Paleobiology, 23(1), 1997, pp. 20–32

Congruence between parsimony and stratigraphy: comparisons of three indices

Rebecca Hitchin and Michael J. Benton

Abstract.—Use of quantitative statistical tests can show that there is generally good congruence between estimated cladistic hypotheses of relationship and observed stratigraphy. A data set of 376 cladograms of fishes, continental tetrapods, and echinoderms was tested using three metrics, Spearman Rank Correlation (SRC), Relative Completeness Index (RCI), and Stratigraphic Consistency Index (SCI), to explore the relationships between the indices and differences in results among the three groups of organisms.

There is a strong relationship between SCI and SRC, since both tests measure the same aspect of the fossil record. There is no relationship between RCI and either SCI or SRC. There is a highly significant relationship, as expected, between SRC coefficients and the number of taxa in a cladogram, but no such relationship for RCI or SCI (except in fishes). There is no significant relationship between any of the indices and either the number of internal nodes or tree balance.

Echinoderms show the best stratigraphic consistency of nodes, while continental tetrapods have the best matching of stratigraphic age and cladistic node order. Fishes have the worst match of age and clade ranks, but they do have the most complete fossil record as measured by the RCI. They are followed by echinoderms, and then continental tetrapods, which have the least complete record. This seems to show that life in an aquatic environment leads, in general, to a more complete fossil record.

Rebecca Hitchin and Michael J. Benton. Department of Geology, University of Bristol, Bristol BS8 1RJ, United Kingdom. E-mail: R.Hitchin@bristol.ac.uk and Mike.Benton@bristol.ac.uk

Accepted: 1 December 1996

Introduction

The fossil record of the history of life is generally assumed to be incomplete (e.g., Darwin 1859; Raup 1972; van Andel 1981; Allison and Briggs 1991). Estimates suggest that the geological record represents only 1–10% of the Earth's history (van Andel 1981), and within that, only a small proportion of all the organisms that existed are preserved as fossils. Assumptions of record quality are usually nonquantitative, based upon expectations of the preservability of different organisms and habitats.

Debates about the use of cladistic analysis of fossil taxa have left the impression that fossils are generally so sparsely represented through time that little can be gained from palaeontological data, though this view seems generally untrue (e.g., Gauthier et al. 1988). However, there are now quantitative techniques that give measures of the quality of the fossil record by assessing congruence between stratigraphic and cladistic data.

In precladistic days, evolutionary schemes were associated with stratigraphy, since systematists often used the first-known occurrence as a measure of the "primitiveness" of a character. With the advent of cladistics, phylogenetics was divorced from stratigraphy as cladists claimed that geological sequence is a poor guide to the polarity of characters (e.g., Hennig 1966; Eldredge and Cracraft 1980; Wiley 1981; Ax 1987) and that the order of branching in a cladogram need not match the order of occurrence in the fossil record.

Cladograms now are therefore essentially independent of geological input. One view (e.g., Patterson 1981; Platnick 1979) holds that the procedure is totally divorced from stratigraphic assumptions, and that the joining of the two would produce an "unholy matrimony," in Fisher's (1994) words, while a more moderate view (e.g., Gauthier et al. 1988; Norell and Novacek 1992b; Novacek 1992; Smith 1994) states that the coding of characters, determination of polarity, and tree rooting depend to a certain extent on stratigraphic information.

The use of stratigraphic indices for congruence testing relies on the premise that there

 0094-8373/97/2301-0002/$1.00

are two essentially independent methods of disentangling the sequence of events in the history of life (Benton 1995):

1. the order of fossils in the rocks (stratigraphic record), and
2. cladograms, based on assessments of morphological or molecular characters, and presenting postulated sequences of branching points.

If we can accept that these two approaches are essentially independent, mutual cross-testing should be possible.

Methods

Three indices were used to assess the quality of the fish, continental tetrapod, and echinoderm fossil records. Stratigraphic data for this study were taken from a recent compilation of range data (Benton 1993) except where greater detail was needed for cladograms containing genera. In these cases, stratigraphic information was taken from the publication containing the cladogram.

Spearman Rank Correlation (SRC)

The fossil record can be tested by comparison of the order of origins of groups from the stratigraphic record with the order of branching as indicated by a cladogram (Gauthier et al. 1988; Norell 1992, 1993; Norell and Novacek 1992a,b). Each taxon is assigned a clade rank, which is the number of nodes by which a terminal taxon is removed from the most recent common ancestor, i.e., the basal node of the cladogram (Fig. 1A).

This method was first used by Gauthier et al. (1988), who arcsine-transformed the age and clade ranks to produce a probability of correlation. In later tests (Norell 1992, 1993; Norell and Novacek 1992a,b; Benton 1994, 1995; Benton and Storrs 1994; Benton and Simms 1995; Benton and Hitchin 1996) these ranks were subjected to a Spearman rank correlation test.

Many published cladograms do not conform to a simple pectinate pattern of branching where all terminal taxa are undivided side branches of a single main stem (Fig. 1A). Topologies may be more complex, with some branches that subdivide further, or with unresolved nodes. In these cases, the cladogram is collapsed to a simple pectinate (Hennigian comb) structure (Fig. 1B), and groups of taxa that meet the main axis at a single point are combined and treated as a single unit. For complex topologies, several analyses may be undertaken as the tree is collapsed in several ways. This was done in one or two cases in the present study. The stratigraphic sequence of clade appearance is assessed from the earliest known fossil representative of sister groups (Fig. 1C).

The clade and age ranks can then be compared using a nonparametric Spearman rank correlation test. This statistical test assesses the probability that one set of data predicts the other better than would be expected by chance (Wilkinson 1989) and produces a correlation coefficient, the significance of which can be assessed. Values of the coefficient can range from -1, a perfect negative correlation, to $+1$, a perfect positive correlation.

Relative Completeness Index (RCI)

A second method of testing the quality of the fossil record is by assessing the relative completeness of stratigraphic records on the basis of independent evidence for the size of gaps (Benton and Storrs 1994, 1996; Benton and Simms 1995; Benton and Hitchin 1996). The ranges of each taxon are hypothetically extended to the same age as the sister group of each taxon, and this assessment of unseen range is known as minimum implied gap (MIG) (Fig. 1C, diagonal shading). MIGs can then be summed for the phylogenetic tree to produce a total MIG value, which is best cited as a measure of relative completeness of the tree.

A relative completeness index is then calculated for each tree with the purpose of comparing the amount of gap in any particular fossil record to the amount of record represented by fossils:

$$\mathrm{RCI} = \left(1 - \frac{\Sigma(\mathrm{MIG})}{\Sigma(\mathrm{SRL})}\right) \times 100\%,$$

where SRL = simple range length for each taxon. Stratigraphic durations are taken from

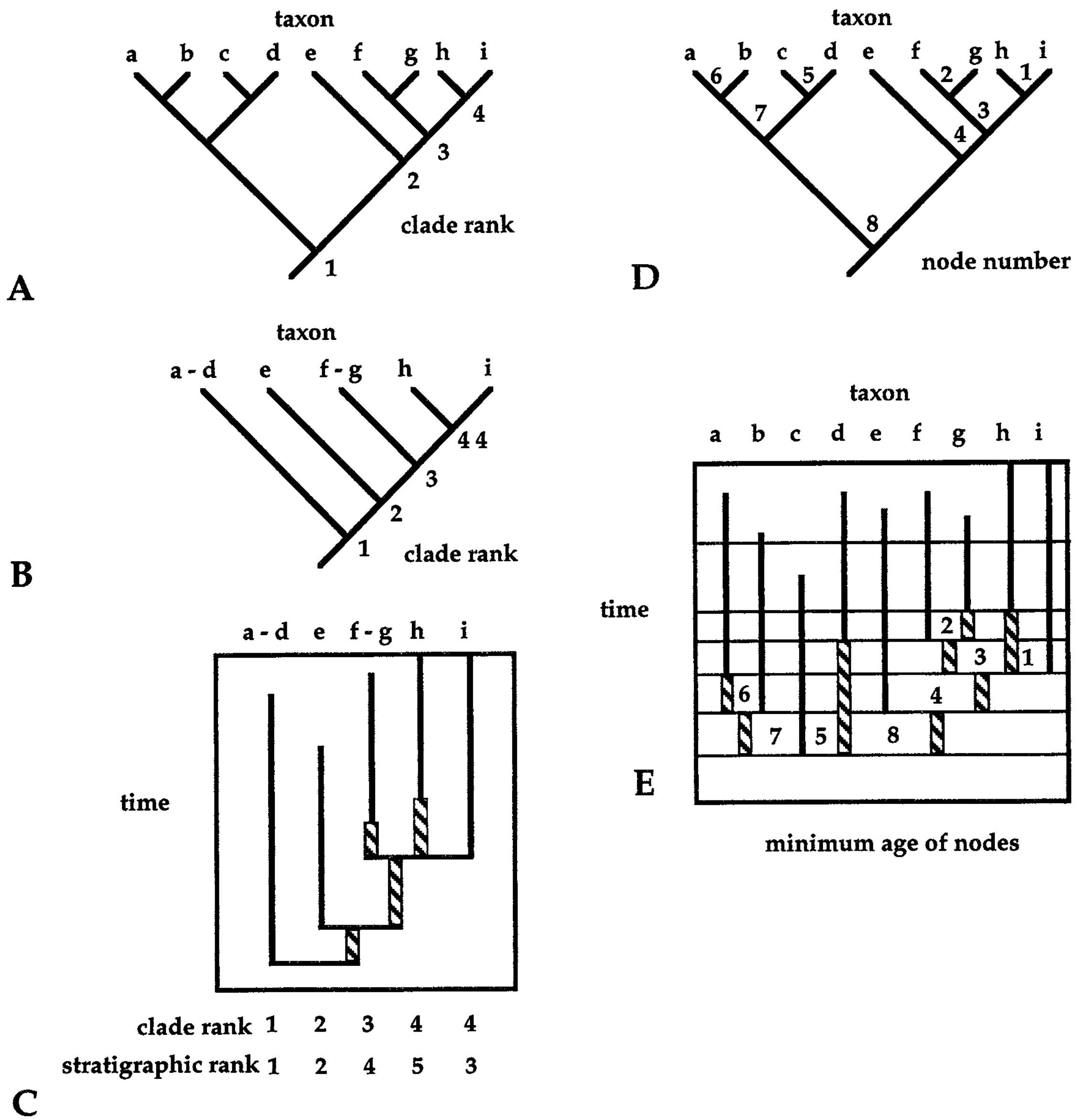

FIGURE 1. Methods for assessing the quality of the fossil record, by comparison of branching order in cladograms (A–E), and by comparison of the relative amount of gap and known record (C). Cladistic rank is determined by counting the sequence of primary nodes in a cladogram (A), the cladogram is reduced to pectinate form for the SRC test (B), and groups that meet the main axis at the same point are combined. The stratigraphic sequence of clade appearance is assessed from the earliest known fossil representatives of sister groups, and clade and age rank may then be compared in an SRC test (C). The minimum implied gap (MIG, diagonal shading) is the difference between the age of the first representative of a lineage and that of its sister, and is used to calculate the RCI. Nodes are ordered from the most distal (D) in the SCI test, and the consistency of each node is calculated.

current standard time scales, and no estimate of uncertainty is included.

Values of RCI can range from an infinitely negative value, where the amount of expected gap exceeds the total sum of proven stratigraphic range, to 100%, where no gaps are evident. The RCI value is a maximum value, which could be lowered by estimates of ancestor-descendant relationships in the cladogram, especially at low taxonomic rank. RCI tests can be performed on collapsed or full cladograms, where the entire cladogram is transposed onto the stratigraphic column, but the latter approach is preferred in order to maximize information.

Stratigraphic Consistency Index (SCI)

The third method used to analyze the quality of the fossil record assesses the stratigraphic fit of a cladogram on a node-by-node basis (Huelsenbeck 1994), and was originally designed for testing which cladogram best fitted the stratigraphic record out of several most-parsimonious trees produced in phylogenetic analyses. This type of method has its origin in the work of Fisher (1994), who suggested that stratigraphic incongruence can help evaluate all viable cladograms produced in an analysis.

Stratigraphic consistency is assessed for each node in a cladogram, starting with the most distal, with respect to the node immediately below (Fig. 1D). The node is said to be consistent if the stratigraphic ages of the taxa above it are younger than, or equal in age to, those of the node below (Fig. 1E; nodes 2, 3, 4, 6). An inconsistent node has older taxa in nodes above it (Fig. 1E; nodes 1, 5, 7).

As we do not possess information on the sister taxon of the root node, its consistency cannot be calculated. However, all internal nodes of a cladogram can be used, and hence there is not the loss of data associated with reduction to a pectinate form. The stratigraphic consistency index is then

$$SCI = C/N,$$

where C = number of stratigraphically consistent nodes and N = number of internal nodes excluding the root. The SCI value can range from 0, where all nodes are inconsistent, to 1, where all nodes show consistency.

Cross-Testing of Indices

Nonparametric statistical methods were used to assess the relationships between the three indices. These nonparametric methods do not assume a normal distribution, and they are less sensitive to distortion by outlying results than are parametric tests.

Spearman rank correlation tests were used to assess the significance of any relationship between the data sets. Analyses were performed to assess whether the indices gave significantly different results in each of the groups. The statistical significance was assessed with the Kolmogorov-Smirnov test. This nonparametric test assesses two distributions and compares their shapes, their mean values, and their standard deviations in order to test the null model that both distributions are the same (Sprent 1989). Test statistics are generated that allow the null hypothesis to be accepted or rejected at a given level of significance; probability values in the range $p < 0.05$ imply that two distributions could be essentially the same, while values in the range $p > 0.05$ imply that the two are different in mean, variance, or skewness, or in some combination of those measures. The Mann-Whitney test was not used in these analyses as it assumes a symmetrical distribution of ranked points (Ryan et al. 1977), which did not occur in the results.

Results

Acanthodian fish and early tetrapods had to be discounted from the following Spearman rank correlation tests as the sample sizes were too small. For any of the tests to have the possibility of showing significance, the sample size had to be greater than four.

Relationship between SRC and RCI

Continental vertebrates showed a significant positive relationship between SRC and RCI ($p < 0.05$) using a Spearman rank correlation test, while the total data set showed no significance (Fig. 2A). When the cladograms with four terminal taxa were discounted, the fishes and echinoderms still did not show any significant relationship, and there were no cladograms with four terminal taxa to discount in the continental tetrapod data.

Two groups in the fish data set showed a significant relationship, the actinopterygians ($p < 0.005$) and agnathans ($p < 0.1$). No other groups showed a significant positive relationship. Significant negative correlation between SRC and RCI values was seen in the sarcopterygians and the chondrichthyans.

Relationship between SCI and RCI

The total data set showed no statistically significant correlation (Fig. 2B). The fish data

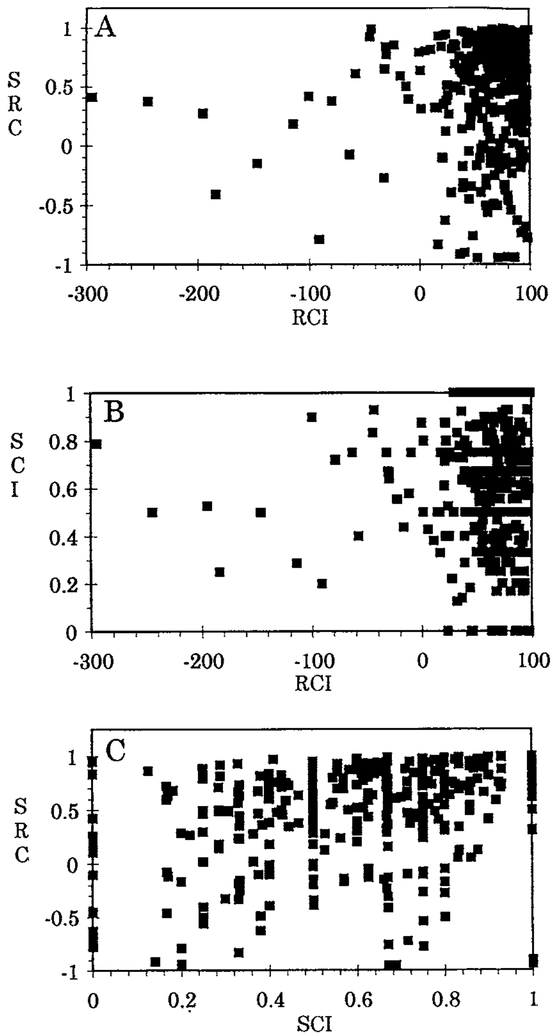

FIGURE 2. A, Relationship between SRC and RCI. Spearman rank correlation of the metrics shows no significant relationship (t = 0.054). B, Relationship between SCI and RCI. Again, Spearman rank correlation tests show no significant relationship (t = 0.081). C, Relationship between SRC and SCI. Spearman rank correlation test shows a highly significant relationship at $p < 0.005$ (t = 0.454).

showed a statistically significant negative correlation between SCI and RCI, whether the four terminal taxa cladograms were counted or excluded. However, the echinoderms showed a correlation at $p < 0.025$, and the continental tetrapod data showed an even higher significance, $p < 0.005$. This correlation was destroyed when the smaller cladograms were removed, at least in the case of the echinoderms.

None of the fish data subsets showed any significant correlations. Significant negative correlations between SCI and RCI were found in the sarcopterygians, agnathans, placoderms, acanthodians, and chondrichthyans.

Relationship between SRC and SCI

Highly significant relationships ($p < 0.0005$) were found between the SRC and SCI for the continental tetrapod and the echinoderm data sets. The significance was lower, $p < 0.05$, for the fish data set, but there was still an obvious relationship. The significance for the total data set was $p < 0.0005$ (Fig. 2C).

Only two of the fish subsets showed any significant relationship between SRC and SCI, the sarcopterygians ($p < 0.0005$) and chondrichthyans ($p < 0.05$). A significant negative correlation between SRC and SCI results was seen in the gnathostome group.

Tests of the Distributions of Values

RCI.—The RCI values have a similar distribution for the fish, echinoderm, and continental tetrapod data, i.e., an increase in frequency of higher values of RCI (Fig. 3). However, the means varied, with the continental tetrapods having the lowest average RCI, 49.8%. The echinoderms had a mean RCI value of 62.3%, and the fish 69.4%. Kolmogorov-Smirnov tests showed that distributions of values for echinoderms and fishes (p = 0.046) and for fishes and tetrapods (p = 0.008) were indistinguishable, but those for echinoderms and tetrapods were very different (p = 0.86).

When the data sets are divided into major habitats (aquatic and continental), a very similar pattern is seen for the two groups (Fig. 3). Both show a large increase at an RCI of approximately 30–40%, and the frequencies climb steadily until an RCI of 60–70%. Then, the aquatic curve continues to climb, while the continental curve begins to fall. The curves were just distinguished by the Kolmogorov-Smirnov test (p = 0.06).

SCI.—The highest mean occurred in the echinoderms, 0.78, followed by the continental tetrapods with a mean of 0.66. The fish data had the lowest average SCI, 0.55. Kolmogorov-Smirnov tests showed that all distributions (Fig. 4) were indistinguishable (echinoderm/

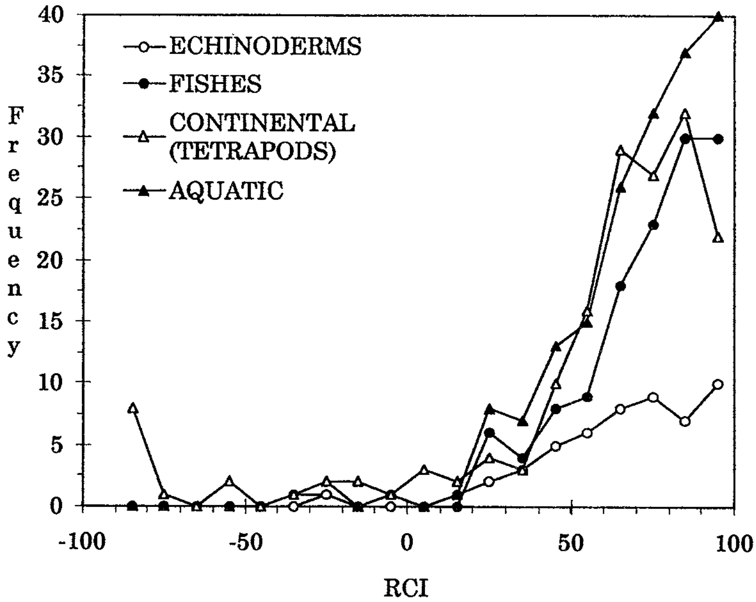

FIGURE 3. Comparison of the RCI by groups and by major habitats. Continental tetrapods and continental habitat give the same results. The echinoderm cladograms (open circles) have a mean RCI = 62.3%, the fish cladograms (filled circles) have a mean RCI = 69.4%, and for continental vertebrates (open triangles) the mean RCI = 49.8%. All three groups have significantly different fossil records, based on normalized data sets (Kolmogorov-Smirnov test, fish/tetrapod $p = 0.023$, fish/echinoderm $p = 0.749$, tetrapod/echinoderm $p = 0.023$). The aquatic habitat (filled triangles) shows a mean RCI = 67.3%, while the continental habitat (open triangles) has a mean RCI = 49.8%. Again, Kolmogorov-Smirnov tests show significant differences in RCI ($p = 0.059$).

fish, $p = 0.001$; echinoderm/tetrapod, $p = 0.001$; fish/tetrapod, $p < 0.0005$).

This similarity of distributions is again seen in the habitat comparison ($p = 0.012$). The SCI habitat curves show considerable similarity to the curve shapes seen in the RCI analyses, i.e., an upward trend until there is a divergence at approximately SCI = 0.8–0.85, when the marine frequency continues to climb and the continental frequency drops.

SRC.—Fishes have the lowest SRC mean (0.31). The continental tetrapods have the second highest SRC mean (0.58), while the echinoderms have the highest (0.68). These values are, however, dependent on cladogram size, so significance values were also assessed.

In terms of significance of correlation between stratigraphic and cladistic data, the frequency distribution (Fig. 5) showed most values in the "not significant" category, followed by trees with a positive correlation at $p < 0.0005$, except in fishes, where a negative correlation was more common. Continental vertebrates showed most cladograms (50%) with a positive correlation ($p < 0.05$) of stratigraphic and cladistic data. The value was lower for echinoderms (40%), and lowest for fishes (26%).

Kolmogorov-Smirnov tests showed that the distributions of SRC values for echinoderms and fishes were indistinguishable ($p = 0.001$), as were those of fishes and tetrapods ($p < 0.0005$), but the echinoderm and tetrapod distributions were seemingly different ($p = 0.35$). There was also no significant difference between the distributions of SRC results for the aquatic and continental habitats ($p = 0.004$).

Discussion of Results

Index Relationships

While there was some evidence for relationships between SRC and RCI, and between SCI and RCI, the most significant relationship was between SRC and SCI ($p < 0.0005$ for the entire data set). This is probably because the SCI and SRC indices look at the fossil record in the same way. While the RCI measures the completeness of the fossil record, the SRC and SCI measure the fit of the age and clade ranks, and

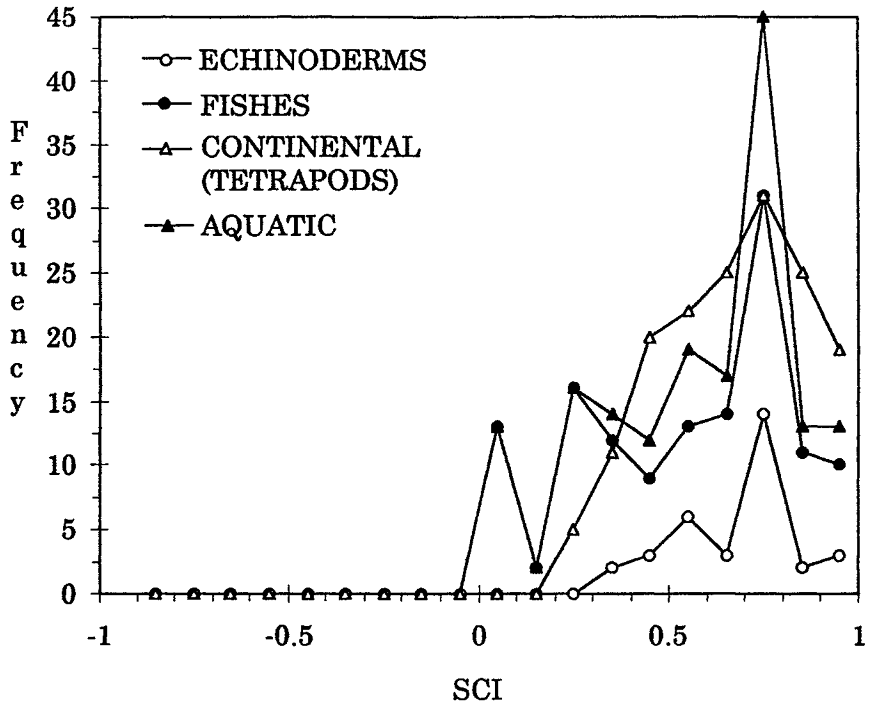

FIGURE 4. Comparison of SCI by groups and by major habitats. Echinoderm cladograms (open circles) have a mean SCI = 0.78, fish cladograms (filled circles) have a mean SCI = 0.55, while continental vertebrates (open triangles) have a mean SCI = 0.66. Kolmogorov-Smirnov tests again show significant differences between all three groups (fish/tetrapod and fish/echinoderm $p = 0.664$, tetrapod/echinoderm $p = 0.309$). Continental cladograms (open circles; mean 0.66) perform slightly better than aquatic cladograms (filled circles). Kolmogorov-Smirnov tests show the distributions are significantly different ($p = 0.664$).

so one might expect some correlation between the two indices. However, what is harder to explain is the lack of correlation in some of the fish groups, i.e., agnathans, actinopterygians, gnathostomes, and placoderms. It is possible that the lack of significant correlation in some cases may be a result of the small sample size, though that seems unlikely as other fish groups of the same sample size showed highly significant relationships.

Evidently, the different metrics of congruence between parsimony and the fossil record may give different results for the same sample of cladograms. This is not surprising, since each metric assesses a different aspect of that congruence. Which index is accepted as giving the "correct" answer must be a product of the original data set and of the original questions posed.

Outliers

Several outliers occurred in the graphs of relationships between the three indices. These could, in part, be explained by the number of terminal taxa in the cladograms. Most of the outliers, at least in the fish and echinoderm data sets, were produced by cladograms with four or five end taxa. However, pruning of these cladograms did not make a significant difference to the results of the relationships. There have been suggestions that these four-taxon cladograms should be dismissed from analyses, but we feel that, while some represent outliers, they increase the size of the data sets, and their exclusion does not affect the results.

Other explanations for the outliers cannot be stated from this study, but possibilities may be incorrect phylogenetic analysis, incorrect stratigraphic ranges, or errors in statistical analysis.

There were two major outliers in the total data set, Forey's (1988) coelacanth cladogram, and a series of cladograms looking at relationships of the Squamata (Etheridge and de Queiroz 1988; Grismer 1988; Caldwell et al.

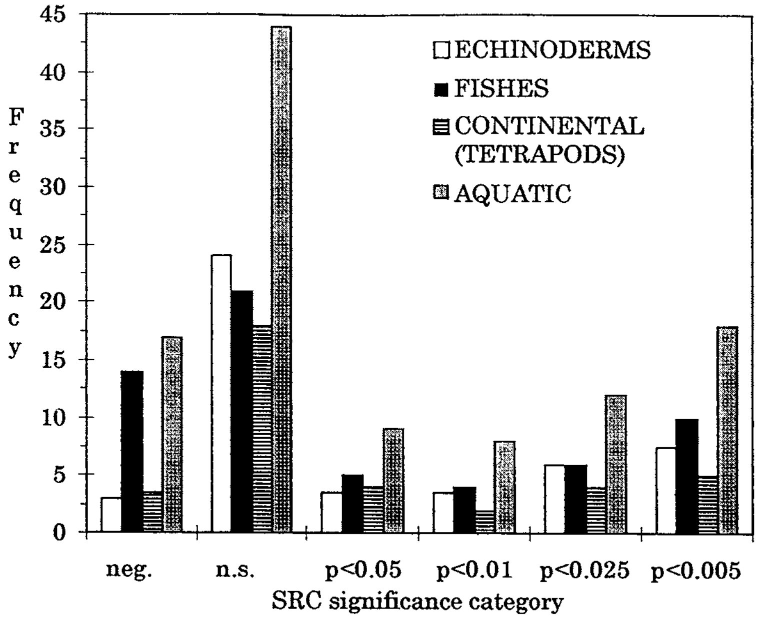

FIGURE 5. Comparison of SRC significance results by groups and major habitats (neg. = negative, n.s. = not significant). Of echinoderm cladograms (white bars), 35% show significance of result, as compared to 25% in fishes (black bars) and 50% in tetrapods (horizontal shading). The probability of matching distributions is low (fish/echinoderm and fish/tetrapod $p = 0.778$, tetrapod/echinoderm $p = 0.223$) and so all groups are regarded as different. Of continental cladograms (horizontal shading), 50% show significance of result, while only 25% of aquatic cladograms (gray shading) show significance. This difference is confirmed by a Kolmogorov-Smirnov test ($p = 0.778$).

1995; Clos 1995; Denton and O'Neill 1995), all of which showed negative RCIs. It is most probable that the outlying nature of the squamate cladograms was caused by a poor fossil record. All had RCIs of under 29, with the largest negative value being −114. These RCI values were coupled with average SRC and SCI values.

Forey's (1988) cladogram can be explained by the nature of the stratigraphic data used in the analyses. A large number of the terminal taxa were described as point occurrences in time, and even though these short ranges were elongated to half stage lengths in the statistical analyses, the amount of record was smaller than the amount of gap. This then produced a negative RCI. A newer cladogram (Forey 1991) showed almost the same taxa in the same situations on the cladogram, and this produced a much higher RCI. It seems that this is explained by the origin of the stratigraphic data. Forey's (1988) cladogram analysis used internal stratigraphic data at generic level, while Forey's (1991) cladogram was analyzed using data from Benton (1993) at family and stage level. This coarsening of the data set probably made the the RCI much higher than for the finer generic data.

Index Biases

SRC.—One problem with the SRC is that, as the test simply orders the times of origin, correlation values may be high where the dates of origin are well spaced out, such as in cladograms that span the whole Phanerozoic. However, if dates of origin mainly occur within one or two stratigraphic stages, such as in many of the placoderm and Tertiary mammal cladograms, correlation is much less likely to be significant (Benton and Storrs 1994). Also, with a shorter fossil record, small gaps or minor dating errors could be enough to introduce serious mismatches, whereas for long timescales, large errors or gaps would be needed to throw out the sequence of origins. In addition, the branching events within a cladogram are not independent, as assumed in the test, but are constrained in sequence by the preceding nodes, and thus a time-series statistical approach, such as is hinted at by the Huelsenbeck SCI test (Huelsenbeck 1994), may be more valid.

Benton and Storrs (1994) noted another problem of the SRC test—and one that is expected—the fact that SRC coefficients are correlated with numbers of terminal taxa in a cladogram ($p < 0.05$). This study again shows such a correlation, this time with $p < 0.0001$. However, this trend does not seem convincing from graphical evidence, or from the correlation coefficient of 0.285. If this relationship is upheld, then the SRC value itself cannot be a perfect test of the fit of stratigraphic and cladistic data on branching points, unless the cladogram size is standardized to allow direct comparison.

Other biases suggested for all the stratigraphic indices are correlations with tree balance, number of internal nodes (NIN) in the cladogram, and number of terminal taxa (NTT) of the tree, as suggested by Siddall (1996) specifically for the SCI metric. However, tree balance, measured by Heard's index of imbalance (Im) (Heard 1992), shows only a doubtful correlation with SRC ($r^2 = 2.7$, $p = 0.07$). There is a positive relationship between SRC and number of internal nodes ($r^2 = 4.0$, $p = 0.016$), and not surprisingly, highly significant correlation between SRC and NTT for fishes ($p < 0.005$). There is no significant relationship between these measures for either continental tetrapods or echinoderms (Hitchin and Benton 1997).

RCI.—The main apparent problem with the RCI is its sensitivity to the magnitudes of time involved and the relative ages of lineage origins. It would then seem that, if all else is equal, the best RCI values would be found for comparisons of ancient long-lasting groups, where the known stratigraphic range will be far longer than any implied gap, e.g. actinopterygian fishes and diapsids. The worst RCI values would be expected in groups that have only a short range length, such as some Tertiary mammal groups and the placoderm fishes. These expectations seem to be fulfilled by the RCI results.

RCI values are therefore sensitive to the precise choice of taxa. A value can be greatly improved by inclusion of one long-lived taxon, or diminished by removal of such a form. In large-sample comparisons, as here, it is legitimate to include a broad range of cladograms as published, since stratigraphic durations and taxa are so variable within each partition of the overall sample. However, for direct comparison of cladograms, taxa must be precisely the same.

Furthermore, the category level at which the cladogram is presented may also affect the RCI value. Durations of species and genera are shorter than those of families and higher-level groupings. Thus some low-level groups may have "good" fossil records by any qualitative measurement, but have a low RCI value as the range lengths approach the stratigraphic acuity of some of the fossil records.

As with the SRC, there is no relationship between the RCI value and tree balance ($r^2 = 0$, $p = 0.949$) or NTT ($r^2 = 0$, $p = 0.814$). The fish and echinoderm data even show negative, though nonsignificant, relationships with NTT (Hitchin and Benton 1997).

SCI.—The SCI can be seriously affected by the position of a single taxon. In particular, a long-lived derived taxon may render the rest of the cladogram fully inconsistent, even though all other nodes are otherwise consistent. This factor is not highlighted by a low SCI value, which could imply broad stratigraphic or cladistic analytical problems in a cladogram or merely a single misplaced taxon. Hence, we propose an additional test, which takes into account errors in stratigraphic range that could otherwise bias the consistency of the cladogram by lowering the SCI value.

The question answered by the test would be, How many stratigraphic ranges would have to be changed to make the cladogram totally consistent? To perform the test, ranges of inconsistent taxa would be theoretically altered so that the cladogram would have an SCI of 1 (Fig. 6). The "least changes" value (LCV) produced as a result would equal

$$\frac{\text{number of changed taxa}}{\text{total number of taxa}}.$$

This LCV could then be quoted in association with the SCI value when expressing the

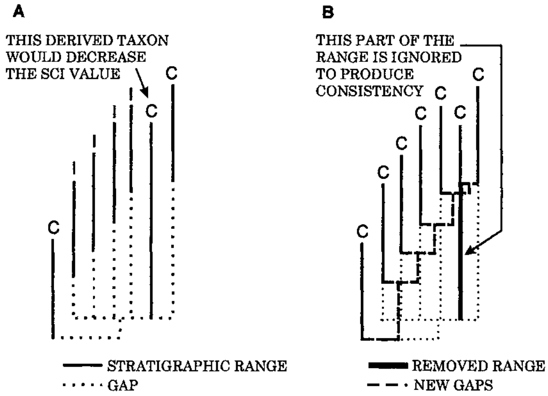

FIGURE 6. Method of calculation of the least changes value (LCV) for the stratigraphic consistency index (SCI). C = consistent node, I = inconsistent node. A, Frequently, one derived taxon has a long early range, which can make all less-derived taxa inconsistent, whereas if that taxon were not present, they would show consistency. B, Removal of the early range segment of this taxon would make the cladogram totally consistent (SCI = 1), and hence the LCV here is one out of seven taxa, or 0.14.

consistency of a cladogram. It would help show why a particular cladogram had a low SCI value. A low LCV would indicate that one taxon, or maybe two or three, was biasing the cladogram towards inconsistency, while a high LCV would show that no such bias was occurring. In conjunction with the LCV, the taxa changed could be noted. In a more elaborate test, the amount of change needed in any stratigraphic range could be taken into consideration in calculating the LCV.

Another problem for the SCI metric arises with small cladograms, especially those with four terminal taxa. As the basal node has to be disregarded in the analysis, that can leave only two nodes resolvable. This can lead to a lack of resolution in the data, especially when the only possible outcomes of the analysis are 0, 0.5, and 1. High acuity is seen only in cladograms with six or more end taxa.

On the basis of the data set of 14 cladograms from Huelsenbeck (1994), Siddall (1996) stated that there was a significant relationship both between SCI and tree balance ($r^2 = 0.723$, $p = 0.002$) and between SCI and NIN ($r^2 = -0.597$, $p = 0.022$). Thus, we suggest that Siddall observed some spurious relationship, maybe due at least partially to the size of his data set, and that there is no relationship between the stratigraphic index and either tree balance or number of internal nodes (Hitchin and Benton 1997).

Congruence between Stratigraphy and Estimated Phylogeny

All three indices show that there are significant differences in the congruence of stratigraphic and cladistic data for fishes, echinoderms, and continental tetrapods. Two of the indices, SRC and SCI, gave highest indications of congruence for echinoderms. This might be expected as there seems to be a strong relationship between SRC and SCI. These two metrics assess a totally different aspect of the fossil record from the RCI index, and so echinoderms may be thought of as having the best fit between clade and age data out of the three groupings. On the other hand, cladograms of fishes suggest the most complete fossil record, as they have the highest average RCI values.

Completeness of the Fossil Record.—The RCI metric shows that the phylogenies of continental tetrapods suggest, in some cases, a poor fossil record. These few cases drag the mean index result down to 12% lower than that of the echinoderm record, and over 19% lower than that of the fish record. The Kolmogorov-Smirnov tests show that this difference is statistically significant, so, contra Benton and Simms (1995), these data, based on a larger sample of cladograms, suggest that continental tetrapods do have a less complete fossil record than marine organisms.

Within the aquatic realm, there seems to be a significant difference between the echinoderm and fish RCI, with the fish having the better mean. This could result from the difference in the amount of paleontological interest in the two groups, but there seem to be more taxonomists at present working on invertebrate groups than on fishes (Gaston and May 1992), so this effect may be not the true reason. Many fish records may be based on scales, otoliths, or teeth, which are often preserved in conditions not suitable for preservation of less sturdy material, and this may lead to a high completeness. However, echinoderm ossicles also have a high preservation potential (Donovan 1991), and some of them can be identified to a high taxonomic level and thus can be used in estimating stratigraphic ranges.

A new paper by Foote and Raup (1996) presents a method that predicts the initial distribution of taxonomic durations for the observed ranges and therefore provides a measure of completeness of the fossil record. It would be useful to examine the congruence between RCI and completeness as measured by this new method.

Fit of Age and Clade Rank.—Fishes had the worst match of clade and age ranks for both the SRC and the SCI tests, with an average of 0.31 (SRC) and 0.55 (SCI). As the echinoderms have the best fit, it is probably not possible to explain these findings on the basis of broad habitat differences.

It may be possible to explain the low fish SRC by the generally low number of terminal taxa contained within the fish cladograms: 6.5, compared with 8.8 in collapsed continental tetrapod cladograms. However, as the echinoderm cladograms contain practically the same number of terminal taxa (6.4 in a collapsed state), this cannot be a true explanation for the lack of clade and age correlation.

There is another possible explanation for the low SCI value. There are several fish taxa, especially when dealing with the lower end of actinopterygian evolution, where some of the age data do not correlate with clade data. However, while this may account for several specific instances of low SCI, it cannot be a universal mechanism for explaining low results. The one thing that may be concluded from these data is that there is a true disagreement between the stratigraphic and cladistic data when it comes to some groups of fish.

When looking at the marine and continental record, it seems that the marine record in general has a higher proportion of very good results for all three metrics. The higher average RCI is probably a result of the differences in preservation potential between the two groups. It has often been assumed that marine organisms have a more complete fossil record because they lived in an aquatic environment, and so had a better chance of being preserved after death. Continental vertebrates, on the other hand, seem to have less chance of being preserved. Shipman (1981) states that there are only a few, specialized environments within continents where fossilization potential is high, and continental vertebrate preservation must be adversely affected.

The RCI may be biased by the number of fossil-bearing strata in which the organisms are preserved. However, this has not been tested for here, since we have considered the stratigraphic range of taxa as a whole and not the intra-range distribution of the fossils. Further expansion of the RCI test could involve analysis of gaps within the range of the taxa to produce a more accurate estimate of the ratio of record to gap.

In the SCI and SRC tests, the higher scores for marine organisms must reflect the high echinoderm results, rather than those of the fishes.

Conclusions

The RCI and the SRC/SCI metrics assess very different aspects of the match of stratigraphic and cladistic data. This difference leads to opposing results for the two types of test, especially in the case of the fishes. Low results in the RCI test tend to suggest a poor fossil record as shown by present-day phylogenies, since this metric focuses on time durations rather than simply on relative rank. Bad results for the SRC/SCI metrics tend to suggest a mismatch of stratigraphy and cladistics—which could be the result of either a poor estimate of phylogeny or a bad fossil record. It is more likely that bad results are the result of a bad fossil record, since this leads to poor sampling, which can in turn lead to a poor estimate of phylogeny (e.g., Huelsenbeck 1991; Lecointre et al. 1993). The solution of this puzzle will be important, as it could help to determine how accurate published phylogenies are.

The index that one should use thus depends on which aspect of the fossil record is to be examined. The RCI should be used when information about the completeness of the record of a cladogram/clade is needed, while the SRC/SCI should be used when information about the fit of stratigraphy and cladistics is needed. The SRC provides a quick, first-approximation test as it simply assesses the fit of clade and age ranks. However, the SCI test may be more informative as it measures the fit of the stratigraphic record to cladograms on a node-by-node basis. The cladogram need not be collapsed to pectinate form, thus minimizing information loss.

Acknowledgments

We thank D. C. Fisher, J. P. Huelsenbeck, and P. J. Wagner for helpful comments on this manuscript, A. B. Yule for help with statistics, and the Leverhulme Trust (individual grants 1988 and Grant F182/AK) for funding of this work. Data used in this paper are available at http://www.gly.bris.ac/research/palaeo/cladestrat.html.

Literature Cited

Allison, P. A., and D. E. G. Briggs. 1991. Taphonomy: releasing the data locked in the fossil record. Plenum, New York.

Ax, P. 1987. The phylogenetic system. Wiley, New York.

Benton, M. J. 1993. The fossil record 2. Chapman and Hall, London.

———. 1994. Palaeontological data, and identifying mass extinctions. Trends in Ecology and Evolution 9:181–185.

———. 1995. Testing the time axis of phylogenies. Philosophical Transactions of the Royal Society of London B 349:5–10.

Benton, M. J., and M. J. Simms. 1995. Testing the marine and continental fossil records. Geology 23:601–604.

Benton, M. J., and G. W. Storrs. 1994. Testing the quality of the fossil record: Paleontological knowledge is improving. Geology 22:111–114.

———. 1996. Diversity in the past: comparing cladistic phylogenies and stratigraphy. Pp. 19–40 *in* M. E. Hochberg, J. Clobert, and R. Barbault, eds. Aspects of the genesis and maintenance of biological diversity. Oxford University Press, Oxford.

Benton, M. J., and R. Hitchin. 1996. Testing the quality of the fossil record by groups and by major habitats. Historical Biology 12:111–157.

Caldwell, M. W., R. L. Carroll, and H. Kaiser. 1995. The pectoral girdle and forelimb of *Carsosaurus marchesetti* (Aigialosauridae), with a preliminary phylogenetic analysis of mosasauroids and varanoids. Journal of Vertebrate Paleontology 15: 516–531.

Clos, L. M. 1995. A new species of *Varanus* (Reptilia: Sauria) from the Miocene of Kenya. Journal of Vertebrate Paleontology 15:254–267.

Darwin, C. 1859. On the origin of species. J. Murray, London.

Denton, R. K., Jr., and R. C. O'Neill. 1995. *Prototeuis stageri* gen. et sp. nov., a new teiid lizard from the Upper Cretaceous Marshalltown Formation of New Jersey, with a preliminary phylogenetic revision of the Teiidae. Journal of Vertebrate Paleontology 15:235–253.

Donovan, S. K. 1991. The taphonomy of echinoderms: calcareous multi-element skeletons in the marine environment. Pp. 241–269 *in* S. K. Donovan, ed. The processes of fossilization. Belhaven, London.

Eldredge, N., and J. Cracraft. 1980. Phylogenetic patterns and the evolutionary process. Columbia University Press, New York.

Etheridge, R., and K. de Queiroz. 1988. A phylogeny of Iguanidae. Pp. 283–368 *in* R. Estes and G. Pregill, eds. Phylogenetic relationships of the lizard families. Essays Commemorating Charles L. Camp. Stanford University Press, Stanford.

Fisher, D. C. 1994. Stratocladistics: morphological and temporal patterns and their relation to phylogenetic process. Pp. 113–171 *in* L. Grande and O. Rieppel, eds. Interpreting the hierarchy of nature. Academic Press, San Diego.

Forey, P. L. 1988. Golden jubilee for the coelacanth *Latimeria chalumnae*. Nature 336:727–732.

———. 1991. *Latimeria chalumnae* and its pedigree. Environmental biology of fishes 32:75–97.

Gardiner, B. G,. and B. Schaeffer. 1989. Interrelationships of lower actinopterygian fishes. Zoological Journal of the Linnean Society 97:135–187.

Gaston, K. J., and R. M. May. 1992. Taxonomy of taxonomists. Nature 356:281–282.

Gauthier, J., A. G. Kluge, and T. Rowe. 1988. Amniote phylogeny and the importance of fossils. Cladistics 4:105–209.

Grismer, L. L. 1988. Phylogeny, taxonomy, classification, and biogeography of eublepharid geckos. Pp. 283–368 *in* R. Estes and G. Pregill, eds. Phylogenetic relationships of the lizard families. Essays Commemorating Charles L. Camp. Stanford University Press, Stanford.

Hennig, W. 1966. Phylogenetic systematics. University of Illinois Press, Urbana.

Hitchin, R., and M. J. Benton. 1997. Stratigraphic indices and tree balance. Systematic Biology (in press).

Huelsenbeck, J. P. 1991. When are fossils better than extant taxa in phylogenetic analysis? Systematic Zoology 40:458–469.

———. 1994. Comparing the stratigraphic record to estimates of phylogeny. Paleobiology 20:470–483.

Lecointre, H. L. V., G. Lecointre, and R. Perasso. 1993. A 28S RNA-based phylogeny of the gnathostomes: first steps in the analysis of conflict and congruence with morphologically-based cladograms. Molecular Phylogeny and Evolution 2:31–51.

Norell, M. A. 1992. Taxic origin and temporal diversity: The effect of phylogeny. Pp. 89–118 *in* M. J. Novacek and Q. D. Wheeler, eds. Extinction and phylogeny. Columbia University Press, New York.

———. 1993. Tree-based approaches to understanding history: comments on ranks, rules, and the quality of the fossil record. Science 293A:407–417.

Norell, M. A., and M. J. Novacek. 1992a. Congruence between superpositional and phylogenetic patterns: Comparing cladistic patterns with fossil records. Cladistics 8:319–337.

———. 1992b. The fossil record and evolution: comparing cladistic and paleontologic evidence for vertebrate history. Science 255:1691–1693.

Novacek, M. J. 1992. Fossils as critical data for phylogeny. Pp. 46–88 *in* M. J. Novacek and Q. D. Wheeler, eds. Extinction and phylogeny. Columbia University Press, New York.

Patterson, C. 1981. Significance of fossils in determining evolutionary relationships. Annual Review of Ecology and Systematics 12:195–223.

Platnick, N. I. 1979. Philosophy and the transformation of cladistics. Systematic Zoology 28:537–546.

Raup, D. M. 1972. Taxonomic diversity during the Phanerozoic. Science 215:1065–1071.

Ryan, T. A., B. L. Joiner, and T. A. Ryan. 1985. Minitab handbook. Duxbury Press, Boston.

Shipman, P. 1981. Life history of a fossil. Harvard University Press, Cambridge and London.

Siddall, M. E. 1996. Stratigraphic consistency and the shape of things. Systematic Biology 45:111–115.

Smith, A. B. 1994. Systematics and the fossil record. Blackwell Scientific, Oxford.

Sprent, P. 1989. Applied nonparametric statistical methods. Chapman and Hall, London.

Van Andel, T. H. 1981. Consider the incompleteness of the geological record. Nature 294:397–398.

Wiley, E. O. 1981. Phylogenetics: the theory and practice of phylogenetic systematics. Wiley, New York.