VIMS Articles

1996

Stratigraphic consistency and the shape of things

ME Siddall
Virginia Institute of Marine Science

Points of View

Stratigraphic Consistency and the Shape of Things

MARK E. SIDDALL

Virginia Institute of Marine Science, School of Marine Science, The College of William and Mary,
Gloucester Point, Virginia 23062, USA; E-mail: mes@vims.edu

Huelsenbeck (1994) identified three unsolved issues regarding the use of temporal information in the fossil record: (1) how goodness of fit between stratigraphy and phylogeny should be determined, (2) how the significance of this fit should be determined, and (3) how those results might be employed other than for description. With respect to goodness of fit, Huelsenbeck (1994) suggested that his stratigraphic consistency index (SCI) was both intuitively simple and not subject to the biases inherent in other stratigraphic indices. With respect to these prior indices (Gauthier et al., 1988; Norell and Novacek, 1992), apparent biases are the result of a logical incompatibility of data types. These indices are simply the nonparametric Spearman correlation between rank stratigraphic age and rank position on a cladogram. The incompatibility stems from the fact that stratigraphic data are inherently linear whereas trees (and the genealogies they represent) are not so constrained. Thus, with respect to the approach of Gauthier et al. (1988), only a maximally unbalanced tree (i.e., a fully pectinate topology) can yield a perfect fit because the nodes, and thus the ranks, are then linearly arranged. Any deviation from this topology will preclude the ordering of nodes in one dimension, yielding a lower maximum goodness-of-fit value; a completely balanced tree is the least likely to yield a good fit. Recognizing this limitation, Norell and Novacek (1992) simply pruned away elements of the tree until a pectinate tree was achieved. Huelsenbeck's SCI was devised explicitly to avoid the logical bias of the Spearman coefficient approach. In this it succeeds. By considering the relative consistency of each node on its own and then taking the ratio of the number of consistent nodes to the total number of nodes, rank stratigraphic data can be conceived of that will yield a perfect or wholly imperfect SCI for various topologies. Given the apparent lack of bias in SCI, Huelsenbeck suggested its use as an optimality criterion for choosing a root for otherwise unrooted most-parsimonious networks or as an optimality criterion that stands apart from the most-parsimonious character distribution (see Fisher [1992] for similar stratocladistic criteria). Huelsenbeck (1994:480) urged caution, indicating that "the behavior of the modified optimality criterion for estimating phylogenetic trees is unknown." My intention here is not to directly address the optimality issue but to investigate the more general premise that the SCI is an unbiased estimator of stratigraphic consistency.

SIZE AND BIAS

A common bias in indices is one relating to size or scale. For example, a systematic bias in Kluge and Farris's (1969) consistency index is a negative correlation with the number of characters and with the number of taxa in an analysis (Farris, 1972; Sanderson and Donoghue, 1989). Klassen et al. (1991) showed convincingly that this bias
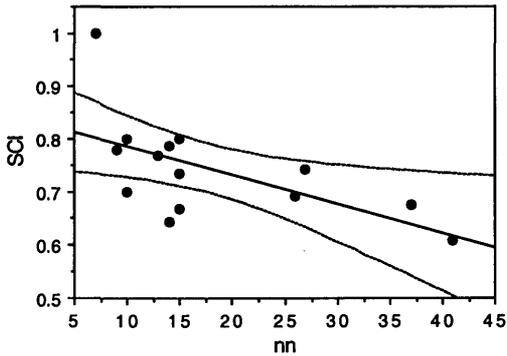
FIGURE 1. Relationship between number of nodes (nn) and SCI for 14 data sets (see Huelsenbeck, 1994). Dashed lines indicate 95% confidence intervals of the mean. $r = -0.597$; $P = 0.022$.

is stronger for increasing numbers of taxa than for increasing numbers of characters. The inclusion of more and more taxa was thought to bias towards more homoplasy in a stochastic fashion relating to the increased probability, as more taxa are included, that a character will be found in a nonhomologous distribution.

Although the logical bias of rank correlation approaches was circumvented by Huelsenbeck (1994), he did not investigate whether or not the SCI suffers from some systematic bias, such as one relating to size or scale. Nonetheless, Huelsenbeck (1994) provided 14 examples of SCI application to data sets for which there are comparable phylogenetic and stratigraphic data. If there is some systematic bias in SCI induced by the size of the comparison being made, one would expect a significant correlation between number of nodes (nn) and SCI. In fact, nn and SCI are significantly ($P = 0.022$) and negatively ($r = -0.597$) correlated (Fig. 1). SCI is apparently biased by the number of nodes (resolved clades) in the tree, yielding lower SCI values for larger trees with more clades. Although one might expect lower SCI because of poorer and poorer fossil retention as maximum age for the clade increases, I have made no such claim here; I suggest only that there is a systematic bias toward lower SCI as the number of taxa increases, irrespective of age.
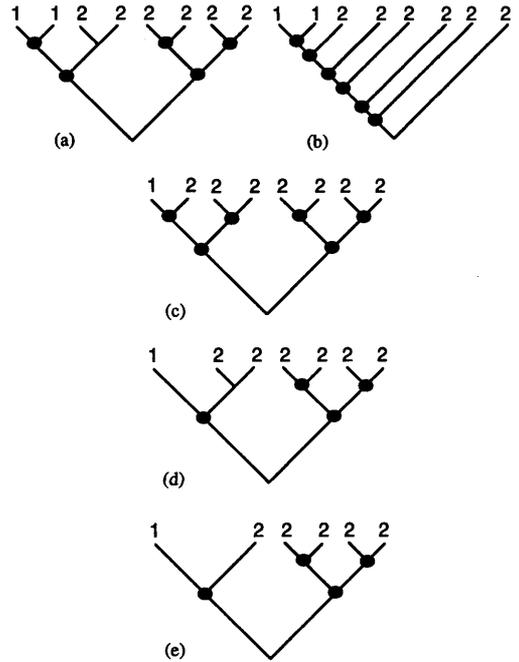

FIGURE 2. Combinations of stratigraphic data in relation to balanced (a) and imbalanced (b) tree shapes and the effect of pruning taxa on number of consistent nodes (c–e). Numbers at terminals indicate the age of each taxon in millions of years. ● = stratigraphically consistent nodes according to Huelsenbeck's (1994) criteria.

CONSISTENCY AND SHAPE

The logical incompatibility between linear fossil data and nonlinear (hierarchical) phylogenetic data applies both to the rank correlation approach and to the SCI. The SCI is far from agnostic as to the topology of the phylogeny with which stratigraphic data are being compared, although it is not as obviously affected as are other indices. Consider the eight-taxon trees presented in Figure 2. Comparison of Figures 2a and 2b for the same ages on two different topologies reveals this problem in terms of clade size. For both trees, all terminals have the same age (2 million years [MY]), except for one pair of sister taxa that is younger than the rest (1 MY). In the balanced tree (Fig. 2a), the SCI is 0.83, whereas for the pectinate tree (Fig. 2b), the SCI is perfect (i.e., 1.00), even though both have a 1 MY–2 MY sister group relationship.
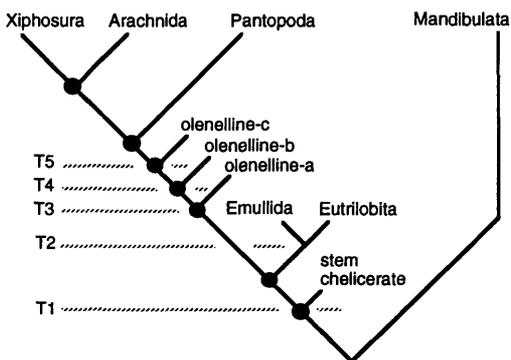
FIGURE 3. Effect on stratigraphic consistency of comparing the extinct monophlyetic group Trilobita (Emulida + Eutrilobita) with the extinct "olenelline" metataxa. T1–T5 denote decreasing stratigraphic age. [Redrawn from Ax (1985).]

FIGURE 4. All N(n) topological types for up to n = 6 taxa. ◆ = sister group relationships in which both groups subtend more than one terminal.

The only difference is that in the balanced tree the 1-MY stratigraphic age is reflecting back on more than one taxon, whereas in the pectinate tree it is reflecting back on only one taxon. The effect on the relative inclusiveness of an analysis is demonstrated in Figures 2c, 2d, and 2e. In each of these trees, there is only one taxon that differs from the rest in age (1 MY). In the first tree (Fig. 2c), there is perfect consistency. Pruning out the 2-MY taxon sister to the younger taxon leads to inconsistency (Fig. 2d) because it now reflects on a clade that has more than one taxon. Pruning of one of these two taxa returns the overall tree to perfect consistency (Fig. 2e).

Lest this be considered too much cladistic gymnastics with little bearing on reality, consider Ax's (1985) treatment of the Chelicerata (Fig. 3). Although few would consider the depicted relationships as inconsistent with the fossil record, the relationship between the Trilobita clade (including two taxa: Emullida and Eutrilobita) and the stem metataxon "olenilline-a" would lead to stratigraphic inconsistency because there is more than one taxon in the Trilobita. If Trilobita had been represented as a single taxon, as were Arachnida, Xiphosura, and Pantopoda, one would not have concluded that there was inconsistency.

SIZE AND SHAPE

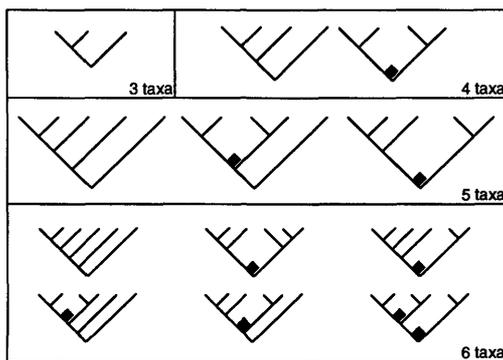Although clearly there is reason to expect a positive relationship between SCI and imbalance, how does this relationship affect the systematic bias apparent in Figure 1? For any given number of taxa (n) there is a finite number of possible bifurcating topologies (N[n]). These topological types are enumerated in Figure 4 for N(3), N(4), N(5), and N(6) to illustrate the relationship between clade size and expectations of imbalance. For n = 3, there is one topology that is completely imbalanced. For n = 4, there are two topologies, one pectinate and one balanced. In the pectinate tree, there are no pairwise sister group comparisons in which both groups have more than one taxon. In the balanced tree there is one such comparison. For n = 5, there are two trees with one such sister group relationship involving two or more taxa in each group and there is the one pectinate tree. For n = 6, there are four trees with one pair of sister groups for which both groups have more than one taxon, one tree for which there are two such sister group comparisons, and the pectinate tree for which there are none. As a tree becomes more balanced, it will have more of these sister group comparisons wherein both groups subtend more than one terminal. As the number of taxa in a tree increases, the number and the relative proportion of topologies with more and more balance (symmetry) increases logarithmically. As expected, then, for Huelsenbeck's (1994) 14 data sets, the correlation between
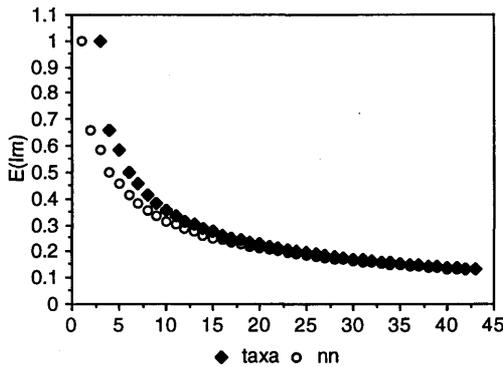
FIGURE 5.   Functional relationship between number of taxa (◆) or number of nodes (○) and the Markovian expectation for Heard's (1992) imbalance index, E(Im).
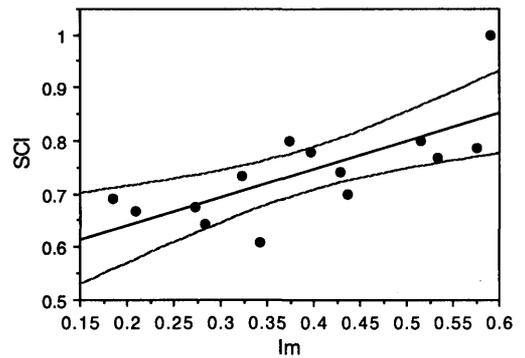

FIGURE 6.   Relationship between Heard's (1992) imbalance index (Im) and SCI for 14 data sets (see Huelsenbeck, 1994). Dashed lines indicate 95% confidence intervals of the mean. $r = 0.723$; $P = 0.002$.

nn and ln(SCI) was stronger ($r = -0.620$, $P = 0.016$) than that for nn and SCI.

The measure of tree balance is not a new concept, and appropriate indices have been derived that are based on the number of nodes between each node and each terminal subtended by that node in a particular tree of a particular shape (see Kirkpatrick and Slatkin [1993] for a review of some of these indices). Heard (1992) suggested the index of imbalance (Im), which is based on the pairwise clade size differences for all nodes in a given tree (Colless's [1995] index of symmetry [I_{sym}] is identical but is a confusing label because the values of Im and I_{sym} are maximal when a tree is most imbalanced [pectinate] and minimal when a tree is most balanced [symmetrical]). The expected value of this index, E(Im), for n taxa assuming a Markovian null model is given by

$$E(\text{Im}) = \frac{2n}{(n-1)(n-2)} \sum_{j=2}^{n/2} \frac{1}{j}$$

when n is even and

$$E(\text{Im}) = \frac{2n}{(n-1)(n-2)} \left(\frac{1}{n} + \sum_{j=2}^{(n-1)/2} \frac{1}{j} \right)$$

when n is odd. Assuming fully bifurcating topologies, Figure 5 illustrates the logarithmically negative relationship of E(Im) to both number of taxa and nn for nn up to 41. However, polytomies present in the 14

trees from which Huelsenbeck calculated SCI values present difficulties in calculating values of Im, which requires fully bifurcating trees. Artificially resolved polytomies would circumvent this problem but would then have implications for the calculation of SCI values (Fig. 2). In my calculation of values of Im for Huelsenbeck's 14 trees (1994: table 1), the two largest clades subtended from a multifurcation were compared to determine the consistency of that node. Also, insofar as the SCI does not consider the outgroup node (Huelsenbeck, 1994), neither did my calculations of Im (Im normally does include this node). Figure 6 illustrates the positive relationship ($r = 0.723$, $P = 0.002$) between Im and SCI for these data. As imbalance in trees increased (as they became more pectinate), SCI was biased towards more favorable values.

These findings have important implications for the suitability and utility of the SCI. In addition to the obvious weakening of its descriptive utility, its use as an optimality criterion either for choosing from among multiple equally parsimonious trees or for choosing trees independent of character parsimony would be biased towards imbalanced trees. For the phylogeny of *Exogyra*-like oysters (Huelsenbeck, 1994: fig. 9), tree B preferred under the stratigraphic consistency optimality criterion was fully pectinate, whereas tree A under

the parsimony criterion was not. Huelsenbeck's (1994:480) cautionary note appears to be well founded: the SCI, although descriptively useful for trees of the same size and shape, can be used neither as an unbiased summary statistic of stratigraphic fit nor as an alternative optimality criterion.

ACKNOWLEDGMENTS

REFERENCES

Ax, P. 1985. Stem species and the stem lineage concept. Cladistics 1:279–287.

COLLESS, D. H. 1995. Relative symmetry of cladograms and phenograms: An experimental study. Syst. Biol. 44:102–108.

FARRIS, S. J. 1972. Inferring phylogenetic trees from distance matrices. Am. Nat. 106:645–668.

FISHER, D. C. 1992. Stratigraphic parsimony. Pages 124–129 in MacClade: Analysis of phylogeny and

Associate Editor: David Baum

character evolution (W. P. Maddison and D. R. Maddison, eds.). Sinauer, Sunderland, Massachusetts.

GAUTHIER, J., A. KLUGE, AND T. ROWE. 1988. Amniote phylogeny and the importance of fossils. Cladistics 4:105–209.

HEARD, S. P. 1992. Patterns in tree balance among cladistic, phenetic, and randomly generated phylogenetic trees. Evolution 46:1818–1826.

HUELSENBECK, J. P. 1994. Comparing the stratigraphic record to estimates of phylogeny. Paleobiology 20: 470–483.

KIRKPATRICK, M., AND M. SLATKIN. 1993. Searching for evolutionary patterns in the shape of a phylogenetic tree. Evolution 47:1171–1181.

KLASSEN, G. J., R. D. MOOI, AND A. LOCKE. 1991. Consistency indices and random data. Syst. Zool. 40: 446–457.

KLUGE, A. G., AND J. S. FARRIS. 1969. Quantitative phyletics and the evolution of anurans. Syst. Zool. 18:1–32.

NORELL, M., AND M. NOVACEK. 1992. The fossil record and evolution: Comparing cladistic and paleontologic evidence for vertebrate history. Science 255:1690–1693.

SANDERSON, M. J., AND M. J. DONOGHUE. 1989. Patterns of variation in levels of homoplasy. Evolution 43:1781–1795.

Imperfect Information and the Balance of Cladograms and Phenograms

STEPHEN B. HEARD[1] AND ARNE Ø. MOOERS

Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

Do cladograms tend to be more imbalanced (less symmetrical) than phenograms? Colless (1982) suggested anecdotally that they do, but more rigorous examination of 208 literature cladograms and phenograms (Heard, 1992; see also Mooers, 1995) revealed no such difference. In a recent note, Colless (1995) returned to this question and concluded that when cladistic and phenetic techniques are used to estimate phylogenies from the same data sets, the cladistic trees are consistently and strikingly more imbalanced than the phenetic trees. Colless (1995:105) explained this difference by claiming that "PAUP (and by extension the phylogenetic methods [parsimony] it seeks to model) is biased towards producing asymmetrical [imbalanced] dendrograms." We argue instead that Colless's results only show differences in the way cladistic and phenetic techniques handle data that are too sparse

[1] Present address: Department of Biological Sciences, University of Iowa, Iowa City, Iowa 52242, USA. E-mail: stephen-heard@uiowa.edu.