

Fossil Record: Quality

Michael J Benton, *University of Bristol, Bristol, UK*

There are divergent opinions about the quality of the fossil record and its ability to give a useful representation of the history of life. Extensive testing, by comparisons of the congruence of morphological and molecular phylogenies, suggests that the fossils do tell the story relatively well.

Introduction

The value of the fossil record in giving a clear account of evolutionary history has been questioned, and doubts have arisen because of questions of quality. Charles Darwin hoped that, over time, more and more fossils would be found which would fill in all the 'missing links' and give a full picture of the history of life. He argued that fossils could be positioned on evolutionary trees, or phylogenies, according to the two main axes of time and morphological evidence of relationships. However, this early hope was to be disappointed. Gould (1983) has argued that the various anniversaries of publication of Darwin's *On the Origin of Species by Means of Natural Selection* provide benchmarks of major changes in the opinion of evolutionary biologists, from rejection of palaeontology, to subservient partnership, to full partnership with the study of modern organisms. **See also:** Darwin, Charles Robert; Fossil record; History of palaeontology

By 1909, all Darwin's expectations that the fossil record would reveal the pattern of evolution had been forgotten. More and more fossils had been found, but they had not yet allowed phylogeny to be read directly. Evolutionary biologists of 1909 barely mentioned fossils or palaeontology in their analysis of the state of Darwinism after 50 years. The newly rediscovered science of genetics, together with other branches of biology, were making all the running. By 1959, the hundredth anniversary of the *Origin*, Gould (1983) pointed to a reintegration of palaeontology into the useful evolutionary sciences. This had been the result of the 'modern synthesis' of the 1930s and 1940s, when biologists had reassessed Darwin's original views, and combined them with the new laboratory discoveries of genetics. At the same time, G.G. Simpson provided a powerful new quantitative approach to palaeontological data, and he showed how evolution over the long term could be analysed from certain fossil sequences at least. Fossils were back in the fold of acceptable tools for the study of evolution, but their role was clearly subservient to information derived from living organisms. Gould (1983) then documented that, by the time he wrote, palaeontology had come into its own again, since most evolutionary biologists accepted that there were patterns and processes of evolution (sometimes grouped together under the heading 'macro-evolution') that took place on long time scales and which

could not be predicted from laboratory experiments. Hence, fossils can offer original insights. **See also:** Simpson, George Gaylord

This change of attitudes has been the inevitable result of research work since 1970 that has shown some of the possibilities of large-scale, analytical palaeobiological studies. These studies have shown how many good-quality sectors of the fossil record there are, case studies where sometimes hundreds of thousands of specimens may be measured, and their patterns of change assessed in well-dated rock sequences over long spans of time. Work of this kind is published widely, in journals such as *Paleobiology*, *Evolution*, *Nature* and *Science*. The debate has re-emerged, with some critics (Peters and Foote, 2001; Smith, 2001) suggesting that much of the shape of the fossil record is driven by sea-level changes. The suggestion is that when sea levels are high, more marine rocks are deposited, and hence more marine fossils are preserved. The corollary, supposedly, is that terrestrial deposition should be somehow the opposite: when sea levels are high, terrestrial deposits (on land, in rivers, in lakes) should be fewer and so there should be a drop in terrestrial fossils. This has been disproved, for the Cretaceous at least, by Fara (2002) that the abundance and diversity of vertebrates on land is in no way linked to sea-level curves, and Smith (2001) indeed accepts that the pattern of diversification of life in the sea in the past 250 Myr is opposite to expectations from sea-level curves: marine diversity goes up dramatically while sea level falls in the long term.

A lesser claim is that fossils should be ignored in the reconstruction of broad-scale patterns of evolution (phylogenies). In particular, cladistic analyses should be based solely on living forms, since the fossils are incomplete morphologically, and they represent an incomplete sample of all the organisms that ever lived. This attitude has been opposed by palaeontologists and by many biologists on the grounds that (1) some evidence is better than none; (2) fossils include a sample of the majority of species that have ever lived; (3) most fossils represent morphologies that are quite unknown today, and they greatly enrich the content of a phylogeny; (4) fossils may be placed more or less pre-

Advanced article

Article Contents

- Introduction
- Preservation Biases
- Study Biases
- Measures of Completeness
- Comparisons of Sectors of the Fossil Record
- Changes in Knowledge through Time

doi: 10.1038/ngp.els.0004144

cisely in time, which provides good cross-evidence for the order and age of branching points of genealogical relationships, and that, for many groups, fossils in practice offer as much morphological data as do museum specimens of modern representatives. **See also:** Cladistics

Preservation Biases

The incompleteness of the fossil record may be ascribed to many factors, of the organisms themselves, of their habitats, of later changes within the lithifying rocks, and of the ways in which palaeontologists work. The biggest problems are usually indicated as biological and preservational.

At the biological level, it is clear that soft-bodied organisms are less likely to be preserved than those with hard parts. The more hard parts an organism has, the more elaborate the skeleton, the more likely it is to survive decay, scavenging and transport. Long-living, rare, organisms are less likely to die and be preserved than short-lived common organisms with skeletons. The commonest fossils are often small *r*-selected organisms that have a high turnover rate, with many generations per year. However, large organisms have a greater preservation potential than small ones, because they can survive incarceration in fine- and coarse-grained sediments, and because they are easier to find. As for habitats, organisms that fly, or live in trees, are less often preserved than those that lurk around ponds and rivers, or live on the sea bed. On the whole, organisms are preserved in their habitats, or close to them, and most sedimentary rocks are laid down in shallow seas or in rivers and lakes. **See also:** Fossils and fossilization

Geological factors are also important. As soon as a carcass enters the sedimentary cycle, the processes of decay, dissolution and destruction take place. Soft tissues, and frequently hard tissues, may be lost to chemical and microbial action in soil or water. Carcasses may be battered and disarticulated by water or wind transport. Even after burial, soft sediments move and recycle through the system, so any enclosed skeletal remains may be further disrupted. After the sediments become cemented into rocks, further chemical and physical transformations may take place, some of them potentially injurious to the embedded microfossils. Pore waters may contain acids that dissolve shells, or the weight of overlying sediment may crush and distort the remains. In the longer term, geological processes have a propensity for causing further damage. Whole areas of the Earth's crust may be subducted (drawn under other crustal plates), metamorphosed (heated or crushed by deep burial), or eroded on exposure to water or the air. The older the rocks in which one seeks fossils, the more likely they are to have suffered some or all of these indignities.

This catalogue of disaster might suggest that the quality of the fossil record is indeed appalling, and that it might be

better to dismiss it out of hand. However, the bulging drawers of palaeontological museum collections around the world, and the wealth of intricate anatomical detail described in the palaeontological journals, show that all is not lost. It has to be assumed that considerably more than 99.99% of all organisms that ever lived are lost to the fossil record. This startling fact is mitigated by the observation that one needs considerably fewer than 0.01% of all organisms that ever lived in order to draw a complete picture of the life of the past. It is not necessary to examine every individual of a species, living or extinct, in order to know the main features of the group of which it is a representative.

The problems associated with assessing the quality of the fossil record may be dissected into three categories of concern, namely: (1) missing species (2) missing phyla (3) missing parts.

The first and the last categories can be resolved, or ignored, to some extent. New species of well-known extinct groups, such as dinosaurs or ammonites, turn up all the time, and, on the whole, these new species are very like other known taxa, and they fit well into the established scheme. If a certain dinosaur or ammonite species has been entirely lost, and there is not a single specimen anywhere in the rock record, then the broad-scale picture of evolution will not be much affected. Likewise, missing parts of ancient organisms need not be too catastrophic. Comparison with modern relatives may allow palaeontologists to make informed guesses about missing soft tissues. For example, it seems certain that dinosaurs had muscles and guts pretty much like modern crocodiles and birds. New discoveries of exceptionally preserved dinosaur specimens (e.g. Del Sasso and Signore, 1998) confirm that dinosaurs had striated muscles, a large intestine, a liver, lungs, a trachea, and so on. **See also:** Dinosaur feeding; Dinosauria (dinosaurs)

Missing phyla are a serious problem. Could there be whole groups of plants and animals of which palaeontologists are entirely unaware? Perhaps there were highly important, entirely soft-bodied creatures that dominated their ecosystems at times in the past, but which have left absolutely no trace. Certainly, studies of the 500 million-year-old Burgess Shale faunas from the Cambrian of Canada, and other parts of the world, have shown evidence for some bizarre animals that were unknown from other, more typical, fossil deposits of the same age. Such sites of exceptional preservation do, however, reveal some of the unusual, otherwise unknown, organisms. These remarkable fossil deposits, scattered randomly through time and space and with no deterioration back in time, provide a corrective, or standard, against which the more typical fossil deposits may be compared. However, palaeontologists can never confidently prove that there was never a phylum of purple-coloured animals that lived in Jurassic seas, moving about on fleshy wheels, and feeding on dinosaurs which they snatched off the beaches using a long flexible trunk. **See also:** Burgess Shale; Cambrian radiation

Study Biases

Human factors are also very important; our knowledge of the fossil record depends critically on the interest people have in particular groups, their geographic location, and their ease of study; these variables have been quantified as 'palaeontologic interest units'. It is self-evident that the fossils of North America and Europe are better known than those of China and Australia. Indeed, it is also clear that many more palaeontologists work on certain time periods and on certain groups. Hence, we know more about the fossils of the past 2 Myr than of the Cambrian, 500 Myr ago. Also, we know more about human fossils and about dinosaurs than about fossil sponges or conifers, since many more people work in the first two fields. These biases can be assessed and corrected to some extent. **See also:** Human evolution: overview

is generally not done. It is more generally done by reference to the condition of the characters in presumed relatives of the group under investigation. **See also:** Phylogeny and stratigraphy comparison

Further, it could be argued that morphological and molecular data are independent of each other: despite the obvious link between genotype and phenotype, it is not clear that genes map directly to phylogenetically informative morphological characters. In addition, each molecule is broadly independent of the others. Thus, it is possible to compare phylogenies based on morphological data, genome data, globins, cytochrome *c*, and other proteins with each other, but more importantly, with independent stratigraphic data. **See also:** Molecular phylogeny reconstruction

Benton and Hitchin (1997) have suggested that this kind of approach offers a unique way of testing history. The congruence-testing approach can be used in specific cases, to assess which phylogenetic hypotheses agree with each other, and which fit the known fossil record best. However, when taken in aggregate, as is done here, and tested as a population sample, more general questions may be addressed, without worrying about minor analytical errors and misidentifications. If the phylogenetic information derived from independent data sets agrees, then the methods are probably finding the correct phylogenies; if there are widespread disagreements, then something is wrong with one or more of the data sets. (Agreement could also mean that all data sources are pointing towards uniformly incorrect conclusions, but that view would be hard to sustain unless it could be shown that the supposedly 'independent' data sources are in fact linked.)

Measures of Completeness

By 1866, some of the first phylogenetic trees based explicitly on evolution were published by the German biologist Ernst Haeckel. Until recently, phylogenetic trees were composed by taking account not only of the morphology of organisms, but also of their place in geological time. Hence it was not possible to test the nature of phylogenies and the fossil record independently. However, the development of cladistic techniques has opened up the possibility of testing the pattern of evolution. Cladograms based on morphological or molecular data involve no direct measure of the age of fossils: fossils can be included as terminal taxa, side by side with living taxa. The divorce of phylogeny reconstruction from stratigraphic evidence opens up exciting possibilities for testing the two sets of evidence against each other, and for moving to fill some of the inevitable gaps in the fossil record. **See also:** Cladistics; History of palaeontology

Congruence-testing methods

There are a variety of metrics for comparing phylogenies and fossil records (**Figure 1**); Spearman rank correlation (SRC), the stratigraphic consistency index (SCI), the relative completeness index (RCI) and the gap excess ratio (GER). 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 (1992). The SRC test, however, is limited in this context, since symmetrical trees have to be collapsed to produce an entirely unbalanced (single stem lineage) tree, as in **Figure 1(b)**.

Independent data sets

A number of authors have argued that phylogenetic data and stratigraphic data are independent of each other, and that they both offer evidence about the historical shape of phylogenetic trees (Gauthier *et al.*, 1988; Norell and Novacek, 1992; Benton and Storrs, 1994; Benton and Hitchin, 1997; Benton *et al.*, 2000). Morphological characters, as used in cladistic reconstruction of phylogeny, are determined solely by inspection of the organisms, whether living or extinct, and their polarities (primitive → derived) are determined by reference to evidence of distribution amongst the taxa being studied and not to stratigraphic age. Indeed, cladistic reconstruction is frequently done without reference to polarity or relative stratigraphic age. Trees could be rooted by choosing the oldest fossil in the analysis, but that

The SCI was proposed by Huelsenbeck (1994) to assess 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.

The RCI was proposed (Benton and Storrs, 1994) to take into account 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.

The gap excess ratio (GER) was developed by Wills (1999) as a modification of the RCI that gives an indication of where the value sits among all possible values. The GER compares, for the cladogram in question, the best possible and worst possible values (i.e. best and worst congruence with stratigraphy), and then places the actual value of the metric in that context. This takes account to a large extent of the confounding influences of tree size and shape upon the possible values, and allows the analyst to focus on the issue of how well phylogeny and stratigraphy match. **See also:** Speciation and the fossil record

Results of congruence testing

Congruence testing has been applied to a variety of questions in phylogeny reconstruction. The first results of congruence testing were encouraging: Norell and Novacek (1992) found that 18 out of 24 test cases of cladograms of vertebrates (75%) gave statistically significant ($P < 0.05$) correlations of clade with age data, using the SRC test, while Benton and Storrs (1994) found significant correlation in 41 of 74 test cases (55%). Subsequent assessments, however, based on a larger sample (384 cladograms, composed of 174 cladograms of tetrapods, 147 cladograms of fishes, and 63 cladograms of echinoderms (Benton and Hitchin, 1997)), provided more disappointing results. For all cladograms in the test sample, 148 out of 384 (38%) showed significant SRC values. These results could indicate poor congruence, or they could simply highlight the fact that the SRC test is rather crude, simply comparing the raw order of points, and taking no account of their actual spacing in time, nor of the degree of mismatch.

Much better results were obtained with the RCI and the SCI metrics, which measure different aspects of cladogram and fossil record quality. For all three groups assessed (tetrapods, fishes, echinoderms), most cladograms have RCI values equal to, or greater than, 50% (Benton and

Hitchin, 1997), with a mean pass rate of 78%. The pass rates are similarly favourable for the SCI measure. In these cases, all three sets of cladograms have significantly more than half their nodes showing stratigraphic consistency than inconsistency, with a mean pass rate of 82% for all cladograms.

Comparisons of Sectors of the Fossil Record

These congruence-testing techniques allow firm quantitative assessments of differences among major groups of plants and animals and among habitats. Many assertions about such differentiation have been made, for example, that the terrestrial fossil record is worse than the marine, that vertebrates have a poorer fossil record than skeletonized marine invertebrates, and so on.

Benton and Hitchin (1997) assessed large samples of cladograms of echinoderms, fishes and tetrapods. For the RCI metric, they found pass rates of 78% for echinoderms, 84% for fishes and 74% for tetrapods, suggesting that tetrapods have the poorest record, and fishes the best, although the differences in values are marginal. The SCI metric indicated pass rates of 95% for echinoderms, 69% for fishes and 87% for tetrapods, a rather different result; here, fishes have the poorest quality records, while echinoderms have by far the best. These tests show no clear differentiation in the quality of the fossil record by major taxonomic groups. The comparisons show that none of them consistently has a better fossil record, or better cladistic resolution, than the others. Each of the animal groups performed best with one of the metrics: tetrapods with SRC, fishes with RCI and echinoderms with SCI.

There is no strong evidence for differentiation by broad habitats either. Benton and Hitchin (1997) 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 SCI metrics. Cladograms of marine organisms scored higher values with the RCI metric. These results 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. Further studies in this area must be carried out before a firm conclusion can be drawn.

Comparisons have also been made between morphological and molecular phylogenies. Tests of a sample of 206 cladograms of mammals have shown (Benton, 1998) that morphological data seem to predict phylogenies that match the known fossil record better than molecular trees. Some published trees showed excellent matching with fossil-record data, others showed almost no correspondence whatsoever. Morphological trees were nearly twice as good

as molecular trees in terms of matching of the rank orders of nodes and oldest fossils (SRC test), while morphological trees were 10% better than molecular in terms of stratigraphic consistency of the nodes (SCI metric). The ratios of ghost range to known range are lower for molecular trees (RCI metric). Among the molecular trees, those based on gene data were considerably better than those based on protein sequences, at least in terms of the rank order of nodes, and the stratigraphic consistency of nodes. Protein trees, however, were best of all in terms of minimizing the proportion of ghost range. These findings probably indicate real phenomena, but the match of molecular trees to the expectations of stratigraphy may improve as the study of molecular phylogeny matures. **See also:** Molecular phylogeny reconstruction

Changes in Knowledge through Time

If the fossil record is a hopelessly biased representation of the true picture of the history of life, new discoveries should randomly modify the perceived patterns. Occasionally, palaeontologists ought to find astonishing new organisms, or fossils of quite unexpected ages, perhaps a human fossil in the Eocene, 50 Myr ago, or a Jurassic penguin, 150 Myr old, or a Permian dinosaur, 270 Myr old.

Studies have been made of how changes in palaeontological knowledge affect perceptions of macroevolutionary patterns. These have shown that, although fossil records change substantially through research time, the macroevolutionary conclusions alter relatively little. Maxwell and Benton (1990) compared the state of knowledge of the fossil record of tetrapods over the past 100 years, the time of the greatest rate of growth in palaeontological knowledge. The gross number of taxa increased as a result of new finds, but other aspects of the stratigraphic range lists changed in nonsystematic and unpredictable ways. Revisions of stratigraphy, revisions of taxonomy at the alpha level and broad-scale cladistic revisions of larger groups all affected the taxon range lists, but some ranges were shortened, some increased, and some remained unaltered. Some families and genera disappeared as a result of taxonomic revision and others appeared, so that there was no overall shift in the results. The main change that was detected in a comparison of a 1967 database (Harland *et al.*, 1967) and one from 20 years later (Benton, 1993) was that familial durations of tetrapods had increased marginally (29.1% of families with unchanged range lengths, 44.8% with increased range lengths, 26.1% with decreased range lengths). Increased range lengths would be predicted by the simple rallying-cry that 'more palaeontological research means collecting more fossils', but the fact that more than one-quarter of the 515 families analysed showed decreased range lengths was more of a surprise. This was the result of cladistic redefinitions of families, and the removal

of suspect ancestral taxa, which lacked autapomorphies of the family, from the bottoms of ranges.

Maxwell and Benton (1990) found, however, that although stratigraphic ranges of 70.9% of the tetrapod families had changed in a span of 20 years of research, the macroevolutionary conclusions derived from the databases altered little. The phases of diversification and of extinction remained the same (Figure 2a). Rates of origination and extinction at particular times also remained in proportion. The main effect to be seen was an increase in overall diversity through much of the past 400 Myr of tetrapod evolution, and slight sharpening up of extinction events.

Sepkoski (1993), in a similar study of the past 10 years of change in his marine animals database, found turnover in 50% of the data: families had been added and deleted, low-resolution stratigraphic data had been improved, and dates of apparent origination and extinction of families had been altered. However, despite these changes, plots of diversification remained nearly identical but for the fact that the 1992 curve lies 13% higher than the older one (Figure 2b). Furthermore, the identity and magnitudes of extinction events remained the same. The main changes were that family range lengths tended to increase, with an approximately equal distribution of earlier originations and later extinctions, and that mass extinctions became slightly sharper, with extinctions shifting closer to the boundaries. Hence, these tests have indicated two facts: (1) palaeontological knowledge is changing (advancing, one hopes) rapidly; and (2) despite this change, broad-scale macroevolutionary patterns remain stable. If the fossil record were hopelessly incomplete, the patterns derived therefrom would probably change wildly as a result of new discoveries.

A more detailed study has been made of palaeontological knowledge at different points in research time. Benton and Storrs (1994) compared the goodness of fit of cladograms of tetrapods against summaries of palaeontological knowledge from 1967 and 1993 (Harland *et al.*, 1967; Benton, 1993). The idea was to find out whether 26 years of palaeontological collecting and systematic study had improved the matching of phylogenetic and stratigraphic data. Two measures of fossil record quality were used: (1) the correlation of clade rank and age rank, measured using the SRC statistic (with quality of correlation assessed at confidence levels of $P < 0.05$ and $P < 0.01$); and (2) the RCI of all test cladograms when plotted as phylogenies against a geological time scale (measured as the proportion of minimum implied gaps (MIGs), indicated by branching points of pairs of sister groups, to simple range lengths (SRLs), based on total range lengths represented by fossils). The results (Figure 3) showed improvements in the goodness of fit of palaeontological and cladistic data between 1967 and 1993. Of the 71 cladograms that could be compared directly, 46 showed an increase in the RCI, one remained constant and 26 showed a reduction. This indicates a sta-

tistically significant (χ^2 test, $P < 0.05$) increase in the quality of the known fossil records of the sample of 71 cladograms of vertebrates in terms of their fit to the control cladograms. However, there was no evidence for change in the degree of correlation of clade rank and age rank data when the 1967 and 1993 data sets were compared for all 71 comparable cladograms; 43 of the 71 comparisons showed no change of state, while 28 changed, but the changes were equally balanced, with 14 'improvements' (i.e. negative to positive correlation; insignificant to significant correlation, at values of $P < 0.05$ or 0.01 ; correlation at $P < 0.05$ to correlation at $P < 0.01$) and 14 deteriorations.

Quality of the fossil record: a perspective from 2004

There are currently two viewpoints about the quality of the fossil record. Although, in 1990 or 1995, most palaeontologists were confident that the fossil record is broadly a biological signal, there are now critics who believe there is a great deal of geological signal to confound us. The degree to which these critics reject the fossil record as a useful source of information on the history of life varies, from apparently outright rejection (Peters and Foote, 2001) to a more measured view (Smith, 2001). Aspects of their evidence have been criticized (Fara, 2002), but more investigation is required. The core question here is: does the fossil record follow changes in sea level closely, or is it good enough to show an evolutionary signal that would allow us to study diversification and extinction patterns?

Benton *et al.* (2000) gave a strong message, from a large-scale study of congruence, that the fossil record is adequate. They summarized the results from a sample of 1000 cladograms of plants and animals, marine and nonmarine, through geological time. The 1000 cladograms were grouped in five time bins, representing divisions of the past 500 Myr, and the congruence levels assessed. If the fossil record diminished in quality backwards in time, congruence should, on average, be poorer for the more ancient time samples. This was not the case; mean congruence levels remained constant through time, and this result held also for individual clades, such as arthropods, fishes and tetrapods. Hence, at the taxonomic scale of families and the stratigraphic scale of stages (5–6 Myr duration), the fossil record apparently shows constant quality through the Phanerozoic.

Published: March 2005

Benton, Michael J (2005) Fossil Record: Quality. In: *Encyclopedia of Life Sciences*. Chichester: John Wiley & Sons, Ltd. <http://www.els.net/> [doi: 10.1038/ngp.els.0004144]

References

- Benton, MJ (1993) *The Fossil Record 2*. London: Chapman & Hall.
- Benton, MJ (1998) Molecular and morphological phylogenies of mammals: congruence with stratigraphic data. *Molecular Phylogenetics and Evolution* **9**: 398–407.
- Benton, MJ and Hitchin, R (1997) Congruence between phylogenetic and stratigraphic data on the history of life. *Proceedings of the Royal Society, London, Series B* **264**: 885–890.
- Benton, MJ and Storrs, GW (1994) Testing the quality of the fossil record: paleontological knowledge is improving. *Geology* **22**: 111–114.
- Benton, MJ; Wills, M and Hitchin, R (2000) Quality of the fossil record through time. *Nature* **403**: 534–537.
- Del Sasso, C and Signore, M (1998) Exceptional soft-tissue preservation in a theropod dinosaur from Italy. *Nature* **392**: 383–387.
- Fara, E (2002) Sea level variations and the quality of the continental fossil record. *Journal of the Geological Society, London* **159**: 489–491.
- Gauthier, J; Kluge, AG and Rowe, T (1988) Amniote phylogeny and the importance of fossils. *Cladistics* **4**: 105–209.
- Gould, SJ (1983) Irrelevance, submission and partnership: the changing role of palaeontology in Darwin's three centennials and a modest proposal for macroevolution. In: Bendall DS (ed.) *Evolution from Molecules to Men*, pp. 347–366. Cambridge, UK: Cambridge University Press.
- Harland, WB; Holland, CH; House, MR *et al.* (1967) *The Fossil Record; a Symposium with Documentation*. London: Geological Society of London.
- Huelsenbeck, JP (1994) Comparing the stratigraphic record to estimates of phylogeny. *Paleobiology* **20**: 470–483.
- Maxwell, WD and Benton, MJ (1990) Historical tests of the absolute completeness of the fossil record of tetrapods. *Paleobiology* **16**: 322–335.
- Norell, MA and Novacek, MJ (1992) The fossil record and evolution: comparing cladistic and paleontologic evidence for vertebrate history. *Science* **255**: 1690–1693.
- Peters, SE and Foote, M (2001) Biodiversity in the Phanerozoic, a re-interpretation. *Paleobiology* **27**: 583–601.
- Sepkoski, JJ Jr (1993) Ten years in the library: how changes in taxonomic data bases affect perception of macroevolutionary pattern. *Paleobiology* **19**: 43–51.
- Smith, AB (2001) Large-scale heterogeneity of the fossil record, implications for Phanerozoic biodiversity studies. *Philosophical Transactions of the Royal Society* **B356**: 1–17.
- Wills, MA (1999) The gap excess ratio, randomization tests, and the goodness of fit of trees to stratigraphy. *Systematic Biology* **48**: 559–580.

Further Reading

- Benton, MJ (1994) Palaeontological data, and identifying mass extinctions. *Trends in Ecology and Evolution* **9**: 181–185.
- Benton, MJ (2003) The quality of the fossil record. In: Donoghue, PCJ and Smith, MP (eds) *Telling the Evolutionary Time: Molecular Clocks and the Fossil Record*, pp. 66–90. New York: CRC Press.
- Donovan, SK and Paul, CRC (eds) (1998) *The Adequacy of the Fossil Record*. New York: Wiley..
- Norell, MA (1992) Taxic origin and temporal diversity: the effect of phylogeny. In: Novacek, MJ and Wheeler, QD (eds) *Extinction and Phylogeny*, pp. 89–118. New York: Columbia University Press.
- Patterson, C (1981) Significance of fossils in determining evolutionary relationships. *Annual Review of Ecology and Systematics* **12**: 195–223.
- Raup, DM (1972) Taxonomic diversity during the Phanerozoic. *Science* **215**: 1065–1071.

Smith, AB (1994) *Systematics and the Fossil Record*. Oxford: Blackwell Scientific.

Accepted date: February 2005

Concept: Method and Techniques

Top Terms: Evolution and Diversity of Life

Keywords: fossils#quality#completeness

Author's name: Benton, Michael J

Glossary

Autapomorphy#A derived character, a unique character possessed by a taxon.

Cladistics#The study of relationships of organisms by the search for synapomorphies.

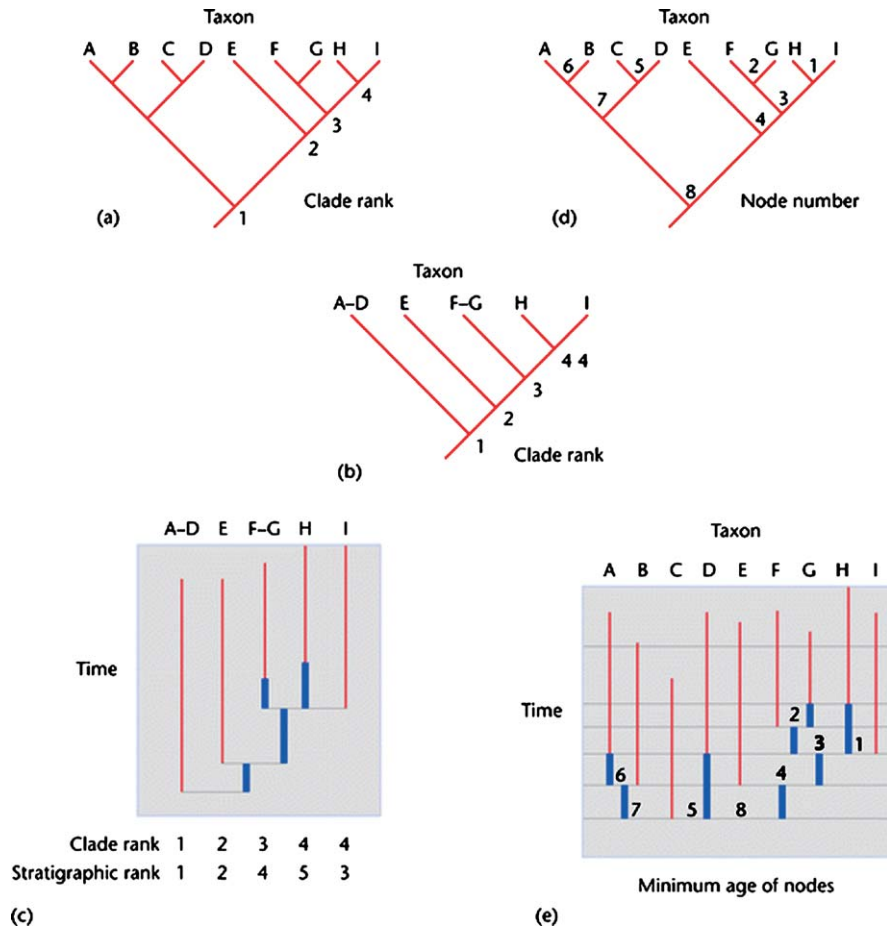
Macroevolution#Large-scale evolution, usually interpreted to mean at and above the species level.

Phylogeny#An evolutionary tree; the study of patterns of evolution.

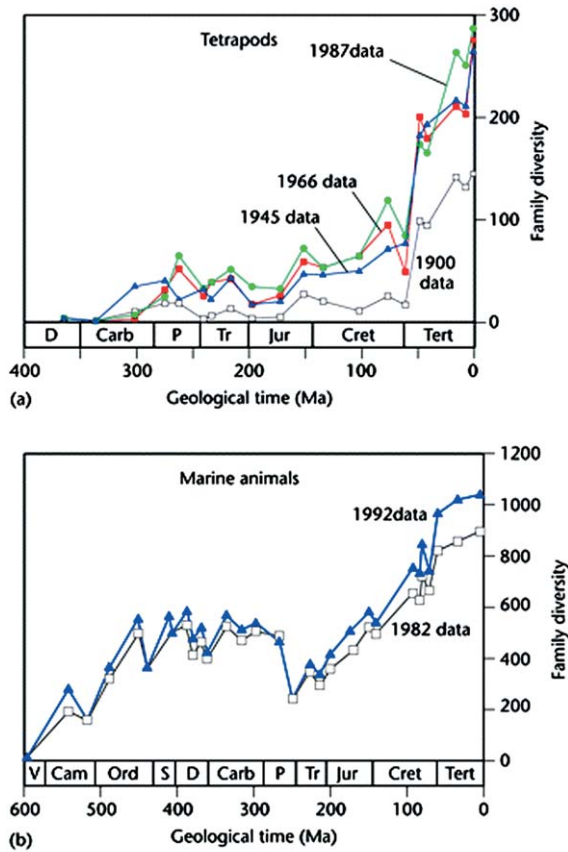
Polarity#The direction of change of a character, from one state to another.

Stratigraphy#The arrangement of rocks in sequence in the Earth's crust; the study of the dating and sequence of rocks.

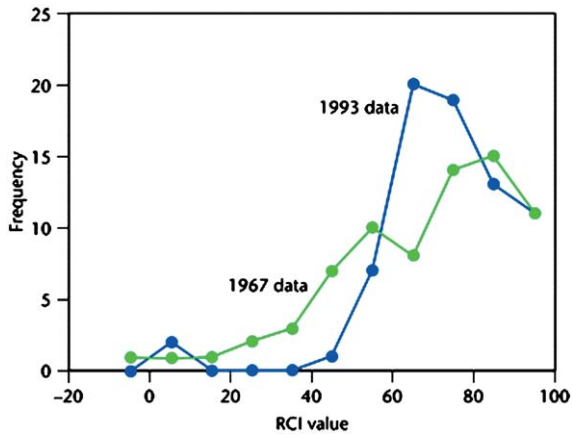
Taxon#A natural group of organisms, such as a species, a genus, a family or a phylum.



004144.f0001 **Figure 1** 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 1 (basal node) upwards to the ultimate node. In cases of nonpectinate 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 $RCI = (1 - (\sum(MIG)) / (\sum(SRL))) \times 100\%$ RCI values may range from 100% (no ghost range) through 0 (ghost range = known range) to high negative values (ghost range \gg 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 with 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).



004144.f0002 **Figure 2** Comparisons of the patterns of diversification of nonmarine tetrapods (a) and marine animals (b) based on data sets of different vintage. The diversity patterns are broadly similar, showing gradual increases in diversity through time, interrupted by various declines, corresponding to mass extinction events (Late Cambrian, Late Ordovician, Late Devonian, end-Permian, late Triassic, end-Cretaceous and Late Eocene). The main change, in both cases, is that overall diversity has increased. Abbreviations: V, Vendian; Cam, Cambrian; Ord, Ordovician; S, Silurian; D, Devonian; Carb, Carboniferous; P, Permian; Tr, Triassic; Jur, Jurassic; Cret, Cretaceous; Tert, Tertiary. Based on data in Maxwell and Benton (1990) and Sepkoski (1993).



004144.F0003 **Figure 3** Relative improvement in our knowledge of the fossil record quality from 1967 to 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 nonparametric signs and Wilcoxon-signed ranks tests). Based on data in Benton and Storrs (1994).