Characters, Computers, and Cladograms: A Review of the Berkeley Cladistics Workshop

Elizabeth A. K. Coombs; Michael J. Donoghue; Ronald J. McGinley

Systematic Botany, Vol. 6, No. 4 (Oct. - Dec., 1981), 359-372.

Stable URL:
http://links.jstor.org/sici?sici=0363-6445%28198110%2F12%296%3A4%3C359%3ACCACAR%3E2.0.CO%3B2-Z

Systematic Botany is currently published by American Society of Plant Taxonomists.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/aspt.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org/
Thu Jun 9 12:45:04 2005

Systematic Botany (1981) 6(4): pp. 359–372
© Copyright 1981 by the American Society of Plant Taxonomists

Characters, Computers, and Cladograms: A Review of the Berkeley Cladistics Workshop

ELIZABETH A. K. COOMBS, MICHAEL J. DONOGHUE, and
RONALD J. MCGINLEY
Department of Biology, Harvard University,
Cambridge, Massachusetts 02138

ABSTRACT. A workshop on cladistic methods was held in Berkeley, 23–28 Mar 1981. It was unusual in that it emphasized analysis of data sets submitted by the participants. This focused attention on the construction of data sets and the differences among computer algorithms rather than on issues of classification or historical biogeography. Like any approach to phylogeny reconstruction, cladistic methods depend heavily on choice of characters and character states, as well as determination of evolutionary polarity and transformation series, and implicit or explicit character weighting. If the group being analyzed is not monophyletic or if it includes hybrids the cladistic relationships of the taxa may be misinterpreted. Minor changes in the choice of characters or the choice of taxa often produced major differences in cladograms. Such problems may be magnified by computer analyses. Different algorithms produce different results because they are based on different assumptions about the likelihood of character state changes. Despite these problems, the great strength of cladistic methods is that they force careful consideration of the nature and distribution of taxonomic characters and in doing so provide us with our best estimates of phylogenetic history.

Since the time of Darwin, systematists have attempted to reconstruct phylogenies by looking at the distributions of character states in organisms. In general they have assumed that those organisms with more character states in common are more closely related and that overall similarity should reflect genealogy. It was not until 100 years after Darwin that Hennig (1950, 1966) emphasized that only synapomorphies (shared derived character states) provide information about phylogenetic relationships; at any one level, similarity based on symplesiomorphies (shared primitive character states) cannot be used to determine relationship. In our view, this is the core of what has come to be known as cladistics. For some workers, however, cladistics is merely the study of the number and sequence of character state changes in the evolution of a group, without regard for synapomorphy (for discussion of this approach see Eldredge 1979, pp. 169–170; Duncan et al. 1981).

Cladistic analysis involves at least two stages. First, taxa are scored for a set of characters; the character states are then polarized (ancestral vs. derived; Stevens 1980). Second, nested sets of synapomorphies are sought in the character data. The result is a cladogram with taxa as end points of branches on which character state changes can be shown (Eldredge and Cracraft 1980, pp. 19–21). Such a cladogram can then form the basis for discussion of the evolution of the study group and a classification may be derived to reflect the structure of the cladogram (Eldredge and Cracraft 1980, Chap. 5). In addition, by comparing cladograms, inferences can be made about historical biogeography (Platnick and Nelson 1978).

Cladogram construction may be done by hand or by computer. Although hand-derived cladograms are common in the zoological literature, use of the computer has become widespread. Indeed, computer-assisted cladistics has become so dominant in the botanical literature that an uninformed person could get the impression that there is no other way to do cladistic analysis. Certain problems with the methods are especially intractable when one is trying to analyze a large group of organisms and is using a computer. In pointing out these problems, though, we do not want to imply that cladistic analysis is useless but merely that computer-assisted cladistics has certain limitations at present.

A distinction must be made between cladistic logic and the application of that logic: there are problems with the latter that do not necessarily reflect on the former. For "clean" data (i.e., with few parallelisms or reversals in character states, or mistakes in polarity assessment or character state coding), all cladistic methods will give identical estimates of phylogeny. Too frequently, however, our data are not so tidy; what concerns us here is the application of cladistic methods to such "messy" data. In what follows it is important to bear in mind that many of the problems with cladistics apply to any attempt at evolutionary inference.

A workshop on the theory and application of cladistic methodology took place in Berkeley, California, 23–28 Mar 1981. It was planned by Thomas Duncan and Tod Stuessy and was sponsored by the National Science Foundation; Duncan, as the local host and coordinator, did a masterful job of organizing the proceedings. The 26 participants were mostly botanists, with a few zoologists, a paleontologist, and a biochemist.

The format of the meetings was unique. Formal presentations were delivered during the morning sessions (see table 1 for list of invited speakers), and the afternoons were primarily devoted to the analysis of the participants' data. About three weeks before the conference each participant submitted one or more data matrices. These had been entered into the computer and subjected to a variety of cladistic analyses and, for comparison, to an unweighted pair group phenetic analysis (UPGMA, cf. Sneath and Sokal 1973), although this was little discussed. Afternoon sessions consisted of interpreting the results of these analyses and performing further manipulations of the data. At the end of the week, participants briefly presented their findings and discussed their impressions of cladistic methods. As Duncan pointed out, more cladistic analyses of plant groups were carried out during the week than had been published in all the time previous to the conference. Therefore, the workshop could have a major influence on the way and the extent to which cladistics is taken up by systematic botanists.

In this paper, we review the workshop proceedings and attempt to

TABLE 1. Invited speakers and titles of talks; cladistics workshop, Univ. California, Berkeley, March 1981.

Peter Ashlock (University of Kansas): Monophyly and phylogeny.
Bernard R. Baum (Biosystematics Research Institute, Ottawa, Canada): Application of compatibility and parsimony methods at the intraspecific, specific, and generic levels in Poaceae.
Daniel Brooks (University of British Columbia): Quantitative parsimony.
George F. Estabrook (University of Michigan): The relationship of cladistic characters to estimates of evolutionary history.
Joseph Felsenstein (University of Washington): The statistical inference approach to inferring evolutionary trees and what it tells us about parsimony and compatibility.
Walter Fitch (University of Wisconsin): Matrix methods in classification.
Vicki Funk (New York Botanical Garden): Patterns of parallelism in *Montanoa.*
David Hull (University of Wisconsin, Milwaukee): Alternative goals for biological classification.
Donald Kaplan (University of California, Berkeley): The concept of homology and its application to problems of plant cladistics.
Arnold Kluge (University of Michigan): The relevance of parsimony to phylogenetic inference.
Ronald McGinley (Harvard University): Overview of Hennigian principles and methodology and their application to studies of the evolution of the Hymenoptera.
Christopher A. Meacham (University of Michigan): Evaluation of data sets using a probability model of character compatibility.
Gareth Nelson (American Museum of Natural History): Cladistics and biogeography.
Raymond B. Phillips (University of Oklahoma): Considerations in formalizing a classification.
Tod Stuessy (Ohio State University): Problems in the estimation of character state trends.
Warren H. Wagner Jr. (University of Michigan): Application of the concepts of groundplan-divergence.
David B. Wake (University of California, Berkeley): Allozymes and phylogeny.
Allan Wilson (University of California, Berkeley): Macromolecules and phylogeny.

convey our impression of ideas that emerged. Because of the workshop's focus on the practical aspects of cladistic analysis, particularly computer-assisted cladistics, most discussion during the week revolved around character data and methods of cladogram construction. Comparatively little was said about the derivation of a classification or about historical biogeography. Therefore, we will briefly review the few comments on these considerations and then focus on characters and cladogram construction.

CLASSIFICATION

The workshop was opened by Hull who described the three major approaches to biological classification—cladistics, evolutionary systematics, and phenetics. Cladistics is concerned with the sequence of branching events in the evolutionary history of groups of organisms; therefore, taxa are united on the basis of shared derived character states. Branching diagrams are constructed to reflect character state transformations in a group; these diagrams are the basis for classification and for discussion of evolutionary history. In evolutionary systematics, branching diagrams are also derived that reflect the evolution of a group, but it differs in that the classifications reflect degrees of divergence as well as branching events. In phenetics, the distinction between ancestral and derived character states is not considered but organisms are clustered according to their aggregate similarities and differences. Later in the week McGinley produced a cladogram of systematists, pointing out that even within the three schools there are sub-divisions.

Phillips attempted to evaluate the relative merits of classifications produced by the three schools. He posed the now standard question: do we always want monophyletic taxa in the sense of Hennig, i.e., what Ashlock (1971) called holophyletic groups, or are paraphyletic taxa sometimes acceptable? Phillips's answer to this was that any non-polyphyletic group is acceptable so long as the author of a classification explicitly presents the data, the cladogram derived from the data, and the methods used in proceeding from the cladogram to the classification. As he pointed out, if classifications include names for paraphyletic groups then a tree could not necessarily be reconstructed from a classification. Hence, names would serve only as references to the taxonomic literature where one could discover what a particular taxon represented. Some of the workshop participants strongly objected, arguing that a classification should reflect one aspect of evolution (branching) unambiguously, rather than representing nothing by trying to incorporate both branching and divergence. This touched off some discussion of the purpose, predictivity, and stability of classifications, but little was resolved.

Biogeography

Nelson was the only speaker to discuss in detail the relationship between cladistics and biogeography. He began by asking what one can and should do with cladograms once they are constructed. His answer was that they should be compared and combined to find common cladistic patterns that then can be related to earth history to yield explanations of biogeographic patterns. Nelson's opening and closing statements, that biologists have heretofore failed to perceive patterns in the distributions of organisms, seem to us to be unfounded. There was, however, general agreement that cladograms are basic to discussions of historical biogeography, so long as geographical considerations do not enter into the assessment of character polarities upon which the cladograms are based. Funk's studies of *Montanoa* (Asteraceae) provided a good example of the relationship between cladistic analysis and biogeographic inference.

Characters

The first step in any systematic analysis involves both the choice of a group and the choice of characters. There are several problems pertaining to characters that are of importance to cladistic analysis, although many of them are not unique to cladistics.

1. *Choice of characters.* Systematists must decide what constitutes a character in each group and whether a character might be better expressed as a function of several others (e.g., leaf shape vs. length and width). There was not a great deal of discussion of this point. Although most of the participants were working with classical morphological characters, Wake and Wilson discussed the use of allozyme and DNA sequence data. Wake stated that allozymes are characters like any others and may be illuminating in some situations, but Wilson took a much stronger position, saying that the true tree could be discerned using DNA sequences, and morphological trees could then be judged against the DNA tree. He did not address the problem of apparent nonsense codons and noisy DNA sequences.

In selecting characters, one must also address the problem of homology. Both Kaplan and Wake discussed this in detail, showing numerous examples of the importance of developmental studies in determining homology. Structures that look very similar in adult organisms may be quite different in development, and internal structure may be independent of external form. Kaplan pointed out that one can never distinguish between homoplasy and homology but can only look for similarities of organization. If organization is different, it indicates homoplasy; if the same, homoplasy is not falsified. We think, however, that establishment of similarity is only a first step; homology can only be determined by reference to a cladogram (Rieppel 1980). (Although taxonomists in the past have used tentative groupings to test hypotheses of homology, this has not been done as rigorously as is possible with a cladogram.) However, the importance of the initial establishment of similarity cannot be overemphasized. The cladogram used to test homology will only be as good as the characters used to construct it; expecting homologies and homoplasies to fall out of a cladistic analysis without careful studies of the kind suggested by Kaplan is naive.

2. *Determining polarity.* Stuessy described nine criteria for determining polarity and said that they all should be considered in character analysis; because all are unreliable, though, they should be used with caution (Crisci and Stuessy 1980). Subsequent speakers disagreed strongly with Stuessy's approach, the consensus apparently being that out-group analysis is the only justifiable way to proceed (see also Stevens 1980; Watrous and Wheeler 1981).

Having reached that conclusion, however, a new problem is raised, particularly if no out-group is known. Two methods were suggested for dealing with this problem. Funk resorted to the functional out-group approach suggested by Watrous and Wheeler (1981). Certain characters of the pale in *Montanoa* are so unusual among composites that she felt secure in calling them derived (out-group comparison). Then using these few characters she produced a cladogram showing the cladistic relationships of three sections of the genus. For each section she used the other two sections as the out-groups. This allowed her to polarize enough characters to produce a fully resolved tree. It is an interesting method, although it rests entirely on the strength of the primary characters.

Michael Donoghue (a participant), working at a different taxonomic level, attempted to discern whether the characters that *Viburnum, Sambucus,* and *Adoxa* have in common can be considered synapomorphies. He used all of the out-groups that had been suggested previously by angiosperm phylogenists to see if, in each case, the three genera shared derived character states. Regardless of the out-group, synapomorphies were found, supporting the hypothesis that the three taxa form a monophyletic group. Thus, although Funk and Donoghue both faced the problem of uncertain out-groups, they were able to proceed with their analyses.

3. *Transformation series.* Systematists tend to link states of multistate characters into logically pleasing transformation series, but it should be obvious that there is no a priori reason for doing this. Using out-group comparisons to determine polarity of a morphocline tells nothing of how the states are ordered beyond the primitive state. Complex transformation series create particular problems in character compatibility analysis. As the number of character states increases, the likelihood of a mistake in interpreting directionality increases; therefore, multistate characters frequently prove to be incompatible with other characters. Many of the workshop participants found this to be true with their data.

4. *Continuous characters.* Although several existing programs can deal with continuous characters, many require binary data. Continuous characters can be coded to give discrete units, but there is a tremendous loss of information, and there is no nonarbitrary way to do it. Even using histograms of data from the sample taxa and looking for breaks may produce artifacts.

Clearly, the way a character is divided may influence the results. A dramatic example of this was provided by Christopher Campbell (a participant), who recoded a single character in his data matrix for a group of *Andropogon* (Gramineae). He divided a continuous character first into two states and then into three and got substantially different cladograms. The coding change moved one taxon from one side of the tree to the other and rearranged several other sister group relationships.

5. *Polymorphic taxa.* Taxa are often polymorphic for a character. Arnold (1981) discussed in detail how this may be misleading in cladistic analysis. If the two states are not present in equal frequency, the minority state may be missed in sampling. Also, the presence of polymorphism in a group increases the chance of a "pseudoreversal", i.e., the loss of the derived state from a previously polymorphic taxon. This will increase the chances for cladistic misinterpretation.

In computer analyses the situation is even more difficult. If a taxon is polymorphic for a character, either the taxon or the character must be left out, or the polymorphic state must be included as an additional step in the transformation series of the character. Felsenstein has several programs (not available at the conference) that can deal with variation of a character within a taxon, but no others are currently available.

6. *Missing values.* Most currently available programs cannot accept characters with missing values—these must be left out of the analysis. Of course, missing values are always a problem, but in hand-done analyses, characters with missing values can still be used to help resolve clades supported by other characters.

7. *Character weighting.* There are a number of problems involved with either explicit or implicit weighting of characters; these are discussed in more detail in a later section.

TAXA

Clearly, there are problems in the selection and coding of characters; there are additional problems in the choice of taxa. Often we are missing a few taxa and/or have mistakenly included a taxon that does not belong in the group, so that the group used for the analysis is not truly monophyletic (sensu Hennig, i.e., any and all descendants of a single ancestor). So long as there is little homoplasy, extraneous taxa should appear as out-groups in the analysis, and missing taxa should not affect the gross structure of the cladogram (but see Arnold 1981). In large computer analyses of messy data sets, though, monophyly does matter. Removal of a taxon causes the computer to reconstruct the tree completely, often giving a very different result. Donoghue found in *Viburnum* that leaving out one taxon substantially changed the trees (fig. 1; see also below). Thus, although monophyly may not be a theoretical requisite for cladistic studies, it is a practical necessity for computer analyses unless the data are particularly clean.

The second problem, especially for botanists, is the recognition of hybridization. Computer cladistic methods result in cladograms that are strictly branching; taxa resulting from hybridization will not be identified as such. Wagner emphasized this but was perhaps too facile in stating that hybrids should simply be removed. He claimed that hybrids would always be apparent because they would be precisely intermediate between two other taxa, an observation that seems to us to be wrong. The problem of recognition of hybrids is clearly critical, yet there is no workable solution in sight.

One way to reduce the chance of including hybrids in the analysis was suggested by Baum in his work on *Avena* and the Triticeae (Gramineae). He restricted his analyses to plants with the same chromosome number and also, when possible, to ones with the same genome. This approach was feasible because he is working on groups that have received some of the most intensive genetic investigation in the plant kingdom. Although such a procedure may be ideal, it may not always be practical.

A final point that came out in the analysis of the participants' data was that the level of the group in the taxonomic hierarchy may determine a group's suitability for cladistic methods. One example of this was Baum's attempt to test character compatibility analysis by analyzing a group of *Avena* cultivars of known pedigree. The results showed no correspondence to these pedigrees. Looking next at species within *Avena,* results were better, in that he got similar trees from several algorithms, and plants with the same genome appeared as monophyletic groups. This suggests that different taxonomic levels may better meet the assumptions of various models because the characters show more hierarchical structure and less likelihood of parallelism or reversal. There are situations in which the only proper response is to admit that one has reached an impasse and that for the time being cladistic analysis intended to give an estimate of phylogenetic history is simply not possible.

Cladograms

Having arrived at an appropriate data set for a group of organisms, one can proceed to cladogram construction. At the conference this was done by computer. In what may be a hold-over from phenetics, there is a tendency to try to analyze data sets with large numbers of taxa and characters, a sort of more-is-better syndrome. Many of the problems we outline are exacerbated by this tendency, especially when the data are problematical. The effect of the following problems can be minimized by using only a few taxa or reducing large groups to a series of three-taxon statements.

A variety of computer algorithms is currently available for cladogram construction. The methods differ in what character state changes they allow and in what they attempt to minimize or maximize in constructing the rooted tree or unrooted network. Felsenstein presented a useful summary of the assumptions of cladistic algorithms. The Camin-Sokal parsimony method (Camin and Sokal 1965) allows characters to change from the ancestral to the derived state but does not allow reversals; the number of steps is minimized. The Dollo parsimony method (Farris 1977) allows each character to change once from the ancestral to the derived state, allows one to many reversals, and minimizes the number of reversals. The Wagner parsimony method (Farris 1970) allows both 0 to 1 and 1 to 0 state changes, and minimizes the total number of transitions. Brooks provided a clear explanation of how Farris's Wagner program proceeds in constructing a cladogram. Character compatibility methods (Meacham 1980), clearly explained by Estabrook and by Meacham, differ from the preceding approaches in that they begin with the tree supported by the maximum number of compatible characters, i.e., those that are completely consistent with each other, showing no apparent parallelisms or reversals. Unresolved portions of the tree are then resolved using secondary sets of characters.

Most of the participants found that different methods produced different cladograms. This is hardly surprising because the assumptions of the models differ. But it was somewhat more disheartening to learn that different cladograms may result from a single program using the same character data. The character compatibility program (CLINCH) often yields several equally large sets of compatible characters (cliques) from each of which a different primary tree is constructed. Furthermore, parsimony programs will not necessarily produce the most parsimonious cladogram. For data sets with more than six taxa, there are over 10,000 possible cladograms (Felsenstein 1978a); a computer cannot search and evaluate every possible branching diagram in a reasonable amount of time, and so current programs resort to approximation procedures comparing only selected diagrams. Also, there may be many equally parsimonious ways to arrange the character state changes on a given branching diagram. Finally, the order in which the taxa are entered into the computer may make a difference; shuffling the order of the taxa often produces different cladograms.

Another concern of the participants was that for many data sets even minor changes in the taxa or characters used, or in the ordering of the character states, can substantially alter the resultant cladograms. This is particularly problematical if the cladograms are then used to generate classifications. Donoghue's analysis of *Viburnum* provides an example of such instability, though these problems were by no means unique to his data (see also the discussion of Campbell's data above). The cladograms in figure 1 were obtained from four different runs of Farris's Wagner '78 program and illustrate the sort of instability seen in cladograms produced by all algorithms.

In the original run the data matrix consisted of 23 taxa (species complexes) scored for 34 characters. The results are shown in figure 1A. In the second run (fig. 1B), two characters were omitted from the data set. These characters concerned the margins and the secondary venation of the leaves. These were eliminated because they showed considerable polymorphism in some taxa and were suspected to be developmentally correlated. The tree in figure 1C was produced when four more characters were removed from the matrix (leaf retention, trichomes on the lower leaf surface, chromosome number, and petiole length). These were removed because of their variation within taxa, uncertainty in the initial polarity assessments, and/or they exhibited considerable homoplasy on the preceding cladograms. Although portions of these three cladograms are identical, notice that the position of the clade H, D, A shifts, and taxa Y and T appear linked to different groups. Inspection of the cladograms will reveal other differences in detail.

The cladogram in figure 1D was obtained when the taxon Y was removed from the data matrix used to construct figure 1C. This taxon was removed because its position on the cladograms varied drastically in all the programs. In fact, in some of the compatibility analyses it appeared

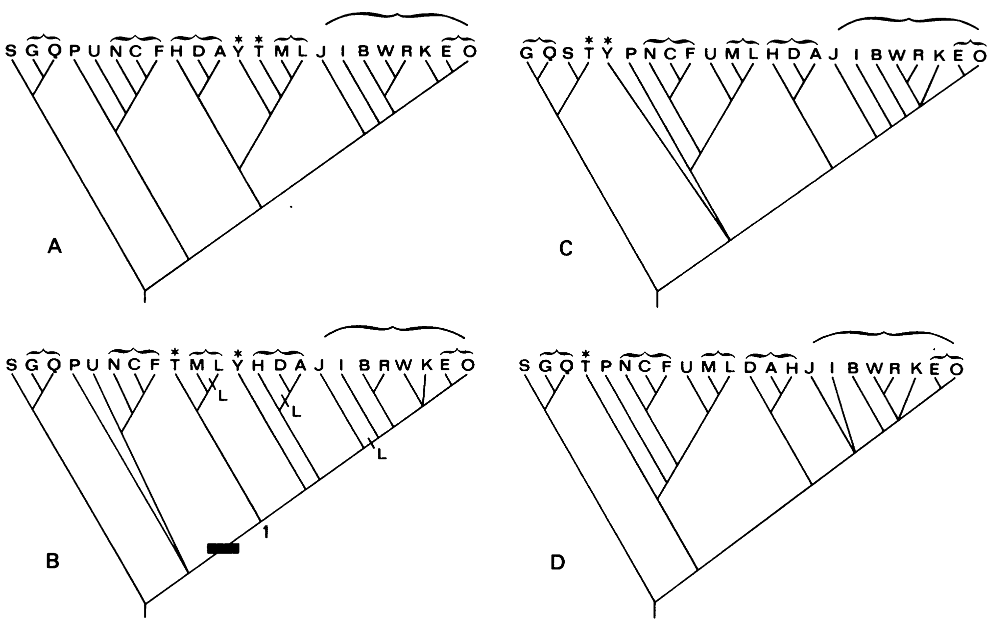

FIG. 1. Cladograms of *Viburnum*. Letters represent species complexes; brackets indicate "stable" clades; asterisks indicate "labile" taxa. See text for discussion. A. All taxa scored for 34 characters. B. Same as A but with two characters removed. Taxa above node 1 are united by the evergreen condition (character 8); this was lost three times. L = lost. C. Same as B but with four characters removed. D. Same as C but with taxon Y removed from the data matrix.

as the ancestral taxon. Also, there are reasons to believe that this complex is of hybrid origin. With its removal, the position of T is shifted, and the structure of the relationships within H, D, A is changed from all previous trees. These examples illustrate that relatively small changes in a data matrix sometimes alter cladograms in unexpected ways.

Another noteworthy observation is that the computer programs may produce clades united by a derived character state lacking in most members of the group. In figure 1B, the taxa above the node labeled 1 are united by the single derived character state, evergreen. However, only five of these 15 taxa are actually evergreen (T, M, Y, H, and J), the evergreen condition having been lost three times. Such groupings would probably not be produced as often if cladograms were constructed by hand, but if they were, we would be painfully aware of them. Computer output can look deceptively tidy; it is absolutely essential to trace character state changes on cladograms to assess the cause(s) of the groupings.

Again, these examples are not so much a reflection on methods of cladogram construction as on the nature of the character data. When there is homoplasy and/or little character support for some nodes in a cladogram then it is apt to change with even minor changes in the data matrix. Inasmuch as real data are often "messy" or "insufficient," it is important to bear in mind the potential instability of cladograms, especially if classifications are to be derived from them.

Clearly, it is possible to obtain many trees by running the same data through different programs or even through the same program. That slight alterations of the data may result in rather different cladograms raises the question of what to do with all of the cladograms that can be obtained.

There are three choices here. (1) Choose from among all cladograms produced by all methods the one matching your prior intuitions of the relationships in the group. If none of the original cladograms are acceptable, then shuffle the order in which taxa are entered or make a few carefully chosen changes in the data matrix; the desired cladogram will probably appear. Unfortunately, these moves defeat the purpose of these analyses and render them as subjective as any intuitive taxonomy. (2) Choose the preferred method and then take whatever cladogram(s) it produces. We will return below to a consideration of the criteria by which methods might be evaluated. (3) Make no choice but use the cladograms as heuristic devices, providing insights into the distribution of character states. Estabrook said that his character compatibility methods were most useful for investigating the structure of the data, rather than for constructing cladograms. In his presentation he retreated from those aspects of cladism relevant to the reconstruction of phylogenies and to classification. Meacham, in an effort to measure the value of characters for phylogenetic inference, has developed a computer program that calculates the probability that two characters will be compatible by chance alone and compares this value to their actual correlation in the data. A high correlation may only indicate that two characters are functionally or developmentally related, but this is also interesting information.

Our opinion is that aspects of (2) and (3) above should be combined; examination and comparison of the various branching diagrams yield information about the characters but also provide some information about relationships of the study organisms. In particular, different cladograms can be examined for areas of correspondence. If certain taxa are always grouped together in spite of changes in the rest of the cladogram, then our confidence that those groups represent actual cladistic relationships is increased. The groups are so well supported by the character data that they are cohesive even when changes are made in algorithms or in the data matrix. For example, in figure 1 several groups of taxa come out together on every cladogram. These robust groups (G, Q; N, C, F; H, D, A; M, L; E, O; J–O) are bracketed on the cladograms.

Now let us return to the question of evaluating the methods. Kluge presented a well-reasoned defense of parsimony as a criterion to choose among competing hypotheses that are equally well supported by the available data. It was pointed out in discussion, and subsequently by Felsenstein and by Fitch in their presentations, that likelihood considerations can bear on which hypothesis we consider to be the most parsimonious. Information about the ease with which one character state can change to another state will affect our parsimony decision. If we know that changes in character *A* are much more likely than changes in character *B*, then we might accept a cladogram with two changes in character *A* over one with one change in character *B*. In other words, we might prefer a cladogram with a few more steps if we have additional information about the likelihood of state changes in the characters.

This critical point leads to the question of the relationships among cladistic methods and the weighting of characters. Felsenstein discussed his maximum likelihood approach, which provides a unifying framework in which to examine all cladistic methods (see Felsenstein 1979). Maximum likelihood techniques are known to possess the statistical property of consistency (Felsenstein 1978b); an estimated value (or cladogram) will converge on the true value (cladogram) as more and more data are accumulated. Consistency seems a desirable property and hence Felsenstein has endeavored to determine when parsimony and compatibility methods will be consistent and when they will not. He found that when reversals are rare and when rates of evolution in different portions of a lineage are not too dissimilar, both parsimony and compatibility algorithms are statistically consistent. As the likelihood of character state changes increases and rates of evolution become more dissimilar in different parts of the tree, then neither parsimony nor compatibility methods have the property of consistency. As a general conclusion, Felsenstein asked that systematists evaluate methods in terms of their statistical properties.

Conclusions

The workshop participants concluded that only the most carefully prepared data sets were useful for cladistic analysis. It was also apparent, however, that even the best data exhibited some homoplasy and could support conflicting cladograms. In such cases some criterion must be employed to choose between the competing hypotheses. We think that parsimony is an absolutely essential methodological criterion for such purposes in systematics and, in fact, in any science. Parsimony considerations underlie both out-group analysis (the method for determining polarity of character states) and also all methods of cladogram construction. The most parsimonious hypothesis accounts for all of the available data in the simplest possible way, that is, it minimizes the number of ad hoc assumptions. The more data we have at hand, the more satisfactorily our most parsimonious hypothesis will account for the phenomenon under study. In the study of phylogeny, the more information we have about the characters of organisms, the more accurately our most parsimonious hypothesis will reflect their genealogical relationships.

One of the principal variables that emerges in the comparison of cladistic methods is the likelihood of character state changes. All characters are not equal with regard to the ease with which character state changes can occur, and hence they are not equally informative about phylogeny. Our estimates of phylogeny would better reflect actual cladistic relationships if we could find ways to assess and quantify such differences and thus weight characters for cladistic analysis (Hecht and Edwards 1976; Arnold 1981). A rigorous solution to the problem of character weighting must lie somewhere in the intersection of genetics, developmental biology, and cladistic analysis.

In the absence of this information, all characters must initially be weighted equally; the cladogram with the fewest character state changes is then the most parsimonious. Such a cladogram, although the best available estimate of phylogeny, will be affected to an unknown extent by the problems described in this paper. As we have seen, unless the data are very good, parts of the cladogram may be unstable. It will undoubtedly contain some interesting information about the characters and the taxa, but as a statement about phylogeny, the cladogram may not reflect reality.

Clearly, if we are ever going to understand evolution and the phylogenetic relationships among organisms, we must go back to careful analysis of characters. Discussion based on real data always has to return from the elegance of theory to the inconsistencies of the data matrix. There is no question that cladistic methods are compelling; the question is simply how to use them responsibly. The cleanest results come with simple three-taxon statements, the structure of which can be determined by a few carefully studied characters. No computers are required for this, only clear thinking. One may then ask whether there is any use for computer-assisted cladistics. Our answer is a qualified yes, in that the computer algorithms (which we believe will continue to be improved) can point to more or less strongly supported regions of an otherwise large and amorphous group and give a framework in which to place small analyses. In this sense then the computer is a tool, and like any other tool it can be used well or badly, its utility depending on the degree to which the data fit the assumptions on which the machine is operating. If we get unsatisfying results from a computer analysis, the problem is probably not with cladistic theory per se; it may be with the algorithm on which the machine is operating, but most likely the problem lies in the data themselves. For satisfactory cladistic analysis we need to return to the organisms.

Acknowledgments. We thank S. and W. Fink and P. F. Stevens for their help throughout the preparation of the manuscript and A. J. Gilmartin for her thoughtful review.

Present address of MJD is Department of Botany, San Diego State University, San Diego, California 92182.

LITERATURE CITED

ARNOLD, E. N. 1981. Estimating phylogenies at low taxonomic levels. Z. Zool. Syst. Evol.-Forsch. 19:1–35.

ASHLOCK, P. D. 1971. Monophyly and associated terms. Syst. Zool. 21:430–438.

CAMIN, J. H. and R. R. SOKAL. 1965. A method for deducing branching sequences in phylogeny. Evolution 19:311–326.

CRISCI, J. V. and T. F. STUESSY. 1980. Determining primitive character states for phylogenetic reconstruction. Syst. Bot. 5:112–125.

DUNCAN, T., R. B. PHILLIPS, and W. H. WAGNER JR. 1981. A comparison of branching diagrams derived by various phenetic and cladistic methods. Syst. Bot. 5:264–293.

ELDREDGE, N. 1979. Cladism and common sense. Pp. 165–198 in *Phylogenetic analysis and paleontology,* ed. J. Cracraft and N. Eldredge. New York: Columbia Univ. Press.

——— and J. CRACRAFT. 1980. *Phylogenetic patterns and the evolutionary process.* New York: Columbia Univ. Press.

FARRIS, J. S. 1970. Methods for computing Wagner trees. Syst. Zool. 19:83–92.

———. 1977. Phylogenetic analysis under Dollo's Law. Syst. Zool. 26:77–88.

FELSENSTEIN, J. 1978a. The number of evolutionary trees. Syst. Zool. 27:27–33.

———. 1978b. Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. 27:401–410.

———. 1979. Alternative methods of phylogenetic inference and their interrelationship. Syst. Zool. 28:49–62.

HECHT, M. K. and J. L. EDWARDS. 1976. The determination of parallel or monophyletic relationships: the Proteid Salamanders—a test case. Amer. Naturalist 110:653–677.

HENNIG, W. 1950. *Grundzüge einer Theorie der Phylogenetischen Systematik.* Berlin: Deutscher Zentralverlag.

———. 1966. *Phylogenetic systematics.* Urbana: Univ. Illinois Press.

MEACHAM, C. A. 1980. Phylogeny of the Berberidaceae with an evaluation of classifications. Syst. Bot. 5:149–172.

PLATNICK, N. J. and G. NELSON. 1978. A method of analysis for historical biogeography. Syst. Zool. 27:1–16.

RIEPPEL, O. 1980. Homology, a deductive concept? Z. Zool. Syst. Evol.-Forsch. 18:315–319.

SNEATH, P. H. and R. R. SOKAL. 1973. *Numerical taxonomy.* San Francisco: W. H. Freeman and Co.

STEVENS, P. F. 1980. Evolutionary polarity of character states. Annual Rev. Ecol. Syst. 11:333–358.

WATROUS, L. and Q. WHEELER. 1981. The out-group comparison method of character analysis. Syst. Zool. 30:1–11.