



Forum

Stratigraphic Fit to Phylogenies: A Proposed Solution

Mark E. Siddall

Museum of Zoology, University of Michigan, Ann Arbor, Michigan, 48109, USA

Accepted 22 December 1997

Three measures intended to assess the fit of stratigraphic age to the fossil record have been suggested previously: the Spearman Rank Correlation (SRC), the Stratigraphic Consistency Index (SCI) and the Relative Completeness Index (RCI). The original formulation of SRC is intractable to all but pectinate trees and the corrective pruning procedure that circumvents this precludes whole-tree estimates of fit. SCI, though it has been claimed otherwise, is strongly biased by tree shape, particularly as one adds more information. RCI is a measure of the amount of gap in the fossil record but has awkward consequences for evolutionary biology when it is maximized. A new approach, the Manhattan Stratigraphic Measure, uses the Manhattan distance between stratigraphic ages to determine fit to a tree. It is not biased by tree shape, it is sensitive to the magnitude of age discrepancy and there is an obvious significance test. © 1998 The Willi Hennig Society

Prior to there being an objective methodology for the empirical evaluation of sister-group relationships, the field of paleontology had a near monopoly on the interpretation of the branching order of taxa. Finding crinoid stems in 500 MY-old Ordovician deposits and finding no frog fossils earlier than the Permian layers was legitimately considered compelling evidence that the common ancestor of Echinodermata predated that of Anura. This paradigm inevitably entailed the

Correspondence to: M. E. Siddall. E-mail: msiddall@umich.edu;
Fax: (313) 663 4080.

assumption that certain fossil taxa necessarily were ancestors of extant forms, and prior to Hennig's (1966) explicit methodology, the relative primitiveness of a trait was determined by its stratigraphic age, not by its concordance with other character information.

Following the ascendancy of cladistic protocols, various authors recognized that corroborating character information was the only test of phylogeny and that ages of fossils alone were unreliable. One does not need a fossil record to construct a well-corroborated cladogram and any putative ancestor can only be designated on the absence of evidence, requiring a monotypic clade without autapomorphies. Fossil taxa still are bearers of important character information and unique combinations of characters (e.g. Crane, 1985; Gauthier et al., 1988; Ramsköld and Werdelin, 1991).

Some have tried to place stratigraphic data in an empirical framework (Gauthier et al., 1988; Norell and Novacek, 1992; Fisher, 1994; Huelsenbeck, 1994; Benton and Hitchin, 1996), each of whom were concerned with how well an independent cladogram fit the stratigraphic placement of taxa. The implications for a good fit or a bad one vary depending on how these data are to be interpreted. Some (e.g. Fisher, 1994; Huelsenbeck, 1994) present arguments for using their measures as phylogenetic optimality criteria. Others (e.g. Norell, 1992, 1996) take the more conservative position that such measures are more of an indication of incompleteness of knowledge as opposed to refutation of cladograms.

SRC, SCI, AND RCI

In general, these measures of stratigraphic fit are constructed in the interests of testing whether or not the age of fossils correlates somehow with their position on a cladogram. Working from the premise that historical theories are testable, which Popper (1980) did not deny, and that stratigraphy is the framework within which to test it, Benton and Hitchin (1996; see also Hitchin and Benton (1997a,b) for a similar treatment) evaluated the available methods in relation to 376 cladograms: the Spearman Rank Correlation (SRC) approach (Gauthier et al., 1988; as modified by Norell and Novacek, 1992), the Stratigraphic Consistency Index (SCI) approach (Huelsenbeck, 1994, but see Siddall, 1995b, 1997), and their own Relative Completeness Index (RCI). Each of these measures, however, is fraught with difficulties. In the original implementation of SRC, Gauthier et al. (1988) measured the Spearman correlation between rank stratigraphic age and clade rank. However, perfect correlations only were possible when clade ranks were non-overlapping; that is, in a pectinate (fully imbalanced) tree. Norell and Novacek (1992) circumvented this problem by performing multiple correlation analyses on pruned subsets of a tree. This has been criticized for discarding information (Huelsenbeck, 1994) which is not actually correct inasmuch as Norell and Novacek (1992) stressed that all trajectories must be passed through. Nonetheless, it is not clear how one can assess the fit of stratigraphic data to the whole tree when it has to be broken up into parts for SRC values. The two SRC values obtained for Gauthier et al.'s (1988) data were 0.538 for sauropsids and 0.978 for synapsids (Norell and Novacek, 1992). These cannot be multiplied for an ensemble SRC of 0.531 because of non-independence of nodes considered in each correlation. Nor can they be averaged for an ensemble SRC of 0.758 because the data are not disjunct and this precludes additivity. Huelsenbeck (1994) claimed that SCI solved this problem by treating each node separately. However, I showed that SCI was, nonetheless, biased by clade shape (Siddall, 1995b).

The primary difficulty in SCI relates to how sister clades, each with more than one taxon, perform relative to sister clades in which one is monotypic. That is,

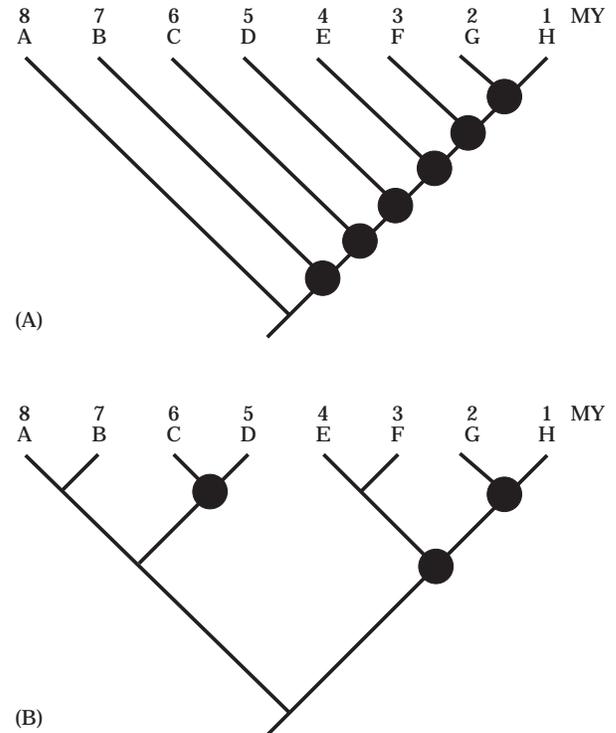


FIG. 1. With the same number of taxa and identical stratigraphic ages, the pectinate tree (A) can yield a value of SCI=1.00, whereas this value is impossible on the balanced tree (B). ● =Stratigraphically consistent nodes.

for example in Fig. 1 where each taxon has a different stratigraphic age, it is possible to achieve SCI=1.00 on the pectinate tree because the monotypic sister taxon to each clade is of greater age (Fig. 1A). In the balanced tree, if one clade is stratigraphically consistent, its sister, by definition, cannot be (Fig. 1B). As is apparent with SRC, SCI is biased in favour of pectinate trees. Also, because a clade's stratigraphic consistency is inseparable from that of others, the values cannot be additive (contra Huelsenbeck, 1994).

RCI is somewhat different in that it tries to assess not the fit of absolute ages per se, but the amount of implied gap in the fossil record (Benton and Hitchin, 1996). Figure 2A depicts the cladistic relationships for four hypothetical taxa. Figure 2B depicts their stratigraphic ranges, so far as they are hypothetically known. This information is combined in Fig. 2C under the logic that sister taxa must be of equal age, and hatched regions represent gaps in stratigraphic information, or minimum implied gap (MIG). In this case, $\sum \text{MIG} = 1.5 \text{ MY}$ [0.5 MY for the ancestor of (C D), and 0.5 MY for the ancestor of (B C D)]. RCI is then

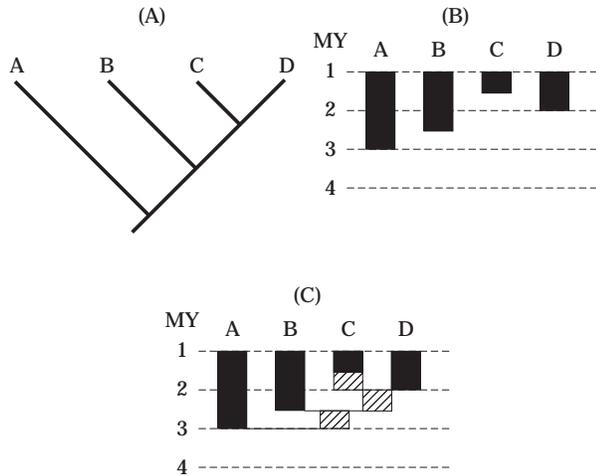


FIG. 2. The Relative Completeness Index (RCI) combines a cladogram (A) with the stratigraphic ranges for the taxa (B) to determine the Minimum Implied Gap (MIG) represented by hatched areas in (C). RCI=1.00 when MIG=0.00, which can only occur if all fossil taxa have their genesis at the same time.

$1 - \frac{\sum \text{MIG}}{\sum \text{SRL}}$ expressed as a percentage, where SRL is the known age range of taxa. Benton and Hitchin (1996: 118) claimed that values of RCI “range from 0% (MIG=SRL) to 100% (MIG=0)”, which in itself reveals the flaw in this method. The only way that MIG could ever equal zero is if all taxa considered were of equal age. Thus the conditions under which the fossil record would *best* corroborate a cladogram under the RCI criterion are restricted to the simultaneous origin of all taxa—a most disturbing proposition.

A PROPOSED SOLUTION—MSM

The Manhattan Stratigraphic Measure (MSM) is intuitively simple and is straightforward in its application. Rather than perceiving the question of stratigraphic fit to be one of “are basally arranged taxa found in older strata?”, which has confounded previous approaches by trying to fit stratigraphic data to a hierarchical tree in a Euclidean framework, MSM fits the Manhattan distance matrix of ages to the tree. The use of Manhattan metrics in phylogenetics is not new (Farris, 1967, 1972; Kluge and Farris, 1969; Farris et al., 1970). Its use lies at the core of the Wagner algorithm for tree construction itself (Farris et al., 1970, Farris, 1972).

The Manhattan stratigraphic matrix is symmetrical and is composed of the absolute difference in ages (D)

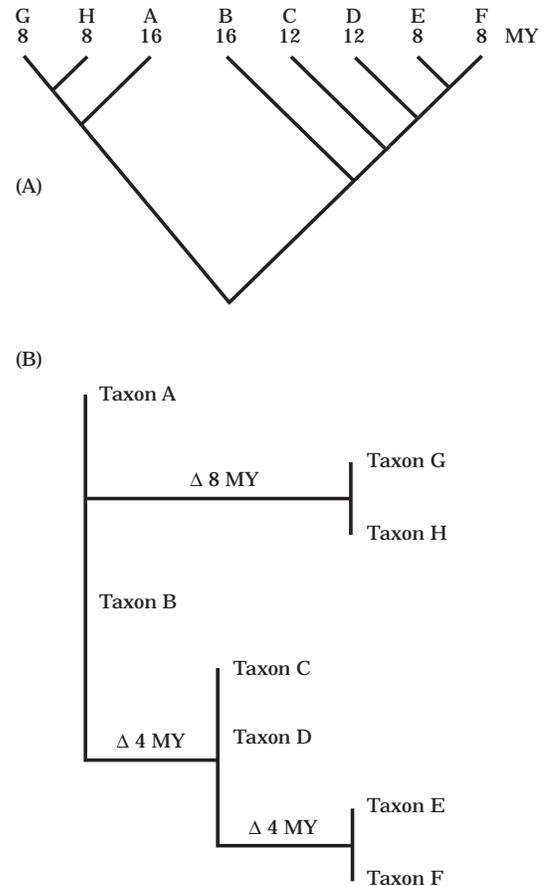


FIG. 3. Huelsenbeck's (1994) figure 2 redrawn (A) and a length optimization of the corresponding Manhattan Stratigraphic matrix for these data on this tree (B).

between each pairwise comparison of taxa ($D(i,j) = |age_i - age_j|$). Using Huelsenbeck's (1994) figure 2 as an example (Fig. 3A), the absolute difference in age between taxon A and taxon B (16-16 MY) is 0 MY, the absolute difference in age between taxon A and taxon C (16-12 MY) is 4 MY, and so on. The Manhattan stratigraphic matrix, then, for the taxa in Fig. 3A is:

Taxon	A	B	C	D	E	F	G	H
Age	16	16	12	12	8	8	8	8
A	0	0	4	4	8	8	8	8
B	0	0	4	4	8	8	8	8
C	4	4	0	0	4	4	4	4
D	4	4	0	0	4	4	4	4
E	8	8	4	4	0	0	0	0
F	8	8	4	4	0	0	0	0
G	8	8	4	4	0	0	0	0
H	8	8	4	4	0	0	0	0

If one takes this matrix and optimizes it on the observed tree as a Sankoff character (Sankoff and Rousseau, 1975; Sankoff et al., 1976) a length (L_o) is obtained (Fig. 3B). However, this length alone is insufficient for our purposes because the meaning of the value is determined by the structure of the data. Data sets with larger numbers of taxa or with relatively older ages will result in longer lengths (L_o) without necessarily implying worse fits than other data sets. In order to be useful, L_o must be compared with how short a length *could* be obtained from these data. By searching for the optimal tree given the Sankoff matrix, this minimum length (L_m) is obtained. MSM then is simply L_m/L_o .

PAUP* (Swofford, forthcoming) permits implementation of this procedure as follows (using the data in Fig. 1 as an example):

- (1) assign each taxon a different character state for one character:

```
begin data;
dimensions ntax=9 nchar=1;
format symbols = "a~z 0~9";
matrix
root a
taxonA a
taxonB b
taxonC c
taxonD d
taxonE e
taxonF f
taxonG g
taxonH h
;
end;
```

Methodological Note #1: The added taxon "root" above is a vector applied to the base of the tree with the oldest of all observed ages. This is required because otherwise one could achieve a perfect value for MSM when the youngest taxon is at the base and the oldest taxon is the most derived with all others graded in between. By applying this root vector, if the oldest taxon is at the base of the tree, there is no added length, whereas if it is not, there is added length. This is appropriate, and is identical to the addition of a root vector in Brooks Parsimony Analyses as argued elsewhere (Siddall, 1995a).

- (2) create a Sankoff matrix character type from the Manhattan stratigraphic matrix using the states defined in step 1:

```
begin assumptions;
usertype strat=8
a b c d e f g h
. 0 4 4 8 8 8 8
0 . 4 4 8 8 8 8
4 4 . 0 4 4 4 4
4 4 0 . 4 4 4 4
8 8 4 4 . 0 0 0
8 8 4 4 0 . 0 0
8 8 4 4 0 0 . 0
8 8 4 4 0 0 0 .
;
end;
```

- (3) Obtain L_m by searching for the optimal tree under this character type:

```
ctype strat: 1;
hsearch;
pscore;
```

- (4) Obtain L_o by optimizing the character on the "observed" tree (but see *Methodological Note 2*):

```
cypte strat: all;
[NOTE: taxon #1="root"]
usertree (1,(((9,8),2),(3,(4,(5,
(6,7))))));
pscore;
```

Methodological Note #2: In applications of Huelsenbeck's (1994) SCI, polytomies are problematic. Huelsenbeck (1994: 472) suggested that polytomies be treated "by examining the age of all lineages radiating above the clade. The oldest age is then taken and used for comparison to the age below the node." In calculation of MSM, if the length is obtained by merely optimizing on the tree (as above), polytomies are treated as "hard" and may inflate values. An appropriate alternative is to treat them as "soft" by using the observed tree as a constraint in a search as follows:

```
ctype strat: all;
[NOTE: taxon #1="root"]
constraint mytree=(1,(((9,8),2),
(3,(4,(5,(6,7))))));
hsearch enforce constraint=mytree;
pscore;
```

If there are no polytomies (as in this case), this is no different from simply fitting the character, if there are polytomies, they will be optimally resolved.

- (5) Calculate $MSM=L_m/L_o$.

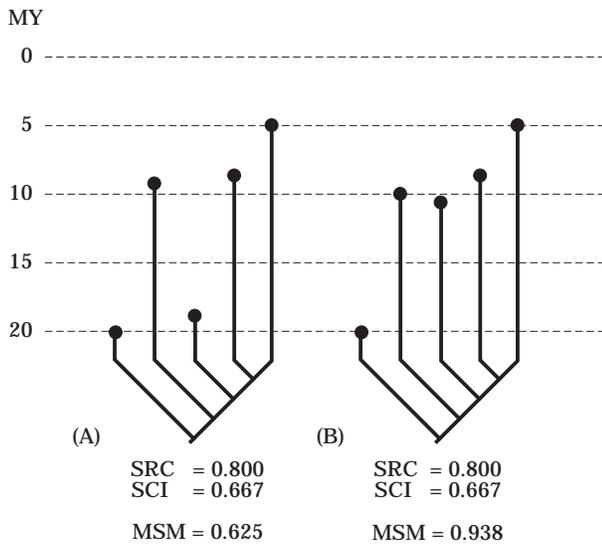


FIG. 4. The Manhattan Stratigraphic Measure (MSM) is sensitive to the absolute differences in ages. Unlike SRC and SCI, if the age discrepancy is larger for one data set (A) than some other (B) the value of MSM will be worse.

MSM as defined above, has certain desirable properties. It is bounded between 0.00 and 1.00. If the fit to the observed tree is the best possible fit, then $MSM=1.00$. Values of exactly 0.00 are not possible, of course, because it is impossible for there to be no fit whatsoever. This, though, is no different than standard Euclidean measures of fit for height and weight, for example, using Pearson's r , or Spearman's S . Unlike previously used methods like SRC (Gauthier et al., 1988; Norell and Novacek, 1992) or SCI (Huelsenbeck, 1994), MSM is appropriately sensitive to the magnitude of the age differences. Compare Fig. 4A, in which a taxon with a stratigraphic age of 9 MY is basal to one that has an age of 18 MY, with Fig. 4B, in which there is a similar arrangement of taxa but the difference in ages of taxa is less marked (10 and 11 MY respectively). Because SRC is concerned with rank order correlation, both of these examples yield the same SRC value (0.800). Similarly, because Huelsenbeck's SCI is concerned that the taxa outside of the clade are merely older than those inside the clade, it too yields identical values (0.667). In contrast, MSM considers the age discrepancy in Fig. 4B ($MSM=0.938$) to be less severe than that in Fig. 4A ($MSM=0.625$). This would appear to be appropriate insofar as the age discrepancy actually is less severe.

Perhaps most importantly, and in consideration of my criticisms of SCI (Siddall, 1995b, 1997), MSM does not suffer from the tree shape biases that SRC and SCI do. In contrast to SCI, MSM shows no shape bias whatsoever (Fig. 5). Applying the same procedure that SCI was subjected to (Siddall, 1997), it is clear that MSM behaves in precisely the same way for the pectinate tree as it does for the symmetrical tree and as it does for random tree topologies. Moreover, and unlike SCI, MSM behaves identically irrespective of the number of possible ages that are distributed across taxa (Fig. 5).

SIGNIFICANCE TEST

It may at first appear to be disturbing that MSM, though not biased by tree shape, still is biased by tree size (Fig. 5). This worry, however, conflates the *magnitude* of a statistic and its *significance*. They are not the same. Consider a correlation of height and weight. For any two points, the correlation for any function is guaranteed to be perfect (i.e. r or $S=1.00$). Moreover, as the number of data points increases, the null expectation of r or S decreases monotonically. MSM is no different. A two-taxon "tree" will always yield an MSM of 1.00. This might be a *good* fit, but it hardly constitutes a *significant* fit. Furthermore, relatively poor fits (e.g. <0.600) might yet be significantly better than chance if there are sufficient degrees of freedom (i.e. taxa). Similar arguments have been raised in relation to supposedly good fits between host and parasite phylogenies (Siddall, 1995a).

In order to construct an appropriate significance test, it must be clear what is being measured. Inappropriate null distributions only render inappropriate P -values. In MSM, the appropriate question is stated as, "How frequently would we expect a stratigraphic fit (MSM) that is as good or better than the observed fit?". As with any correlative analysis, one can determine the exact distribution of possible outcomes for a given set of data. For n stratigraphic age assignments, there are $n!$ permutations across the terminals. For the hypothetical example in Fig. 3, there are 40,320 possible assignments of ages to the taxa. To calculate L_0 for all of these in this exact manner, using PAUP* operating with a 100MHz Sirex 5x86 processor and 24 megabytes of RAM, took 40 minutes. For 10 taxa, this would take

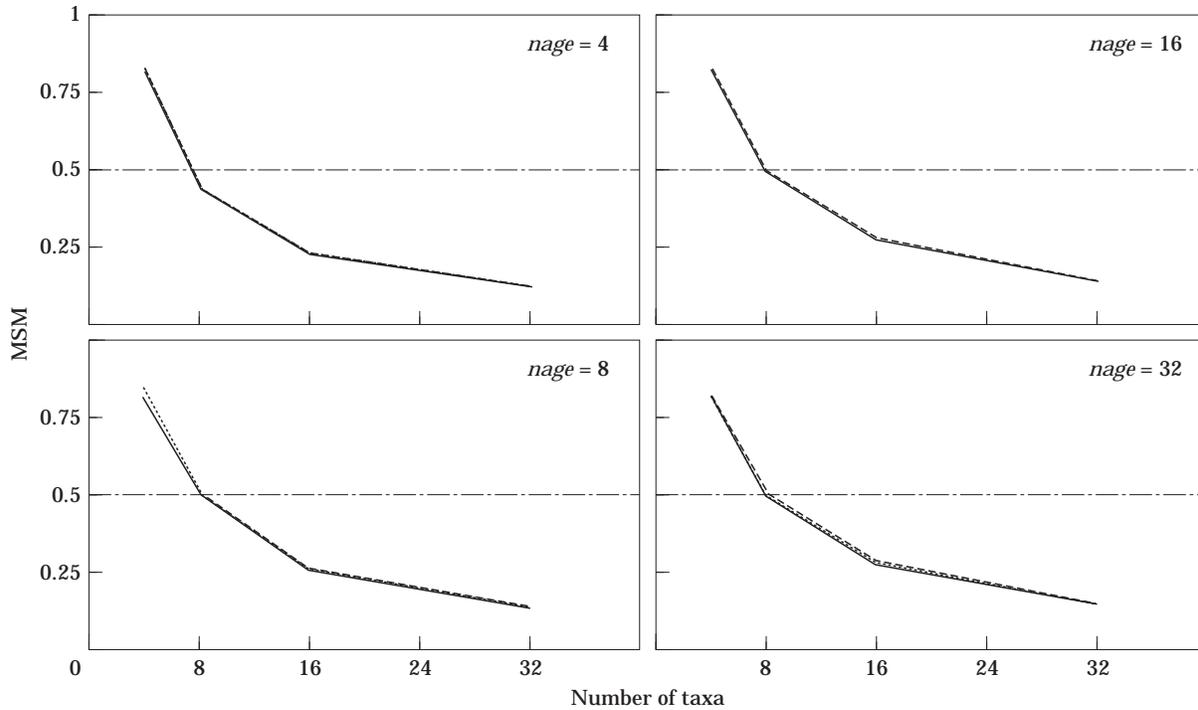


FIG. 5. Results of simulations designed to assess tree shape bias of MSM in relation to wholly meaningless distributions of stratigraphic information across terminal taxa when varying the number of taxa as well as the number of possible strata (*nage*) for pectinate trees (dotted lines), balanced trees (dashed line) and Markovian expectations (solid line).

90 hours. An appropriate alternative to the exact method, and one that is more tractable, is “approximate randomization”; in this case by repeated random assignment of observed ages to taxa and obtaining L_o from the same tree. Considering the data matrix defined in step 1 of the methodological protocol above, this simply entails reassigning character states to the taxa in a manner identical to Archie’s (1989; see also Faith and Cranston, 1991). Unlike the application of PTP, however, trees are not recalculated from the permuted data. Because the shortest tree will always be of the same length (L_m) irrespective of how the states are permuted, only the observed lengths from the original data (L_o) and from the permuted data need be calculated [L_i for all $i < R$, where R is large (e.g. 1000)]. The frequency with which values of L_i are found to be less than, or equal to, L_o is the tail distribution and one-sided P -value for MSM. The “permute” command in PAUP* does this nicely such that the entire procedure of calculating L_m , L_o , as well as $p(L_i \leq L_o)$ can be accomplished by executing the following procedure (using the example in Fig. 3):

```
#NEXUS
begin data;
```

```
dimensions ntax=9 nchar=1;
format symbols="a~z 0~9";
matrix
root a
taxonA a
taxonB b
taxonC c
taxonD d
taxonE e
taxonF f
taxonG g
taxonH h
;
end;
begin assumptions;
usertype strat=8
a b c d e f g h
. 0 4 4 8 8 8 8
0 . 4 4 8 8 8 8
4 4 . 0 4 4 4 4
4 4 0 . 4 4 4 4
8 8 4 4 . 0 0 0
8 8 4 4 0 . 0 0
8 8 4 4 0 0 . 0
```

```

8 8 4 4 0 0 0 .;
end ;
begin paup;
log file=huels.log;
set autoclose;
ctype strat: 1;
hsearch noenforce;
pscore; [ <-- Lm],
constraints mytree= [NOTE: taxon #1 =
"root" ]
(1,(((9,8),2),(3,(4,(5,(6,7))))));
hsearch enforce constraints=mytree;
pscore; [ <-- Lo]
outgroup 1;
permute randomize=ingroup nreps=1000;
[ <-- P - value]
log stop;
quit;
end;

```

Methodological Note #3: Although not stated explicitly in the PAUP block, the permutations are carried out in the context of the same constraint tree as was the search immediately preceding the “permute” command. Thus, it is imperative that the “noenforce” search be conducted prior to the “enforce constraints” search, otherwise the resulting P-value will be invalid.

Methodological Note #4: The “outgroup 1” and “randomize=ingroup” directives retain the root vector (see Methodological Note #1 above) at the base of the tree during permutations in keeping with the premise that the oldest possible age should be there.

Execution of this file in PAUP* results in an unconstrained search length (L_m) of eight steps and an observed optimized length (L_o) of 16 steps (i.e. MSM=0.50); the results of permutation are as follows:

Tree length	Number of replicates
12	10
16*	63
20	201
24	385
28	341

* = length for original (unpermuted) data. $P = 0.073$.

Thus, the fit is not terribly strong [contra Huelsenbeck’s (1994: figure 2) SCI=1.00] and, depending on

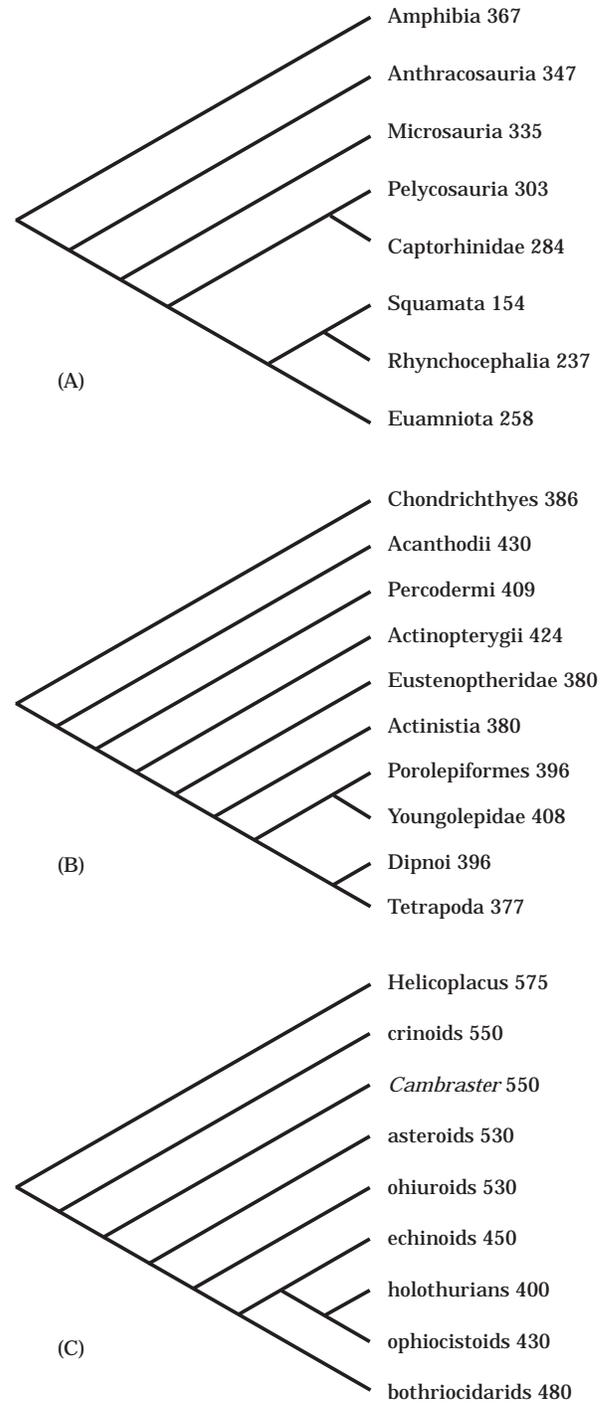


FIG. 6. Phylogenetic hypotheses and associated stratigraphic ages for tetrapods (A), fishes (B), and echinoderms (C) used to examine the utility of MSM.

what level of significance is desired, is not particularly different from what would be expected from ages being randomly distributed across taxa.

As with any kind of significance test, it is impossible to state precisely what cut-off level of the tail distribution is acceptable. Insofar as we are not building bridges or manufacturing toothbrushes, the notions that 0.00001% of cars being blown off a bridge is a limit of acceptability and that producing only five defective toothbrushes in 100 is satisfactory, have no meaning. For that matter, with fewer than seven taxa, a cut-off level of 0.05 would seem to be beyond the power of the test, because there would be less than 1000 possible permutations anyway.

EMPIRICAL EXAMPLES

Figure 6 illustrates three applications of MSM to real data taken from Benton and Hitchin's (1996) study. One relates to tetrapods (Gardiner, 1982), another to fishes (Gardiner, 1984) and a third to echinoderms (Smith, 1988). The ages applied to the terminals were provided by M. J. Benton (pers. comm.). These ages (in MY) are absolute minimum ages of the fossil taxa.

Following the methodological protocol for each of these, the results obtained were values of $MSM=0.833$ ($P=0.002$) for the tetrapod data, $MSM=0.350$ ($P=0.448$) for the fish data, and $MSM=1.000$ ($P=0.001$) for the echinoderm data. Thus, it would seem that there is a reasonably good and significant fit of stratigraphic placement and phylogeny for tetrapods and echinoderms, but not for the fishes.

ACKNOWLEDGEMENTS

I thank Jim Clark for encouraging me to keep thinking about a solution to this issue as well as M. J. Benton and R. Hitchin for providing me with some of their data sets. Jennifer Ast, Arnold Kluge, and Michael Sorenson offered constructive comments on earlier drafts. An anonymous reviewer provided insightful comments which improved the final manuscript. A portion of this work was made possible through a grant-in-kind from the Cladistics Institute in Harbor Springs, Michigan in the summer of 1997.

REFERENCES

Archie, J. W. (1989). A randomization test for phylogenetic information in systematic data. *Syst. Zool.* **38**, 239–252.
 Benton, M. J., and Hitchin, R. (1996). Testing the quality of the fossil record by groups and by major habitats. *Hist. Biol.* **12**, 111–157.
 Crane, P. (1985). Phylogenetic analysis of seed plants and the origin of angiosperms. *Ann. Missouri Bot. Gard.* **72**, 716–793.
 Faith, D. P., and Cranston, P. S. (1991). Could a cladogram this short have arisen by chance alone?: On permutation tests for cladistic structure. *Cladistics* **7**, 1–28.

Farris, J. S. (1967). The meaning of relationship and taxonomic procedure. *Syst. Zool.* **16**, 44–51.
 Farris, J. S. (1972). Estimating phylogenetic trees from distance matrices. *Am. Nat.* **106**, 645–668.
 Farris, J. S., Kluge, A. G., and Eckardt, M. J. (1970). A numerical approach to phylogenetic systematics. *Syst. Zool.* **19**, 172–189.
 Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**, 401–410.
 Fisher, D. C. (1994). Stratocladistics: Morphological and temporal patterns and their relation to phylogenetic process. In "Interpreting the Hierarchy of Nature". (L. Grande, and O. Rieppel, Eds.) pp. 133–171. Academic Press, New York.
 Gardiner, B. G. (1982). Tetrapod classification. *Zool. J. Linn. Soc.* **74**, 207–232.
 Gardiner, B. G. (1984). The relationships of the palaeoniscid fishes, a review based on new specimens of *Mimia* and *Myothomasi* from Upper Devonian of Western Australia. *Bull. Brit. Mus. (Nat. Hist.), Geol. Ser.* **37**, 173–428.
 Gauthier, J., Kluge, A. G., and Rowe, T. (1988). Amniote phylogeny and the importance of fossils. *Cladistics* **4**, 105–209.
 Hennig, W. (1966). "Phylogenetic Systematics". University of Illinois Press, Urbana.
 Hitchin, R., and Benton, M. J. (1997a). Congruence between parsimony and stratigraphy: Comparisons of three indices. *Paleobiology* **23**, 20–32.
 Hitchin, R., and Benton, M. J. (1997b). Stratigraphic indices and tree balance. *Syst. Biol.* **46**, 563–569.
 Huelsenbeck, J. P. (1994). Comparing the stratigraphic record to estimates of phylogeny. *Paleobiology* **20**, 470–483.
 Kluge, A. G., and Farris, J. S. (1969). Quantitative phyletics and the evolution of anurans. *Syst. Zool.* **18**, 1–32.
 Norell, M. A. (1992). Taxic origin and temporal diversity: The effect of phylogeny. In "Extinction and Phylogeny". (M. J. Novacek, and Q. D. Wheeler, Eds.) pp. 89–118. Columbia University Press, New York.
 Norell, M. A. (1996). Ghost taxa, ancestors, and assumptions: A comment on Wagner. *Paleobiology* **22**, 453–455.
 Norell, M., and Novacek, M. (1992). The fossil record and evolution: Comparing cladistic and paleontologic evidence for vertebrate history. *Science* **255**, 1690–1693.
 Popper, K. (1980). Evolution. *New Scientist* **21 August**, 611.
 Ramsköld, L., and Werdlin, L. (1991). The phylogeny and evolution of some phacopid trilobites. *Cladistics* **7**, 29–74.
 Sankoff, D., Cedergren, R. J., and Lapalme, G. (1976). Frequency of insertion-deletion, transversion, and transition in the evolution of 5S ribosomal RNA. *J. Mol. Evol.* **7**, 133–149.
 Sankoff, D., and Rousseau, P. (1975). Locating the vertices of a Steiner tree in arbitrary space. *Math. Prog.* **9**, 240–246.
 Siddall, M. E. (1995a). Phylogenetic covariance probability: confidence and historical associations. *Syst. Biol.* **45**, 48–66.
 Siddall, M. E. (1995b). Stratigraphic consistency and the shape of things. *Syst. Biol.* **45**, 111–115.
 Siddall, M. E. (1997). Stratigraphic indices in the balance: A reply to Hitchin and Benton. *Syst. Biol.* **46**, 569–573.
 Smith, A. B. (1988). Fossil evidence for the relationships of extant echinoderm classes and their times of divergence. In "Echinoderm Phylogeny and Evolutionary Biology" (C. R. C. Paul, and A. B. Smith, Eds), pp. 85–97. Clarendon Press, Oxford.
 Swofford, D. L. (forthcoming). "PAUP*—Phylogenetic Analysis Using Parsimony. version 4.0.0d55 (PAUP.EXE for MS-DOS)". Not available.