Syst. Zool., 35(3):422–426, 1986

Outgroup Analysis, Homoplasy, and Global Parsimony: A Response to Maddison, Donoghue, and Maddison

CURTIS CLARK[1] AND DANIEL J. CURRAN[2]

[1]*Biological Sciences Department, California State Polytechnic University, Pomona, California 91768; and*
[2]*Department of Biology, University of California, Los Angeles, California 90024*

The "outgroup method" for determination of character-state polarity, although explicitly set forth only five years ago (Watrous and Wheeler, 1981), is now widely employed. Farris (1982) pointed out that the outgroup method can be interpreted as a special case of parsimony analysis, and Maddison et al. (1984) showed that outgroup analysis as commonly practiced, often using but a single outgroup, may not always give the most-parsimonious solution.

Maddison et al. (1984) considered multiple types of analysis, ranging from those where the relationships among several outgroups are well known, through analyses using outgroups of uncertain relationship, to unconstrained analyses with few a priori assumptions about the data. They also dealt with two-step analyses as an alternative to simultaneous ones. We find no flaws in their presentation. However, our initial impression was that they favored constrained, two-step analyses and found unconstrained simultaneous analysis to be less useful; their treatment of the former was much more extensive than of the latter. While they pointed out to us that this was not their intent (Donoghue, pers. comm.; Maddison and Maddison, pers. comm.), our discussions with colleagues convince us that we are not the only ones to misconstrue their work. In this paper we will argue that unconstrained simultaneous analysis is superior to other techniques because it embodies fewer a priori assumptions.

TABLE 1. Sample data, with taxon J as the root, to demonstrate problems with a priori constraints.

	Character					
Taxon	1	2	3	4	5	6
A	a	a	a	a	a	a
B	a	a	a	a	a	a
C	a	a	a	a	a	a
D	a	a	a	b	b	b
E	a	a	a	b	b	b
F	b	a	a	b	b	b
G	b	b	a	b	b	b
H	b	b	b	a	a	a
I	b	b	b	a	a	a
J	b	b	b	a	a	a

CONCEPTUAL BACKGROUND

It is important to assess character polarity (distinguish apomorphic and plesiomorphic states) for two reasons. First, monophyletic groups can be diagnosed only by apomorphy. Second, the evolution of a given character can only be understood when one can infer the historical sequence of character states (the "transformation series"). For taxonomists who are not concerned with both of these, assessing polarity is superfluous.

Parsimony analysis is a method for dealing with homoplasy in phylogenetic data. We think that it is a powerful tool for phylogenetic reconstruction, but do not believe that it is necessarily the best tool in every case. Other arguments for monophyly, even unsubstantiated hunches, are useful investigatory tools, as long as: (1) the investigator uses them explicitly; and (2) a subsequent objective analysis (such as one based on parsimony) is used to confirm or at least explore the results. Possibly, the investigator may choose to abandon the most-parsimonious tree in favor of a less-parsimonious alternative; we accept this as well, provided that the reasons for doing so are spelled out.

Maddison et al. (1984) use the term "global" in the mathematic sense of referring to an entire "universe" under consideration; "local" then refers to a subset of that universe. In their sense, the universe over which global parsimony operates consists of the ingroup plus the outgroups and, thus, is circumscribed by the scope of the analysis.

We would like to use the same terms "global" and "local" in a different analogy. In computer programming, a global variable holds the same value in all parts of the program, while a local variable is assigned a value in a specific procedure or subroutine, but may have a different value in another one. Thus, there can be many levels of "local," but there is only one "global"—an absolute universe pertaining to the entire program.

We contend that phylogenetic analysis has such an absolute universe. In an early draft of this manuscript we referred to that universe as "all life," but we specifically mean all organisms grouped into species capable of cladogenesis. Phylogenetic analysis is concerned with reconstructing the sequence of lineage-splitting and, thus, has no direct explanatory value for other types of evolution (or non-evolution). Excluded from this universe, for example, are rocks, individuals in sexually-reproducing

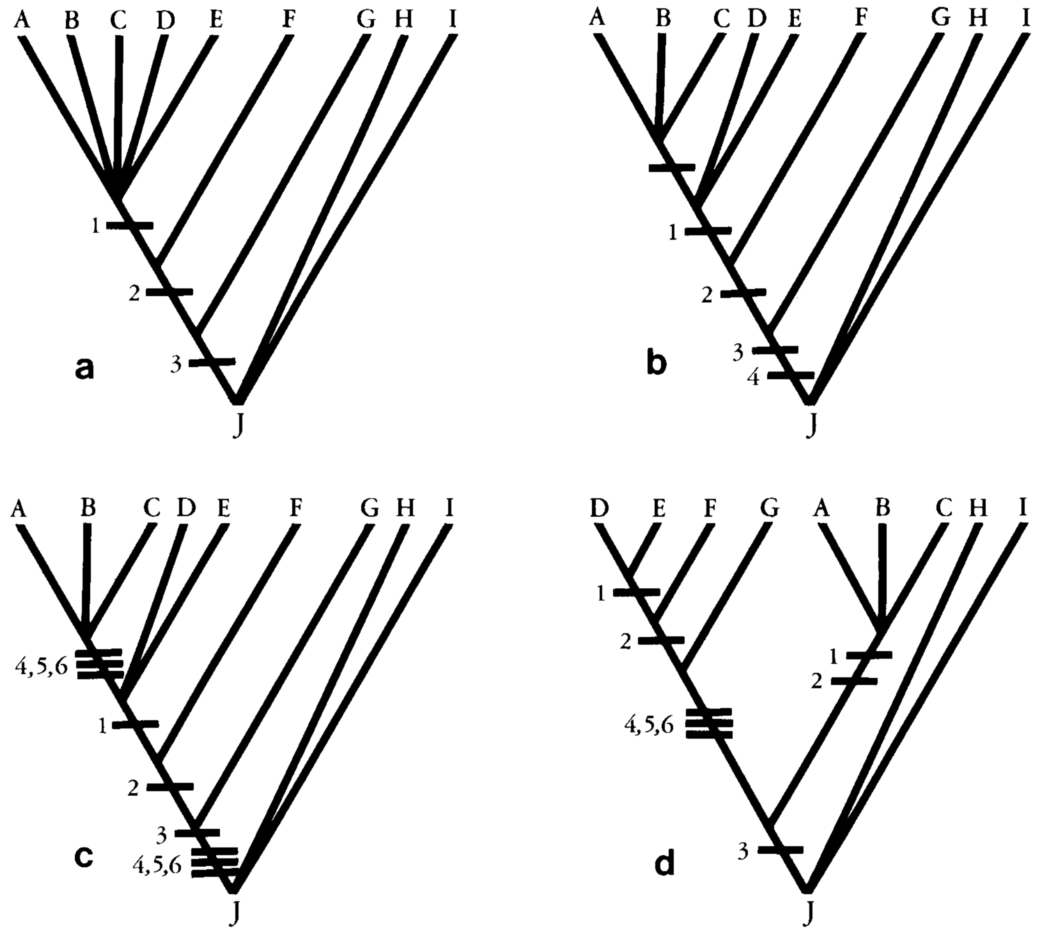

FIG. 1. Minimum-length cladograms based on data from Table 1. Horizontal bars represent character changes. (a) Characters 1, 2, and 3 used to constrain ingroup monophyly and outgroup structure. (b) Character 4 added to the constrained analysis. (c) Characters 5 and 6 added to the constrained analysis. (d) An unconstrained analysis of all characters.

biparental species, and genetic populations.

By the definition of Maddison et al. (1984), an analysis of all species of organisms would be only locally parsimonious, for want of an outgroup, and yet such an analysis would encompass the entire universe of phylogenetic analysis. We believe it is important to understand that every analysis of a portion of this universe is relatively local, no matter what its scope.

UNCONSTRAINED ANALYSIS

Constrained, two-step analyses, of the sort considered at length by Maddison et al. (1984), require several a priori assumptions. First, one must hypothesize that the ingroup is monophyletic. This implies that the root of the entire analysis is located basal to the ingroup (we mention this because the converse is not true; the root can be as easily located basal to a paraphyletic ingroup). Constraints of outgroup structure involve additional hypotheses of monophyly. All of these hypotheses are based on information from character distributions, but they are not tested by subsequent analysis with additional characters, since they are not allowed to change.

A simple data set (Table 1) will illustrate some of the problems with such an analysis. The first three characters taken alone establish the monophyly of the ingroup (ABCDE) and provide structure to the outgroups (Fig. 1a). A fully constrained, two-step analysis begins with these two assumptions met. Note that these first three characters are congruent.

Character 4 varies within the ingroup and, thus, could be expected to provide information about the ingroup structure. A simple outgroup analysis using only the sister group F would suggest that 4b is the plesiomorphic state, hence A, B, and C—sharing 4a—constitute a clade. The approach of Maddison et al. (1984:88) yields the same result; although H has the state 4b, G possesses 4a, and the "first doublet" rule tells us to infer that as the plesiomorphy. The resulting cladogram (Fig. 1b) has five steps.

Suppose two additional characters share the distribution of character 4. In both cases, the plesiomorphy can be inferred to be state b. The resulting cladogram (Fig. 1c) has nine steps, and the additional characters support the monophyly of ABC. But if we analyze all six characters simultaneously, without constraints, we get a very different cladogram (Fig. 1d), with eight steps. In this cladogram, the monophyly of the ingroup ABCDE is not evident; instead, DEFG and ABC are sister groups.

An unconstrained simultaneous analysis of the ingroup and outgroups requires only the single a priori assumption that the root is basal to the ingroup, an assumption that it shares with the constrained, two-step approach. If one were to perform both types of analyses on the same data, one would expect two possible outcomes. If the additional assumptions of monophyly required by the constrained analysis are supported by all the characters, the resulting cladograms should be identical. If the characters used in the second step of the constrained analysis do not support these assumptions, the unconstrained analysis will always give a more parsimonious solution. We concur with Farris (1982:331) that "direct use of parsimony [allows] analyses at different levels to be evaluated simultaneously, rather than requiring that they be performed sequentially." Given that unconstrained, simultaneous analysis will never provide a less-parsimonious solution to the same data than two-step analysis, and that the former procedure in fact provides a method to test the assumptions of monophyly required by the latter, unconstrained analysis is clearly the method of choice.

Although using an unrooted tree to establish character polarity and monophyly may seem unusual, it is implicit in the broadest analysis of all—that of all species. As we pointed out above, life lacks a sister group. Thus, an analysis of all species would be an unrooted tree. However, the position of the root can be inferred by methods other than parsimony analysis and, if the inference is correct, every convex group not containing the root is monophyletic. At the other extreme, an analysis that includes only the ingroup, of the sort advocated by Colless (1985) and Meacham (1984), insures that the ingroup contains the root (Maddison et al. [1984] amply discussed the dangers of such an analysis). In any case, these considerations apply to two-step analyses as well, since the assumption of ingroup monophyly makes it axiomatic that the ingroup does not include the root.

In summary, the parsimony method with the fewest a priori assumptions and the greatest likelihood of finding the globally most-parsimonious ingroup cladogram involves simultaneous unconstrained analysis of the ingroup and the outgroups to produce an unrooted tree. One might ask whether, by supporting this method, we are abandoning outgroup analysis. In the broad sense, the method is consistent with the "operational rule" of Watrous and Wheeler (1981:5): "For a given character with two or more states within a group, the state occurring in related groups is assumed to be the plesiomorphic state." Groups supported by parsimony analysis will adhere to this principle. Our disagreement is with applications of this rule that attempt to deal with homoplasy, yet require a priori formation of "ingroups" and "outgroups" based on only a portion of the available data.

PRACTICAL CONSIDERATIONS AND CONCLUSIONS

It is clear that simultaneous unconstrained analysis is not greatly limited in actual practice. It may be used in situations where outgroup relationships are basically unresolved, since it makes no assumptions about those relationships. Indeed, the method may help to resolve them. Of course, as the portion of interest expands to fill the total analysis, its reliability decreases, and additional outgroups should be included. Thus, the method is useful in many of the same situations as outgroup substitution (Donoghue and Cantino, 1984), although the latter may be more effective in dealing with cases where two different, distantly related outgroups have been advanced as possible sister taxa to the ingroup (G. Levin, pers. comm.). Even if some actual outgroups are unknown and not included, the method will provide the most-parsimonious resolution of those that remain; the situation is no different from that generally encountered in the analysis of groups with extinct members.

If the character states of potential outgroup taxa are not well known, unconstrained analysis can still be employed, provided the parsimony algorithm being used has the ability to handle missing states. The danger here is that the analysis of those characters in the ingroup that are unknown in the outgroups will not be parsimonious over all the included taxa, but this may not be immediately apparent if the analysis itself is being relied upon to form or substantiate those groups.

Gould (1986:66) correctly pointed out that cladistics is "a 'pure' method for defining historical order and rigidly excluding all other causes of similarity." One of the advantages of cladistics in this sense is its reliance on a minimum of a priori assumptions for its historical reconstruction. Hypotheses of monophyly should be the result of cladistic analysis, rather than part of the data. Although the outgroup method in all its incarnations provides a useful tool for determining character polarity in many situations, it relies on a priori assumptions of monophyly that can often be discarded with no loss of information.

ACKNOWLEDGMENTS

David Swofford discussed with us many of the concepts involved in unconstrained analysis. James S. Farris provided a thoughtful critique of an early draft of this paper. Michael Donoghue, as well as Wayne and David Maddison, patiently disagreed with us through subsequent drafts as well.

REFERENCES

COLLESS, D. H. 1985. On the status of outgroups in phylogenetics. Syst. Zool., 34:364–366.

DONOGHUE, M. J., AND P. D. CANTINO. 1984. The logic and limitations of the outgroup substitution approach to cladistic analysis. Syst. Bot., 9:192–202.

FARRIS, J. S. 1982. Outgroups and parsimony. Syst. Zool., 31:328–334.

GOULD, S. J. 1986. Evolution and the triumph of homology, or why history matters. Am. Sci., 74:60–69.

MADDISON, W. P., M. J. DONOGHUE, AND D. R. MADDISON. 1984. Outgroup analysis and parsimony. Syst. Zool., 33:83–103.

MEACHAM, C. A. 1984. The role of hypothesized direction of characters in the estimation of evolutionary history. Taxon, 33:26–38.

WATROUS, L. E., AND Q. D. WHEELER. 1981. The outgroup comparison method of character analysis. Syst. Zool., 30:1–11.

Received 18 February 1986; accepted 30 May 1986.