

Testing the time axis of phylogenies

M. J. BENTON

Department of Geology, University of Bristol, Bristol, BS8 1RJ, U.K.

SUMMARY

Fossils are not perfect materials for phylogenetic analysis because of problems of missing characters and missing taxa. However, fossils have three major advantages: (1) they give the only direct evidence of the order of acquisition of characters, (2) they frequently present character combinations not found in modern forms, and (3) they may allow the coding of characters that have been overwritten by subsequent evolution within a clade. There are three independent sources of evidence about sequences of historical events in evolution – morphological, molecular and stratigraphic – and these may be mutually cross-tested. Tests of the quality of the fossil record against morphological cladistic data show that (1) age and clade data on branching sequences generally agree, (2) knowledge of the fossil record has improved by 5% over the past 26 years of research, and (3) the fossil record of continental vertebrates is as good as that of (marine) echinoderms. Hence, systematists and evolutionary biologists may use fossil data with confidence in phylogeny reconstruction and to calibrate the time axis of phylogenies.

1. INTRODUCTION

Phylogenetic trees produced from morphological or molecular information are generally represented with a horizontal axis that represents clade diversity, and a vertical axis that represents character-state change. The nodal order of branching may be calibrated as a time axis by the addition of stratigraphic information about the occurrence of fossils in geological time. There is no other, independent, source of information for calibrating the time of cladogenetic events in phylogenies, and yet many doubts have been expressed about the trustworthiness of the fossil record. This subject will be explored further, and a case will be made for the use of stratigraphic and palaeontological data in producing and in testing phylogenies.

Two temporal aspects of phylogenies are of interest: the implied order of nodes in a phylogenetic tree, and the absolute time represented. Both of these aspects of phylogenies may be tested by direct comparison between cladograms, or molecular trees, and stratigraphic data. In particular, it is shown that (1) stratigraphic data on group appearances match cladistic data on branching patterns; (2) palaeontological data have improved by about 5% over the past 26 years; and (3) marine and continental, and invertebrate and vertebrate, fossil records may be equally complete.

2. THE VALUE OF PALAEOLOGICAL DATA

(a) *Incompleteness of the fossil record*

The fossil record of the history of life is incomplete (Darwin 1859; Raup 1972; Allison & Briggs 1991). Clearly, very few individual organisms that have ever

existed produced a fossil. A number of substantial filters must be crossed between the death of an organism and its identification as a fossil: (1) habitat and ecology (short-lived benthic marine organisms with hard skeletons are more likely to be buried than long-lived arboreal organisms that lack skeletons), (2) sedimentation patterns (many organisms live in areas where deposition is not taking place), (3) subsequent geological history (older rocks may be metamorphosed, subducted, eroded, and suffer all kinds of similar indignities), (4) serendipity and human factors (fossils do not exist as unique taxa until they have been collected by an interested person, studied, and described).

Two specific criticisms of the fossil record have been made: (1) fossils provide far less character information than modern taxa, and so should take only a secondary role in phylogeny reconstruction (Patterson 1981), and (2) the order of occurrence of fossils is so mixed up by the patchiness of representation that the fossil record reveals only a crude outline of what happened in the past. These criticisms, focusing on (1) missing characters and (2) missing taxa deserve consideration and testing.

(b) *Missing characters?*

A number of systematists have made the assertion that fossil taxa are much inferior to living taxa for the reconstruction of phylogeny, and that they should be ignored, or accorded much lower value in the assessment of branching patterns (Hennig 1966; Nelson & Platnick 1981; Patterson 1981). Specific criticisms of fossils were: (1) much character information is missing because of the loss of soft tissues; and (2) it is impossible to include fossil taxa in molecular phylogenies.

No one can deny that fossils often lack soft-part data, and it is possible to show that fossil taxa are typically associated with more missing data in taxon/character data matrices than are extant taxa. However, it is important to recall that incompleteness is not the same as an absence of information. Even if only one or two characters are available for a particular taxon, these may be the attributes that determine the nature of a particular node in the cladogram.

Events have now overtaken the second criticism. Proteins and DNA have been recovered from a variety of fossils, and molecular phylogenetic trees, incorporating fossil taxa, have been published. When problems of analysis and interpretation are resolved, such studies may become commonplace. Fossils may yield molecular sequences that assist in phylogeny reconstruction, especially when long terminal branches are present (Smith and Littlewood 1994).

Numerous studies (see Smith & Littlewood 1994) have now contradicted the view that fossils are of secondary value in phylogeny reconstruction. Unique properties of fossils (Smith 1994) are that (1) they give the only direct evidence of the order and precise date of the acquisition of characters and character complexes, (2) they may allow the coding of characters that have been overwritten by subsequent evolution within a clade (such as the teeth of turtles and birds, which are known only in fossil forms), and (3) they frequently present character combinations not found in modern forms.

The last point, that fossils may provide unique character combinations, is critical. Even when fossil taxa are rare or incomplete, they may offer character data critical for resolving particular nodes. Fossil taxa may assist in distinguishing synapomorphy from homoplasy (e.g. convergence), and this is particularly so in cases where there are long terminal branches; the fossil taxa may divide these long branches up. Fossils may also assist in polarity determination in cases where there has been a change of character state within a clade (a postulated synapomorphy may be found to occur in early members of the supposed outgroup, which shows that its distribution was once more general, and that it is in fact a symplesiomorphy).

(c) *Missing taxa?*

Some palaeontologists (e.g. Harper 1976; Szalay 1977) have argued that stratigraphic occurrence may be used as a test of character polarity, that the character state found in the oldest fossil is the ancestral state. However, this simplistic assumption that early = primitive has been criticized by Eldredge & Cracraft (1980), Nelson & Platnick (1981) and many others. The equation is probably valid in many cases, but there is no guarantee that, even in well-sampled groups, the order of occurrence of fossils is correct. A cladogram must be constructed from character data to establish the hierarchical branching pattern.

This debate, about the use of fossils in polarity determination, as well as the wider debates between cladists and traditional systematists, have left an impression that fossil taxa are generally so sparsely

represented in time that little may be gained from palaeontological data. Until recently, the only replies that could be made by palaeontologists were qualitative: simple assertions that the nature of the fossil record had been misrepresented, and that it was not so bad as people had suggested. There are now quantitative techniques that give absolute measures of the quality of the fossil record, and their results are favourable to palaeontological data. These techniques depend upon an assumption that cladograms and molecular trees are independent of stratigraphic data.

3. THREE INDEPENDENT SOURCES OF EVIDENCE ABOUT PHYLOGENY

It will be asserted here that there are three independent sources of data on series of events in phylogeny: cladistic morphological, molecular, and stratigraphic. There are probably more than three, if it can be maintained that the nucleic acids evolve independently of proteins such as myoglobin, haemoglobin and cytochrome *c*, and indeed it may be that unrelated proteins offer a further selection of independent sources of data on the order of phylogenetic events.

Cladistic analyses are generally based on morphological characters alone. The hierarchical structure of the cladogram is discovered by means of analysis of a matrix of characters coded across a number of taxa. At times, fossils have been proposed as the arbiters of two equivalent techniques in cladistics: (1) character polarity and (2) tree rooting, and if these methods relied upon the age of the fossils, cladistic methods would not be independent of stratigraphic data.

Fossils have been used on occasion to determine character polarity, but that method is discredited (see above). The technique of choice for polarity determination is outgroup comparison, and this method treats all organisms under consideration, whether living or extinct, as equivalent terminal taxa. In any case, most current analytical methods require no assumptions about character polarity, and the issue may be avoided completely. The modern techniques identify homologies, but they do not indicate the direction of change. This is discovered by the process of tree rooting.

Tree rooting is carried out by identifying one or more outgroups, which fix the shape of the cladogram and thereby determine the direction of character change. Fossil taxa are often chosen as the outgroups, because they illustrate a more plesiomorphous set of characters (Smith 1994). As an example, in attempting to resolve a cladogram of lungfish, coelacanth, and tetrapods, which diverged about 400 Ma (million years) ago, it is more helpful to use one or two Devonian fishes as outgroups than to select a modern cod or seahorse, because they have built up a 400 Ma overprint of their own character transformations, which obscures many of the relevant character states.

Molecular phylogenies are compiled either (1) from comparisons of relative similarity between proteins or nucleic acids (DNA, RNA) of different species, or (2) by comparisons of sequences of amino acids in proteins, or of bases in nucleic acids. The quantitative data on

similarities among taxa, or on precise differences between sequences, are converted into trees by a variety of multivariate techniques. The only point at which fossils enter the technique is in calibrating the time scale of the trees. The fossils do not affect the shape of the tree, and hence there is no mixing of molecular and stratigraphic data in the identification of the hierarchy of branching points.

If stratigraphic, morphological cladistic, and molecular techniques provide independent evidence about branching patterns, then those patterns may be mutually tested in an attempt to check the performance of each technique. In addition, such cross-testing of historical patterns of phylogeny acts as a test of evolution in general. Evolution, meaning simply the patterns of organic change through time (Mayr 1982), has been seen as a theory founded on such a broad accumulation of observations and hypotheses that it cannot be tested in any simple way. And yet, classic evolutionists (Darwin 1859; Ridley 1993) have indicated that the kind of overwhelming fact that would confound their acceptance of evolution would be the discovery of a fossil in a wholly anomalous stratigraphic situation. J. B. S. Haldane once said that his faith in evolution would be destroyed by the discovery of a fossil rabbit in Precambrian rocks. Creationists have striven to supply such anomalous fossils: trilobites preserved in the base of human sandal-prints, and dinosaur and human footprints occurring together. So far, all such finds have turned out to be hoaxes.

4. TESTING THE QUALITY OF THE FOSSIL RECORD

(a) Order of branching

The order of branching may be compared readily among different hypotheses for the phylogeny of a particular group. The basic approach (figure 1) is to enumerate the order of nodes on each phylogenetic tree and to compare them with stratigraphic data for mismatches of ordering. An appropriate simple statistical test is Spearman Rank Correlation (SRC), which assesses the probability that one predicts the other better than would be expected by chance. The method works only for samples of more than four taxa, because there must be more than three nodes (number of nodes = number of taxa - 1) for the test to have any meaning.

To deal with real cases the method must be amplified (Norell 1992, 1993; Norell and Novacek 1992*a,b*; Benton 1994; Benton and Storrs 1994, 1995). It is only possible to assess node order in a pectinate (unbalanced, Hennigian comb) type of cladogram (figure 1*b*), because there is no evidence in a more balanced cladogram (figure 1*a*) for the relative ordering of nodes on both major branches: should the branching point between A and B be numbered 2, 3, or 4? Hence, one branching stem must be collapsed at each equivocal point, and that stem is then treated as a single taxon, equivalent to all the others (figure 1*b*). It is possible to collapse balanced cladograms in

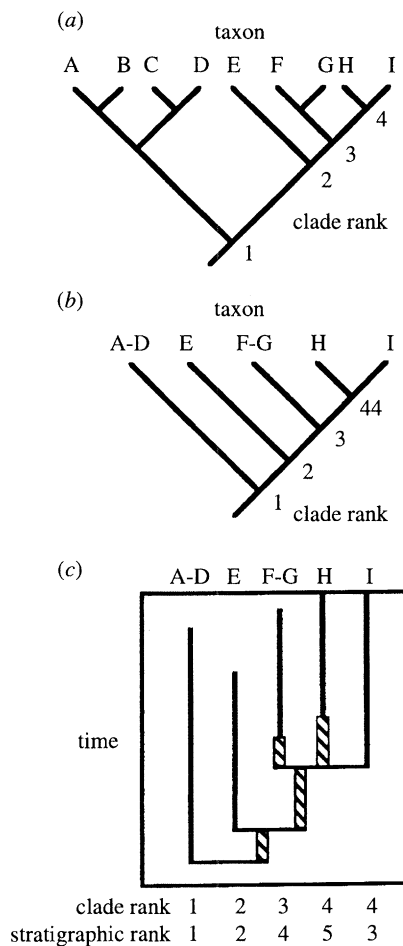


Figure 1. Methods for assessing the quality of the fossil record, by comparing branching order in cladograms (*a-c*) with stratigraphic data, and by comparing the relative amount of gap and known record (*c*). Cladistic rank is determined by counting the sequence of primary nodes in a cladogram (*a*). In non-pectinate cladograms (*a*), the cladogram is reduced to pectinate form (*b*), and groups of taxa that meet the main axis at the same point are combined and treated as a single unit. The stratigraphic sequence of clade appearance is assessed from the earliest known fossil representative of sister groups, and clade rank and stratigraphic rank may then be compared (*c*). The minimum implied gap (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. MIG is a minimum estimate of stratigraphic gap, as the true age of lineage divergence may lie well before the oldest known fossil.

different ways, and it is sometimes informative to test one version with the left-hand half collapsed, and then to test a second version with the right-hand half collapsed.

Tests of relative ordering of branching points in morphological cladograms with known stratigraphic ordering of first occurrences of the same taxa have demonstrated three facts: (1) clade data and age data agree; (2) improvements in palaeontological knowledge do not necessarily improve the fit between clade and age data; and (3) continental vertebrates show equivalent levels of matching between clade and age data as do (marine) echinoderms.

Clade data and age data generally agree (figure 2*a*).

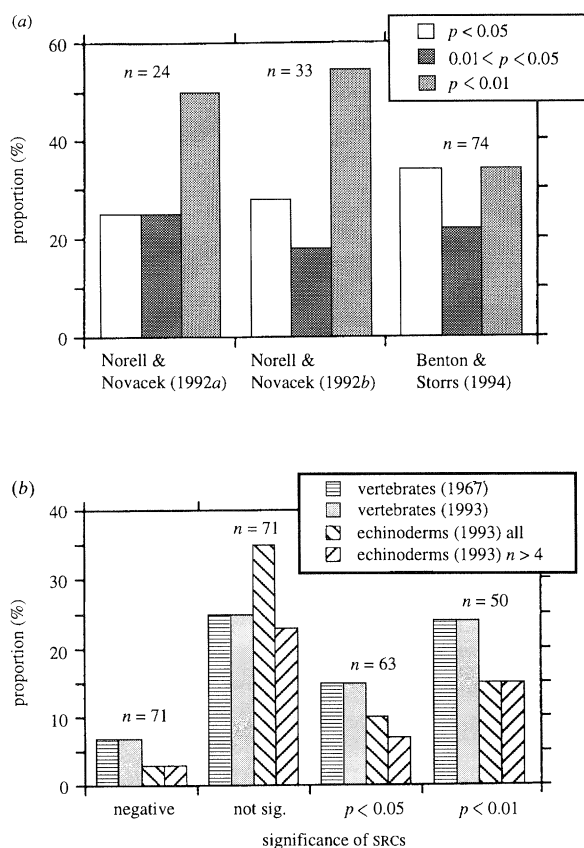


Figure 2. Comparison of measures of completeness of the fossil record of vertebrates and echinoderms. (a) Assessments of the statistical significance of Spearman Rank Correlation (SRC) tests for three studies of cladograms of vertebrates (Norell and Novacek 1992a,b; Benton & Storrs 1994). In all three cases, most cladograms show statistically significant ($P < 0.05$) correlation of clade order and age order, but the pass rate declines with larger data sets. (b) Comparison of the statistical significance of SRC tests for four sets of fossil data, two of vertebrates (Harland *et al.* 1967; Benton 1993) and two of echinoderms (both with stratigraphic data from Benton (1993)), one the set of all echinoderm cladograms, and one the set of echinoderm cladograms with more than four terminal taxa. The statistical significance of SRC matching does not change from 1967 to 1993 for vertebrates, but vertebrates show more statistically significant matches ($P < 0.05$ and $P < 0.01$) than does either of the echinoderm data sets.

Norell and Novacek (1992a) found that, with the SRC test, 18 out of 24 test cases of cladograms of vertebrates (75%) gave statistically significant ($P < 0.05$) correlations of clade and age data. In larger samples, Norell and Novacek (1992b) found significant correlation in 24 of 33 test cases (73%), and Benton and Storrs (1994) found significant correlation in 41 of 74 test cases (55%). Both pairs of investigators found the same results in independent analyses of the same cladograms, and the larger sample of Benton and Storrs (1994) shows lower levels of correlation possibly because of the addition of some less well resolved cladograms. All these studies, based only on vertebrates, confirm none the less that clade rank and age rank match.

It might be expected that the addition of new fossil finds and reanalysis of older ones would improve the fit

of age data to a fixed sample of cladograms, by the filling of gaps and by corrections of former taxonomic assignments. However, in a comparison of a 1967 data set (Harland *et al.* 1967) and one from 1993 (Benton 1993), Benton & Storrs (1994) found no change at all in the proportions of cladograms that showed statistically significant ($P < 0.05$ and $P < 0.01$) matching of clade and age order (figure 2b), although there had been a change in the status of 28 of the 71 cladograms compared (39%). In other words, as a result of 26 years of work, new discoveries and reassignments had improved the fit in 20% of cases, but caused mismatches of clade and age data in a further 20% of cases. Sometimes a new fossil does not fill a gap, but creates additional gaps on other branches of a cladogram.

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

If cladograms of vertebrates are generally matched by the correct order of appearance of the fossils, it might be expected that marine animals, with potentially richer fossil records, might show better results. However, in a study of cladograms of echinoderms, Benton and Simms (1995) found correlation in only 24 out of 63 test cases (38%). This disappointing result may arise from the fact that many (21%) of the echinoderm cladograms consisted of only four taxa. When these were excluded, 23 out of 50 (46%) cladograms showed a statistically significant ($P < 0.05$) correlation of clade and age rank – somewhat better, but still worse than the results for continental vertebrates (figure 2b).

(b) Relative completeness

Tests of the order of branching provide a great deal of information, but they are highly sensitive to perturbation, especially when the time intervals over which several branching events occurred is small. In such cases, a trivial change in the relative dates of some basal taxa could entirely overturn the stratigraphic rank order. In addition, the significance of SRC values may correlate with cladogram size, for vertebrates at least (Benton & Storrs 1994), although this was not found for echinoderms (Benton & Simms 1995). The tests of rank-order correlation take no account of the amount of time involved, and an additional test is required.

A number of statistics have been proposed that measure the relative amount of time represented by fossils and unrepresented by fossils. A cladistic measure of absolute completeness is based on the identification of minimum implied gaps (MIGs), or ghost ranges, for each cladogram (figure 1c) (Norell 1992, 1993; Norell & Novacek 1992a,b; Weishampel & Heinrich 1992; Benton 1994; Benton & Storrs 1994, 1995). MIGs may

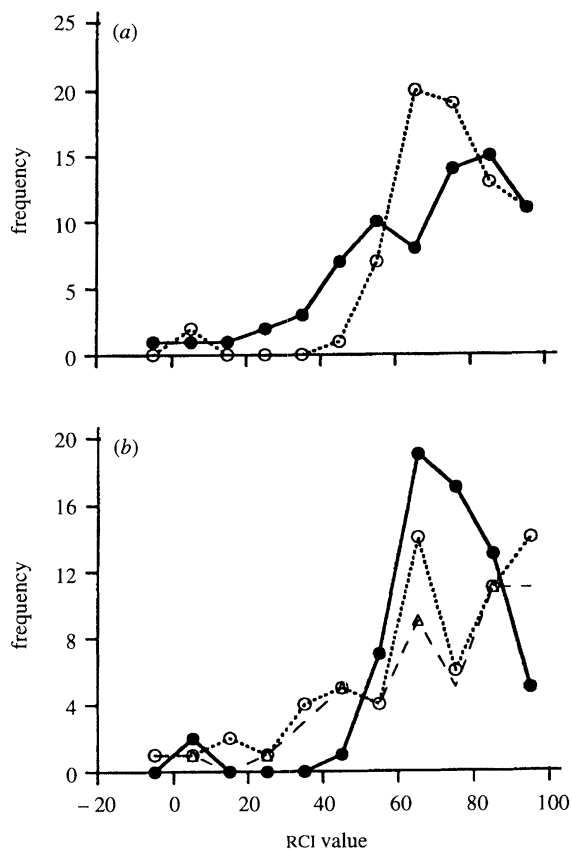


Figure 3. Comparisons of relative completeness of different fossil records, measured by the relative completeness index (RCI). (a) Relative improvement in fossil record quality as determined from Harland *et al.* (1967) (filled circles) and Benton (1993) (open circles). There is a statistically significant 5% shift of the distribution of RCI values to the right from 1967 to 1993, indicating an improvement in palaeontological knowledge. (b) Comparison of the fossil records of continental vertebrates and of echinoderms (all cladograms, and all cladograms with more than four terminal taxa). The sample of cladograms of all echinoderms ($n = 63$) (open circles, dotted line) has a mean RCI = 66%; for echinoderm cladograms with more than four terminal taxa ($n = 50$) (open triangles, dashed line), the mean RCI = 70%; and for continental vertebrate cladograms ($n = 63$) (filled circles, solid line), the mean RCI = 70%. Continental vertebrates do not have a significantly different fossil record than do echinoderms, based on the all-echinoderm distribution (Kolmogorov–Smirnov test, $P < 0.01$), and the distributions are the same for continental vertebrates and echinoderms, based on the larger ($n > 4$) echinoderm cladograms (Kolmogorov–Smirnov test, $P < 0.01$).

be measured and summed for any phylogenetic tree (a cladogram plotted against known stratigraphic taxon ranges) giving total MIG value. This is best cited as a measure of *relative* completeness of the tree, as the quality of any particular fossil record will depend upon the numbers of taxa involved and their known stratigraphic ranges.

Benton & Storrs (1994) presented a relative completeness index (RCI), calculated by comparing the amount of gap in a particular fossil record (assessed as MIG) to the part of the record represented by fossils:

$$\text{RCI} = [1 - (\sum(\text{MIG}) / (\sum(\text{SRL})))] \times 100\%$$

SRL is the simple range length for each taxon (the total time between first and last known appearance). RCI, being a measure of total inherent gap to known lineage duration, theoretically may vary from a negative value, where the amount of expected gap exceeds the sum total of proven stratigraphic range lengths, to 100%, where no gaps are evident. Examples of total gap being greater than total known record appear to be rare for analyses involving numerous taxa, although Benton & Storrs (1994) found one example among the vertebrates, and Benton & Simms (1995) found one among the echinoderms.

In a study of 74 cladograms of vertebrates, Benton & Storrs (1994, 1995) found that 71 (96%) had RCI values in excess of 50% (figure 3a). In other words, all but three of the 74 cladograms represented phylogenies in which more than half the range is represented by fossil specimens. Indeed, the mean RCI value is 72.3%, indicating that ghost ranges make up just over one quarter of all the total stratigraphic ranges.

The RCI tests on cladograms of vertebrates also showed a statistically significant improvement over the past 26 years of research. In a comparison of the RCI values implied by the 1967 and 1993 data sets, assessed across a fixed sample of 73 cladograms, the mean RCI value shifted from 67.9% to 72.3%, a statistically significant difference according to a paired t test ($P = 0.045$). A comparison between the two data sets (figure 3a), using a Wilcoxon signed ranks test, a non-parametric analogue of the t test, also indicated a statistically significant ($P = 0.026$) difference between 1967 and 1993 values. In other words, comparisons of the relative completeness of cladograms shows a significant *improvement*, by about 5%, in knowledge of the fossil record over the past 26 years of research. Hence, new fossil discoveries, and reassignments of older ones, do positively affect the amount of ghost range.

The comparison of rank-order correlation of clade and age data on cladograms of echinoderms and vertebrates (Benton & Simms 1995) suggested that both had equivalent matching (see above). The RCI comparisons are different when all echinoderm cladograms are considered: 78% of echinoderm cladograms have RCI values of over 50%, whereas the figure is 95% for cladograms of continental vertebrates (figure 3b). Mean RCI values are 66% for echinoderms and 70% for continental vertebrates. When cladograms of echinoderms with four taxa are excluded from the comparisons, 80% have RCI values over 50%, and the mean RCI value increases to 70%, the same as for continental vertebrates. Hence, comparisons of relative completeness indicate that echinoderms have similarly well-represented, or somewhat poorer, fossil records than do continental vertebrates.

The finding that continental vertebrates have an equivalent fossil record to echinoderms suggests two observations. (1) The relative abundance of specimens at individual fossil localities is no indicator of the completeness of their fossil record on a large scale: this depends on the number of stratigraphic horizons that have yielded fossils, and on the packing of those horizons in time. (2) The fossil record of vertebrates

has probably been more intensively studied than has that of echinoderms. Hence, our knowledge of the vertebrate fossil record is placed higher on the collector curve (numbers of taxa versus effort), and may be assumed to approach closer to the level of complete sampling and full knowledge of all fossil taxa that exist in the rocks.

5. DISCUSSION

The time axis of phylogenies is important, and it can be calibrated, with increasing confidence, by the use of stratigraphic data about geological dates of occurrence of basal taxa within clades. Recent studies have shown a good concordance between stratigraphic information about the fossil record, and morphological cladistic data on branching patterns. In addition, it has also been shown that knowledge of the fossil record of vertebrates at least has improved by 5% in the past 26 years of research. There is no reason to doubt that such a measure of improvement might apply to other segments of the fossil record, nor that such improvements might continue into the future. Additional analyses, both of additional groups of organisms, and after further increments of research input, may confirm or refute these findings. In addition, further revisions of cladograms and molecular trees may have a bearing on the quality of matching between clade and age data.

In the above discussions, mutual testing among the independent sources of data on the order of phylogenetic events, from morphological cladistics, molecular phylogenetics, and the stratigraphic distribution of fossils, has been aimed at testing the quality of the fossil record. However, because none of these three forms of information can be said *a priori* to be more reliable in all cases than the other two, mutual testing can extend to assessing the validity of cladograms derived from molecular and morphological data. In a case where a number of equally most parsimonious trees (MPTs) are identified, the validity of one scheme over the others could be tested by a probabilistic statement based on calculation of the correlation between clade and age data (SRC), and of the RCI implied by each cladogram. The cladogram that gave the most significant SRC metric of correlation, and the highest RCI value (i.e. smallest amount of ghost range), has the greatest probability of being correct, all other things being equal.

I thank Derek Briggs, Andrew Smith, and Matthew Wills for helpful comments on earlier drafts of this paper. This work was funded in part by Grant F/182/AK from the Leverhulme Trust.

REFERENCES

- Allison, P.A. & Briggs, D.E.G. 1991 *Taphonomy: releasing the data locked in the fossil record*. New York: Plenum.
- Benton, M.J. 1993 *The fossil record 2*. London: Chapman & Hall.
- Benton, M.J. 1994 Palaeontological data, and identifying mass extinctions. *Trends Ecol. Evol.* **9**, 181–185.
- Benton, M.J. & Simms, M.J. 1995 Testing the marine and continental fossil records. *Geology*, **23**. (In the press.)
- Benton, M.J. & Storrs, G.W. 1994 Testing the quality of the fossil record: paleontological knowledge is improving. *Geology* **22**, 111–114.
- Benton, M.J. & Storrs, G.W. 1995 Diversity in the past: comparing cladistic phylogenies and stratigraphy. In *Phylogeny and biodiversity* (ed. M. Hochberg). Oxford: Oxford University Press. (In the press.)
- Darwin, C. 1859 *On the origin of species by means of natural selection*. London: John Murray.
- Eldredge, N. & Cracraft, J. 1980 *Phylogenetic patterns and the evolutionary process*. New York: Columbia University Press.
- Harper, C.W. 1976 Phylogenetic inference in paleontology. *J. Paleontol.* **50**, 180–193.
- Hennig, W. 1966 *Phylogenetic systematics*. Urbana: University of Illinois Press.
- Mayr, E. 1982 *The growth of biological thought: diversity, evolution and inheritance*. Cambridge, Massachusetts: Harvard University Press.
- Nelson, J.S. & Platnick, N.I. 1981 *Systematics and biogeography. Cladistics and vicariance*. New York: Columbia University Press.
- Norell, M.A. 1992 Taxic origin and temporal diversity: the effect of phylogeny. In *Extinction and phylogeny* (ed. M.J. Novacek & Q.D. Wheeler), pp. 89–118. New York: Columbia University Press.
- Norell, M.A. 1993 Tree-based approaches to understanding history: comments on ranks, rules, and the quality of the fossil record. *Am. J. Sci. A* **293**, 407–417.
- Norell, M.A. & Novacek, M.J. 1992a The fossil record and evolution: comparing cladistic and paleontologic evidence for vertebrate history. *Science, Wash.* **255**, 1690–1693.
- Norell, M.A. & Novacek, M.J. 1992b Congruence between superpositional and phylogenetic patterns: comparing cladistic patterns with fossil records. *Cladistics* **8**, 319–337.
- Patterson, C. 1981 Significance of fossils in determining evolutionary relationships. *A. Rev. Ecol. Syst.* **12**, 195–223.
- Raup, D.M. 1972 Taxonomic diversity during the Phanerozoic. *Science, Wash.* **215**, 1065–1071.
- Ridley, M. 1993 *Evolution*. Oxford: Blackwell Scientific.
- Smith, A.B. 1994 *Systematics and the fossil record*. Oxford: Blackwell Scientific.
- Smith, A.B. & Littlewood, D.T.J. 1994 Paleontological data and molecular phylogenetic analysis. *Paleobiology* **20**, 259–273.
- Szalay, F.S. 1977 Ancestors, descendants, sister groups and testing of phylogenetic hypotheses. *Syst. Zool.* **26**, 12–18.
- Weishampel, D.B. & Heinrich, R.E. 1992 Systematics of Hypsilophodontidae and basal Iguanodontia (Dinosauria: Ornithopoda). *Hist. Biol.* **6**, 159–184.