FORUM

CHOOSING AMONG MULTIPLE EQUALLY PARSIMONIOUS CLADOGRAMS

James M. Carpenter[1]

[1] *Museum of Comparative Zoology, Harvard University,
Cambridge MA 02138, U.S.A.*

With the realization that multiple equally parsimonious cladograms might exist for a given data set (Mickevich, 1978), construction of classifications in such instances became a problem. Consensus trees (Adams, 1972; Nelson, 1979), originally developed for producing a "compromise classification" (Adams, 1972) between cladograms produced from *different* data sets, is one possible solution. The consensus tree, representing the information on grouping shared by all the competing cladograms, might be viewed as a "conservative" classification. Such a viewpoint has been advocated by proponents of distance techniques in the case of multiple trees of near-optimal fit (e.g. Prager and Wilson, 1978). In this case, such a procedure is a misapplication. The consensus tree, being less resolved than any of the cladograms from which it is calculated, has less explanatory power (Farris, 1983) than any of them, and so *any* of the competing cladograms would be a better choice as a phylogenetic hypothesis/classification. Mickevich and Farris (1981) and Miyamoto (1985) make similar points. The question is then how to choose among the cladograms, if this can be done.

Recently, Brooks et al. (1986) have addressed the issue of choice among multiple equally parsimonious cladograms. However, their discussion is deficient. They considered three statistics: the consistency index (Kluge and Farris, 1959), F-ratio (a simple function of Farris' (1972) f statistic) and the D measure, an application of the Shannon entropy statistic from Gatlin (1972). The consistency index does measure evidential support for phylogenetic hypotheses, but this is not true for the other measures. The f statistic, developed for application to distance analyses, measures the pairwise homoplasies for a given tree, and, to quote Farris (1983:22), "the pairwise homoplasies are not independent". This measure is thus not closely related to the concept of evidential support underlying phylogenetic analysis (Farris, 1983), and, in fact, a tree optimizing f (hence the F-ratio) may not be most parsimonious for a given data set. Brooks et al. (1986) make that last observation, but do not then provide any rationale for using the F-ratio. Similar comments apply to their use of D, which they characterize as measuring the "*information content of the constraints*" (Brooks et al., 1986:572). They provide an example (their fig. 5) where D is optimal for a non-parsimonious cladogram, which is reason enough to reject this measure. Brooks et al. argue that D can distinguish between autapomorphies and synapomorphies, whereas the consistency index cannot. Their discussion of the consistency index is misled. Whereas this index is inflated by inclusion of invariant and autapomorphic characters, plainly it should be calculated with such features excluded in a proper cladistic analysis—a point pheneticists have made (Colless, 1983). Thus, the example shown in their fig. 6a–c has an identical consistency index for all three cladograms calculated over all the characters. But calculated only on informative characters, their cladogram 6b

has a higher index than others (.73 vs. .67). This is because characters 8 and 9 are synapomorphies on cladogram 6b but uninformative on the others, either autapomorphic or invariant. Brooks et al. state (1986:576) that:

> "A comparison of CI values alone will not indicate that one person has found two autapomorphies, thus corroborating the monophyly of that taxon, but adding no support to the proposed genealogical groupings, while the other has found two synapomorphies corroborating the monophyly of the entire clade."

This is only true if the consistency indices are calculated naïvely. The consistency index *is* capable of distinguishing between the cladograms in this example in terms of the informativeness of their underlying data sets. The rationale of Brooks *et al.* for employing *D* is therefore specious.

For one data set, the consistency index will be the same for each most parsimonious cladogram, whether uninformative characters are included or not, and so offers no grounds for choice among the cladograms. I suggest use of a different technique for selecting among cladograms; application of successive approximations character weighting (Farris, 1969). This character weighting method is developed directly from the concept of "cladistic reliability" (Farris, 1969), and thus *is* related to evidential support. In the method, consistency indices are determined for each character on an initial cladogram, and are termed unit character consistencies (Farris, 1969). In the case of multiple equally parsimonious cladograms, they are average values. These consistencies are then used to weight the initial character matrix, and another cladistic analysis is performed on the weighted data set. New unit character consistencies are calculated for the resulting cladogram (or cladograms), the characters re-weighted, another cladistic analysis performed and the process continued until the cladograms on successive iterations are identical. The technique thus selects the cladogram (or cladograms) based on the set of most consistent characters. Such characters are "best" in that they provide the strongest evidence. The rationale for weighting at all is most succinctly stated by Farris (1983:11): "No one supposes . . . that characters in general all deserve the same weight—that they all yield equally strong evidence." The successive approximation approach, by performing the weighting a posteriori in context of all characters, avoids the subjectivity inherent in both a priori (e.g. Hecht and Edwards, 1977; Felsenstein, 1981; Neff, 1986[1]) and arbitrary a posteriori methods (e.g. Davies, 1981). The compatibility approach to phylogenetic inference could possibly be characterized as an extreme form of a priori character weighting, but is obviously inferior to this finely graded technique (Farris and Kluge, 1979), which is an extension of parsimony.

I have used this technique extensively in analysis of data sets assembled in the course of studies of vespid wasps. In the PHYSYS system (Mickevich and Farris, 1984) the method may be implemented via a single command loop as follows:

```
/DIAGNOSE,"cladograms","characters",,"weights1";
WEIGHT;
WAGNER.S;
DIAGNOSE,,,,"weights2";
WEQUAL,"weights1","weights2";
SWAP,"weights1","weights2";
REPEAT/.
```

[1] Actually, Neff's paper merely argues for careful homology decisions, which should be given.

Here, "cladograms" and "characters" refer to the initial cladograms and data. The DIAGNOSE command calculates the mean unit character consistencies and places them in the weight data set "weights1". The WEIGHT command weights the initial character matrix by simply multiplying the character scores by these weights, and the WAGNER.S performs another cladistic analysis with branch-swapping on the weighted data. The next DIAGNOSE command calculates a set of unit character consistencies for the new cladogram or cladograms and places them in a second weight data set ("weights2"). The WEQUAL command tests for equality of the two weight data sets. If they are identical the procedure terminates, since the cladograms are identical; if not, the SWAP command interchanges the second weight data set with the first and the REPEAT command loops through the command sequence again using the second set of weights. Successive approximations can also be done as a series of several steps in the PAUP program (Swofford, 1985), which does not have a similar recursive capability. Unit character consistencies for the cladograms from an initial analysis can be obtained with the CHGLIST command, mean values calculated manually, and these values input as character weights using the WEIGHTS command. These steps can then be repeated until inspection shows the cladograms on successive iterations to be identical.

Note that the weighting is a simple linear function. Farris (1969) originally used four other functions. Defining the proportion $p(i)$ for a binary character i as $((t-1)c(i))^{-1}$, where t is the number of nodes on the cladogram and $c(i)$ is the unit character consistency for character i, these weighting functions were "linear", $1 - p(i)$; "convex", $1 - (p(i))^k$, $k > 1$; "concave and bounded", $(1 - p(i))^k$, $k > 1$; and "concave and unbounded", $(p(i))^{-k} - 1$, $k \geq 1$. In simulations, his concave functions worked better than the others at finding a correct cladogram in the presence of unreliable characters. Most effective of all was the most drastic (unbounded) concave function. This is a much stronger weighting function that the one employed here. In my applications of successive approximations it has not always converged on a single cladogram, although in these cases it has always reduced the number of cladograms to be considered. Use of one of the stronger functions might be more effective, as might a weight proportional to the consistency index raised to a power greater than 1. Further investigation of this point is highly desirable.

Felsenstein (1981) suggested that, for large values of parameter k, weighting binary data with Farris' bounded concave function could yield a compatibility technique. The concave functions weight strongly the highly consistent characters, which in the extreme might approach the all-or-none weighting of compatibility methods. But this then suggests that in such cases successive approximations might itself converge on a cladogram which is not parsimonious for the unweighted data. This could possibly occur for some data sets with very low consistency indices for most characters, in which case the technique might have failed but a less parsimonious cladogram based on the "better" characters may as well be considered. Obviously this would be very different from the situation where the f and D statistics achieve their optima for non-parsimonious cladograms; the selection in this case would be choice among characters and so would still be related to evidential support. Thus, successive approximations may only be a partial solution to ambiguous data sets, but by restricting attention to those cladograms based on the most consistent characters it focuses on the best evidence. The necessity of reanalysis (or introduction of new data) remains in such cases, but at least the process of "checking, correcting and rechecking" (Hennig, 1966) might be usefully directed.

It should be noted that when applying this technique to multistate characters additive

Table 1
Stenogastrine data integer-coded.

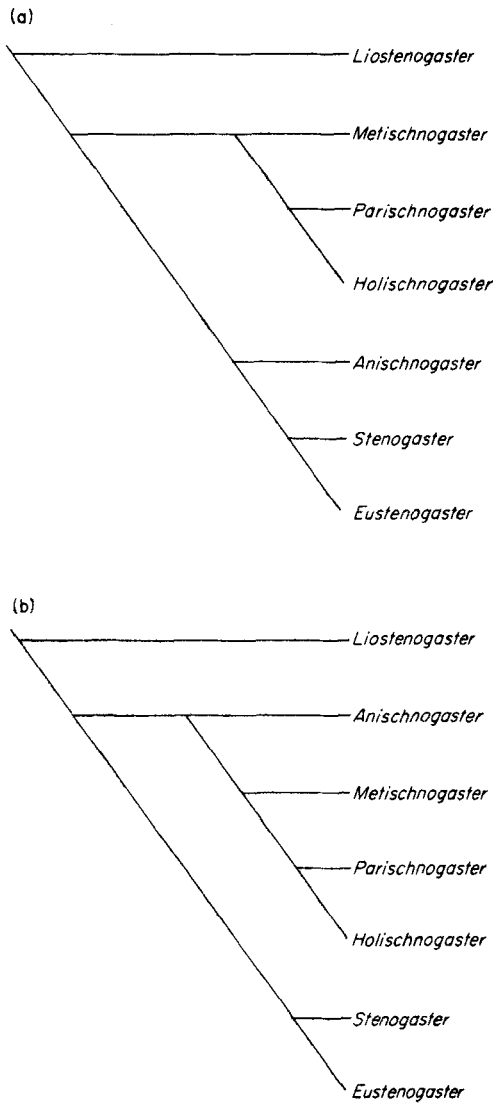|	
000000000	*Liostenogaster*
101011111	*Parischnogaster*
100011111	*Holischnogaster*
100001110	*Metischnogaster*
100100010	*Anischnogaster*
112200000	*Stenogaster*
111200000	*Eustenogaster*

(a)

(b)

Fig. 1. Alternative cladogram (a) and (b) for the data of Tables 1 and 3.

Table 2
Weights for characters in Table 1.

1	2	3	4	5	6	7	8	9
1.000	1.000	0.667	0.833	1.000	1.000	1.000	0.750	1.000

binary coding rather than integer coding should be used. This is because higher integers may result in greater weight being given to a particular character simply as a function of scoring, whereas additive binary coding avoids this by evaluating each state of each character individually. Farris (1969:382) recommends this procedure: "Use of additive binary coding and separate evaluation of steps is also one way to extend the successive weighting technique to multistate characters." A simple example illustrates this point. Table 1 is a data set drawn from a study of the genera of hover wasps (Carpenter, 1988). The characters are integer-coded. The branch-and-bound routine in PHYSYS (XWAGNER) produces two most parsimonious cladograms, shown here as Figs 1a and b. The weights for the characters of Table 1 derived from these cladograms are shown in Table 2; these are simply the mean consistency indices for the characters scaled from 0 to 1. Successive approximations character weighting selects the first cladogram (Fig. 1a).

Table 3
Stenogastrine data recoded in additive binary form.

> 00000000000 *Liostenogaster*
> 10100011111 *Parischnogaster*
> 10000011111 *Holischnogaster*
> 10000001110 *Metischnogaster*
> 10001000010 *Anischnogaster*
> 11111100000 *Stenogaster*
> 11101100000 *Eustenogaster*

Table 4
Weights for characters in Table 3.

1	2	3	4	5	6	7	8	9	10	11
1.000	1.000	0.500	1.000	0.750	1.000	1.000	1.000	1.000	0.750	1.000

When the data are recoded into additive binary form (Table 3) and a new set of weights calculated (Table 4), successive weighting results in both cladograms being retained. Because character 4 in Table 1 (5 and 6 in Table 2) has "2" as one of its states, it achieves greater weight than character 8 in Table 1 (10 in Table 2), which is binary. Conflict between these two characters is solely responsible for the ambiguity in the data sets, as the cladograms differ only in the placement of *Anischnogaster*, which shares apomorphies with two different clades. It has the state of "1" for both characters, and when character 4 is recoded in additive binary form, weighting does not converge to one of the cladograms.

Acknowledgments

REFERENCES

ADAMS, E. N. 1972. Consensus techniques and the comparison of taxonomic trees. Syst. Zool. 21: 390–397.

BROOKS, D. R., R. T. O'GRADY AND E. O. WILEY. 1986. A measure of the information content of phylogenetic trees, and its use as an optimality criterion. Syst. Zool. 35: 571–581.

CARPENTER, J. M. 1988. The phylogenetic system of the stenogastrinae (Hymenoptera: Vespidae). J. N.Y. Ent. Soc. 96: 140–175.

COLLESS, D. H. 1983. Wagner trees in theory and practice. In Felsenstein, J. (ed.), Numerical Taxonomy. NATO ASI Series G. Springer-Verlag, Berlin, pp. 259–278.

DAVIES, R. G. 1981. Information theory and character selection in the numerical taxonomy of some male Diaspididae (Hemiptera: Coccoidea). Syst. Ent. 6: 149–178.

FARRIS, J. S. 1969. A successive approximations approach to character weighting. Syst. Zool. 18: 374–385.

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

FARRIS, J. S. 1983. The logical basis of phylogenetic analysis. In Platnick, N. I. and V. A. Funk (eds), Advances in Cladistics 2. Proceedings of the second meeting of the Willi Hennig Society. Columbia University Press, New York, pp. 7–36.

FARRIS, J. S. AND A. G. KLUGE. 1979. A botanical clique. Syst. Zool. 28: 400–411.

FELSENSTEIN, J. 1981. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. Biol. J. Linn. Soc. 16: 183–196.

GATLIN, L. L. 1972. Information theory and the living system. Columbia University Press, New York.

HECHT, M. K. AND J. S. EDWARDS. 1977. The methodology of phylogenetic inference above the species level. In Hecht, M. K., P. C. Goody and B. M. Hecht (eds), Major patterns of vertebrate evolution. NATO ASI Series 14. Plenum Press, New York, pp. 3–51. New York, pp. 3–51.

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

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

MICKEVICH, M. F. 1978. Taxonomic congruence. Syst. Zool. 27: 143–158.

MICKEVICH, M. F. AND J. S. FARRIS. 1981. The implications of congruence in Menidia. Syst. Zool. 30: 351–370.

MICKEVICH, M. F. AND J. S. FARRIS. 1984. PHYSYS (phylogenetic analysis system). Documentation for VAX/VMS installation at Harvard University, Cambridge.

MIYAMOTO, M. M. 1985. Consensus cladograms and general classifications. Cladistics 1: 186–189.

NEFF, N. A. 1986. A rational basis for a priori character weighting. Syst. Zool. 35: 110–123.

NELSON, G. 1979. Cladistic analysis and synthesis: principles and definitions, with a historical note on Adanson's "Familles des plantes" (1763–1764). Syst. Zool. 28: 1–21.

PRAGER, E. M. AND A. C. WILSON. 1978. Construction of phylogenetic trees for proteins and nucleic acids: empirical evaluation of alternative matrix methods. J. Mol. Evol. 11: 129–142.

SWOFFORD, D. L. 1985. PAUP (phylogenetic analysis using parsimony). Documentation for version 2.4. Illinois Natural History Survey, Champaign.