

Chapter 4

The quality of the fossil record

Michael J. Benton

ABSTRACT

Ever since the days of Charles Darwin, palaeontologists have been concerned about the quality of the fossil record. New concerns have arisen from two themes: (1) the finding that molecular dates of origin of certain major clades are often twice as old as the oldest fossils, and (2) the discovery that much of the variation in diversity, origination, and extinction signals from the fossil record can be explained by sampling. The molecular age-doubling phenomenon may be real, or it could be explained by either major gaps in the fossil record or by the inability of molecular techniques to discount unequivocally the possibility of rapid clock rates during times of divergence. The rock record certainly controls much of the fine detail of diversity and extinction plots, but mass extinctions, and the overall rise in diversity through time, may be real. Comparison of molecular and morphological phylogenies with the order and spacing of events in the rock record shows congruence, and hence suggests that much of the biotic signal in the fossil record is not misleading.

Introduction

The quality of the fossil record is a focal issue in current debates about the timing of origins of major groups. Some molecular estimates place the origins of Metazoa (animal phyla), green plants, angiosperms, and modern orders of birds and mammals at points up to twice as old as the oldest representative fossils (e.g. Hedges *et al.* 1996; Wray *et al.* 1996; Cooper and Penny 1997; Kumar and Hedges 1998; Heckman *et al.* 2001; Nei *et al.* 2001; van Tuinen and Hedges 2001; Wray 2001). The range of molecular estimates for the origin of metazoans is 600–1200 Ma (million years ago), with most estimates closer to 1000 Ma than 600 Ma. The range of molecular estimates for the origin and basal splitting of placental mammals, and of modern birds, is 130–70 Ma, again with more estimates nearer 120 Ma than 70 Ma. The first fossils date, respectively, from around 600 and 70 Ma.

The finding that molecules tend to give estimates of origin twice as old as the fossils in these five cases (animals, green plants, angiosperms, birds, mammals) almost seems like a rule, and perhaps a rule that will be found in other cases. Bleiweiss (1999) has already alluded to this when he linked the molecular age-doubling phenomenon in these broad-scale examples to an identical finding for species of birds in the Quaternary (Klicka and Zink 1997). The interesting question to be explored in

the next ten years is whether molecular age-doubling is an issue of method or of reality.

Palaeontologists have long debated the quality of the fossil record. Since the time of Darwin, and before, a serious theme has been 'the incompleteness of the fossil record', and authors have repeatedly emphasized the obvious fact that only a tiny fraction of those organisms that have ever lived are preserved as fossils, and only a tiny fraction of those fossils will ever become objects of scientific scrutiny (Raup 1972; reviewed in Donovan and Paul 1998). Currently, palaeontologists fall into two camps, those who are content that the fossil record is adequate to show the broad outlines of the history of life (e.g. Sepkoski *et al.* 1981; Benton 1995, 1999a,b; Foote 1997; Miller 1998; Benton *et al.* 2000), and those who believe that sampling problems overwhelm the signal in rocks older than perhaps 20 or 30 Ma (e.g. Alroy *et al.* 2001; Smith 2001; Peters and Foote 2001, 2002). For example, Smith (2001; Smith and Peterson 2002) and Peters and Foote (2002) have demonstrated that diversity signals from the fossil record vary with the amount of exposed rock and with sea level change, and hence may not contain much of an original biological signal.

The purpose of this chapter is to consider approaches to estimating sampling in the fossil record, and whether they can assess quality. The apparent mismatch of fossil and molecular dates for the radiation of major clades is explored and, for some groups at least, a rapprochement appears to have taken place. The proposition that the fossil record contains more of a sampling than a biotic signal will be considered. Finally, it will be suggested that the fossil record is adequate, and robust in the face of sampling problems, at certain scales, based on semi-independent phylogenetic investigations.

Molecular age doubling and error

If the molecular age-doubling phenomenon is real, there are two possible explanations for the apparent mismatch between molecular and fossil dates for the origin of major clades. One is based on the assumption that the fossil record is adequate, the other that it is not. If it is assumed that the fossil record is good enough, then the first half of the history of many (most?) major clades has evidently been cryptic. The organisms remained small, soft-bodied, or restricted geographically for a long span of time, before they finally flourished and became detectable as fossils. This is the 'phylogenetic fuse' idea of Cooper and Fortey (1998). A simpler assumption is that the fossil record is inadequate (early fossils too small or delicate to be preserved, appropriate rocks absent or in parts of the world that have not been sampled). This has been the general view of many molecular analysts (e.g. Hedges *et al.* 1996; Wray *et al.* 1996; Cooper and Penny 1997; Kumar and Hedges 1998; Eastal 1999; Heckman *et al.* 2001; Nei *et al.* 2001; van Tuinen and Hedges 2001; Wray 2001; Hedges, Chapter 2).

If the molecular age-doubling phenomenon is questionable, then one has to consider the methods of assigning dates in molecular phylogenetic analysis. Fossil and molecular evidence are both subject to error. Where fossils are correctly identified, palaeontological evidence will always underestimate the maximum age of a clade, whereas molecular evidence can both over- and underestimate ages. Several authors (e.g. Vermeij 1996; Benton 1999b; Foote *et al.* 1999; Lee 1999; Conway Morris 2000) have stressed the propensity of molecular methods to overestimate the timing of

origin of major clades since they do not fully take account of the possibility that molecular rates speed up enormously during times of major diversification ('adaptive radiation'). Extrapolating with a constant-rates model over a time of enhanced rates means that the point of origin is projected too far back in time.

Current molecular clock techniques take account of rate variation across the tree (e.g. maximum likelihood techniques, the quartet method, and the use of multiple calibration points), but it is not clear that they can yet assess the validity of molecular age doubling at the radiation of major clades.

(1) Maximum likelihood techniques (Cavalli-Sforza and Edwards 1967; Huelsenbeck and Rannala 1997; Whelan *et al.* 2001) may be used to calculate differing rates of evolution in extant lineages, but there are many available models, and arbitrary choices among possible models have to be made (Siddall and Whiting 1999). Differing rates between lineages within a clade may be detected, but a decisive test between a model that posits explosive diversification of all lineages in a clade at one time, and a model that does not include such a dramatic change of rate, cannot be made (Huelsenbeck and Rannala 1997).

(2) The quartet method (Rambaut and Bromham 1998) compares subsets of four taxa extracted from a tree. Those quartets that show significant rate variation between the two pairs of taxa are rejected. Surviving quartets then provide numerous estimates of the date of a common basal node. The method can then allow analysts to calculate the amount of error associated with such a basal date, but it cannot take account of a situation where molecular rates were *all* uniformly faster during a time of explosive diversification (Smith and Peterson 2002).

(3) Multiple calibration points allowed Springer (1997) to detect changes in rates of molecular evolution across the tree of placental mammals. If the confirmed dates, based on fossil data, are scattered densely enough across a molecular phylogeny, they can provide constraints on other, undated branching points. However, extrapolating from multiple dates high in a tree to determine dates of branching low in a tree still does not address the possibility of a uniform explosive rate of evolution early in a diversification event. The difficulty in establishing calibration points that tightly bracket a clade radiation is illustrated by Paton *et al.* (2002). To date the origin of modern birds, they use one distant low date, the split of the bird and crocodile line in the Triassic (245 Ma), one distant high date, the split of the emu and cassowary (35 Ma, based on the oldest emu fossil of 25 Ma), and one 'close' date, the Galloanserae divergence (85 Ma). These three dates give a range of estimates, from 108–155 Ma, for the radiation of modern birds. Each of the three reference dates can be criticized, the first for being too distant, the second for being based on an arbitrary addition of 10 Ma to a known fossil date, and the third for being itself a molecular estimate that might involve similar error to the date being assessed.

Hence, it is frustratingly clear that none of these approaches can test unequivocally whether or not certain past diversification events were marked by rates of molecular change that speeded up dramatically for a short time across a whole clade. Comparison with well-sampled outgroups might supply a yardstick for comparison, but when such living outgroups are phylogenetically distant, as for example with birds and mammals, such comparisons may be impossible.

Sampling methods

Two main approaches have been used by palaeontologists in assessing the completeness of their fossil records: confidence intervals and group sampling. Both methods may suffer from circularity in reasoning – if the input sampling distributions are incomplete, the estimates too will be incomplete.

Confidence intervals

Estimation of confidence intervals is an intuitive approach. It is based on the assumption that, if fossils are known from many geological horizons within a known stratigraphic range, then it is likely that very much older (or younger) fossils will not be found. Potential range extensions, at a particular probability level, will be small. If, on the other hand, fossils are scattered sparsely through a known range, it is likely that unknown fossils may occur far below the known oldest fossil (or far above the known youngest fossil, for an extinct group). The method of gap analysis is a statistical expression of this intuitive assumption, where the probability, P , is the confidence level (say 0.95) that a hypothetical interval added to a known range will include the true stratigraphic range,

$$P = 1 - (1 + a)^{-(n-1)}$$

where a is the confidence interval expressed as a fraction of the observed stratigraphic range, and n is the number of known fossiliferous horizons (Strauss and Sadler 1989). The method was developed to deal with local rock sections, but it may be applied to global examples of this kind (Marshall 1990), providing that the distribution of known fossiliferous horizons within the overall range is random and independent. If the distribution of potentially fossiliferous rocks is not random and independent, then the appropriate statistical tests, generalized confidence intervals (Marshall 1997), must be applied, but these are statistically much less powerful.

In an example of this approach, Bleiweiss (1998) looked at the fossil records of three bird groups, the Strigiformes (owls), Caprimulgiformes (goatsuckers), and Apodiformes (swifts, hummingbirds), and documented all known fossils in each order. His purpose was to compare fossil and molecular evidence for the origin of those orders. The oldest fossils are dated at 58–54.5 Ma for each of the groups, definitively within the Tertiary, whereas molecular estimates (Hedges *et al.* 1996; Cooper and Penny 1997) placed modern bird ordinal origins at 80–100 Ma, well down in the Cretaceous. Bleiweiss (1998) found that fossils in each of the orders have been reported from some 20–30 separate horizons from the date of these oldest fossils to the present day (Figure 4.1). He tested for randomness, and calculated that the maximum possible range extensions, based on the known fossil records of the owls, goatsuckers, and swifts, would hardly even take these orders into the latest Cretaceous, let alone the mid-Cretaceous. The 95 per cent confidence intervals estimated for the base of each of the three ranges were 62 Ma for swifts, 67 Ma for goatsuckers, and 63 Ma for owls. When all three groups were combined, producing a more densely sampled composite record, the range extension, at 95 per cent confidence, was back to only 61 Ma.

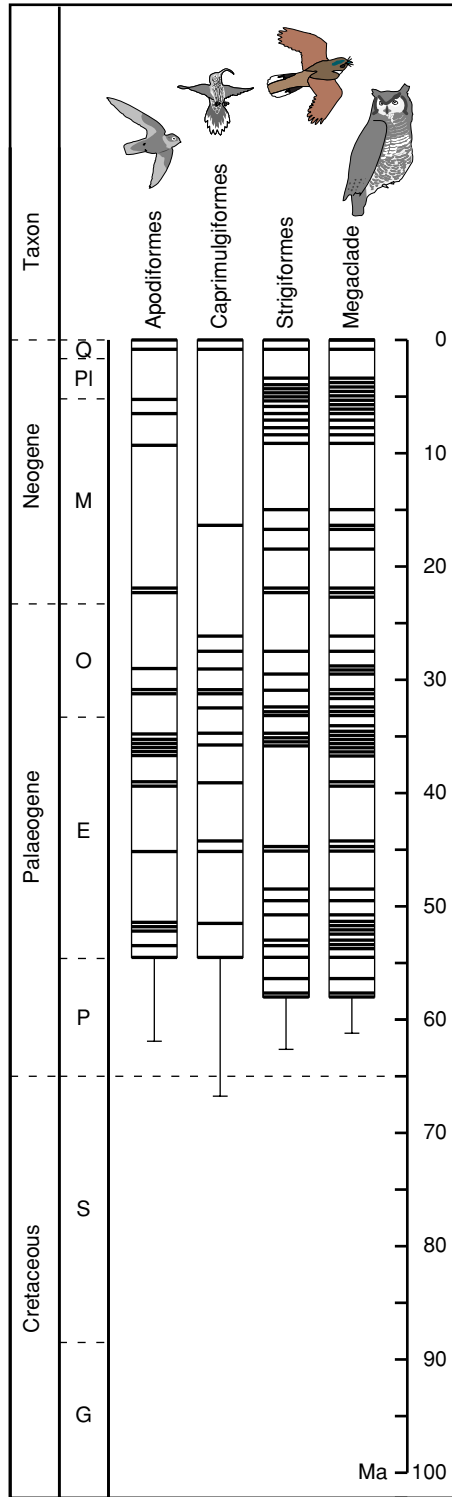


Figure 4.1 The known fossil record of modern bird groups gives no indication that fossil finds can be expected in the Cretaceous. Based on the distribution in time of known fossils of three modern bird orders, the apodiforms (swifts and hummingbirds), caprimulgiforms (goatsuckers) and strigiforms (owls), confidence intervals (95 per cent) are confined mainly to the Palaeogene (early Tertiary), with just one barely creeping into the latest Cretaceous. The 'megacalade' consists of all three orders summed together. The levels of known finds are indicated by horizontal bars (which may represent one find or many dozens of the same age). The more closely spaced the known finds, the shorter the confidence interval. Abbreviations: G, Gallia; S, Senonian; P, Palaeocene; E, Eocene; O, Oligocene; M, Miocene; Pl, Pliocene; Q, Quaternary. Based on data in Bleiweiss (1998).

Corroborating evidence (Bleiweiss 1999) comes from morphological studies that suggest rapid divergence of modern bird orders. Earliest Tertiary forms may be assigned to modern orders, but they generally have only one or two synapomorphies of those orders (Dyke 2001), and the full suite of distinguishing characters of the modern orders was acquired piecemeal through the early Tertiary. Indeed some of the earliest modern bird fossils have traditionally been very hard to assign to orders (Feduccia 1996). Had the orders been divergent for 40 myr or so before the first fossils occur, they might be expected to have accumulated much more character difference by the beginning of the Tertiary.

In a response to the study by Bleiweiss (1998), Marshall (1999) suggested that the confidence intervals method falls down if the global abundance and diversity of a group varies through time. Hence, he suggests, if the three bird orders in question had a long early history when they were rare, then the confidence intervals would expand. Using his generalized confidence interval approach, Marshall (1999) found that the 95 per cent confidence intervals estimated by Bleiweiss as 4–10 myr grew to 40 myr. This insight, that early parts of clades are less likely to produce fossils than later parts of clades, provided the basis for a model used by Tavaré *et al.* (2002) to estimate true points of origin for the primates from the fossil record. Using a logistic diversification model, and taking information from the modern diversity of a group, mean species duration, and the distribution of known fossil finds, the calculated point of origin of a group then falls well before the first fossil. How far before depends on the input relative sampling intensity (the lower the value, the longer the range extension) which depends on the shape of the logistic curve and on assumptions about early members of a clade being small, rare, and geographically restricted.

Corroborating evidence in favour of this view is that most molecular evidence suggests long internodes between at least some basal bird groups (e.g. Cooper and Penny 1997; Harlid and Arnason 1999), although Bleiweiss *et al.* (1994, 1995) found the opposite, based on DNA hybridization studies. Also, the dataset used by Bleiweiss (1998) was geographically biased, consisting of 71 sites in western Europe, and only 23 from elsewhere. The bias was inevitable since this represents the current knowledge of fossils of the bird groups in question. But the critic can claim that Bleiweiss (1998) has produced predictions only for future finds in Europe, where there is a demonstrable bias in the rock record (Smith and Peterson 2002), even though he used the best currently available data.

If Bleiweiss (1998) is right, then the gap analysis approach has predicted rather short 95 per cent confidence intervals, and that the fossil record is adequate. On the other hand, if Marshall (1999) is right, then the potential range extension for the bird orders is very large, and completely in keeping with the molecular estimates, and the fossil record of birds is evidently not adequate to show major features of their early history. The method of Tavaré *et al.* (2002) would presumably produce a figure somewhere between the two extremes. Both viewpoints contain assumptions that are hard to test. The best evidence in favour of Marshall's (1999) view would be finds of definitively mid- and Late Cretaceous birds of modern type. There is no equivalent test of the view expressed by Bleiweiss (1998), since the absence of such finds supports his view, but of course 'absence of evidence is not evidence of absence'. This imbalance in the possibility of confirming evidence was noted by Benton (1999b): one fossil find could confirm the 'early origins' view, but there is no such simple confirmation of the 'late origins' view.

Group sampling

Foote and Raup (1996) developed a simple method to derive an empirical estimate of sampling at group level, which they termed FreqRat. This depends on a knowledge of the distribution of frequencies of species or genera of particular durations within a larger clade, and follows the formula:

$$R = [f(1)f(3)] / [f(2)^2]$$

where R is the probability that a taxon will be preserved at least once in a time unit, and f(1), f(2), and f(3) are the recorded frequencies of taxa spanning one, two, and three equal-length intervals, respectively. This is a simplification of a much more complex set of equations that take account of relative extinction probabilities of different taxa, distributions of occurrences within ranges, and other factors, but empirically the relationship works for exponential ('hollow curve') distributions, where there are relatively large numbers of taxa with short ranges, and rapidly falling numbers of taxa with longer durations. Foote (1997) developed the method further for continuous (rather than discrete) ranges, and for situations where there might be a sample-size bias, but the FreqRat formula is a good approximation for most typical cases.

Foote and Raup (1996) found values in the range of 60–90 per cent for the completeness of different groups – the proportions of species of trilobites, bivalves, and mammals, and the proportions of genera of crinoids preserved. They confirmed that incompleteness of these readily fossilizable groups was a result of the loss of fossiliferous rock rather than the failure of species to enter the fossil record in the first place. Foote and Sepkoski (1999) presented a wider array of estimates of the probability of preservation of genera of different animal groups, ranging from 5 per cent for polychaete worms, to 40–50 per cent for sponges, corals, crinoids, gastropods, bivalves, and ostracods, to essentially 100 per cent for brachiopods.

Foote *et al.* (1999) applied their technique to the fossil record of mammals in North America, to assess whether molecular estimates for the origin of the orders (130–70 Ma) were more or less likely than fossil estimates (oldest fossils, 70–50 Ma). They modelled typical patterns of branching evolution, and then applied imaginary filters to cut out species. In other words, they decreased the value of R, the preservation probability, until all fossils disappeared over a set span of time, the situation implied by the molecular age-doubling hypothesis for the initial radiation of modern mammals. The preservation probability of North American Cenozoic mammal species is 0.25 per 0.7 Ma interval, corresponding to a completeness of 58 per cent (Foote and Raup 1996), whereas values predicted for the complete or virtual absence of modern mammals in the mid- to Late Cretaceous are two orders of magnitude lower, a level that Foote *et al.* (1999) find to be lower than any other calculated preservation probabilities for any taxa, and hence most unlikely.

Is this a valid test? Smith and Peterson (2002) argue that there is a major flaw, that Foote *et al.* (1999) were mistaken to calculate preservation probabilities from a sampling of the fossil record that was overwhelmingly dominated by the Campanian and Maastrichtian record of North America. Indeed, Foote *et al.* (1999) included only limited evidence about mammalian faunas from other parts of the world, and it would

be worthwhile to repeat their experiment, but with fuller documentation. It is wrong to assume that all we know about continental vertebrates in the Late Cretaceous comes from North America. Sea levels were high worldwide, but continental units with fossil vertebrates are known (Weishampel 1990) from around the Mediterranean (Portugal, Spain, France, Romania, Middle East), from Asia (Uzbekistan, Tadzhikistan, Kazakhstan, Russia, India, Mongolia, China, Japan, Laos), from South America (Colombia, Bolivia, Peru, Brazil, Uruguay, Chile, Argentina), and from Africa (Morocco, Algeria, Egypt, Niger, Kenya, South Africa, Madagascar). Mammal fossils are known from all these areas, with spectacular examples from numerous horizons in the Late Cretaceous of Uzbekistan, Kazakhstan, Mongolia, China, Argentina, and Madagascar. Expanding the sampling worldwide might not in fact invalidate the findings by Foote *et al.* (1999).

Oldest fossils

The current literature about origins of major clades often includes discussion about the oldest relevant fossils. Does close scrutiny of the fossil evidence indicate rapprochement between molecular early dates and palaeontological late dates of origin? This does not seem to be the case for the earliest metazoans, green plants, angiosperms, or birds, but there is now good agreement for the basal radiation of mammals.

Mismatch

The first metazoan fossils are generally accepted to date from the earliest Cambrian, the great burst of expansion of skeletonized animal groups from about 545 Ma, long after the molecular date of around 1000 Ma. Fossil evidence for metazoans is known from the Precambrian, but nothing before about 600 Ma. This evidence includes possible sponges and cnidarians from the Ediacara faunas, exquisitely preserved fossil embryos, and simple creeping trails (Valentine *et al.* 1999; Conway Morris 2000).

The first vascular land plants are found as fossils in the Silurian, and earlier evidence from spores extends the range back to the Ordovician (475 Ma; Kenrick and Crane 1997), considerably younger than a molecular estimate of 700 Ma (Heckman *et al.* 2001). A similar gap exists for angiosperms, with the oldest generally accepted fossils being from the Early Cretaceous, pollen records dated at about 130 Ma, and abundant pollen and macroplant fossils from 120 Ma onwards (Crane *et al.* 1995). Older putative fossil angiosperms, from the Jurassic, and even from the Triassic, have not been generally accepted. DNA sequence evidence places the divergence of angiosperms in the Mid-Jurassic (175 Ma; Wikström *et al.* 2001). However, genealogical evidence actually suggests a much more ancient date of origin, back in the Carboniferous at 290 Ma (Kenrick 1999), if it turns out that the sister group of angiosperms is the gymnosperms.

In the case of the origin of modern birds, many supposed Cretaceous representatives of modern bird orders have been cited (e.g. Cooper and Penny 1997; van Tuinen *et al.* 2000; Paton *et al.* 2002), but all have been disputed, mostly because the fossils are isolated elements (Dyke 2001). Hence, the oldest uncontroversial fossils of modern bird orders date from the Palaeocene (60 Ma), much younger than most molecular estimates of origins, at 70–120 Ma.

Rapprochement

The dating of the radiation of modern placental (= eutherian) mammals seemed to be similarly fraught until a year or two ago. The traditional, palaeontological view (e.g. Carroll 1988; Benton 1990) was that placentals split from marsupials some time in the Early Cretaceous (144–99 Ma), and modern orders split at the end of the Cretaceous and in the early Tertiary (70–55 Ma). The first molecular dates (Hedges *et al.* 1996; Kumar and Hedges 1998; Eastal 1999) seemed much older: origin of eutherians in the Late Jurassic (c. 150 Ma), split of major placental groups in the Early Cretaceous (c. 110–120 Ma), and split of modern placental orders in the mid- to Late Cretaceous (c. 80–100 Ma). It seemed there was a major problem.

Since 1996, however, there has been a rapprochement, and palaeontological and molecular evidence now seem to agree. The change has happened because of a better understanding of what the fossils show, and because molecular age estimates have been revised upwards.

First, the oldest fossils of modern placental mammals are not entirely basal Tertiary (Archibald 1996; Archibald *et al.* 2001), but it is important to distinguish group membership. Modern placental mammals are divided into 18 orders, and these fall into four larger superorders, the Xenarthra (edentates), Afrotheria (elephants, hyraxes, sirenians, tenrecs, golden moles, and the armadillo), Glires plus Euarchonta (rodents, rabbits, flying lemurs, tree shrews, primates), and Laurasiatheria (insectivores, bats, pangolins, carnivores, perissodactyls, artiodactyls, whales). An extraordinary Late Cretaceous locality in Uzbekistan, dated at 85–90 Ma, has yielded no representatives of modern placental *orders*, but has produced specimens that are assigned to basal parts of placental *superorders*, the Glires and Ungulatomorpha (part of Laurasiatheria) (Archibald *et al.* 2001). The recent discovery of a beautifully preserved basal placental mammal from the Early Cretaceous (c. 125 Ma) of China (Ji *et al.* 2002) could point either way. It could be a late survivor of a split that happened in the Jurassic, but it is just as likely that it confirms the split of placentals and marsupials in the Early Cretaceous.

Since 1997, the molecular estimates have been revised upwards (Hedges and Kumar 1999; Eizirik *et al.* 2001; Murphy *et al.* 2001). Superordinal diversification is dated at 64–104 Ma (mean 84 Ma), and ordinal diversification at 50–75 Ma, entirely in line with the fossils. Note that different probability modelling approaches either support the new consensus (Foote *et al.* 1999; Archibald and Deutschmann 2001), or dispute it by suggesting the possibility of much older ordinal divergence dates (Tavaré *et al.* 2002). Is the mammal example an indicator of possible rapprochement in the future over other disputed dates of origination?

Evidence for a poor-quality fossil record

Heterogeneity

It is clear that the fossil record contains a biotic and an abiotic signal. The distribution of fossils in the rocks consists of a combination of the record of the history of life and the vicissitudes of the history of the rocks. Most palaeontologists would like to believe that the history of life is a robust enough signal to stand out from the

background smearing. Critics, however, believe that sampling overwhelms the biotic signal; the heterogeneity of the temporal and geographical distribution of rocks masks the real story.

Smith (2001) and Peters and Foote (2001, 2002) have argued that the distribution of sedimentary rocks controls the preservation of fossils and that much (?most) of the standard plots of diversifications and extinctions from the fossil record (e.g. Sepkoski 1984, 1996; Benton 1995) are artefacts. For example, in a study of the marine fossil record of the post-Palaeozoic, Smith (2001) found that outcrop area and sea level changes correlated with some aspects of diversity change, and Peters and Foote (2001, 2002) made the same observation for the whole of the Phanerozoic. Small-scale changes in diversity and in origination rate were related to the surface area of outcrop, and these authors stress that it would be foolhardy to interpret every rise and fall in the global diversity, extinction, and origination signals as biologically meaningful.

Mass extinctions represent a particular issue. Smith (2001) found that most peaks in extinction did not correspond to changes in outcrop area, but two occurred towards the culmination of stacked transgressive system tracts and close to system bases. One of these, falling at the Cenomanian–Turonian boundary, and representing a well-known postulated mass extinction in the sea (e.g. Raup and Sepkoski 1984; Hallam and Wignall 1997), may then be truly an artefact of sampling and sea level change (Smith *et al.* 2001). Peters and Foote (2001, 2002) found that all such ‘lesser’ global extinction events disappeared when the effect of sampling was taken into account. Most of the ‘big five’ mass extinctions also seem to be equivocal, or at least to be exaggerated by sampling. These are rather startling findings.

A criticism of the work by Smith (2001) and Peters and Foote (2001) could be that they use limited datasets on sampling. Smith (2001) used map areas of rocks of different age from Britain and France only for comparison with the global biodiversity signal, while Peters and Foote (2001) used a lexicon of numbers of named stratigraphic formations in North America. However, both studies, using such different samples of sampling, came to the same conclusion, and Peters and Foote (2002) have taken a broader sample of named marine stratigraphic units from around the world, and the results are the same.

Both Smith (2001; Smith *et al.* 2001) and Peters and Foote (2001, 2002) tested for the relative roles of abiotic and biotic factors, and they found that changes in rock surface area explained most of the variance. So, the results indicate that rock outcrop area drives the record of the diversification of life and that extinction events are largely artefacts of the appearance and disappearance of rock rather than of organisms. Alternatively, as both teams stressed, an additional factor, perhaps sea level change, could drive both signals, that marine rock area rises and falls as sea level rises and falls, and marine biodiversity expands and contracts as does the volume of the sea. Either way of course the fossil record is much weakened as an accurate document of the history of life.

The finding that short-term rises and falls in marine biodiversity mirror the rock record is convincing. Most palaeontologists already knew that most of the postulated Mesozoic ‘mass extinctions’ required by the hypothesis of periodicity in mass extinction (Raup and Sepkoski 1984), such as those in the Early and Mid-Jurassic, the Jurassic–Cretaceous extinction, and the Early Cretaceous event were artefactual

to a greater or lesser extent (Hallam 1986; Benton 1995; Little and Benton 1995). Does any biotic signal survive?

The dramatic diversification of life in the sea, and the even more dramatic diversification of life on land over the past 250 Ma has been noted by many authors (Sepkoski 1984, 1996; Benton 1995, 1997, 2001). Smith (2001) found that the overall rise in marine generic and familial diversity through the Mesozoic and Cenozoic could not be explained by sea level change, and hence was probably real. Peters and Foote (2001) question this, however, suggesting that even the massive diversification of marine life in the past 250 Ma could be an artefact of low turnover and the pull of the Recent, confirming Raup's (1972) earlier suggestion that marine life diversified dramatically early in the Palaeozoic, and has remained at a constant level ever since (see below, 'Bias').

The studies so far have focused on life in shallow seas. An interesting further test would be to compare the results with the deep sea, where habitats are less heterogeneous and sedimentation is more continuous, and with continental settings, where habitats are diverse and sedimentation is often supposed to be even more sporadic than in shallow seas. Smith (2001) argued for a strong bias in the case of terrestrial tetrapods, and Smith and Peterson (2002) indeed show a strong correlation between the rock record in western Europe and the number of bird families recorded through time. On a broader scale, Peters and Foote (2001) noted a correlation of numbers of terrestrial animal families and terrestrial formations in North America. However, the correlation was weaker than for marine animal families and genera, contrary to expectations. In a more detailed study, Fara (2002) actually found no evidence for a correlation for continental tetrapods: as sea levels fell, and continental areas expanded, there is no evidence for a matching expansion in the diversity of land animals.

What about the 'big five' mass extinctions? Smith (2001) does not question the reality of the end-Permian and K-T events, although he notes an interaction of a biotic (rapid extinction) and abiotic (major sea-level change and reduction in surface area of preserved onshore facies) signal that must be disentangled. In this case, correlation of the biotic and abiotic signals need not indicate that the first is an artefact of the second, but that both are part of the global cataclysms associated with times of mass extinction. Peters and Foote (2002) leave the issue of mass extinctions much more open. By their modelling approach, they argue that all of the big five mass extinctions *could* be explained as artefacts of sampling. The first two, in the Late Ordovician and Late Devonian, might be real, but only if global generic extinction rates are modelled as constant, rather than declining. The end-Permian mass extinction, the largest ever, does appear above the noise if extinction rates are modelled as declining, but it is swamped by sampling in a constant-extinction-rate model. The same is true of the end-Triassic mass extinction. Only the K-T mass extinction shines through in all models where sampling is accounted for. Peters and Foote (2002) say that this result could indicate one of two things, either that mass extinctions are merely artefacts of variations in available rock volume, or that mass extinctions and reductions in rock volume are associated with an additional common factor, such as major sea-level change, as Smith (2001) suggested.

If the method can reject mass extinctions, perhaps there is a problem with the method. Peters and Foote (2002) meant to challenge conventional assumptions, and they are

careful to outline potential pitfalls in their data and their models. The explanation of minor variations in the biotic signal as abiotic artefacts (Peters and Foote 2001, 2002; Smith 2001; Smith *et al.* 2001) makes a great deal of sense. However, there is so much geological, geochemical, and palaeontological evidence, in addition to the broad-brush diversity plots that are under scrutiny, for the end-Permian, end-Triassic, and K–T crises that these events can probably be accepted as real. Hence if a statistical method is capable of rejecting their reality, one has to look closely at the statistical method since it may be too crude.

This new work on heterogeneity gives mixed messages about scaling of time and taxa. Smith (2001) stressed that the key geological driver is at the level of major sequence stratigraphic cycles of 20–50 myr, not 1–10 myr. Peters and Foote (2001) showed a close linkage of the biotic and geological signals at the level of epochs (2–42 myr, mean 19 myr), but in their later paper (Peters and Foote 2002), the scaling was at stage level (2–20 myr, mean 7.1 myr). Is the proposal that scaling is fractal, and every biotic signal can be shown to follow a geological signal slavishly? Or can palaeontologists expect that observations on certain timescales may be free of geological control? Care is required in seeking fossil versus rock correlations: if genera are sampled, and the time bins are too broad, then each occurrence is effectively a single point, and it is then most likely that the number of fossils will be controlled by the rock area or volume, but the linkage would be largely an artefact of the method. Finer time divisions will allow true ranges of genera to be assessed. We have found (Fara and Benton 2000) that known gaps in the continental rock record of the Cretaceous are bridged by new discoveries on either side, if the taxonomic scale is appropriate to allow spanning (we chose to look at families and stratigraphic stages). So, at family level, the biotic signal is robust to global stratigraphic gaps. Had we chosen genera, then they could never have spanned the known gaps (stages, 2–13 myr, mean 6.4 myr), and then rock area could have been said to drive the biotic signal.

Sea level changes can clearly produce artificial extinctions, but could they hide a diversification for tens of millions of years, as postulated by the molecular age-doubling observation? This seems most unlikely because of timescale considerations. Known habitat shifts and hiatuses in the global rock record account for a few million years at most. Such heterogeneity could not delete 30 or 40 Ma of the history of a group. So sea level changes can create false extinction events, but it is hard to see how they alone could hide real diversifications.

Bias

A related, but more extreme, view has been that a combination of factors render the fossil record poorer and poorer the further back in time one goes. Raup (1972) argued that the fossil record suffers from a number of biases, such as the evident loss in volume of rocks, and the exposed area of rocks, as a result of the cumulative effects of burial, metamorphism, subduction, erosion, and covering through geological time. In addition, fossils from ever older rocks belong to groups that are less and less like modern forms, and are hence harder and harder to identify to species level. Practically speaking also, relatively fewer palaeontologists work on more ancient rocks and fossils than more recent ones.

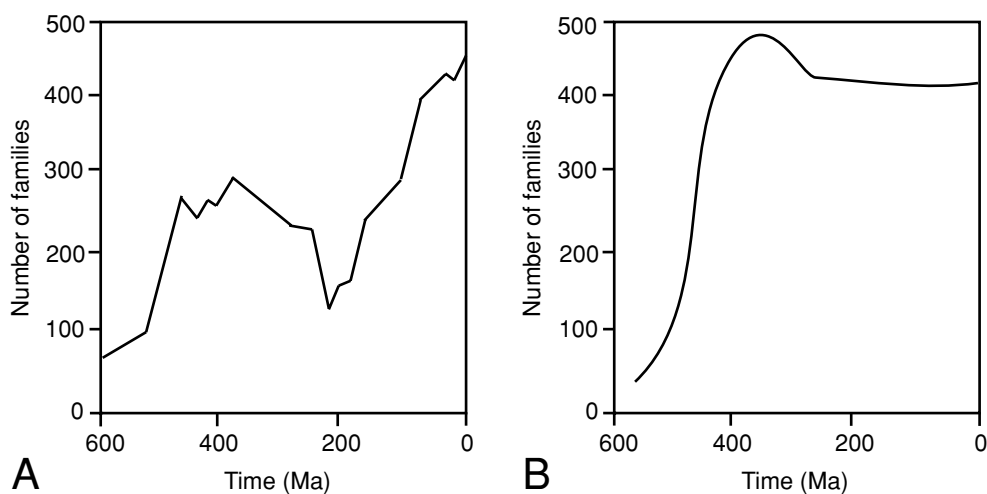


Figure 4.2 Comparison of empirical (A) and bias-simulation models (B) for diversification of well-skeletonized marine invertebrates through the Phanerozoic. The empirical pattern (A) is a literal reading of changes in diversity of families, and the bias-simulation model (B) is a theoretical construct that purports to show the true pattern of diversification after corrections for the poorer Palaeozoic fossil record and lower levels of study of such materials. Based on data in Valentine 1969 (A) and Raup 1972 (B).

Raup (1972) presented his view of the likely true shape of the diversification of life in the sea (Figure 4.2A), his so-called 'bias-simulation model', and contrasted it with the 'empirical' model of Valentine (1969). Based on theoretical considerations, deriving from competition studies in ecology and the MacArthur and Wilson's (1967) *Theory of Island Biogeography*, Raup (1972) argued that, following the Cambrian explosion, the sea filled up with families and species, and reached its carrying capacity within a geologically short span of time. After an overshoot, and some adjustment, a dynamic equilibrium level was achieved, which is the present diversity of life in the sea, and this level has been sustained for some 500 Ma.

Further consideration of these polarized views in the 1970s led to a reconciliation in which the empirical model was considered to be nearer the truth than Raup's (1972) bias-simulation model (Sepkoski *et al.* 1981) based on a comparison of a number of independently compiled datasets. Since then, palaeontologists have felt that they could legitimately study diversification and extinction on the basis of global-scale compilations of data on the fossil record, and that the broad patterns were correct (e.g. Raup and Sepkoski 1982, 1984; Sepkoski 1984, 1996; Benton 1985, 1995, 1997, 2001; Niklas *et al.* 1983; Miller 1998).

Alroy *et al.* (2001) have now suggested that Raup (1972) might well have been correct. In the first publication from the Palaeobiology Database (PD) project, in which a sample-based approach is adopted, global levels of diversity extrapolated from fossil samples from the Palaeozoic appear to be comparable with those from the Cenozoic. The empirical finding by all authors that diversity increased dramatically at the levels of families and genera through the past 250 myr (e.g. Valentine 1969; Sepkoski *et al.* 1981; Sepkoski 1984, 1996; Benton 1995; Miller 1998) must then be

explained as an artefact of dramatically improved sampling during that interval, and sampling that improves steadily from the Triassic to the present. As Alroy *et al.* (2001) make clear, their preliminary results are based on the sample of fossil collection data that has been accrued in the database so far, and it cannot yet be assessed whether that sample might include Palaeozoic collections that exaggerate apparent generic diversity (taxa oversplit, samples based on large 'localities', high levels of time-averaging, localities with sparse faunas omitted) when compared with the Cenozoic collections. Broadly put, the rarefaction approach adopted by the PD team requires unbiased environmental sampling through time, an objective that will be hard to achieve. For these reasons, Jackson and Johnson (2001) urge caution in the use of such a database based on random samples instead of comprehensive databases.

Failure of statistical approaches?

On the face of it, current standpoints on the quality of the fossil record could not be more extreme, and resolving these differences might seem an insurmountable problem. Available statistical approaches such as confidence intervals and group sampling are based on internalized assessments of the data which are being assessed, so they are not true statistical tests, where the data would be compared with an external standard. Smith and Peterson (2002) argue that these approaches cannot test whether temporal and geographical heterogeneity in the distribution of sedimentary rocks are not controlling patterns in the fossil record, but does this mean that the techniques should be abandoned?

Probably not. The critique of the confidence intervals approach is clearly correct: Marshall (1997, 1999) has argued that case already, and his generalized confidence intervals method can deal with heterogeneous preservation probability. However, the group sampling approach should not be rejected simply because it assumes homogeneous preservation probability. Foote (1997) showed that his methods can be misled by heterogeneity in the distribution of rocks, and in extreme cases, when suitable rocks are largely absent, the preservation probability is overestimated. However, this failure applies only in extreme cases, and Foote *et al.* (1999, note 31) claim that fluctuating preservation rate, associated with changes in sea level and other factors, is not likely to distort substantially either the overall probability of species preservation or estimates of preservation rate by the FreqRat and associated methods. Foote and Sepkoski (1999) and Foote *et al.* (1999) make a strong case that their methods are valid for estimating general broad-scale fossil record quality.

None the less, rock-record heterogeneity clearly causes problems for all statistical approaches to fossil sampling. Is there an alternative approach that might allow palaeontologists to escape from the risk of circularity in using internal measures of the fossil record to assess the quality of that same fossil record?

Clade versus age techniques

Independence

A sideways leap provides a partial answer, and that is to use a source of data on the history of life that is independent of the rocks, namely phylogenetics. It has been argued

(Platnick 1979; Patterson 1982; Smith and Littlewood 1994; Benton and Hitchin 1996, 1997) that phylogenetic data, whether from the cladistic analysis of morphological characters or from molecular phylogenetic reconstruction, are essentially independent of stratigraphic (geological age, rock distribution) evidence.

In cladistics, characters are determined and polarized (primitive → derived) according to their distribution among a group of organisms, living and extinct, and without reference to geological age. Many analysts these days do not even polarize the characters, so they include no directional information prior to the analysis. Molecular phylogeny reconstruction is even more obviously divorced from stratigraphy in that all the organisms under investigation are extant, and characters are generally not polarized. The implied history of the group is then subtended from the present-day with no reference to fossil taxa.

Trees, whether cladistic or molecular, cannot entirely escape from time-related aspects. A small input of stratigraphic bias may be associated with the choice of outgroups (comparator standards), but outgroups can readily be changed and analyses re-run. In addition, there are certain unavoidable temporal biases in phylogenies that relate to their geometry and the completeness of taxon sampling (Wagner 2000a) and to the relative timings of acquisition of apomorphies and homoplasies (Wagner 2000b). These issues do not substantially modify the geometry of a tree and the relative order of branching points, and hence independence between stratigraphy and tree shape is sustained (Benton 2001; Wills 2002). More serious though is that the age versus clade measures cannot detect major hiatuses in the rock record. So, if all early members of a clade are unknown as fossils, the relative order of appearance of lineages within that clade then becomes meaningless.

The hypothesis behind all age versus clade comparisons is that congruence indicates the true pattern. It is accepted of course that the fossil record is subject to bias, as are the techniques of cladistics and molecular phylogenetics. But the biases that might affect these three approaches are clearly different, and unlikely to reinforce each other. So, if a phylogenetic tree is congruent with the order of fossils in the rocks, it is most likely that both are correct. If incongruence is found, then it cannot be said whether the tree or the fossils, or both, are incorrect. The test for congruence does not mean that the entire tree is congruent with stratigraphy, but that it is more congruent than random, and certainly not significantly incongruent.

Metrics

Trees and fossil sequences may be compared in various ways, and several metrics have been proposed (reviewed in Benton *et al.* 1999). First attempts to compare clade and age data concentrated simply on the rank order of first fossils and branching points. The age and clade rank orders could then be compared by the use of simple correlation measures, such as Spearman rank correlation. Rank-order approaches are crude, however, since they can cope only with single runs of digits, and they cannot code the more complex patterns of typical cladograms with their multiple branches except by dividing them. Furthermore, in cases where numerous fossils occur close together in time, or where many branching events happened in a short time, the rank order is hard to sort out. For these reasons, we have abandoned use of this approach (Benton and Hitchin 1996; Hitchin and Benton 1997; Wagner 1998).

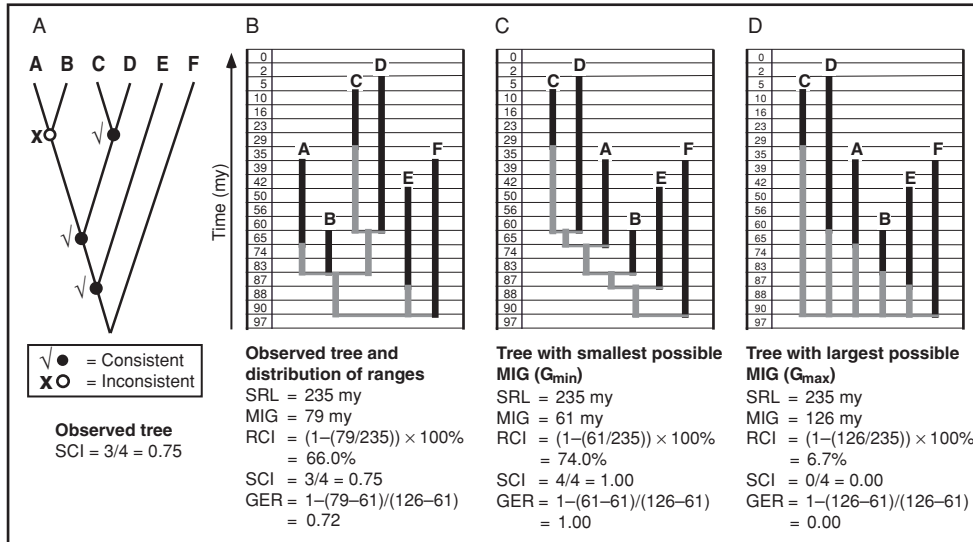


Figure 4.3 Calculation of the three congruence metrics for age versus clad comparisons, the stratigraphic consistency index (SCI), the relative completeness index (RCI), and the gap excess ratio (GER). SCI is the ratio of consistent to inconsistent nodes in a cladogram. RCI is,

$$RCI = 1 - \left[\frac{\sum MIG}{\sum SRL} \right] \times 100\%$$

where MIG is the minimum implied gap, or ghost range, and SRL is the standard range length, the known fossil record. GER is,

$$GER = 1 - \frac{(MIG - G_{min})}{(G_{max} - G_{min})}$$

where G_{min} is the minimum possible sum of ghost ranges and G_{max} the maximum, for any given distribution of origination dates. A, The observed tree with SCI calculated according to the distribution of ranges in B. B, The observed tree and observed distribution of stratigraphic range data, yielding an RCI of 54.6 per cent. GER is derived from G_{min} and G_{max} values calculated in C and D. C, The stratigraphic ranges from B rearranged on a pectinate tree to yield the smallest possible MIG or G_{min} . D, The stratigraphic ranges from B rearranged on a pectinate tree to yield the largest possible MIG or G_{max} .

We have used three measures to assess age versus clad congruence (Figure 4.3): the stratigraphic consistency index (SCI; Huelsenbeck 1994), the relative completeness index (RCI; Benton and Storrs 1994), and the gap excess ratio (GER; Wills 1999). The first measure looks at the branching points (nodes) in a cladogram and their relation to each other. A minimum date is assigned to each branching point by assessment of the oldest known fossils of each of the subtended sister taxa. The consistency of each node is then assessed by determining whether it is younger than, or the same age as, the node immediately below. The SCI is the ratio of consistent to inconsistent nodes, and it can range from 0 to 1.0 in a fully pectinate (unbalanced) tree, but the minimum value lies between 0 and 0.5 in balanced trees (Siddall 1996; Wills 1999).

The RCI and GER depend on numerical age estimates of the branching points on a cladogram, and the calculation of 'ghost ranges'. The ghost lineage (Norell 1992) is the implied missing evolutionary line indicated by the difference in age between the oldest known fossils of two sister taxa, and that missing span of time is termed the ghost range, or the minimum (cladistically) implied gap (MIG; Benton and Storrs 1994). It is based on the observation that a node in a cladogram represents a single point in time, but that the oldest fossil representatives of the two lineages branching from that node are most often not of the same age (Smith and Patterson 1988). The RCI is assessed as the ratio of the sum of ghost ranges to the sum of recorded fossil ranges in any cladogram. The GER focuses solely on the estimated dates of origin of groups, and compares the sum of actual ghost ranges in a cladogram with the theoretical minimum and maximum ghost ranges if the various branches in the cladogram are rearranged. Values for the GER range from 0.0 (no congruence) to 1.0 (perfect congruence), while values for the RCI range from 0–100 per cent where $MIG < SRL$. However, the RCI can range to $-\infty$ when the known ranges (SRL) are point occurrences, and the sum of ghost ranges (MIG) is large.

Many additional clade versus age congruence metrics have been proposed, and all are related to one or other of the metrics we use. Norell (1992) proposed the Z statistic, one minus the ratio of the sum of ghost ranges divided by the number of taxa to the sum of known ranges. Smith and Littlewood (1994) proposed the implied gap (IG) metric, the ratio of summed ghost ranges to summed ghost ranges plus summed observed ranges in a cladogram. Weishampel (1996) used sums of ghost lineage durations (GLDs). Siddall (1998) presented his Manhattan stratigraphic measure (MSM), which uses Manhattan distances between stratigraphic ages. Brochu and Norell (2000) proposed SMIG, the sum of minimum implied gaps (i.e. ghost ranges).

Results

Based on various samples of published trees, age versus clade comparisons have yielded a number of results. Norell and Novacek (1992) found that 75 per cent of their sample of trees of mammals showed congruence, confirmation of the validity of the tree-making methods and of the fossil record. Later studies, based on larger samples of trees, and for wider groups of organisms, found lower proportions of congruent trees based on Spearman rank correlation (36–50 per cent; Benton and Hitchin 1997), but the rather more realistic metrics outlined above confirmed that roughly three-quarters of published trees *are* congruent with the fossil record (Benton *et al.* 1999, 2000).

Comparisons of different sectors of the data (the current data set stands at 1000 trees; <http://palaeo.gly.bris.ac.uk/cladestrat/cladestrat.html>) showed no clear bias. For example, marine and continental organisms show equivalent levels of congruence between trees and the fossil record (Benton and Simms 1995; Benton and Hitchin 1996; Benton 2001). Different taxonomic groups on the whole also show equivalence, with no clear evidence that plants or animals, invertebrates or vertebrates, or whatever, are uniformly better preserved, or have uniformly better analysed cladograms than any other (Benton and Simms 1995; Benton and Hitchin 1996; Benton 2001). Wills (2001) showed that arthropods on the whole show poorer age–clade congruence than vertebrates, but he could not readily explain this observation.

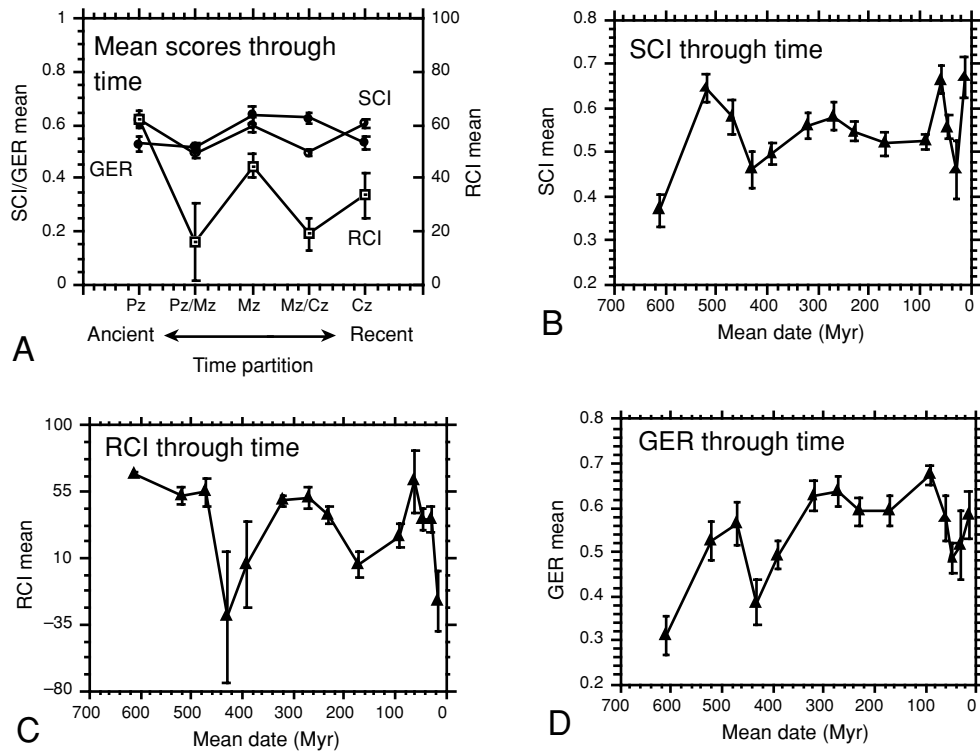


Figure 4.4 No change in fossil record quality through time, mean scores of the age versus clade metrics for finer-scale divisions of geological time. A, Stratigraphic consistency index (SCI), the relative completeness index (RCI), and the gap excess ratio (GER) for five time partitions of the data set of 1000 cladograms, namely cladograms with origins solely in the Palaeozoic (Pz), cladograms with origins spanning the Palaeozoic and Mesozoic (Pz/Mz), cladograms with origins solely in the Mesozoic (Mz), cladograms with origins spanning the Mesozoic and Cenozoic (Mz/Cz), and cladograms with origins solely in the Cenozoic (Cz). B–D, Age versus clade metrics for cladograms partitioned into geological periods and epochs showing temporal variations in the SCI (B), RCI (C), and GER (D). The age versus clade metrics are explained in the caption to Figure 4.3. There is no statistically significant secular trend for the broad-scale time divisions (A), nor for the period-by-period assessments by the SCI (B) or RCI (C). The GER values (D) do improve through time ($0.5 > P > 0.025$), but the regression becomes non-significant if the low Vendian value (based on 34 trees) is omitted.

Comparisons of the change in palaeontological knowledge through time show a statistically significant reduction in ghost ranges over a 25 year sample period (Benton and Storrs 1994). This confirms that new fossil discoveries do not extend fossil ranges in unexpected ways, but they tend to fill predicted gaps.

Finally, there is no evidence for a decline in the quality of the fossil record back through geological time (Benton *et al.* 2000). The sample of 1000 cladograms was divided into various time bins, and the age versus clade metrics assessed. The expectation was that Palaeozoic trees would show poorer congruence values than, say, Cenozoic trees. Actually, the two time-independent metrics (SCI, GER) showed essentially no change through time (Figure 4.4), while the RCI worsened through time

(but that was expected, since the RCI is a ratio of ghost range to known range, and known stratigraphic ranges for groups originating in the Palaeozoic may be much longer than those originating in the Cenozoic).

Are the age versus clade methods subject to the problem of temporal and geographical heterogeneity of the rock record, as are the confidence interval and group sampling methods? The answer is no, where heterogeneity is at 'normal' scales:

(1) Heterogeneity in the rock record is purely a geological issue, and it is related to all the other biases outlined by Raup (1972). Therefore, in assessing congruence of the order of fossils in the rocks with the patterns of cladograms, there is no linkage.

(2) The available set of published cladograms includes trees for soft-bodied organisms and those with hard parts alike. There is no evidence for a substantial difference in expectations of congruence between groups with readily fossilizable parts (such as vertebrates with their bones, or echinoderms with their calcite skeletons) and those with less robust skeletons (e.g. plants, arthropods, with thinner skeletons or organic cuticles). Of course, the comparison cannot be extended to entirely soft-bodied organisms since, in the absence of any fossils, it is impossible to make any age versus clade comparisons. However, there is no reason to assume that entirely soft-bodied organisms were any more or less abundant in the past than they are now, and hence that the pattern of evolution of groups with hard parts can be accepted as a proxy for the evolution of groups lacking such hard parts (Valentine 1969).

In the case of large-scale heterogeneity, however, the methods cannot function (Andrew Smith, Natural History Museum, London, pers. obs.). For example, if there were truly a gap of 30–40 myr in the Late Cretaceous where no bird fossils are found, then any calculations of age versus clade metrics based on a cladogram of major bird lineages plotted against time would be meaningless. I question, however, how often such vast gaps in the record actually occur.

Peters and Foote (2001, pp. 597–8) have stressed that the age versus clade metrics cannot be used to assess absolute or overall completeness of the fossil record, a point made also by Benton *et al.* (2000). The metrics can only compare known and postulated parts of the record that exist in the rocks, termed the intrinsic completeness of the fossil record. Peters and Foote (2001) contrast these two aspects as 'global' and 'local', but their terms have common geographical meanings, and perhaps 'absolute' and 'intrinsic' are preferable.

Conclusion

The fossil record is, as ever, under close scrutiny. Current viewpoints range from gung-ho to abject despair. The gung-ho view is that things have never been better: the fossil record may be read like a book that documents every nuance of the history of life. The despairing position is that the fossil record can never say much about the history of life since it is so riddled with bias and error, much of which can never be estimated and corrected.

Our age versus clade studies may be read to confirm something midway between these two stances, but definitely tending to the gung-ho end of the scale. The two key results have been that most trees are congruent with the fossil record, and that there

is no evident large-scale time bias through the past 500–600 myr or so, at the scale of eras (65–300 myr) and periods (40–80 myr). The latter result is counter-intuitive, since it is evident that many factors must act as time-related biases: rock volume, rock area, metamorphism, erosion, study levels. However, these biases evidently affect things only at the lower level of focus, when one considers individual specimens or species in particular localities at fine-scale stratigraphic divisions. At the taxonomic level of families and above, and the stratigraphic level of geological periods, the patterns may well be sound.

In support of the molecular age doubling found for modern orders of birds and mammals, Hedges *et al.* (1996), Cooper and Penny (1997), Kumar and Hedges (1998), Easteal (1999) and others have suggested three reasons why the fossils have not been found, but I doubt these:

(1) Ancestral forms were cryptic, or did not display all synapomorphies. This idea is that somehow molecular and morphological evolution are uncoupled, and that molecular divergence between major clades could happen tens of millions of years before full morphological differentiation. There is no evidence for such substantial uncoupling of molecular and morphological evolution, indeed rather the opposite (Omland 1997), and it is hard to see how the suggestion could ever be tested.

(2) Ancestors were unpreservable (too small, soft-bodied). This might be true for basal metazoans in the late Precambrian, but the ancestors of modern bird and mammal groups were most unlikely to have been unpreservable: dozens of localities through the Early and Late Cretaceous have yielded tiny, delicate skeletons of birds and mammals, but none of them pertains to extant orders (Benton 1999b; Fara and Benton 2000).

(3) Ancestors lived in hitherto unexplored parts of the world, such as the southern continents. This might be true, but current work is opening up richly fossiliferous sites in the Cretaceous of Gondwana, in South America, southern Africa, Madagascar, and India, and not a hint of a modern bird or mammal has been found. Similarly, for the early origins of Metazoa case, palaeontologists are working actively in hitherto palaeontologically unexplored parts of the world (China, Australia, Africa), so the chances that the fossils required by the age-doubling molecular argument will be found are diminishing. I termed this the ‘living mastodon’ argument (Benton 1999b), after expectations in the 18th century that mastodons, represented by abundant fossils from North America, might yet be found living in the Wild West.

Acknowledgements

Supported by Leverhulme Grant F/182/AK. Many thanks to Andrew Smith for his critical comments and advice on the manuscript.

References

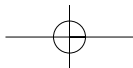
- Alroy, J., Marshall, C.R., Bambach, R.K., Bezusko, K., Foote, M., Fürsich, F.T., Hansen, T.A., Holland, S.M., Ivany, L.C., Jablonski, D., Jacobs, D.K., Jones, D.C., Kosnik, M.A., Lidgard, S., Low, S., Miller, A.I., Novack-Gottshall, P.M., Olszewski, T.D., Patzkowsky, M.E., Raup, D.M., Roy, K., Sepkoski, J.J., Sommers, M.G., Wagner, P.J. and Webber, A. (2001)

- 'Effects of sampling standardization on estimates of Phanerozoic marine diversification', *Proceedings of the National Academy of Sciences, USA*, 98: 6261–6.
- Archibald, J.D. (1996) 'Fossil evidence for a Late Cretaceous origin of "hoofed" mammals', *Science*, 272: 1150–3.
- Archibald, J.D. and Deutschmann, D.H. (2001) 'Quantitative analysis of the timing of the origin and diversification of extant placental orders', *Journal of Mammalian Evolution*, 8: 107–24.
- Archibald, J.D., Averianov, A.O. and Ekdale, E.G. (2001) 'Late Cretaceous relatives of rabbits, rodents, and other extant eutherian mammals', *Nature*, 414: 62–5.
- Benton, M.J. (1985) 'Mass extinction among non-marine tetrapods' *Nature*, 316: 811–4.
- (1990) *Vertebrate Palaeontology*, London: Unwin Hyman.
- (1995) 'Diversification and extinction in the history of life', *Science*, 268: 52–8.
- (1997) 'Models for the diversification of life', *Trends in Ecology and Evolution*, 12: 490–5.
- (1999a) 'The history of life, large data bases in palaeontology', in D.A.T. Harper (ed.) *Statistical Methods in Palaeobiology*, London: Wiley, pp. 249–83.
- (1999b) 'Early origins of modern birds and mammals: molecules vs. morphology', *BioEssays*, 21: 1043–51.
- (2001) 'Biodiversity on land and in the sea', *Geological Journal*, 36: 211–30.
- Benton, M.J. and Hitchin, R. (1996) 'Testing the quality of the fossil record by groups and by major habitats', *Historical Biology*, 12: 111–57.
- (1997) 'Congruence between phylogenetic and stratigraphic data on the history of life', *Proceedings of the Royal Society, London*, B264: 885–90.
- Benton, M.J. and Simms, M.J. (1995) 'Testing the marine and continental fossil records', *Geology*, 23: 601–4.
- Benton, M.J. and Storrs, G.W. (1994) 'Testing the quality of the fossil record, paleontological knowledge is improving', *Geology*, 22: 111–4.
- Benton, M.J., Hitchin, R. and Wills, M.A. (1999) 'Assessing congruence between cladistic and stratigraphic data', *Systematic Biology*, 48: 581–96.
- Benton, M.J., Wills, M.A. and Hitchin, R. (2000) 'Quality of the fossil record through time', *Nature*, 403: 534–7.
- Bleiweiss, R. (1998) 'Fossil gap analysis supports early Tertiary origin of trophically diverse avian orders', *Geology*, 26: 323–6.
- (1999) 'Fossil gap analysis supports early Tertiary origin of trophically diverse avian orders, Comment and Reply', *Geology*, 27: 95–6.
- Bleiweiss, R., Kirsch, J.A.W. and Lapointe, F.-J. (1994) 'DNA–DNA hybridization-based phylogeny for higher nonpasserines, reevaluating a key portion of the avian family tree', *Molecular Phylogenetics and Evolution*, 3: 248–55.
- Bleiweiss, R., Kirsch, J.A.W. and Shafi, N. (1995) 'Confirmation of a portion of the Sibley and Ahlquist "tapestry"', *Auk*, 112: 87–97.
- Brochu, C.A. and Norell, M.A. (2000) 'Temporal congruence and the origin of birds', *Journal of Vertebrate Paleontology*, 20: 197–200.
- Carroll, R.L. (1988) *Vertebrate Paleontology and Evolution*, San Francisco: W. H. Freeman.
- Cavalli-Sforza, L.L. and Edwards, A.W.F. (1967) 'Phylogenetic analysis, models and estimation procedures', *Evolution*, 21: 550–70.
- Conway Morris, S. (2000) 'The Cambrian "explosion", slow-fuse or megatonnage?', *Proceedings of the National Academy of Sciences, USA*, 97: 4426–9.
- Cooper, A. and Fortey, R. (1998) 'Evolutionary explosions and the phylogenetic fuse', *Trends in Ecology and Evolution*, 13: 151–6.
- Cooper, A. and Penny, D. (1997) 'Mass survival of birds across the Cretaceous–Tertiary boundary, molecular evidence', *Science*, 275: 1109–13.

- Crane, P.R., Friis, E.M. and Pedersen, K.R. (1995) The origin and early diversification of angiosperms', *Nature*, 374: 27–33.
- Donovan, S.K. and Paul, C.R.C. (eds) (1998) *The Adequacy of the Fossil Record*, New York: Wiley.
- Dyke, G.J. (2001) 'The evolutionary radiation of modern birds, systematics and patterns of diversification', *Geological Journal*, 36: 305–15.
- Easteal, S. (1999) 'Molecular evidence for the early divergence of placental mammals', *BioEssays*, 21: 1052–8.
- Eizirik, E., Murphy, W.J. and O'Brien, S.J. (2001) 'Molecular dating and biogeography of the early placental mammal radiation', *Journal of Heredity*, 92: 212–9.
- Fara, E. (2002) 'Sea-level variations and the quality of the continental fossil record', *Journal of the Geological Society, London*, 159: 489–91.
- Fara, E. and Benton, M.J. (2000) 'The fossil record of Cretaceous tetrapods', *Palaaios*, 15: 161–5.
- Feduccia, A. (1996) *The Origin and Evolution of Birds*, New Haven: Yale University Press.
- Foote, M. (1997) 'Estimating taxonomic durations and preservation probability', *Paleobiology*, 23: 278–300.
- Foote, M. and Raup, D.M. (1996) 'Fossil preservation and the stratigraphic ranges of taxa', *Paleobiology*, 22: 121–40.
- Foote, M. and Sepkoski, J.J. Jr. (1999) 'Absolute measures of the completeness of the fossil record', *Nature*, 398: 415–7.
- Foote, M., Hunter, J.P., Janis, C. and Sepkoski, J.J. Jr. (1999) 'Evolutionary and preservational constraints on origins of biologic groups, divergence times of eutherian mammals', *Science*, 283: 1310–4.
- Hallam, A. (1986) 'The Pliensbachian and Tithonian extinction events', *Nature*, 319: 765–8.
- Hallam, A. and Wignall, P.B. (1997) *Mass Extinctions and their Aftermath*, Oxford: Oxford University Press.
- Harlid, A. and Arnason, U. (1999) 'Analyses of mitochondrial DNA nest ratite birds within the Neognathae, supporting a neotenus origin of ratite morphological characters', *Proceedings of the Royal Society, London*, B266: 305–9.
- Heckman, D.S., Geiser, D.M., Eidell, B.R., Stauffer, R.L., Kardos, N.L. and Hedges, S.B. (2001) 'Molecular evidence for the early colonization of land by fungi and plants', *Science*, 293: 1129–33.
- Hedges, S.B. and Kumar, S. (1999) 'Technical comments: divergence times of eutherian mammals', *Science*, 285: 2031a.
- Hedges, S.B., Parker, P.H., Sibley, C.G. and Kumar, S. (1996) 'Continental breakup and the ordinal diversification of birds and mammals', *Nature*, 381: 226–9.
- Hitchin, R. and Benton, M.J. (1997) 'Congruence between parsimony and stratigraphy, comparisons of three indices', *Paleobiology*, 23: 20–32.
- Huelsenbeck, J.P. (1994) 'Comparing the stratigraphic record to estimates of phylogeny', *Paleobiology*, 20: 470–83.
- Huelsenbeck, J.P. and Rannala, B. (1997) 'Phylogenetic methods come of age, testing hypotheses in an evolutionary context', *Science*, 276: 227–32.
- Jackson, J.B.C. and Johnson, K.G. (2001) 'Measuring past biodiversity', *Science*, 293: 2401–4.
- Ji, Q., Luo, Z., Yuan, C.-X., Wible, J.R., Zhang, J.-P. and Georgi, J.A. (2002) 'The earliest known eutherian mammal', *Nature*, 416: 816–22.
- Kenrick, P. (1999) 'The family tree flowers', *Nature*, 402: 358–9.
- Kenrick, P. and Crane, P.R. (1997) 'The origin and early evolution of plants on land', *Nature*, 389: 33–9.
- Klicka, J. and Zink, R.M. (1997) 'The importance of Recent ice ages in speciation, a failed paradigm', *Science*, 277: 1666–9.

- Kumar, S. and Hedges, S.B. (1998) 'A molecular timescale for vertebrate evolution', *Nature*, 392: 917–20.
- Lee, M.S.Y. (1999) 'Molecular clock calibrations and metazoan divergence dates', *Journal of Molecular Evolution*, 49: 385–91.
- Little, C.T.S. and Benton, M.J. (1995) 'Early Jurassic mass extinction: a global long-term event', *Geology*, 23: 495–8.
- Macarthur, R.H. and Wilson, E.O. (1967) *The Theory of Island Biogeography*, Princeton: Princeton University Press.
- Marshall, C.R. (1990) 'Confidence intervals on stratigraphic ranges', *Paleobiology*, 16: 1–10.
- (1997) 'Confidence intervals on stratigraphic ranges with nonrandom distribution of fossil horizons', *Paleobiology*, 23: 165–73.
- (1999) 'Fossil gap analysis supports early Tertiary origin of trophically diverse avian orders, Comment and Reply', *Geology*, 27: 95.
- Miller, A.I. (1998) 'Biotic transitions in global marine diversity', *Science*, 281: 1157–60.
- Murphy, W.J., Eizirik, E., O'Brien, S.J., Madsen, O., Scally, M., Douady, C.J., Teeling, E., Ryder, O.A., Stanhope, M.J., de Jong, W.W. and Springer, M.S. (2001) 'Resolution of the early placental mammal radiation using Bayesian phylogenetics', *Science*, 294: 2348–51.
- Nei, M., Xu, P. and Glazko, G. (2001) 'Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms', *Proceedings of the National Academy of Sciences, USA*, 98: 2497–502.
- Niklas, K.J., Tiffney, B.H. and Knoll, A.H. (1983) 'Patterns in vascular land plant diversification', *Nature*, 303: 614–6.
- Norell, M.A. (1992) 'Taxic origin and temporal diversity, the effect of phylogeny', in M.J. Novacek and Q.D. Wheeler (eds) *Extinction and Phylogeny*, New York: Columbia University Press, pp. 89–118.
- Norell, M.A. and Novacek, M.J. (1992) 'The fossil record and evolution, comparing cladistic and paleontologic evidence for vertebrate history', *Science*, 255: 1690–3.
- Omland, K.E. (1997) 'Correlated rates of molecular and morphological evolution', *Evolution*, 51: 1381–93.
- Paton, T., Haddrath, O. and Baker, A.J. (2002) 'Complete mitochondrial DNA genome sequences show that modern birds are not descended from transitional shorebirds'. *Proceedings of the Royal Society, London*, B269: 839–46.
- Patterson, C. (1982) 'Morphological characters and homology', in C. Patterson, (ed.) *Problems of Phylogenetic Reconstruction*, Systematics Association Special Volume 21, London: Academic Press, pp. 21–74.
- Peters, S.E. and Foote, M. (2001) 'Biodiversity in the Phanerozoic, a reinterpretation', *Paleobiology*, 27: 583–601.
- (2002) 'Determinants of extinction in the fossil record', *Nature*, 416: 420–4.
- Platnick, N.L. (1979) 'Philosophy and the transformation of cladistics', *Systematic Zoology*, 28: 537–46.
- Rambaut, A. and Bromham, L. (1998) 'Estimating divergence dates from molecular sequences', *Molecular Biology and Evolution*, 15: 442–8.
- Raup, D.M. (1972) 'Taxonomic diversity during the Phanerozoic', *Science*, 177: 1065–71.
- Raup, D.M. and Sepkoski, J.J. Jr. (1982) 'Mass extinctions in the marine fossil record', *Science*, 215: 1501–3.
- (1984) 'Periodicity of extinctions in the geologic past', *Proceedings of the National Academy of Sciences, USA*, 81: 801–5.
- Sepkoski, J.J. Jr. (1984) 'A kinetic model of Phanerozoic taxonomic diversity. III. Post-Paleozoic families and mass extinctions', *Paleobiology*, 10: 246–67.
- (1996) 'Patterns of Phanerozoic extinction, a perspective from global data bases', in O.H. Walliser (ed.) *Global Events and Event Stratigraphy*, Berlin: Springer, pp. 35–51.

- Sepkoski, J.J. Jr., Bambach, R.K., Raup, D.M. and Valentine, J.W. (1981) 'Phanerozoic marine diversity and the fossil record', *Nature*, 293: 435–7.
- Siddall, M.E. (1996) 'Stratigraphic consistency and the shape of things', *Systematic Biology*, 45: 111–5.
- (1998) 'Stratigraphic fit to phylogenies: A proposed solution', *Cladistics*, 14: 201–8.
- Siddall, M.E. and Whiting, M.F. (1999) 'Long-branch abstractions', *Cladistics*, 15: 9–24.
- Smith, A.B. (2001) 'Large-scale heterogeneity of the fossil record, implications for Phanerozoic biodiversity studies', *Philosophical Transactions of the Royal Society, London*, B356: 1–17.
- Smith, A.B. and Littlewood, D.T.J. (1994) 'Paleontological data and molecular phylogenetic analysis', *Paleobiology*, 20: 259–73.
- Smith, A.B. and Patterson, C. (1988) 'The influence of taxonomic method on the perception of patterns of evolution', *Evolutionary Biology*, 23: 127–216.
- Smith, A.B. and Peterson, K.J. (2002) 'Dating the time of origin of major clades, molecular clocks and the fossil record', *Annual Reviews in Ecology and Systematics*, 30: 65–88.
- Smith, A.B., Gale, A.S. and Monks, N. (2001) 'Sea-level change and rock record bias in the Cretaceous: a problem for extinction and biodiversity studies'. *Paleobiology*, 27: 241–53.
- Springer, M.S. (1997) 'Molecular clocks and the timing of the placental and marsupial radiations in relation to the Cretaceous–Tertiary boundary', *Journal of Mammalian Evolution*, 4: 285–302.
- Strauss, D. and Sadler, P.M. (1989) 'Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges', *Mathematical Geology*, 21: 411–27.
- Tavaré, S., Marshall, C.R., Will, O., Soligo, C. and Martin, R.D. (2002) 'Using the fossil record to estimate the age of the last common ancestor of extant primates', *Nature*, 416: 726–9.
- Valentine, J.W. (1969) 'Patterns of taxonomic and ecological structure of the shelf benthos during Phanerozoic time', *Palaeontology*, 12: 684–709.
- Valentine, J.W., Jablonski, D. and Erwin, D.H. (1999) 'Fossils, molecules and embryos, new perspectives on the Cambrian explosion', *Development*, 126: 851–9.
- van Tuinen, M. and Hedges, S.B. (2001) 'Calibration of avian molecular clocks', *Molecular Biology and Evolution*, 18: 206–13.
- van Tuinen, M., Sibley, C.G. and Hedges, S.B. (2000) 'The early history of modern birds inferred from DNA sequences of nuclear and mitochondrial ribosomal genes', *Molecular Biology and Evolution*, 17: 451–7.
- Vermeij, G.J. (1996) 'Animal origins', *Science*, 274: 525–6.
- Wagner, P.J. (1998) 'Phylogenetic analyses and the quality of the fossil record', in S.K. Donovan and C.R.C. Paul (eds) *The Adequacy of the Fossil Record*, New York: Wiley, pp. 165–87.
- (2000a) 'The quality of the fossil record and the accuracy of phylogenetic inferences about sampling and diversity', *Systematic Biology*, 49: 65–86.
- (2000b) 'Exhaustion of morphologic character states among fossil taxa', *Evolution*, 54: 365–86.
- Weishampel, D.B. (1990) 'Dinosaurian distribution', in D.B. Weishampel, P. Dodson and H. Osmólska (eds) *The Dinosauria*, Berkeley: University of California Press, pp. 63–139.
- (1996) 'Fossils, phylogeny, and discovery, a cladistic study of the history of tree topologies and ghost lineage durations', *Journal of Vertebrate Paleontology*, 16: 191–7.
- Whelan, S., Liò, P. and Goldman, N. (2001) 'Molecular phylogenetics: state-of-the-art methods for looking at the past'. *Trends in Genetics*, 17: 262–72.
- Wikström, N., Savolainen, V. and Chase, M.W. (2001) 'Evolution of the angiosperms, calibrating the family tree', *Proceedings of the Royal Society, London*, B268: 2211–20.
- Wills, M.A. (1999) 'The gap excess ratio, randomization tests, and the goodness of fit of trees to stratigraphy', *Systematic Biology*, 48: 559–80.



90 Michael J. Benton

- (2001) 'How good is the fossil record of arthropods? An assessment using the stratigraphic congruence of cladograms', *Geological Journal*, 36: 187–210.
- (2002) 'The tree of life and the rock of ages, are we getting better at estimating phylogeny?', *BioEssays*, 24: 203–7.
- Wray, G.A. (2001) 'Dating branches on the Tree of Life using DNA', *Genome Biology*, 3: 1–7.
- Wray, G.A., Levinton, G.S. and Shapiro, L.H. (1996) 'Molecular evidence for deep Precambrian divergences among metazoan phyla', *Science*, 274: 568–73.

