

Majority Does Not Rule: The Trouble with Majority-Rule Consensus Trees

Michael J. Sharkey* and Jason W. Leatherst†

*Department of Entomology, University of Kentucky, S-227 Ag. Sci. N., Lexington, Kentucky 40546-0091; and

†Department of Entomology, Oregon State University, Corvallis, Oregon 97331-2907

Accepted May 20, 2001

The use of majority-rule consensus trees as a means of resolving ambiguity in phylogenetic analyses is investigated. It is shown to be an inappropriate method for this purpose. © 2001 The Willi Hennig Society

Accuracy has been the prime motivation for the development of character-weighting techniques (summarized in Kitching *et al.*, 1998). More recently it has become rather common practice to use majority-rule consensus (MRC) as a method of weighting clades to resolve ambiguous strict consensus trees (e.g., Swofford, 1991; Candall and Fitzpatrick, 1996; Colgan and Flannery, 1995; Lutzoni, 1997; Titus and Larson, 1996). The purpose of this paper is to illustrate what we perceive to be a critical weakness in these applications of majority-rule consensus.

MRC is a form of consensus that retains all clades that are found in more than 50% of the fundamental cladograms (Margush and McMorris, 1981). The 50% rule ensures that all included clades are compatible. A more relaxed modification of MRC includes clades that are consistent with the basic MRC topology. The clades that are repeated most often in the fundamental set of minimum-length trees are preferred. This preference is based on the implicit assumption that all fundamental

cladograms are independent and equally likely to be the correct tree.

An example will show the lack of justification for this assumption. The cladograms A through C (Fig. 1) are the three minimum-length cladograms generated from the data set of Table 1. This data set is modified from the empirical data set of *Hoplicnema* species (Pakaluk, 1987; Sharkey, 1989, 1994). The MRC and the strict consensus trees are illustrated in Fig. 2. Examination of the fundamental minimum-length cladograms (Fig. 1) shows that there are two basal topologies. The first is found in cladograms A and C, both of which place *darlingtoni* as the sister-group of *spiniventer* + *aquilonaria* + *cubensis* + *thomasi*. The second basal topology is found in cladogram B, which places *darlingtoni* in an unresolved trichotomy with *woldai* and a large clade composed of eight species.

Using the criterion of MRC the basal topology of cladograms A and C is preferred over the alternative basal topology represented by cladogram B. An examination of cladograms A and C reveals that the clade composed of *spiniventer* + *aquilonaria* + *cubensis* + *thomasi* is resolved in two different ways in the two cladograms, but otherwise, they are identical. Each of the two parsimonious reconstructions of this restricted clade supports the basal topology that they share. The alternative basal topology of cladogram B results in a

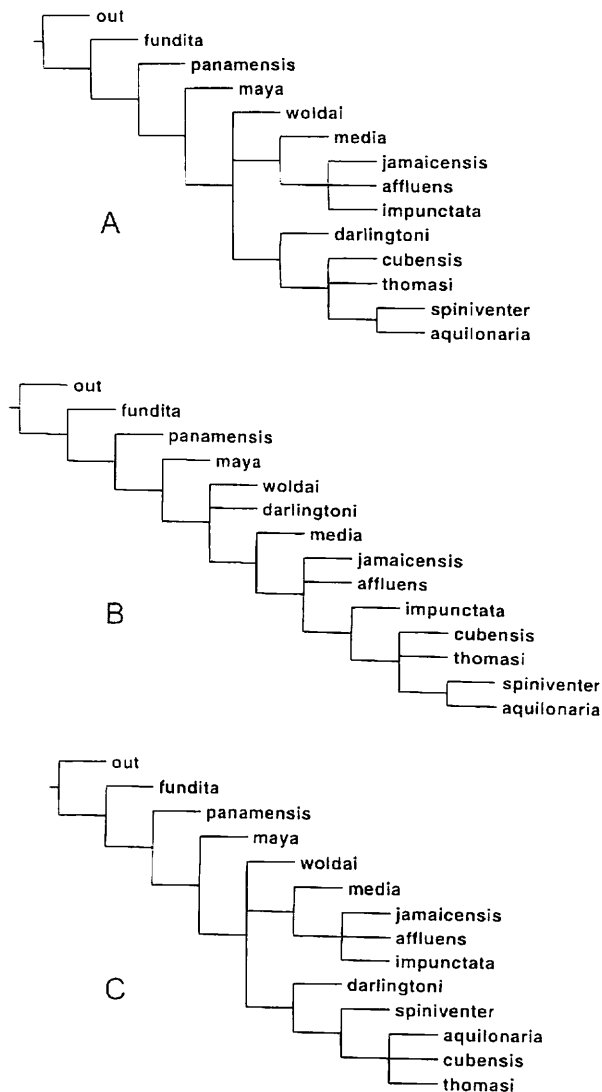


FIG. 1. (A–C) The three minimum-length trees generated (using NONA) from the data set of Table 1.

topology that is internally unambiguous, including the clade composed of *spiniventer* + *aquilonaria* + *cubensis* + *thomasi*. This definitive resolution results in only one tree supporting its basal topology. In cladograms A and C, it is ambiguity that causes the basal topology to be repeated and therefore preferred by MRC. The more ambiguity there is, the more multiple equally parsimonious reconstructions there are. Using MRC as a criterion for selecting trees equates reliability with ambiguity. The more ambiguous the clades nested in a basal topology are, the more that the topologies basal to the ambiguity are repeated in minimum-length trees.

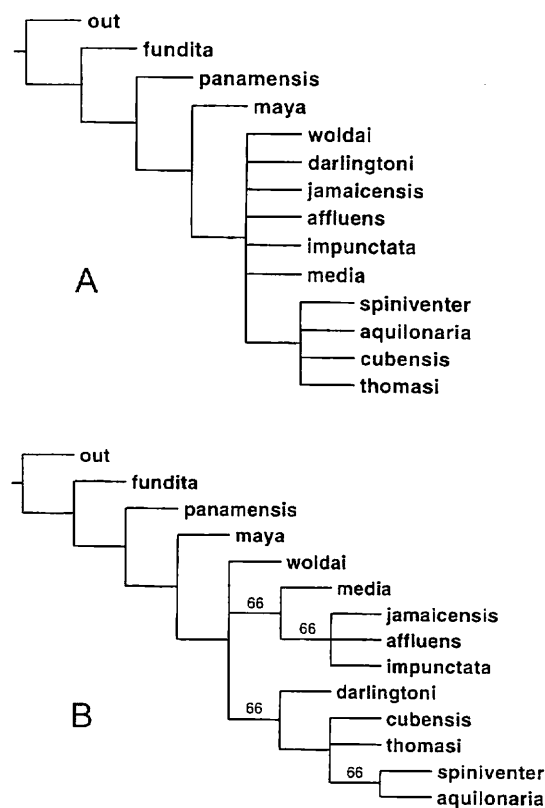


FIG. 2. (A) The strict consensus tree and (B) the majority-rule consensus tree generated (using NONA) from the data set of Table 1.

It is difficult to find a convincing rationale for using this as a measure of support and until one is offered the practice of using MRC to increase cladistic resolution

TABLE 1
Data Set of *Hoplicnema* Species (Modified from Pakaluk, 1987)

Taxa	Characters
Out	00000000
<i>fundita</i>	00010000
<i>panamensis</i>	01000000
<i>maya</i>	01000100
<i>woldai</i>	01000110
<i>darlingtoni</i>	01100110
<i>spiniventer</i>	01111110
<i>aquilonaria</i>	11111110
<i>cubensis</i>	11101110
<i>thomasi</i>	11101111
<i>jamaicensis</i>	00001100
<i>affluens</i>	00001110
<i>impunctata</i>	10001110
<i>media</i>	00000110

should be avoided. Although we have illustrated only a single example, almost any set of trees where the majority-rule consensus differs from the strict consensus will illustrate the problem.

Reweighting characters using goodness-of-fit statistics, such as the consistency, retention, and rescaled consistency indices, suffers from the same weakness as outlined for MRC when one averages the statistics over all preferred trees. PAUP offers several reweighting options including using maximum-, minimum-, and average-fit values. When the average-fit values are used, one is implicitly assuming that the trees are independent. It is probably not a coincidence that Hennig86 uses maximum-fit values for weighting using the successive approximations approach.

This paper was initiated as a reaction to four presentations that one of us (M.J.S.) attended at the 1999 meetings of the Society of Systematic Biologists in Madison, Wisconsin, all of which employed MRC as a method of obtaining increased cladistic resolution. Doubtlessly there were more. It is the objective of this paper to advertise at least one serious problem with the procedure.

ACKNOWLEDGMENTS

Thanks are extended to the editor and an anonymous reviewer. This is Paper 99-08-114 of the Kentucky Agricultural Experimental Station. This research was supported by NSF Grant DEB9972024.

REFERENCES

- Candall, K. A., and Fitzpatrick, J. F., Jr. (1996). Crayfish molecular systematics: Using a combination of procedures to estimate phylogeny. *Syst. Biol.* **45**, 1–26.
- Colgan, D. J., and Flannery, T. F. (1995). A phylogeny of Indo-west Pacific Megachiroptera based on ribosomal DNA. *Syst. Biol.* **44**, 209–220.
- Kitching, I. J., Forey, P. L., Humphries, C. J., and Williams, D. L. (1998). "Cladistics: The Theory and Practice of Parsimony Analysis." Oxford Univ. Press, Oxford.
- Lutzoni, F. M. (1997). Phylogeny of lichen- and non-lichen-forming omphalinoid mushrooms and the utility of testing for combinability among multiple data sets. *Syst. Biol.* **46**, 373–406.
- Margush, T., and McMorris, F. R. (1981). Consensus n-trees. *Bull. math. Biol.* **43**, 239–244.
- Pakaluk, J. (1987). Revision and phylogeny of the neotropical genus *Hoplicnema* Matthews (Coleoptera: Corylophidae). *Trans. Am. Entomol. Soc.* **113**, 73–116.
- Sharkey, M. J. (1989). A hypothesis-independent method of character weighting for cladistic analysis. *Cladistics* **5**, 63–86.
- Sharkey, M. J. (1994). Discriminate compatibility measures and the reduction routine. *Syst. Biol.* **43**, 526–542.
- Swofford, D. L. (1991). When are phylogenetic estimates from molecular and morphological incongruent? In "Phylogenetic analysis of DNA sequences" (M. M. Miyamoto and J. Cracraft, Eds.), pp. 295–333. Oxford Univ. Press, Oxford.
- Titus T. A., and Larson, A. (1996). Molecular phylogenetics of desmognathine salamanders (Caudata: Plethodontidae): A reevaluation of evolution in ecology, life history, and morphology. *Syst. Biol.* **45**, 451–472.