nature*, 379, 628-630.
- 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. 1984. A kinetic model of Phanerozoic taxonomic diversity. III. Post-Paleozoic families and mass extinctions, *Paleobiology*, 10, 246-267.
- Sepkoski, J. J., Jr. 1993. Ten years in the library: how changes in taxonomic data bases affect perception of macroevolutionary pattern, *Paleobiology*, 19, 43-51.

Sepkoski, J. J., Jr. 1996. Competition in macroevolution: the double wedge revisited, in *Evolutionary paleobiology*, (Ed. D. Jablonski, D. H. Erwin, and J. H. Lipps), University of Chicago Press, Chicago, pp. 211-255.

Skelton, P. 1993. *Evolution. A biological and palaeontological approach*, Addison-Wesley, Wokingham, 1064 pp.

Stauffer, R. C. 1975. *Charles Darwin's natural selection. Being the second part of his big species book written from 1856-1858.*, Cambridge University Press, New York, 694 pp.

Valentine, J. W. 1969. Patterns of taxonomic and ecological structure of the shelf benthos during Phanerozoic time, *Palaeontology*, 12, 684-709.

Wagner, P. J. 1995. Stratigraphic tests of cladistic hypotheses. *Paleobiology*, 21, 153-178.

Walker, T. D. and Valentine, J. W. 1984. Equilibrium models of evolutionary species diversity and the number of empty niches, *American Naturalist*, 124, 887-899.