Points of View

Syst. Biol. 47(3): 495–496, 1998

Archiving Molecular Phylogenetic Alignments as NEXUS Files

BERNARD L. COHEN,[1] JONATHAN A. SHEPS,[1] AND MARK WILKINSON[2]

[1]*University of Glasgow, Institute of Biomedical and Life Sciences, Division of Molecular Genetics, Pontecorvo Building, 56 Dumbarton Road, Glasgow G11 6NU, Scotland. E-mail: b.l.cohen@ bio.gla.ac.uk*
[2]*University of Bristol, School of Biological Sciences, Bristol BS8 1UG, and Department of Zoology, Natural History Museum, London SW7 5BD, England.*

Many researchers working on the phylogenetic analysis of nucleotide or amino acid sequences find it necessary, at some time, to build upon prior work. When this involves importing an alignment (and masks) that have been described and sometimes illustrated in a published paper but for which the sequences have not been deposited in a public database, there may be no alternative but to painstakingly retype the data, with all the possibilities that this offers for error. More commonly, the unaligned sequences may be available from a sequence database, but the awkward task of manual alignment reconstruction remains. Most conveniently, alignments may be available as computer files (1) on disk, by application to the corresponding author, (2) from an ftp or http site maintained by the corresponding author or his institution, or (3) from an alignment database. However, when received by these routes, an alignment (and masks) may well be in a file format that requires tedious, error-prone word-processing adjustment to make it readable by a sequence alignment editor or phylogenetic analysis program. Moreover, of the three approaches, the first two are potentially evanescent and the third suffers from underuse and lack of file-format standardization (Stoesser, pers. comm.). Masks and weighting information, when presented PHYLIP-style (Felsenstein, 1993) as a string of digits, offer a particular difficulty; few alignment editors or phylogenetic analysis programs that we have encountered can import them successfully. Given these difficulties, we propose that it should become standard practice in molecular phylogenetics for alignments, masks, and supplementary information to be made available in NEXUS file format by use of the EMBL alignment database.

NEXUS file format is an ASCII text-file format that was "designed to make sharing of data between programs as easy and flexible as possible" (Maddison and Maddison, 1992: p.145; Maddison et al., 1997), Importantly, NEXUS file format provides unlimited flexibility, by the use of character partition and assumption blocks, for the identification of special categories of data such as masks, weights, secondary structure features, etc., and comments can be freely added. NEXUS files are written automatically by some alignment editors, including GDE (Smith et al., 1994), MacClade (Maddison and Maddison, 1992) and SeqApp (Gilbert, 1993), and most such programs also provide file-format interchange facilities. Alternatively, NEXUS files may readily be constructed by hand from published descriptions (Swofford, 1993; Maddison et al., 1997). Many phylogeneticists already use or have access to programs that read and write NEXUS files, and many more are likely to do so after publication of the new multicapable PAUP* (Swofford, pers. comm.).

The existence of a sequence alignment database separate from the GenBank (Benson et al., 1998), EMBL (Stoesser et al., 1998), or other databases for individual sequences is little known, but would certainly be advantageous if it were widely used and if its use were stan-

dardized. The possibility of such standardization has been discussed by the EMBL database staff, but these discussions and input from external users and experts in the field suggest a wide spectrum of opinions concerning possible alignment formats, and some feel that depositing alignments may not be worth the effort since sequences get updated frequently and alignments get out of date (Stoesser, pers. comm.). Nevertheless, in our view, the phylogenetic community should itself make an approach to standardization by adopting the NEXUS format. Because alignment and masking are still somewhat subjective procedures and because existing routes for the dissemination of aligned sequences are both unstable and tend to restrict access to all but the most determined, the effects of alternative alignments and masks are rarely explored by anyone other than the original author(s). It would be healthier science (though perhaps less comfortable) if alignments were more readily available, as they would be if routinely archived as we suggest.

The EMBL alignment database may be accessed as follows:

— EBI FTP server: by anonymous FTP from FTP.EBI.AC.UK in directory /pub/databases/embl/align
— EBI file server: by sending an e-mail message to: netserv@ ebi.ac.uk that includes the line HELP ALIGN or GET ALIGN: DSXXXXX.DAT where DSXXXXX is an alignment accession number
— EBI WWW server: URL ftp://ftp.ebi.ac.uk/pub/databases/embl/align/

ACKNOWLEDGMENTS

We are grateful to Gunther Stoesser, EMBL Nucleotide Sequence Database, for discussion and for ascertaining that our proposal is acceptable to the database staff. A. B. Gawthrop made a valued background contribution. Our work is supported by the Taxonomy Initiative of the UK Natural Environment Research Council (GST02/832).

REFERENCES

BENSON, D. A., M. BOGUSKI, D. L. LIPMAN, J. OSTELL, AND B. F. F. OUELLETTE. 1998. GenBank. Nucleic Acids Res. 26:1–7.

FELSENSTEIN, J. 1993. PHYLIP: Phylogeny inference package. Distributed by the author, Department of Genetics, Univ. Washington, Seattle.

GILBERT, D. 1993. SeqApp. Available by FTP from Molecular Biology Software Archive, Univ. Indiana, Bloomington.

MADDISON, D. R., D. L. SWOFFORD, AND W. P. MADDISON. 1997. NEXUS: An extensible file format for systematic information. Syst. Biol. 46:590–621.

MADDISON, W. P., AND D. R. MADDISON. 1992. MacClade. Sinauer, Sunderland, Massachusetts.

SMITH, S. W., R. OVERBEEK, C. R. WOESE, W. GILBERT, AND P. M. GILLEVET. 1994. The Genetic Data Environment, an expandable GUI for multiple sequence analysis. Cabios 10:671–675.

STOESSER, G., M. A. MOSELEU, J. SLEEP, M. MCGOWRAN, M. GARCIA-PASTOR, AND P. STERK. 1998. The EMBL nucleotide sequence database. Nucleic Acids Res. 26:9–15.

SWOFFORD, D. L. 1993. PAUP: Phylogenetic analysis using parsimony, version 3.1. Smithsonian Institution, Washington, D.C.

Received 29 October 1997; accepted 18 March 1998
Associate Editor: D. Cannatella

Syst. Biol. 47(3): 497–508, 1998

Properties of Matrix Representation with Parsimony Analyses

OLAF R. P. BININDA-EMONDS[1] AND HAROLD N. BRYANT[2,3]

[1]*Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom; E-mail: olaf.bininda@ zoology.oxford.ac.uk*
[2]*Provincial Museum of Alberta, Edmonton, Alberta T5N 0M6, Canada;*

Baum (1992) and Ragan (1992b) independently devised a method that uses additive binary coding and parsimony to combine trees derived from different data sets, a procedure Ragan termed matrix representation with parsimony analysis (MRP). Because the method utilizes the topology of source trees rather than the original data, (1) trees derived from different data types (e.g., molecular sequences, morphological characters, pairwise distances) and analyzed by different clustering techniques (e.g., maximum parsimony, maximum likelihood, neighbor joining) can be combined, and (2) the source patterns are evaluated on a more-or-less equal basis, so that the phylogenetic signal from data matrices with a smaller number of characters is not swamped by those with a larger number (see Miyamoto, 1985; Hillis, 1987). The method is also unusual in that (1) trees with different terminal taxa can be combined, a feature that among consensus methods characterizes only the supertree method (Gordon, 1986; Steel, 1992) and the modified semistrict algorithm of Lanyon (1993), and (2) it is less sensitive to conflict among source trees than are most conventional consensus techniques so that resolution is not necessarily lost as increasing numbers of conflicting trees are analyzed (see also Purvis, 1995b).

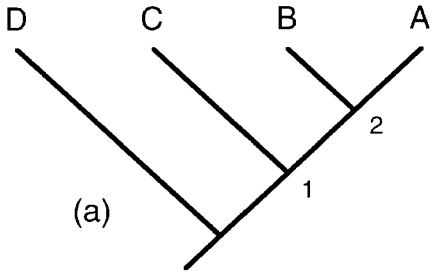
Although the appropriateness of MRP to phylogenetic inference has been discussed (Baum and Ragan, 1993; Rodrigo, 1993, 1996; Bruneau et al., 1995) and modifications to the method have been proposed (e.g., Purvis, 1995b; Ronquist, 1996), its properties, mechanics, and biases have not been considered in sufficient depth. We discuss some of the properties of MRP, show how MRP differs from standard consensus techniques, and explore some modifications to the method. Although other clustering methods, such as compatibility (Ragan, 1992a; Purvis, 1995b; Rodrigo, 1996), have also been suggested as methods for generating composite trees by using matrix representation, we will not discuss them.

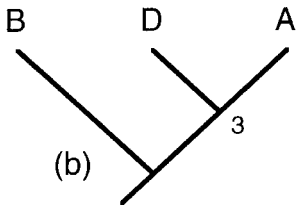
THE BASIC PROCEDURE AND SUGGESTED MODIFICATIONS

MRP uses additive binary coding (Farris et al., 1970) to represent the hierarchical structure of trees as a series of "matrix elements" (Baum and Ragan, 1993:637). Each node (i.e., component; sensu Wilkinson, 1994) on each source tree is represented by a binary matrix element, with terminal taxa in the clade delimited by that node scored as 1 and all other taxa scored as 0. Taxa that are missing from an individual source tree are coded as missing for elements that represent that tree. Trees are rooted either by an all-zero outgroup (Ragan, 1992b; Purvis, 1995b) or by using a taxon common to all source trees (Baum, 1992). Parsimony analysis of the element matrix produces a tree or trees (hereinafter, the composite tree[s]; Purvis, 1995a) that most parsimoniously synthesize(s) the hierarchical information in the source trees (for details, see Baum, 1992; Ragan, 1992b). Analyses that generate multiple most-parsimonious composite trees (MPCTs) are summarized by using strict consensus to generate a consensus composite tree (CCT).

Purvis (1995b) argued that the topology of particular source trees can unduly influence that of the composite tree (Fig. 1). He attributed this to the lack of independence among elements derived from a source tree, which adds redundant information to the matrix. He removed this apparent redundancy by coding taxa that are in neither the clade delimited by the node nor its sister taxon as ? rather than 0. As with

[3] Present address and address for reprint requests: Royal Saskatchewan Museum, Regina, Saskatchewan S4P 3V7, Canada, E-mail: hbryant@ gov.sk.ca





(c)

	Components		
Taxon	1	2	3
A	1	1	1
B	1	1	0
C	1	0	?
D	0	0	1

FIGURE 1. Purvis's (1995b) conflicting source trees (a, b) (with components numbered), and the combined element matrix (c). MRP analysis of the matrix results in tree a. Purvis's method results in an unresolved CCT.

unmodified MRP (Ragan, 1992b), parsimony analysis of the elements derived from one source tree recovers the correct topology (Purvis, 1995b).

Ronquist (1996) demonstrated that this modification to the coding procedure is flawed. He showed that the information content of matrices generated with Purvis's (1995b) method is less than that generated by standard additive binary coding and demonstrated that Purvis's method does not always achieve its goals. Also, because of the specific manner by which Purvis's coding adds missing data to the matrix, the relative positional stability of taxa is altered (Ronquist, 1996; his Fig. 3) so that the position of a taxon on the composite tree is influenced more by source trees on which it is further from the base. We would add that although the zeros replaced under Purvis's method are not strictly informative because they denote the lack of membership of taxa in components, they provide essential, restrictive information regarding the position of a taxon on its source tree that might become important when its elements are combined with those from other source trees.

Ronquist (1996) concluded that the bias described by Purvis (1995b) was associated not with redundant information but with the relative sizes of the source trees. Purvis's method proportionately reduces the influence of larger trees because they contribute a proportionately larger number of missing data points to the element matrix. Ronquist argued that the difference in the amount of information contributed by each source tree could be removed by inversely weighting each tree according to its number of internal branches (i.e., nodes). However, Ronquist favored weighting based on the support for nodes as measured by Bremer's decay index (Bremer, 1988) or the bootstrap (Felsenstein, 1985), both of which he implied would also compensate for any size bias.

Ronquist's (1996) analyses focused largely on the ability of various coding and weighting options to represent the information in a single source tree (and the original data set), rather than the ability of MRP to appropriately combine the information provided by multiple source trees in a single topology. Our discussion focuses more on the latter.

PROPERTIES OF MRP

Matrix Elements versus Characters

We have referred to the coded components of source trees as "matrix elements" rather than as "characters" because the two are not equivalent (Baum and Ragan, 1993). Characters are attributes of organisms. In contrast, a matrix element refers to a component of a tree and is a membership criterion. Matrix elements also differ from characters in that groups of elements representing a single source tree are necessarily congruent, forming a clique of elements. Conflicts between matrix elements from different source trees often involve other elements from their respective source trees, with

members of each clique of elements supporting one another.

Although characters are also occasionally derived by using additive binary coding, the requisite use of additive binary coding in MRP results in nonindependence among matrix elements. This nonindependence implies that, compared to standard character matrices, goodness-of-fit indices should be interpreted differently (e.g., the CI would have a higher minimum value and would presumably measure the agreement among source trees) and some statistical methods may be inappropriate (e.g., bootstrap analysis; see Felsenstein, 1985; Purvis, 1995b).

Does MRP Combine Nodes or Trees?

Although MRP is described as a method for combining trees (Baum, 1992; Ragan, 1992b; Purvis, 1995b; Ronquist, 1996), matrix elements represent the nodes on those source trees. As a result, source trees can contribute different amounts of information; trees with more nodes (due to having more taxa or greater resolution or both) contribute more elements to the matrix and therefore generally have a greater influence on the topology of the composite tree. Thus, the claim that MRP eliminates the effect of data matrix size (Baum, 1992; Ragan, 1992b) applies to the number of characters but not to the number of taxa.

However, the claim that MRP favors larger source trees (Ronquist, 1996) is inaccurate. Despite the difference in size of the two source trees in Figure 2, clade (A, B, C) is unresolved in the CCT (Fig. 2c) because each source tree provides one (conflicting) piece of information regarding its resolution. The relative number of matrix elements provided by the source trees determines the resolution of regions of conflict ([A, B, C] in this example). The placement of taxa D–H is determined by the larger tree, which provides the only information concerning their positions. In other words, a size bias occurs only when the "missing nodes" (missing taxa or polytomies) are located within the region of conflict among the source trees. In Purvis's (1995b) example (Fig. 1), the region of conflict coincides with the entire tree.

This bias towards trees with more information (nodes) in regions of conflict may or may not be perceived as a problem. Trees with more nodes possess more hierarchical clustering information; this provides the basis for the argument that these trees should have a greater contribution to that region of the composite tree. In Figure 1, MRP results in the topology of the larger tree because both elements from Figure 1a support (A, B)D and therefore over-

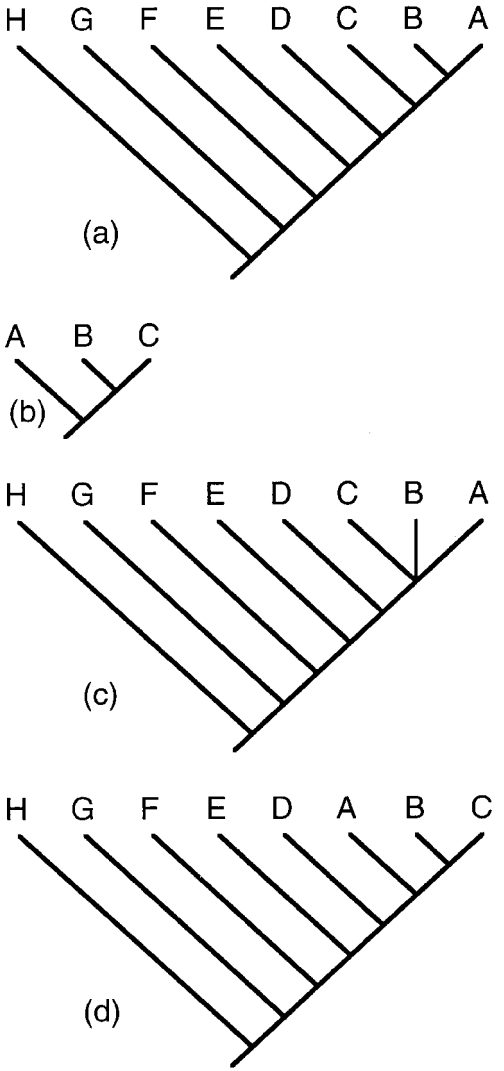


FIGURE 2. Demonstration of the localized nature of the size bias in MRP. Although tree a is much larger than tree b, the CCT (c) is unresolved in the region they share (A, B, C), reflecting the equal information content of the two trees in this region. (d) The intuitively erroneous result that results from inversely weighting the source trees according to their number of nodes.

rule the single conflicting element from Figure 1b. From this node-based perspective, whereby trees are viewed as solely the sum of their nodes, any size bias associated with MRP is appropriate.

From a tree-based perspective, however, each source tree is seen as a holistic entity that should have equal input into the topology of the composite tree. Purvis (1995b) noted that no placement of C on Figure 1b yields agreement with Figure 1a; as a result, he argued that the composite tree should be unresolved and that the bias of MRP toward the larger tree was inappropriate. Purvis's (1995b) argument, and this perspective as a whole, tacitly assumes that the addition of C to Figure 1b will not alter other relationships on that tree (see also Arnold, 1981; Donoghue et al., 1989; Lecointre et al., 1993), possibly to a pattern more similar to that on Figure 1a.

The use of MRP under a tree-based perspective requires correction for the bias toward trees with more nodes in regions of conflict. Inversely weighting elements based on the number of nodes on the source tree so that the total weight of each tree is equal (Ronquist, 1996) fails when the region of conflict forms only part of one or more of the source trees; it ignores the local nature of the size bias. For example, if the nodes on the trees in Figure 2 are inversely weighted, the composite tree includes A(B, C) (Fig. 2d); in contrast, unweighted MRP leaves (A, B, C) unresolved on the CCT, the intuitively correct result. Thus, weighting must be applied only to the conflicting regions between source trees, which becomes increasingly complex as more source trees are combined.

Differentiation between node- and tree-based perspectives is relevant methodologically only when the source trees have different terminal taxa or distributions of resolved nodes. Until recently, techniques that summarized multiple trees on a single topology dealt with multiple equally most-parsimonious trees (MPTs) derived from a single data set, among which differences in the number of nodes arose only from differences in resolution. With the development of MRP and other methods that combine trees with different terminal taxa, the question of whether a tree is equal solely to the sum of its nodes (see Adams [1986] for a discussion of this issue within a different context) has become an issue.

Novel Components

In the discussion of his composite tree synthesizing previous phylogenetic hypotheses concerning extant primates, Purvis (1995a: 414) claimed that "because all of the information on which it is based has been published previously, the composite tree cannot contain any clades that have not been implied by any previous study." Although this statement was intended to apply only to his modified coding method (A. Purvis, pers. comm.), it is also true of most consensus methods, which simply accept or reject components on the basis of agreement among the source trees. One exception is Adams consensus (Adams, 1972), which resolves disagreement among source trees by placing taxa of uncertain position as part of a polytomy at the least inclusive common node (Wilkinson, 1994). In contrast, the use by MRP of parsimony to produce the composite tree provides the potential that incongruence among the matrix elements may generate novel clades. This potential may be increased by the ability of MRP to combine trees with different terminal taxa.

In Figure 3, the CCT (Fig. 3c) includes a clade (marked with a solid circle) that is not present in either source tree (Figs. 3a, 3b). The CCT resembles Figure 3a except that *Pteronura* clusters with *Lutrogale*, as on Figure 3b. *Pteronura*'s membership in three components on Figure 3b appears to outweigh the evidence for a more basal position (Fig. 3a). The overall resemblance of the CCT to Figure 3a reflects the polytomies (lower information content) in Figure 3b.

The creation of novel clades appears to be uncommon. In the results of Bininda-Emonds et al. (in review), only 8 of the 198 nodes (4.0%) on the 13 composite trees occurred on none of the 274 source trees. The apparent rarity of novel clades may be related to the congruence among the matrix elements derived from each source tree, which may reduce the ability of individual elements from different source trees to interact in new combinations to form novel components. Most novel clades found in our analyses occur on only a fraction of a set of

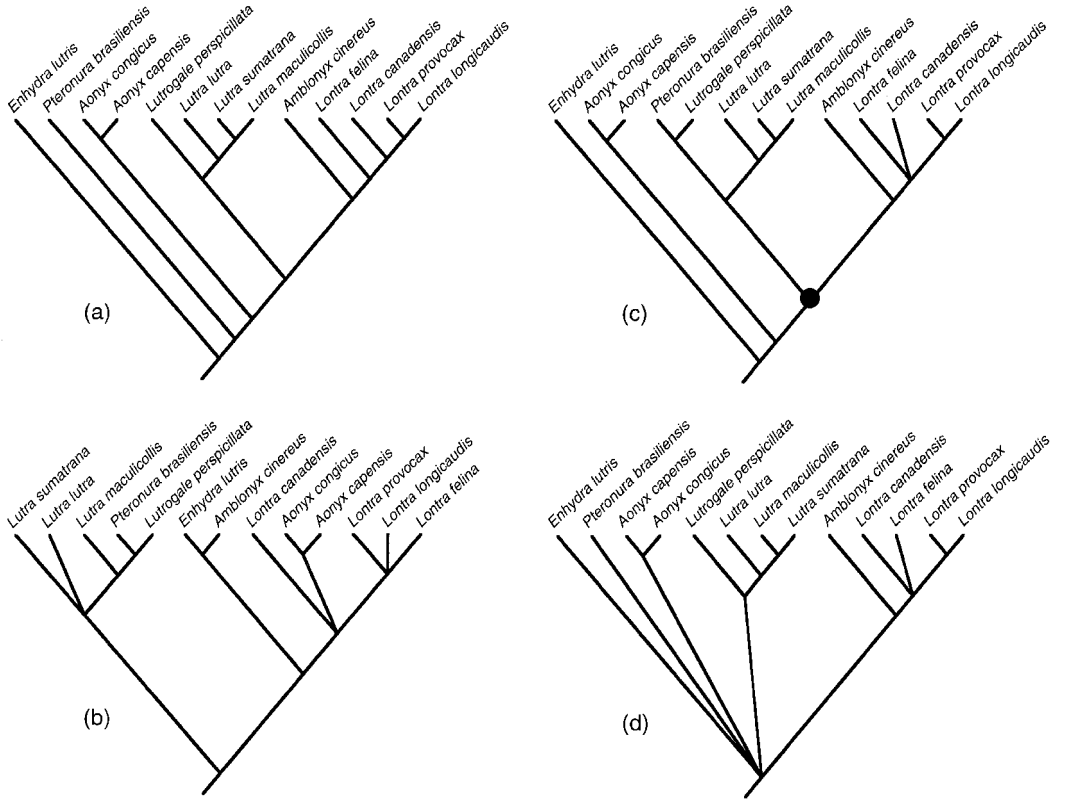


FIGURE 3. Creation of novel components using MRP (lutrine data from van Zyll de Jong, 1987). (a, b) Two source trees. (c) CCT with a component (●) that is not found in either tree a or tree b. (d) CCT generated when reversals are prohibited. The topology is similar to that of an Adams consensus tree.

MPCTs and therefore are subsumed within a polytomy when the MPCTs are summarized by use of a strict CCT.

Is MRP a Consensus Technique?

Although both conventional consensus techniques and MRP combine source trees based on their nodes, there are fundamental differences between them. Most consensus techniques look for the common occurrence of (agreement among) constituent nodes among source trees; conflict usually results in a polytomy (exceptions: majority rule and other consensus trees of the M_l family [McMorris and Neumann, 1983], which retain components found on a certain percentage of the trees). Within source trees, nodes are treated in isolation; individual components are either accepted or rejected (based on information from other source trees), and support for less-inclusive nodes by more-inclusive ones consists only of allowing those nodes with which they are congruent to occur on the consensus tree. Thus, although standard consensus techniques look for agreement among components, they are tree-based, in that source trees are combined equally, regardless of their size.

In contrast, in MRP, elements representing more-inclusive nodes directly support those of less-inclusive ones. For example, in Figure 1a the grouping of A and B to the exclusion of D is supported by both nodes on the tree: (A, B, C)D and (A, B)C, D. When that tree is combined with the smaller tree, (A, D)B, the composite tree includes (A, B). With standard consensus techniques, the contradiction of (A, B) by the second tree results in A and B forming part of a polytomy. The latter also occurs when using Lanyon's (1993) modified semistrict consensus

algorithm, which can handle trees with different terminal taxa. This feature of MRP results from it being node-based and arises through the use of additive binary coding to produce the element matrix and the use of parsimony to resolve the incongruence among elements from different source trees.

MRP has often been considered a consensus technique for combining the information in multiple data sets (e.g., DeSalle, 1994; Williams, 1994; Bruneau et al., 1995; de Queiroz et al., 1995), particularly because it eliminates the effect of differences in character number. Although both MRP and consensus techniques appear superficially to combine trees, clustering them on this basis conceals their different mechanics. Also, the ability of MRP to incorporate information about signal strength in the source matrix (see below) sets it still further apart from consensus techniques. Given the fundamental differences in how MRP combines source trees, we argue that MRP is not a consensus technique.

Other Possible Modifications to MRP

Prohibiting Reversals

The use of parsimony algorithms that allow reversals entails that clades in the composite tree can be supported in whole or in part by 0s in the element matrix. In the results of Bininda-Emonds et al. (in review), between 39 (19.7%) and 81 (40.9%) of the 198 nodes (DELTRAN vs. ACCTRAN optimization, respectively) were supported by one or more 0s. A few nodes were supported by more 0s than 1s, particularly under ACCTRAN optimization (Fig. 4).

Clustering on the basis of 0s seems inappropriate in MRP because support for a clade is based in part on taxa that share a lack of membership in the components on the source trees. Unlike conventional character data, in which transformation in either direction between character states can be considered potential evidence for clustering taxa, in MRP only the 1s in the element matrix represent membership in components and therefore seem appropriate for grouping taxa. This suggests that the parsimony algorithm used in MRP should not allow reversals.

To test the effect of prohibiting reversals, we reanalyzed 19 recent "total evidence" studies

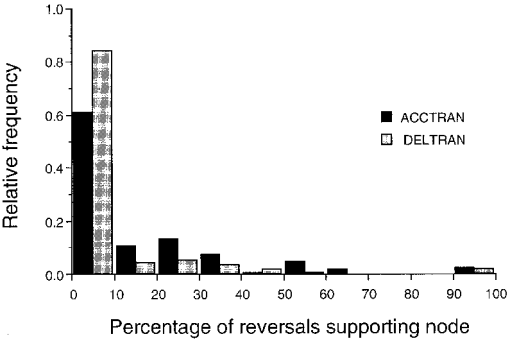


FIGURE 4. Frequency histogram showing the percentage of reversals supporting each of the 198 nodes of Bininda-Emonds et al. (in review) under both ACCTRAN and DELTRAN optimizations. Most nodes were not supported by any 0s.

(sensu Kluge, 1989; Table 1), using MRP to combine the topologies of the process partitions. Analyses were conducted with and without reversals, and differences in topology between the (consensus) composite trees and the total evidence tree were quantified by using the symmetric difference metric (Penny and Hendy, 1985). Despite the variation among the 19 studies in partition size, number of taxa, and disagreement among the partition trees, the effect of prohibiting reversals was usually minor. In eight cases, MRP with and without reversals produced the same CCT; in five cases, prohibiting reversals yielded a CCT that was more similar to the total evidence tree; and in six cases, allowing reversals produced a CCT that was closer to the total evidence tree. The topologies with and without reversals were markedly different from each other in only three instances, and in only two of these (Bininda-Emonds and Russell, 1996; Bremer, 1996) did the source trees conflict strongly.

The example in Figure 5 may suggest why prohibiting reversals does not necessarily produce the better result. Although allowing reversals produces the intuitively correct result in this example, the reversal (Fig. 5c) involves only taxon B (i.e., it does not support a clade) and simply represents the incongruence of the position of B on the CCT with that on one of the source trees (Fig. 5a). This example suggests that MRP might perform better if it were based on an algorithm that prohibited reversals on internal branches but allowed them on terminal ones.

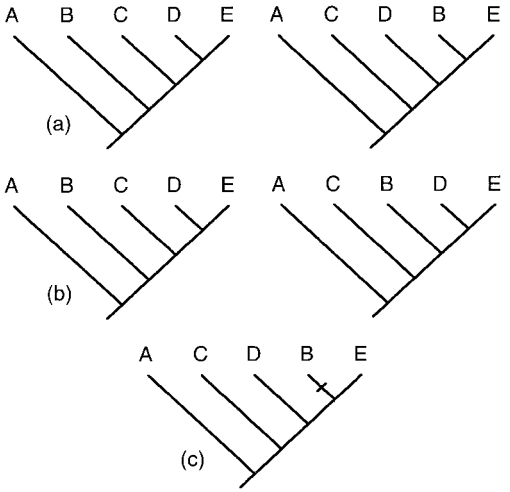


FIGURE 5. Comparison of standard MRP and MRP without reversals. (a) Source trees. (b) Two equally parsimonious composite trees generated by MRP without reversals. (c) Additional tree produced with reversals (location marked with a bar). The three MPCTs based on standard MRP seem to better cover the range of possible positions of taxon B on the source trees.

The question of whether to allow reversals in MRP analyses requires further study. Our sample shows that prohibiting reversals usually produces only minor differences in topology. Prohibiting reversals also does not markedly alter the number of novel clades (which may be supported largely by 0s); reanalysis of the source trees in Bininda-Emonds et al. (unpubl.) with reversals prohibited increased the number of novel clades from 8 to 10. These reanalyses also suggest that MRP without reversals is somewhat conservative, frequently producing a CCT that, like Adams consensus trees, places incongruent components (e.g., Fig. 3d) or particular members of those components (e.g., taxon B in Fig. 5) in basal positions.

Increasing the Informativeness of Source Tree Polytomies

The number of nodes on a source tree, and consequently the number of matrix elements associated with the tree, is reduced by polytomies. Thus, for a given number of terminal taxa, trees with more polytomies have relatively less influence on the pattern of the composite tree in regions of conflict than do completely dichotomous trees. The use of additive binary coding in MRP makes this method unable to distinguish between polytomies that are considered "hard" (representing putative multiple speciation events) and those that are considered "soft" (representing either a lack of resolution or conflicting resolutions) (Maddison, 1989); all polytomies are considered unresolved. While this is appropriate for soft polytomies that do represent a lack of resolution, it is inappropriate for polytomies that are purported to be hard (and therefore fully resolved), and for soft polytomies that represent conflicting solutions, such as those on a strict consensus tree.

We have no suggestions for counteracting any perceived bias this causes against hard polytomies. Nonetheless, the problems in identifying hard polytomies makes this largely a nonissue. For soft polytomies that represent conflicting resolutions, the goal is to retrieve those resolutions of the polytomy that occur among the MPTs. Although a case can be made for using the consensus solution on practical grounds when there are large numbers of MPTs (Ragan, 1992b), ideally, the more-resolved topologies of the MPTs and not that of the less-resolved consensus should contribute to the element matrix.

This can be accomplished by coding each unique component on any one or more trees in a set of MPTs as an element in the combined matrix (see Fig. 6). These elements can be handled in at least two ways. Weighting each component according to its frequency among the MPTs is the equivalent of Ragan's (1992b) suggestion of individually coding and including each MPT in the element matrix. In this procedure, the influence of a clade on the pattern of the composite tree depends on its frequency among the MPTs: clades that occur on more of the MPTs will have a greater influence, and any clade occurring on only some MPTs will have less weight in the analysis than will clades supported by data sets producing only one MPT. Alternatively, one of us (H.N.B.) argues that the above weighting is inappropriate. Because each of the shortest trees obtained from a single data set is equally parsimonious, overall evidence in the data set for each clade on any one or more of the MPTs is equal and so they should all receive equal weight.

In either instance, weighting of these elements relative to those based on other source trees is not necessary. Although the number of elements derived from data sets that produce more than one MPT will usually be larger than that associated with a single MPT with the same number of taxa, either the frequency-dependent weighting of, or the incongruence among, elements representing alternative resolutions of polytomies on consensus trees negates the increased influence their increased number might have on the topology of the composite tree (see Fig. 6).

Weighting Elements Based on Evidential Support

Because MRP generates composite trees based solely on the topologies of the source trees, there is no inherent consideration of either the overall support for the topology of a source tree or for the differential support for the nodes on that tree (Rodrigo, 1993; Galtier and Gouy, 1994; Bandelt, 1995; Bruneau et al., 1995). Although bypassing the original data is necessary in some instances, it has been argued that the differential support for the overall topologies of, or different nodes on, the source trees could or should be used in deriving the composite tree whenever possible (Purvis, 1995b; Ronquist, 1996).

Potential measures of support for entire source trees include goodness-of-fit indices (Farris, 1989; Baum, 1992), PTP values (Faith and Cranston, 1991), or total support (Källersjö et al., 1992). Potential means of weighting based on differential support for individual clades on source trees include bootstrap frequencies (Felsenstein, 1985), Bremer support (Bremer, 1988), the number of unique synapomorphies (Kluge, 1989), or T-PTP probabilities

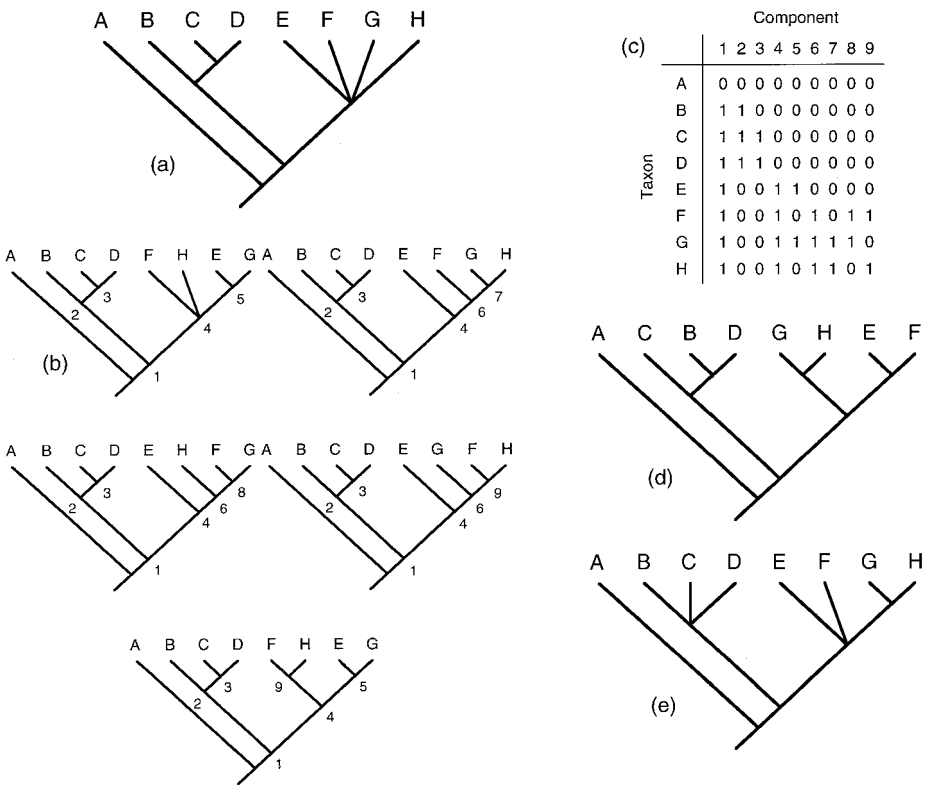


FIGURE 6. One method of coding source trees with polytomies. (a) Consensus tree. (b) Five MPTs (the nine unique components are numbered). (c) Matrix consisting of one element for each of the numbered unique components. (d) Conflicting source tree. (e) CCT resulting from unweighted MRP analysis of the matrix c and the elements derived from the conflicting source tree d.

(Faith, 1991). Ronquist (1996) showed that weighting elements from single source trees based on either Bremer support or bootstrap values improves the correlation between tree lengths obtained from the element matrix and those obtained from the original character data. Weighting by Bremer support also causes the decay values to be reproduced exactly (Ronquist, 1996). Thus, these methods apparently solve a major criticism of MRP, namely, that it fails to incorporate relative support for nodes into the analysis.

The use of any of these measures of support requires that two conditions be met. First, the chosen metric must be available for all source trees. Because source trees lacking the metric should not be ignored, this requirement may in some instances preclude the use of weighting based on evidential support. The use of multiple metrics might be feasible in some of these cases; however, the necessity that they all yield equivalent, standardized information will probably prevent this. Second, the values of the chosen metric must provide a comparable measure of the relative support for a given node across studies, regardless of the characteristics of the original data and the algorithm that produced the source tree. Bremer support, total support, and the number of unique synapomorphies are influenced by the number of characters in the original data matrices and need to be standardized across data sets. Bremer support cannot be used when components on multiple MPTs, rather than those on their consensus, are coded because the Bremer support values for components that do not occur on all MPTs are zero and so their associated elements would have a weight of zero. The bootstrap may be less influenced by the differential characteristics of data sets (i.e., values are probably more standardized), but this issue requires further study.

Because all these weighting schemes based on support for entire trees or individual nodes operate on the elements in the matrix, they do not offset the inherent size bias of MRP (contra Ronquist, 1996). The problems associated with weighting the element matrix to make MRP tree-based are only compounded if weighting based on evidential support in the original data is also desired.

Closing Statements

MRP is unique in that it combines source trees by using additive binary coding to convert the hierarchical information within them into an element matrix and using parsimony to derive the composite tree; these mechanics clearly differentiate MRP from standard consensus techniques, despite being associated with them by many authors. MRP is inherently node-based. The influence of individual source trees on the composite tree depends on their size and resolution, and the matrix elements derived from a single source tree directly support one another. As a result, source trees are not combined equally. In contrast, consensus techniques are tree-based. Although they also operate at the node level, components are treated in isolation and are simply accepted or rejected based on the agreement among the source trees. Therefore, all trees have an equal vote regarding the topology of the consensus tree. This difference is fundamental, and is the basis for our conclusion that MRP is not a consensus technique. The difference in both the mechanics and results of MRP, as compared to those of consensus techniques, may require a shift in current thinking as to appropriate methods for combining source trees. At the very least, MRP is providing a synthesis different from consensus techniques of the information in a set of source trees.

MRP has been promoted as a "total evidence approach" (Purvis, 1995b:253) for data sets that are not amenable to standard character congruence methods (e.g., Kluge, 1989). The data in Table 1 suggest that MRP often falls short of this goal. In only 3 of 19 cases does the topology produced by MRP match the total evidence tree, and in several instances the differences between the results of the two methods are marked. MRP tends to collapse clades found on the total evidence tree, producing more polytomies, but taxa are also occasionally placed in different positions. There is no obvious relationship between the number of partitions and the ability of MRP to match the total evidence tree.

The use of parsimony allows for weighting of the matrix elements to adjust for inherent biases in the method or to incorporate additional information. To date, attempts to adjust

TABLE 1. Comparison of MRP with and without reversals in 19 total evidence studies. The CCTs obtained from the respective MRP analyses of the partitions were compared with the total evidence topology and each other using Penny and Hendy's (1985) symmetric difference metric.

Study	No. taxa	No. partitions	With reversals vs. total evidence	Without reversals vs. total evidence	With vs. without reversals
Kluge, 1989	11	2	3	3	0
Vane-Wright et al., 1992	10	2	1	3	2
Cundall et al., 1993	18	3	13	13	0
Eernisse and Kluge, 1993	5	7	1	0	1
Wheeler et al., 1993	26	3	15	17	12
Kim and Jansen, 1994	7	4	2	0	2
Lundrigan and Tucker, 1994	12	3	0	0	0
Omland, 1994	9	2	1	1	0
Vrana et al., 1994	31	2	9	7	4
Yoder, 1994	13	2	7	5	2
Littlewood and Smith, 1995	45	3	9	10	1
Paterson et al., 1995	18	3	6	7	1
Tehler, 1995	5	2	2	1	1
Zhang, 1995	8	2	0	0	0
Bininda-Emonds and Russell, 1996	27	7	22	32	20
Bremer, 1996	33	2	12	16	8
Friesen et al., 1996	25	2	14	14	0
Sites et al., 1996 (*Enyalioides* outgroup)	14	3	4	4	0
Sites et al., 1996 (*Oplurus* outgroup)	14	3	0	0	0

for any perceived bias toward more-informative (i.e., larger, more resolved, or both) regions of source trees (e.g., Purvis, 1995b; Ronquist, 1996) to make the method tree-based have been unsuccessful. The appropriateness of correcting for this bias is arguable; however, without this correction, MRP should not be used if a tree-based result is desired. With certain limitations, weighting provides a means of incorporating the differential support for entire trees or individual clades present in the original data into the analysis, while still allowing the composite tree to be based primarily on the hierarchical information in the source trees. These modifications might allow MRP to more closely approximate a "total evidence" result.

Baum (1992) noted that detailed study of the properties of MRP, empirical testing of its results, and comparisons with standard consensus techniques had not yet been conducted. Subsequent analyses (Baum and Ragan, 1993; Rodrigo, 1993; Purvis, 1995b; Ronquist, 1996; this study) have considered some of these issues. Issues requiring further study include the appropriateness of allowing reversals within matrix elements (on either all or only terminal branches) and the ability of MRP to replace total evidence analyses when the data are not suitable for the latter. These studies are essential to assess the potential contribution of MRP and its variants to phylogenetic inference.

ACKNOWLEDGMENTS

The first author thanks Andy Purvis for introducing him to MRP and for many fruitful initial discussions regarding the technique, and acknowledges Alberta Heritage, the United Kingdom's Overseas Research Scholarship plan, and a Natural Sciences and Engineering Research Council of Canada postgraduate scholarship for funding the project. Andy Purvis also pointed out the counterexample used in Figure 2. Fredrik Ronquist kindly supplied us with a preprint copy of his manuscript. David Cannatella, Paul Harvey, Andy Purvis, Alan Rodrigo, and Fredrik Ronquist made constructive comments on an earlier version of this paper.

REFERENCES

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

ADAMS, E. M., III. 1986. N-trees as nestings: Complexity, similarity, and consensus. J. Classif. 3:299–317.

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

BANDELT, H.-J. 1995. Combination of data in phylogenetic analysis. Pages 355–361 *in* Systematics and evolution of the Ranunculiflorae (U. Jensen and J. W. Kadereit, eds.). Plant Syst. Evol. Suppl. 9, Vienna.

BAUM, B. R. 1992. Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. Taxon 41:3–10.

BAUM, B. R., AND M. A. RAGAN. 1993. Reply to A. G. Rodrigo's "A comment on Baum's method for combining phylogenetic trees." Taxon 42:637–640.

BININDA-EMONDS, O. R. P., AND A. P. RUSSELL. 1996. A morphological perspective on the phylogenetic relationships of the extant phocid seals (Mammalia: Carnivora: Phocidae). Bonn. zool. Monogr. 41:1–256.

BREMER, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution 42:795–803.

BREMER, B. 1996. Combined and separate analyses of morphological and molecular data in the plant family Rubiaceae. Cladistics 12:21–40.

BRUNEAU, A., E. E. DICKSON, AND S. KNAPP. 1995. Congruence of chloroplast DNA restriction site characters with morphological and isozyme data in *Solanum* sect. *Lasiocarpa*. Can. J. Bot. 73:1151–1167.

CUNDALL, D., V. WALLACH, AND D. A. ROSSMAN. 1993. The systematic relationships of the snake genus *Anomochilus*. Zool. J. Linn. Soc. 109:275–299.

DE QUEIROZ, A., M. J. DONOGHUE, AND J. KIM. 1995. Separate versus combined analysis of phylogenetic evidence. Annu. Rev. Ecol. Syst. 26:657–681.

DESALLE, R. 1994. Flies and congruence. Am. J. Phys. Anthropol. 94:125–141.

DONOGHUE, M. J., J. A. DOYLE, J. GAUTHIER, A. G. KLUGE, AND T. ROWE. 1989. The importance of fossils in phylogeny reconstruction. Annu. Rev. Ecol. Syst. 20:431–460.

EERNISSE, D. J., AND A. G. KLUGE. 1993. Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology. Mol. Biol. Evol. 10:1170–1195.

FAITH, D. P. 1991. Cladistic permutation tests for monophyly and nonmonophyly. Syst. Zool. 40:366–375.

FAITH, D. P., AND P. S. CRANSTON. 1991. Could a cladogram this short have arisen by chance alone?: On permutation tests for cladistic structure. Cladistics 7:1–28.

FARRIS, J. S. 1989. The retention index and the rescaled consistency index. Cladistics 5:417–419.

FARRIS, J. S., A. G. KLUGE, AND M. J. ECKHARDT. 1970. A numerical approach to phylogenetic systematics. Syst. Zool. 19:172–191.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791.

FRIESEN, V. L., A. J. BAKER, AND J. F. PIATT. 1996. Phylogenetic relationships within the Alcidae (Charadriiformes, Aves) inferred from total molecular evidence. Mol. Biol. Evol. 13:359–367.

GALTIER, N., AND M. GOUY. 1994. Molecular phylogeny of Eubacteria: A new multiple tree analysis method applied to 15 sequence data sets questions the monophyly of Gram-positive bacteria. Res. Microbiol. 145:531–541.

GORDON, A. D. 1986. Consensus supertrees: The synthesis of rooted trees containing overlapping sets of labeled leaves. J. Classif. 3:31–39.

HILLIS, D. M. 1987. Molecular versus morphological approaches to systematics. Annu. Rev. Ecol. Syst. 18:23–42.

KÄLLERSJÖ, M., J. S. FARRIS, A. G. KLUGE, AND C. BULT. 1992. Skewness and permutation. Cladistics 8:275–287.

KIM, K.-J., AND R. K. JANSEN. 1994. Comparisons of phylogenetic hypotheses among different data sets in dwarf dandelions (*Krigia*, Asteraceae): Additional information from internal transcribed spacer sequences of nuclear ribosomal DNA. Plant Syst. Evol. 190:157–185.

KLUGE, A. G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). Syst. Zool. 38:7–25.

LANYON, S. M. 1993. Phylogenetic frameworks: Towards a firmer foundation for the comparative approach. Biol. J. Linn. Soc. 49:45–61.

LECOINTRE, G., H. PHILIPPE, H. L. VÂN LÊ, AND H. LE GUYADER. 1993. Species sampling has a major impact on phylogenetic inference. Mol. Phylogenet. Evol. 2:205–224.

LITTLEWOOD, D. T. J., AND A. B. SMITH. 1995. A combined morphological and molecular phylogeny for sea urchins (Echinoidea: Echinodermata). Philos. Trans. R. Soc. Lond. B 347:213–234.

LUNDRIGAN, B. L., AND P. K. TUCKER. 1994. Tracing paternal ancestry in mice, using the Y-linked, sex-determining locus, *Sry*. Mol. Biol. Evol. 11:483–492.

MADDISON, W. P. 1989. Reconstructing character evolution on polytomous cladograms. Cladistics 5:365–377.

MCMORRIS, F. R., AND D. NEUMANN. 1983. Consensus functions defined on trees. Math. Social Sci. 4:131–136.

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

OMLAND, K. E. 1994. Character congruence between a molecular and a morphological phylogeny for dabbling ducks (*Anas*). Syst. Biol. 43:369–386.

PATERSON, A. M., G. P. WALLIS, AND R. D. GRAY. 1995. Penguins, petrels, and parsimony: Does cladistic analysis of behavior reflect seabird phylogeny? Evolution 49:974–989.

PENNY, D., AND M. D. HENDY. 1985. The use of tree comparison metrics. Syst. Zool. 34:75–82.

PURVIS, A. 1995a. A composite estimate of primate phylogeny. Philos. Trans. R. Soc. Lond. B 348:405–421.

PURVIS, A. 1995b. A modification to Baum and Ragan's method for combining phylogenetic trees. Syst. Biol. 44:251–255.

RAGAN, M. A. 1992a. Matrix representation in reconstructing phylogenetic relationships among the eukaryotes. Biosystems 28:47–55.

RAGAN, M. A. 1992b. Phylogenetic inference based on matrix representation of trees. Mol. Phylogenet. Evol. 1:53–58.

RODRIGO, A. G. 1993. A comment on Baum's method for combining phylogenetic trees. Taxon 42:631–636.

RODRIGO, A. G. 1996. On combining cladograms. Taxon 45:267–274.

RONQUIST, F. 1996. Matrix representation of trees, redundancy, and weighting. Syst. Biol. 45:247–253.

SITES, J. W., JR., S. K. DAVIS, T. GUERRA, J. B. IVERSON, AND H. L. SNELL. 1996. Character congruence and phylogenetic signal in molecular and morphological data

sets: A case study in the living iguanas (Squamata, Iguanidae). Mol. Biol. Evol. 13:1087–1105.

STEEL, M. 1992. The complexity of reconstructing trees from qualitative characters and subtrees. J. Classif. 9:91–116.

TEHLER, A. 1995. Morphological data, molecular data, and total evidence in phylogenetic analysis. Can. J. Bot. 73:S667–S676.

VAN ZYLL DE JONG, C. G. 1987. A phylogenetic study of the Lutrinae (Carnivora; Mustelidae) using morphological data. Can. J. Zool. 65:2536–2544.

VANE-WRIGHT, R. I., S. SCHULZ, AND M. BOPPRÉ. 1992. The cladistics of *Amauris* butterflies: Congruence, consensus and total evidence. Cladistics 8:125–138.

VRANA, P. B., M. C. MILINKOVITCH, J. R. POWELL, AND W. C. WHEELER. 1994. Higher level relationships of the arctoid Carnivora based on sequence data and "total evidence." Mol. Phylogenet. Evol. 3:47–58.

WHEELER, W. C., P. CARTWRIGHT, AND C. Y. HAYASHI. 1993. Arthropod phylogeny: A combined approach. Cladistics 9:1–39.

WILKINSON, M. 1994. Common cladistic information and its consensus representation: Reduced Adams and reduced cladistic consensus trees and profiles. Syst. Biol. 43:343–368.

WILLIAMS, D. M. 1994. Combining trees and combining data. Taxon 43:449–453.

YODER, A. D. 1994. Relative position of the Cheirogaleidae in strepsirhine phylogeny: A comparison of morphological and molecular methods and results. Am. J. Phys. Anthropol. 94:25–46.

ZHANG, Z.-Q. 1995. A cladistic analysis of Trombidiidae (Acari: Parasitengona): Congruence of larval and adult character sets. Can. J. Zool. 73:96–103.

Received 22 July 1996; accepted 10 September 1997
Associate Editor: D. Cannatella

Syst. Biol. 47(3): 508–519, 1998

Why Morphometrics Is Not Special: Coding Quantitative Data for Phylogenetic Analysis

DONALD L. SWIDERSKI,[1] MIRIAM LEAH ZELDITCH,[2] AND WILLIAM L. FINK[3]

[1]Museum of Zoology, [2]Museum of Paleontology, and [3]Department of Biology and Museum of Zoology, University of Michigan, Ann Arbor, Michigan 48109, USA; E-mail: dlswider@ umich.edu (D.L.S.), zelditch@ umich.edu (M.L.Z.), wfink@ umich.edu (W.L.F.)

Systematists often use qualitative descriptions of shape in phylogenetic analyses, but several biologists object to phylogenetic analyses using quantitative descriptions of those same shapes (Pimentel and Riggins, 1987; Felsenstein, 1988; Mickevich and Weller, 1990; Garland and Adolph, 1994). In a previous paper (Zelditch et al., 1995), we argued that the problem with phylogenetic analysis of quantitative shape data lies in the particular methods traditionally used to quantify shapes, not in quantification per se. In addition, we demonstrated that some of the more serious objections to using morphometric data in phylogenetic analyses are removed by using landmark-based morphometric methods developed by Bookstein (1991). Although we demonstrated that phylogenetic analysis of quantified shape variables is valid in theory, some practical problems remain. In this paper, we address the major remaining problem, that of coding: specifically, the problem of recognizing divergent character states.

Even a brief survey of the literature shows that coding is a complicated task in which several obstacles must be overcome (see Mickevich and Weller, 1990; Mabee and Humphries, 1993; Wilkinson, 1995). In this paper, we focus on one particular obstacle: evaluation of the diversity of a feature to determine which sets of taxa are similar in that feature. These judgements of similarity (and differences) are the foundations on which inferences of homology and monophyly are based. If these judgements employ inappropriate criteria, then those inferences are apt to be misled, and the resulting phylogeny is likely to be wrong.

Several biologists have argued that there can be no valid criteria for dividing quantitative data into discrete states because quantitative traits are inherently continuous (Pimentel and Riggins, 1987; Felsenstein, 1988; Garland and Adolph, 1994). In fact, they claim that coding quantitative data introduces artificial distinctions even if the observed distribution is discontinuous. This claim has even been parlayed

into arguments against cladistic parsimony (Felsenstein, 1988; Garland and Adolph, 1994), or against phylogenetic analysis of all morphometric data (Pimentel and Riggins, 1987; Mickevich and Weller, 1990). Therefore, we begin this paper by showing that the arguments against coding quantitatively described traits are not supported by theory. Rather, the obstacles posed by continuity are only practical problems and are not unique to quantitative data.

In the remainder of this paper, we address the practical problems of recognizing different states. Several systematists have proposed methods of coding that are designed to recognize states despite the lack of discontinuities between taxa. We review some commonly used methods and show that the criteria most methods employ to delimit states are not appropriate for phylogenetic analysis. Consequently, the character states produced by these methods do not support hypotheses of homology. We did find one method that is suitable for phylogenetic analysis, which we illustrate by using it to code features of adult body shape in six species of piranha. We prefer this method because it does not rely on arbitrary distance criteria or on statistical hypotheses that are irrelevant to the inference of homology.

Continuity

Thiele (1993) suggested that some objections to coding quantitative data can be removed by making a distinction between terms that indicate how the trait was described and terms that indicate how the trait varies. Four terms, (qualitative, quantitative, continuous, and discontinuous) indicate how a trait is described. As did Wiley (1981), Thiele argued that quantitative should mean only that the trait was described by a numerical scale, i.e., by counting or measurement. In contrast, qualitative should mean only that the trait was described by using words. Continuous characters are a subset of quantitative characters, specifically those described by using an infinitely divisible numerical scale (e.g., real numbers). Discontinuous refers to the subset of quantitative characters that are described by using a numerical scale that is not infinitely divisible (e.g., integers). Used in this way, these terms imply nothing about how a trait varies. They imply nothing about biology because they refer only to our measurement scales.

Objections to coding morphometric data are not really concerned with the use of a continuous quantitative scale, but with what Thiele called "overlapping;" i.e., the range of variation of a trait in one taxon contains values that are also within the range of variation of that trait in another taxon. The contrasting pattern is disjunct, meaning that none of the values within the range of one taxon lie within the range of the other taxon. The words overlapping and disjunct can be applied whether the trait is described quantitatively or qualitatively; however, the comparison of ranges of qualitatively described features is necessarily subjective.

Morphometric data are usually reported as though the measurements were taken on a continuous scale. In reality, the scale of any instrument is discontinuous (e.g., 0.01, 0.02, 0.03, . . .), reflecting the limit of the resolving power of the instrument. A report utilizing the instrument's discontinuous scale is interpreted as an approximation to a continuous scale, not as an indication of steplike behavior of the character. Traits that are customarily reported on a discontinuous scale are counts for which fractional values are excluded conceptually (e.g., the number of teeth, for which incompletely formed teeth are either counted or not counted).

Thiele's discussion of the semantic issues clarifies the point that chains of overlapping ranges are the primary obstacle to coding morphometric traits. However, Thiele, as did Stevens (1991), also pointed out that this problem is not unique to quantitative data. In fact, Thiele and Stevens argued that this is one of many of the problems that are the same for quantitative and qualitative data. For example, one issue that must be resolved for every feature is the comparability of that feature across all taxa in the study; another is the recognition of distinct conditions of the feature. Thus, Thiele and Stevens argued for applying the same criteria to quantitative and qualitative data, and for making the criteria explicit for all data.

Pimentel and Riggins (1987) were quite explicit. They argued that features with overlapping ranges should not be coded as having distinct states. This position is evident from

their definition of a cladistic character as "a feature of organisms that can be evaluated as a variable with two or more *mutually exclusive* and ordered states" (p. 201, emphasis added). It is also clear that this definition was meant to apply to all kinds of data, because Pimentel and Riggins stated it at the beginning of the paper, and again, in their discussion of quantitative data (p. 207). In the latter context, they elaborated on the requirement that characters have mutually exclusive states and argued that the only valid basis for coding any character is a gap, a hiatus in the distribution of a character, such that no individuals are observed to have those values. In their view, the ideal case would be a gap between ranges (Thiele's "disjunction"). Pimentel and Riggins did allow coding if a few taxa contain individuals that are on each side of the gap (these taxa would be polymorphic), but no taxon can have individuals within the gap. For Pimentel and Riggins, the gap is absolutely required for coding because it unambiguously demarcates mutually exclusive sets of values, without any statistical or mathematical manipulation. They characterized these gaps as "natural" (p. 207), implying that any distinction which is not based on disjunction in the raw data is artificial.

Felsenstein (1988) agreed that division of overlapping ranges into separate states creates artificial distinctions. However, he argued that coding based on observed gaps also imposes artificial distinctions. Felsenstein claimed this argument is supported by theoretical predictions that polygenic characters will exhibit gradual, incremental change. Thus, even if a trait evolves rapidly, it still passes from one value to the next with no values skipped (i.e., saltation does not occur). This implies that a descendant population will overlap the immediately ancestral population. From this implication, Felsenstein inferred that disjunction of terminal taxa represents missing data (e.g., unrecovered fossils), because if all the ancestral populations were known they would form links in an unbroken chain connecting the terminals. Felsenstein argued that observed gaps between terminals are not real and should not be used as the foundation for any coding scheme.

Felsenstein's argument about the reality of gaps overextends a legitimate theory. That theory describes anagenetic change, transformation in a single unbranched lineage. However, lineages branch (speciation occurs), and the branches are genetically and evolutionarily independent. Because they are independent, the chains representing the descendant lineages will eventually become distinct from each other, as individuals within the lineages acquire novelties. This divergence is simply a consequence of independent evolution within separate lineages. The unbroken links of the chain connect ancestors to descendants, not terminal taxa to each other.

We conclude that there is no obstacle in theory to coding taxa with overlapping ranges. In fact, Felsenstein's argument provides grounds for us to argue that phylogenetic systematists need an approach to coding that does not require gaps. The ranges of populations representing nascent branches can be expected to overlap each other and the range of their common ancestor. Obviously, a gap would be useful, but a lesser amount of differentiation can also indicate evolutionary independence. The goal of phylogenetic systematics is to infer evolutionary independence (branching) from evidence of divergence. When divergence is relatively small and ranges overlap, the real obstacle to coding is distinguishing between differences due to poor sampling and differences due to evolution. We address this problem below.

Methods of Coding

In this section, we review some of the most widely used methods of coding. For each method, we focus on the criterion used to divide a series of taxa with overlapping ranges into smaller groups, and on the validity of that criterion as a basis for inferring homology.

We begin with gap coding (Mickevich and Johnson, 1976), both because it is one of the oldest methods of coding and because most newer methods of coding are intended to improve on gap coding. This method is illustrated with the five hypothetical populations shown in Figure 1. The smallest mean is assigned state 0. The next largest mean is assigned a new state only if the difference between means is greater than the value of the pooled standard deviation (s_p). Then the third mean is compared to the second, and so

on, until all pairs of adjacent means have been evaluated. In this context, "gap" refers to the difference between means, not the disjunction between ranges.

The principal problem with gap coding is that it provides a small amount of unreliable information from which to judge the similarity of taxa. The information is the similarity of means, as indicated by s_p. This information is unreliable because variances of taxa are often dissimilar, making s_p a poor indicator of the actual overlap between two taxa. Some will overlap more than expected; others, less. Gap coding also misrepresents the amount of overlap when distributions are skewed or otherwise deviate from normality. For these reasons, we do not recommend gap coding as a basis for inferring similarity of taxa.

Most critiques of gap coding focus on other problems (e.g., Thorpe, 1984; Archie, 1985; Chappill, 1989). One common complaint is that taxa may be quite different but still be assigned the same state because they are ends of a long series of closely spaced taxa. This problem is illustrated in Figure 1 by taxa B, C, and D. The distance between B and C is small ($< s_p$), as is the distance between C and D. Consequently, all three taxa are assigned the same state, even though the distance between B and D is large ($> s_p$).

To solve this problem, Archie (1985) proposed a method of defining subsets of taxa and a method of coding overlapping subsets. To define a subset, Archie used the mean of a taxon (x_i) and s_p to define an interval (x_i to $x_i + s_p$). The subset includes all taxa with means in that interval. For Figure 1, the subsets would be {A}, {B, C}, {C, D}, {D}, and {E}. The method of coding begins by deleting any subset that is completely included in another, such as {D}. Then, state 0 is assigned to the subset that includes the lowest mean, {A}. Codes increment by 1 at the beginning or end of a subset (or by 2 if the subsets are disjunct). Thus, the Archie coding for Figure 1 would be A = 0, B = 2, C = 3, D = 4, E = 6.

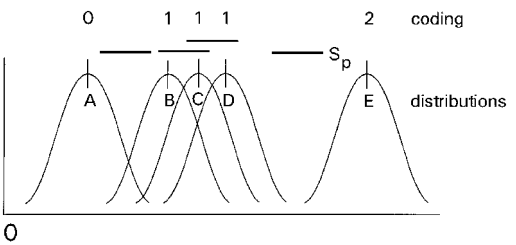


FIGURE 1. Gap coding of five idealized populations (A–E).

Archie's methods solve the problem of long series, but still rely on the dubious information provided by the mean and s_p. More importantly, Archie's methods have a serious problem of their own. Farris (1990) criticized Archie's methods because subsets are defined by some criterion in the first step, but then that criterion is ignored in the second step. In the example above, B and C are placed in the same subset because they are not different from each other, but are assigned different states because C is not different from D. The inconsistent logic is particularly clear when subsets are defined by statistical analyses, as in homogeneous subset analysis (Simon, 1983; Farris, 1990). Using Archie coding on these subsets would assign different states to taxa despite statistical tests showing that their means are not significantly different.

A somewhat different solution to the problem of long series is incorporated in methods proposed by Colless (1980), Thorpe (1984), and Chappill (1989). In these methods, the morphometric distance between the most widely separated means or individuals is divided into two or more equal segments. The segments are numbered in order, and the code assigned to a mean or individual is the number of the segment in which it is located (Fig. 2). In effect, segment coding rescales the original measurements to a smaller number of larger increments.

Segment coding solves the problem posed by long chains of closely spaced means. However, it replaces that problem with a more fundamental one. It distorts the similarities and differences among the taxa. Means or individuals near the limits of a segment may be more similar to those in the adjacent segment than they are to the ones in their own segment. In Figure 2, C is no closer to B than it is to D, but B and C are assigned the same state and D is assigned a different state. Because segment coding does not reflect similarity, it cannot be used as a basis for inferring homology.

Archie coding and segment coding create bigger problems than the one they solve. This

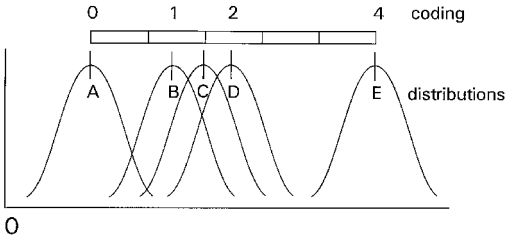


FIGURE 2. Segment coding of five idealized populations (A–E).

is because these methods are designed to solve the wrong problem. That problem, failure to distinguish between taxa at the ends of a long series, does not arise from a defect of gap coding; rather, gap coding reveals the reality that sometimes intermediate taxa bridge the difference between the ends. This problem does not have a solution. Any consistent method that evaluates ranges of variation will encounter cases in which intermediate taxa form a bridge between taxa that would otherwise be considered different. Such cases may bring attention to the criterion used to judge whether ranges are similar or different, but they are not grounds for replacing evaluation of similarity with computation of arbitrary distance metrics.

Farris (1990) argued that the criterion of similarity used in gap coding (s_p) should be replaced with explicit statistical tests of the differences between means. Then, a series of pairwise tests (with appropriate adjustments for multiple comparisons) is used to construct homogeneous subsets: groups in which no two sample means are statistically significantly different (Simon, 1983). Because homogeneous subset coding uses an explicit statistical test of similarity, rather than an unreliable indicator of overlap, it eliminates one source of error that affects gap coding. In addition, this method eliminates some ambiguity by using a statistical test, rather than a proxy for a statistical test. Taxa in mutually exclusive subsets can be assigned different codes because the mean for each taxon in one set *is* significantly different from the mean of each taxon in the other set. Equally obvious, homogeneous subsets that intersect (share taxa) cannot be assigned different states because the means of the shared taxa cannot be distinguished from the means of any taxa in either set.

The improvements incorporated in homogeneous subset coding are important, but the flaw it retains is more important. Like gap coding, homogeneous subset coding uses a minimal description of the variability within each taxon: the mean and standard deviation. Consequently, both methods of coding are prone to errors when the observed distribution within a taxon departs from the expected normal distribution. This is not a trivial or purely formal objection. Several factors may account for deviation from normality, and many of them are commonly encountered in systematic studies (e.g., allometry, geographic variation, sexual dimorphism, biased collecting methods). Additional sources of biased distributions may be encountered when multiple species are combined into higher taxa (e.g., in studies of evolutionary trends or differential extinction). Given these common sampling problems, it is crucial that a method of coding uses as much information as possible about the distribution of individuals within each taxon.

Almeida and Bisby (1984) also recognized that coding should be based on more information than a comparison of means and standard deviations. They used box plots to show the entire range of each species divided into quartiles. Figure 3a shows box plots for five hypothetical taxa similar to those in Figure 1. Almeida and Bisby used the box plots to find regions where there was no overlap (Fig. 3a, zone a) and regions where only the outer quartiles of the taxa overlap (Fig. 3a, zone b). These regions delimit the sets of taxa that can be assigned the same character state code for that trait.

Almeida and Bisby's use of quartiles is an improvement over the other methods because it conveys some information about deviations from normality. However, it produces a coarse-grained analysis, in which taxa that overlap as much as 25% can be assigned different character states. Almeida and Bisby were uncomfortable with allowing this much overlap (p. 408), as are we. This problem could be remedied by using a different cutoff (e.g., outer 5th percentiles), but a more important problem would remain: the lack of any a priori justification for applying a fixed standard to all comparisons. Unfortunately, the use of a fixed standard is probably unavoidable when the distribution

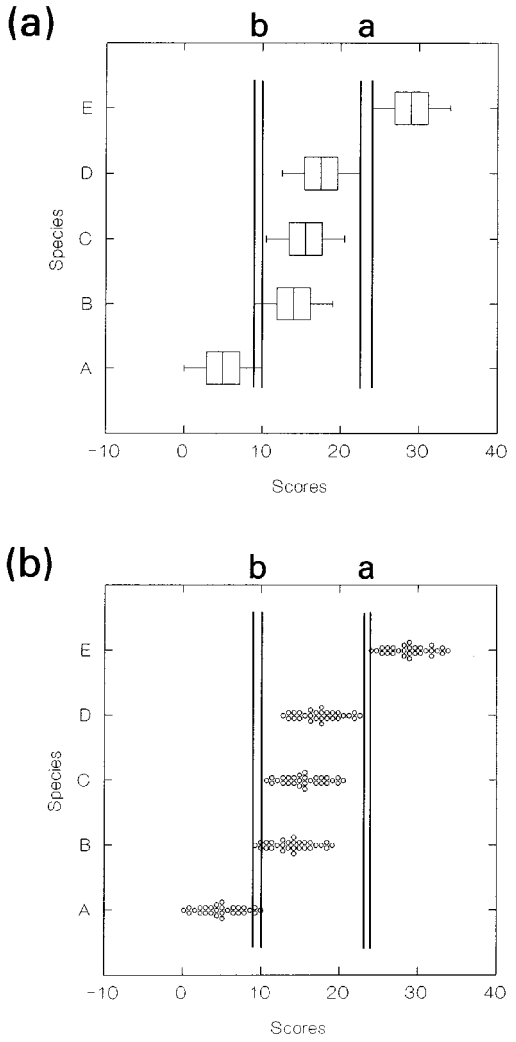


FIGURE 3. Coding based on overlap of taxon ranges. (a) Box plots dividing ranges into quartiles. (b) Dot plots of individual scores. The paired lines denote: a) the gap between taxa D and E, b) the overlap between taxa A and B.

is represented by a range bar divided into categories, as done by Almeida and Bisby. Such graphs omit the number and distribution of individuals in each sample, forcing systematists to base coding decisions on the number and size of the overlapping categories. Without a rule for these decisions, systematists are likely to base coding on subjective impressions of the pattern of overlap among all taxa, which is even less justified (Gift and Stevens, 1997).

To improve Almeida and Bisby's method, we suggest using dot plots (Fig. 3b), which are essentially symmetrical histograms with large numbers of small intervals. These diagrams illustrate the spread of individuals rather than the clustering shown by conventional histograms. Dot plots produce finer-scale descriptions, which allow coding decisions to be based on analysis of the individuals in the study, not on the numbers of individuals within coarse classes.

Almeida and Bisby's approach, with our modification, puts the coding of quantitative data on the same footing as traditional analysis of qualitative data: The diversity within each group is evaluated and then is compared with the membership of other groups to see if there is overlap. Then, the overlap is evaluated to determine whether a hypothesis of evolutionary transformation is justified. Our preference is to determine whether the density of individuals decreases near the edge of the observed range of a taxon, which would be an indication that the edge of the observed range is close to the edge of the actual range. If overlap involves only individuals from these fringes, then we would recognize different states. Other systematists may prefer different criteria; one advantage of the dot plots is that readers can apply their own criteria to the same data.

One problem that is not solved by using dot plots is the one caused by intermediate taxa overlapping the ranges of taxa that do not overlap each other. In Figure 3b, the ranges of taxa A and C do not overlap, but the range of taxon B overlaps both. This problem cannot be solved by any method that consistently applies a criterion for recognizing differentiation. If B cannot be distinguished from either A or C, the character should be considered phylogenetically uninformative for those taxa.

Some systematists may prefer a method of coding that incorporates a rigid, automatic criterion for recognizing different states. We have not proposed any such rules, because none can be realistically applied to all cases. Several of the methods discussed above represent attempts to employ rules; their failures demonstrate that the rules do not apply universally. We see no reason to obey rules to code quantitatively described traits when we would not obey those rules to code the same traits if they

were qualitatively described. Instead, we suggest that the coding of each trait should be decided on its own merits, by examining the distribution of individuals in each taxon.

Analysis of Piranhas

Methods

Below, we illustrate overlap coding of data from real populations rather than the hypothetical constructions used above. Most of these data come from analyses of the ontogeny of piranha shape (Fink and Zelditch, 1995; Zelditch and Fink, 1995). Descriptions of the morphometric methods, including a selection of landmarks, are presented in those papers (see also, Bookstein, 1989, 1991; Zelditch et al., 1992; Swiderski, 1993). Because our purpose is to demonstrate the coding method, not the morphometric methods, here we present only a brief description of the morphometric methods, highlighting departures from previous studies or details that are particularly relevant to coding.

Fink and Zelditch (1995) analyzed ontogenetic shape change in five species: *Pygopristis denticulata*, *Serrasalmus gouldingi*, *Pygocentrus cariba*, *Pygocentrus nattereri*, and *Pygocentrus piraya*. In this study, we use the adults from that study, and add new data by including the adults of a sixth species, *S. elongatus*. We define adults as specimens with centroid size > 100 (corresponding to a standard length > 75 mm, which is approximately the size at which the juvenile phase of growth ends (centroid size is defined by Slice et al., 1996). We restricted this study to adults because many studies are limited to adults, given the difficulties of obtaining juveniles, and because the description of ontogenies adds several problems that are beyond the scope of this demonstration (cf., Zelditch et al., 1992; Mabee and Humphries, 1993; Fink and Zelditch, 1995).

Shape was described by using the thin-plate spline analysis, which can be implemented with either of the following programs: F. J. Rohlfs TPSPLINE or J. M. Humphries JSPLINE (both are available at http://life.bio.sunysb.edu/morph/).

Each adult in this study was compared to the same starting form, an average juvenile of the outgroup, *Pygopristis denticulata*. Even in studies of allometry, comparison of adults to the juvenile of an outgroup is unusual, but we do it here because the starting form defines the variables used in the morphometric analysis (principal warps). By using the same starting form that was used in the ontogenetic studies, we insure that our descriptions of adult shape can be compared to the descriptions of shape ontogenies.

Principal warps differ from conventional measurements in many ways (see Bookstein 1991; Zelditch et al., 1992, 1995; Swiderski, 1993), but one difference that is particularly relevant here is that principal warps are two-dimensional variables. The observed values, called *partial* warps, reflect not only the magnitude of shape change, but also its direction with respect to the organism. Partial warps commonly are reported as x, y coordinate pairs, representing amounts of change in two directions of an orthogonal grid system. We aligned the starting form so that x is the anteroposterior axis and y is the dorsoventral axis.

The results of the spline analysis, the partial warps scores, are presented in two formats for coding. The first format is one-dimensional, and describes the anteroposterior and dorsoventral components separately. In this format, the distributions of individuals are displayed by dot plots, as suggested above. The second format is two-dimensional, in which the dot plots are replaced with scatter plots showing the distributions of the anteroposterior (x) and dorsoventral (y) components jointly. (Other methods of coding can also be adapted for use with two-dimensional data, by computing a set of ellipses or computing an appropriate multivariate test statistic. The logic of our argument favoring overlap coding was not contrived so as to favor the only method that could be applied to two-dimensional data.)

Results

Figure 4a shows the pattern of landmark displacements for the largest-scale principal warp of the starting form. In this pattern, landmarks near the middle move in one direction, and landmarks near the ends move in the opposite direction. This pattern is illustrated with an arbitrarily chosen $+y$ multiplier to show the proportions of relative displacements that this

component would represent. This pattern of landmark displacement can be described verbally as a change in dorsal convexity. Figure 4b shows the observed scores for this dorsoventral component of shape change for each individual in this study. A score of zero indicates that, in this component, the specimen is not different from the average juvenile *Pygopristis denticulata*. Figure 4b indicates that the ranges for these six species are very similar. Based on their broad overlaps, we infer that there has been no differentiation of this feature in the dorsoventral direction.

The scores in Figure 4b represent only one component of the changes described by the pattern in Figure 4a. The same pattern oriented in the anteroposterior direction (all arrows rotated 90° clockwise) represents a graded pattern of change in which one end of the body is expanded relative to the other end. Positive scores represent anterior elongation; negative scores, posterior elongation. Figure 4c shows the distribution of scores for this feature. There is a broad overlap between *Pygocentrus* and *Serrasalmus*. Line A marks the edge of the *Pygocentrus* range, and a third or more of each *Serrasalmus* species is on the left of this line. *Serrasalmus* also overlaps *P. denticulata*: Two specimens of *S. elongatus* are on the right of line B and two *Pygopristis denticulata* are on the left of the line. This distribution could be interpreted as indicating two evolutionary transitions, one across each line, but neither line unambiguously demarcates two sets of taxa. In both cases, the source of ambiguity is uncertainty about the limits of the *Serrasalmus* species. For example, at line B the density of *Pygopristis denticulata* specimens drop abruptly, and the two individuals to the left of the line can be reasonably interpreted as lying on the fringe of the distribution. In contrast, the two *S. elongatus* on the right of the line cannot be interpreted as lying on the fringe of their distribution, because the sample size for *S. elongatus* (13), is too small to reliably infer the distribution of within-group variation. The two *S. elongatus* on the right of the line may only appear to be unusual because variation within that species is not adequately described. Consequently, we do not feel that these distributions justify an inference of separate character states for this shape feature.

Figure 4d shows the two-dimensional distribution of shape changes described by the pattern in Figure 4a. A specimen with coordinates (0, 0) would not be different from the average juvenile *Pygopristis denticulata* in this feature in any direction. Coordinates of $(+x, +y)$ would indicate that the specimen differs from the juvenile *Pygopristis denticulata* in both greater anterior elongation and greater dorsal convexity. As in the previous plots, the two-dimensional plot shows that this component of shape varies almost entirely in the anteroposterior direction, and that the species ranges overlap too broadly to recognize separate states. The ranges of the three *Pygocentrus* species are nearly identical. Most specimens of the two *Serrasalmus* species, and all specimens of *Pygopristis denticulata* lie outside of the *Pygocentrus* range, but several specimens of both *Serrasalmus* species lie within the *Pygocentrus* range. More importantly, some *Serrasalmus* are found near dense clusters of *Pygocentrus* individuals, well beyond the edge of the *Pygocentrus* range. Few *Serrasalmus* are found near clusters of *Pygopristis denticulata*, but the widely scattered *S. elongatus* surround much of the *P. denticulata* range. The relatively sparse distribution of *S. elongatus* suggests that additional samples should be expected to have individuals that fall within the *P. denticulata* range. Consequently, we conclude that *S. elongatus* bridges the gap between *Pygopristis* and *Pygocentrus*, and that this shape feature is not phylogenetically informative.

Figure 5a shows the pattern of landmark displacement for a small-scale feature localized to part of the head. The dorsoventral scores for this feature (Fig. 5b) show broad overlaps for all species. The anteroposterior scores (Fig. 5c) indicate that there is some differentiation between *Pygopristis denticulata* and the other species along this axis (relative length of the snout and jaws). The two-dimensional plot (Fig. 5d) provides better evidence of a shape change. All *Pygopristis denticulata* but one are on the left of the line, whereas all *Serrasalmus* and all *Pygocentrus* but one are on the right of the line. In addition, a sparsely populated space lies to the right of this line. Based on the space between the two groups and the small number of specimens that have crossed that space, we recognize two states for this shape feature (one unique to *Pygopristis denticulata*).

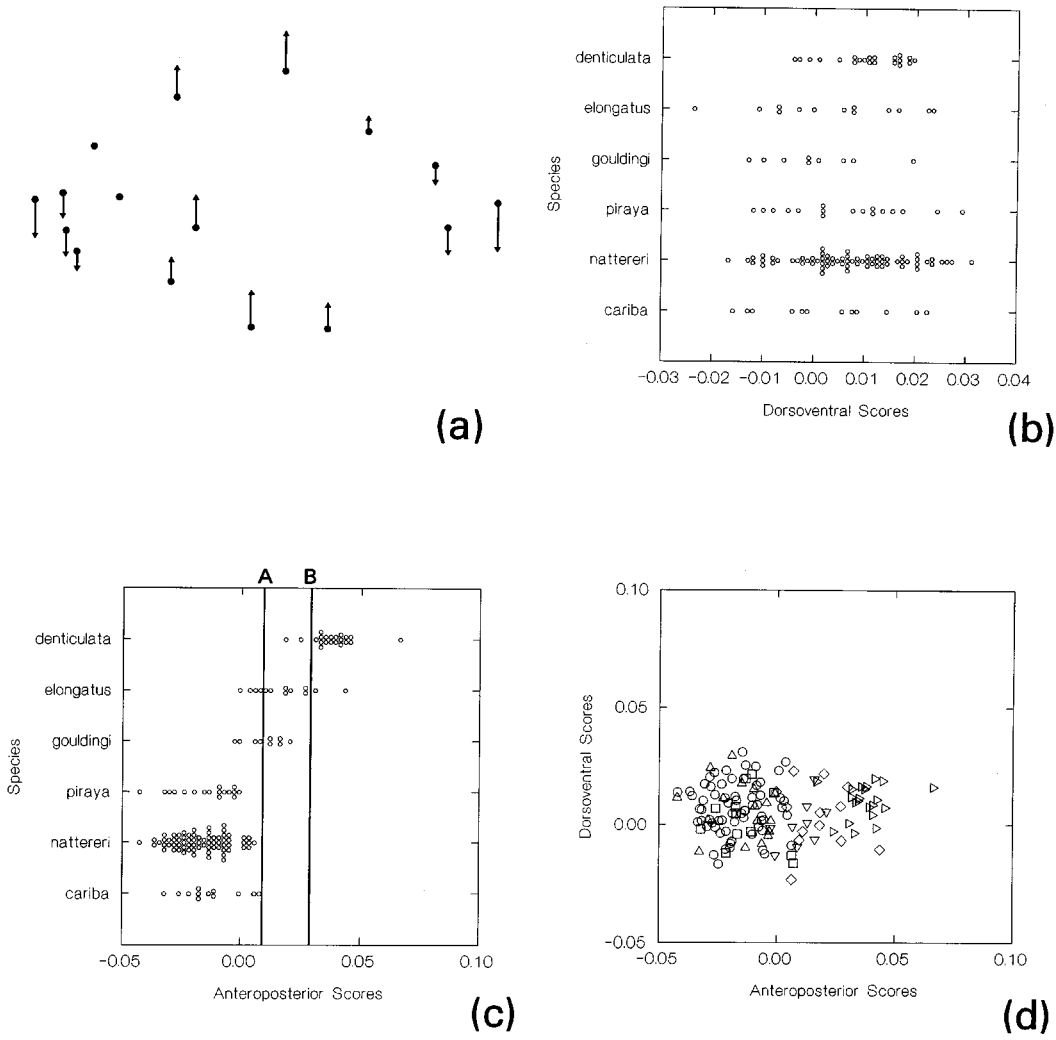


FIGURE 4. Variability of a large-scale principal warp. ▷ = *Pygopristis denticulata*, ◇ = *S. elongatus*, ▽ = *S. gouldingi*, ○ = *Pygocentrus. nattereri*, △ = *Pygocentrus piraya*, □ = *Pygocentrus cariba*. (a) Pattern of landmark displacement. (b) Partial warp scores for displacement along the dorsoventral axis. (c) Partial warp scores for displacement along the anteroposterior axis. A marks one edge of the *Pygocentrus* range; B separates most *Pygopristis denticulata* from most *Serrasalmus*. (d) Bivariate plot of anteroposterior and dorsoventral partial warp scores.

These two states could not be recognized in the one-dimensional plot because the direction of change is not aligned with the anatomical axes.

Figure 6a illustrates a small-scale feature that describes changes in the region extending from the base of the dorsal fin through the caudal peduncle at the base of the tail fin. The one-dimensional plots for this feature have been omitted; the two-dimensional plot (Fig. 6b) again indicates a change that is not aligned with the anatomical axes (across line B). However, the main reason we show this feature is because it appears to have transformations in two different directions. The ranges of *Serrasalmus* and *Pygocentrus* are completely disjunct. Both overlap the range of *Pygopristis denticulata*, but from different sides, and neither overlap is enough to prevent recognition of distinct states. Based on these distributions, we infer two independent character changes.

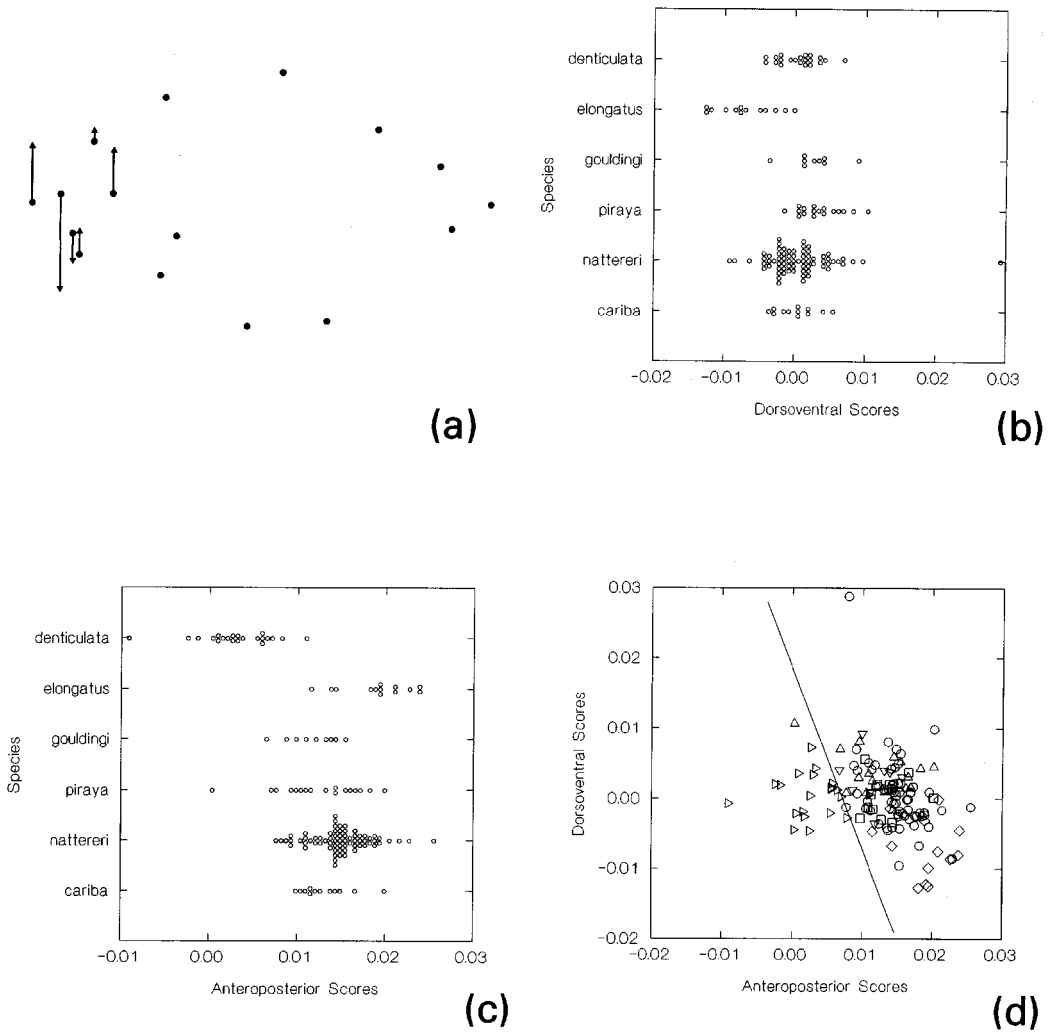


FIGURE 5. Variability of a small-scale principal warp. (a) Pattern of landmark displacement. (b) Partial warp scores for displacement along the dorsoventral axis. (c) Partial warp scores for displacement along the anteroposterior axis. (d) Bivariate plot of anteroposterior and dorsoventral partial warp scores. The line delimits the sets of taxa inferred to have different character states. Symbols as in figure 4.

The anteroposterior transformation of *Serrasalmus* is primarily a relative elongation of the region between the dorsal fin and the caudal peduncle. In contrast, the dorsoventral transformation of *Pygocentrus* is primarily a relative thickening of the caudal peduncle.

There are several other features we could show, but these three are sufficient to demonstrate the approach we advocate, for both one-dimensional and two-dimensional characters. Our preliminary analysis of the distribution of 14 shape features indicates there may be more than 10 shape transformations among these six species. Six of the changes are in features that underwent two changes. Some of the changes may be autapomorphies, but at least half are potentially informative for resolving phylogenetic relationships.

SUMMARY

Thorpe (1984) and Chappill (1989) argued that selection of a coding method should be

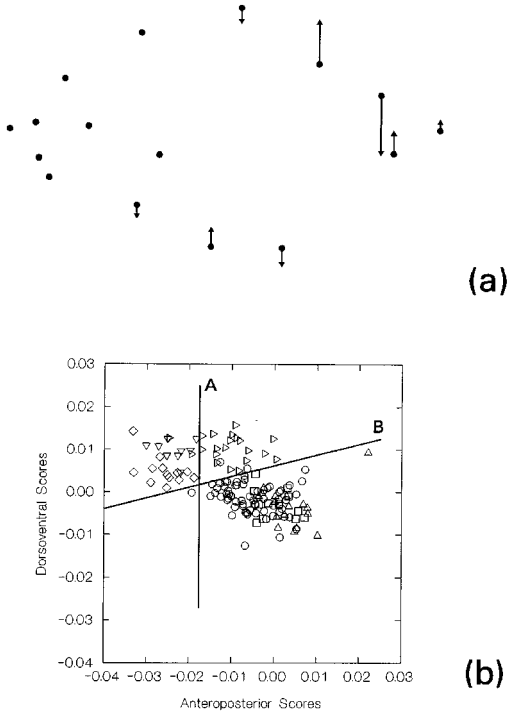


FIGURE 6. Variability of another small-scale principal warp. (a) Pattern of landmark displacement. (b) Bivariate plot of anteroposterior and dorsoventral partial warp scores. Lines A and B delimit sets of taxa inferred to have different character states. Symbols as in Figure 4.

based on the purpose of coding. In our view, the purpose of coding is dictated by the principles of phylogenetic systematics. The foundation of phylogenetic systematics is the observation that monophyletic groups can be recognized if homologous character states, shared evolutionary novelties, can be identified (Hennig, 1966). Unfortunately, characters do not have labels indicating their homology. Instead, a systematist must propose a hypothesis of homology and evaluate its congruence with independent hypotheses based on other traits. In this context, the purpose of coding is to represent those hypotheses.

The major obstacle to coding is that the a priori groups under analysis (i.e., taxa) often have ranges of variation that overlap to some degree. This is true whether traits are described qualitatively or quantitatively. One advantage of quantitative description is that it permits a more detailed analysis of how much the ranges of variation overlap. It may seem appropriate to use statistical methods to summarize the amount of overlap and even to decide objectively (on a priori grounds) whether taxa are similar or different. Above, we demonstrated some of the problems resulting from these uses of statistical analysis. In our view, the most important problem is the implication that similarity of the feature across taxa is the basis for inferring homology. The similarity that is relevant to phylogenetic analysis is not proximity in morphospace, but shared novelty. Statistical methods can describe proximity, but they cannot recognize novelty.

The method of coding we recommend uses graphical displays of individual values. Coding decisions are based on all of the individuals in each taxon, not on summaries derived from models of expected distributions. Then, the evidence for inferring divergence is independently evaluated for each pair of overlapping taxa. Coding decisions are not based on a priori rules that have no bearing on recognition of evolutionary novelty. This is the same approach that is used to code qualitatively described traits.

ACKNOWLEDGMENTS

We thank Sharon Jansa, Barb Lundrigan, and two anonymous reviewers for comments on previous drafts of this paper. Piranha studies were supported by NSF DEB-9509195 to (M.L.Z. and W.L.F.).

REFERENCES

ALMEIDA, M. T., AND F. A. BISBY. 1984. A simple method for establishing taxonomic characters from measurement data. Taxon 33:405–409.

ARCHIE, J. W. 1985. Methods for coding variable morphological features for numerical taxonomic analysis. Syst. Zool. 34:326–345.

BOOKSTEIN, F. L. 1989. Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Trans. Pattern Anal. Mach. Intell. 11:567–585.

BOOKSTEIN, F. L. 1991. Morphometric tools for landmark data: Geometry and biology. Cambridge Univ. Press, New York.

CHAPPILL, J. A. 1989. Quantitative characters in phylogenetic analysis. Cladistics 5:217–234.

COLLESS, D. H. 1980. Congruence between morphological and allozymic data for *Menidia*: A reappraisal. Syst. Zool. 29:288–299.

FARRIS, J. S. 1990. Phenetics in camouflage. Cladistics 6:91–100.

FELSENSTEIN, J. 1988. Phylogenies and quantitative characters. Annu. Rev. Ecol. Syst. 19:445–471.

FINK, W. L., AND M. L. ZELDITCH. 1995. Phylogenetic analysis of ontogenetic shape transformations: A reassessment of the piranha genus *Pygocentrus* (Teleostei). Syst. Biol. 44:343–360.

GARLAND, T., JR., AND S. C. ADOLPH. 1994. Why not to do two species comparative studies: Limitations on inferring adaptation. Physiol. Zool. 67:797–828.

GIFT, N., AND P. F. STEVENS. 1997. Vagaries in the delimitation of character states in quantitative variation— an experimental study. Syst. Biol. 46:112–125.

HENNIG, W. 1966. Phylogenetic systematics. Univ. Illinois Press, Urbana.

MABEE, P. M., AND J. HUMPHRIES. 1993. Coding polymorphic data: Examples from allozymes and ontogeny. Syst. Biol. 42:166–181.

MICKEVICH, M. F., AND M. F. JOHNSON. 1976. Congruence between morphological and allozyme data. Syst. Zool. 25:260–270.

MICKEVICH, M. F., AND S. J. WELLER. 1990. Evolutionary character analysis: Tracing character change on a cladogram. Cladistics 6:137–170.

PIMENTEL, R. A., AND R. RIGGINS. 1987. The nature of cladistic data. Cladistics 3:201–209.

SIMON, C. 1983. A new coding procedure for morphometric data with an example from periodical cicada wing veins. Pages 378–382 *in* Numerical taxonomy. (J. Felsenstein, ed.). NATO ASI Series G, Volume 1. Springer-Verlag, Berlin.

SLICE, D. E., F. L. BOOKSTEIN, L. F. MARCUS, AND F. J. ROHLF. 1996. Appendix I. A glossary for geometric morphometrics. Pages 531–551 *in* Advances in morphometrics. (L. F. Marcus, M. Corti, A. Loy, G. J. P. Naylor, and D. E. Slice, eds.). NATO ASI Series A, Volume 284. Plenum, New York.

STEVENS, P. F. 1991. Character states, morphological variation, and phylogenetic analysis: A review. Syst. Bot. 16:553–583.

SWIDERSKI, D. L. 1993. Morphological evolution of the scapula in tree squirrels, chipmunks and ground squirrels (Sciuridae): An analysis using thin-plate splines. Evolution: 47:1854–1973.

THIELE, K. 1993. The Holy Grail of the perfect character: The cladistic treatment of morphometric data. Cladistics 9:275–304.

THORPE, R. S. 1984. Coding morphometric characters for constructing distance Wagner networks. Evolution 38:244–355.

WILEY, E. O. 1981. Phylogenetics: The theory and practice of phylogenetic systematics. J. Wiley and Sons, New York.

WILKINSON, M. 1995. A comparison of two methods of character construction. Cladistics 11:297–308.

ZELDITCH, M. L., F. L. BOOKSTEIN, AND B. L. LUNDRIGAN. 1992. Ontogeny of integrated skull growth in the cotton rat *Sigmodon fulviventer*. Evolution 46:1164–1180.

ZELDITCH, M. L., AND W. L. FINK. 1995. Allometry and developmental integration of body growth in a piranha, *Pygocentrus nattereri* (Teleostei: Ostariophysi). J. Morphol. 223:341–355.

ZELDITCH, M. L., W. L. FINK, AND D. L. SWIDERSKI. 1995. Morphometrics, homology and phylogenetics: Quantified characters as synapomorphies. Syst. Biol. 44:179–189.

Received: 10 April 1997; accepted 20 December 1997
Associate Editor: D. Cannatella

Syst. Biol. 47(3): 519–537, 1998

Systematic Bias in Phylogenetic Analysis: Is the Strepsiptera Problem Solved?

JOHN P. HUELSENBECK

Laboratory of Molecular Systematics, Smithsonian Museum Support Center, 4210 Silver Hill Road, Suitland, Maryland, 20746, USA; E-mail: johnh@ onyx.si.edu

A phylogenetic method is inconsistent if it converges to an incorrect tree as characters (e.g., the columns or sites in a DNA sequence data matrix) are added to a phylogenetic problem. Inconsistency was first identified as a potential problem in phylogenetics by Felsenstein (1978) who showed that parsimony and compatibility methods could become inconsistent for four taxa under a restricted set of circumstances. However, the inconsistency problem was later shown to occur under less stringent conditions by Hendy and Penny (1989) who demonstrated the inconsistency of parsimony for trees of more than four species when the data obeyed a molecular clock. Other methods of phylogenetic estimation were later shown to be inconsistent under some conditions (DeBry, 1992; Huelsenbeck and Hillis, 1993; Gaut and Lewis, 1995; Huelsenbeck, 1995a; Waddell, 1995; Yang, 1996). For example, when the assumptions of distance and maximum-likelihood methods are

violated, they too can become inconsistent. In this article, I discuss the conditions under which parsimony and maximum likelihood may become inconsistent and ask whether the phylogenetic tree of insects recently produced by Whiting et al. (1997) suffers from systematic bias introduced by long-branch attraction.

Inconsistency of Phylogenetic Methods

For the parsimony method, the inconsistency problem is best understood for the simple case of four taxa (Fig. 1; an example drawn from Swofford et al., 1996). Although inconsistency is a potential problem for many types of character data, I will discuss inconsistency for methods used on DNA sequences. When two opposing peripheral branches of a model four-taxon tree are very long and the remaining branches short ($\nu_1 = \nu_3 \gg \nu_2 = \nu_4 = \nu_5$), the parsimony method may converge to a tree that connects the two long branches (hence, the maxim that long branches attract). Why is this the case? Consider the tree shown in Figure 1. For the tree of Figure 1, the chance that a change occurs along the three small branches (ν_2, ν_4, ν_5) is very small compared to the chance that a change occurs along the two long branches (ν_1, ν_3). In fact, as $\nu_2 = \nu_4 = \nu_5$ becomes very small, we can ignore the event that a change occurs along the small branches, in which case the same nucleotide state will be observed at nodes n_2 and n_4. Under these conditions, a site only changes along the long branches of the tree (ν_1, ν_3). What possible character patterns can we expect at the tips of the long branches (n_1 and n_3)? One possibility is that no changes occur along the long branches, in which case the character pattern that would be observed at the tips of the tree (n_1, n_2, n_3, and n_4) would be *xxxx* (where *x* is the nucleotide assigned to tips n_1, n_2, n_3, and n_4, respectively, and is either A, C, G, or T). Another possibility is that only one change would occur along the long branches, in which case we would observe the character pattern *xyyy* or *xxyx*. Finally, a change could occur along both of the long branches. If the changes are to different nucleotides, then the resulting pattern of nucleotides at the tips of the tree would be *xyzy*. If, on the other hand, the two independent changes are to the same nucleotide, then the pattern of nucleotides at the tips of the tree would be *xyxy*. For parsimony using Fitch (1971) character optimization, only the patterns *xxyy*, *xyxy*, or *xyyx* are informative and distinguish among possible trees. The site patterns that are possible for the case when $\nu_1 = \nu_3 \gg \nu_2 = \nu_4 = \nu_5$ include *xxxx*, *xyyy*, *xxyx*, *xyzy*, and *xyxy*; only the site pattern *xyxy* is informative for the parsimony method, and this site pattern is informative for the incorrect phylogeny: $((n_1, n_3), n_2, n_4)$.

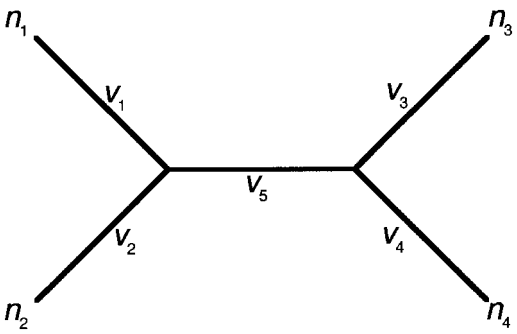


FIGURE 1. A model four-taxon (n_1 to n_4) tree with branch lengths ($\nu_1, \nu_2, \ldots, \nu_5$) specified.

Figure 2 shows the estimates of branch lengths that would be expected if the true tree had two long branches separated by short branches (a four-taxon tree for which parsimony and maximum likelihood assuming a Jukes-Cantor [1969] model are inconsistent; the actual process of substitution generating the sequences is the Jukes-Cantor [1969] model with gamma-distributed rate variation among sites [with shape parameter $\alpha = 0.5$]). The estimated branch lengths leading to species n_1 and n_3 are long for all three possible trees. For the maximum parsimony and maximum likelihood criteria, the best estimate of phylogeny is the one that (incorrectly) places the two long branches together. It has been suggested that when an estimated tree has two taxa connected by long branches, that long-branch attraction may be a problem (i.e., this is a pattern that would be expected if a method were incorrectly grouping long branches; Carmean and Crespi, 1995). Yet, another possible explanation of such a pattern is that the long branches do, in fact, belong together and that there was a single increase

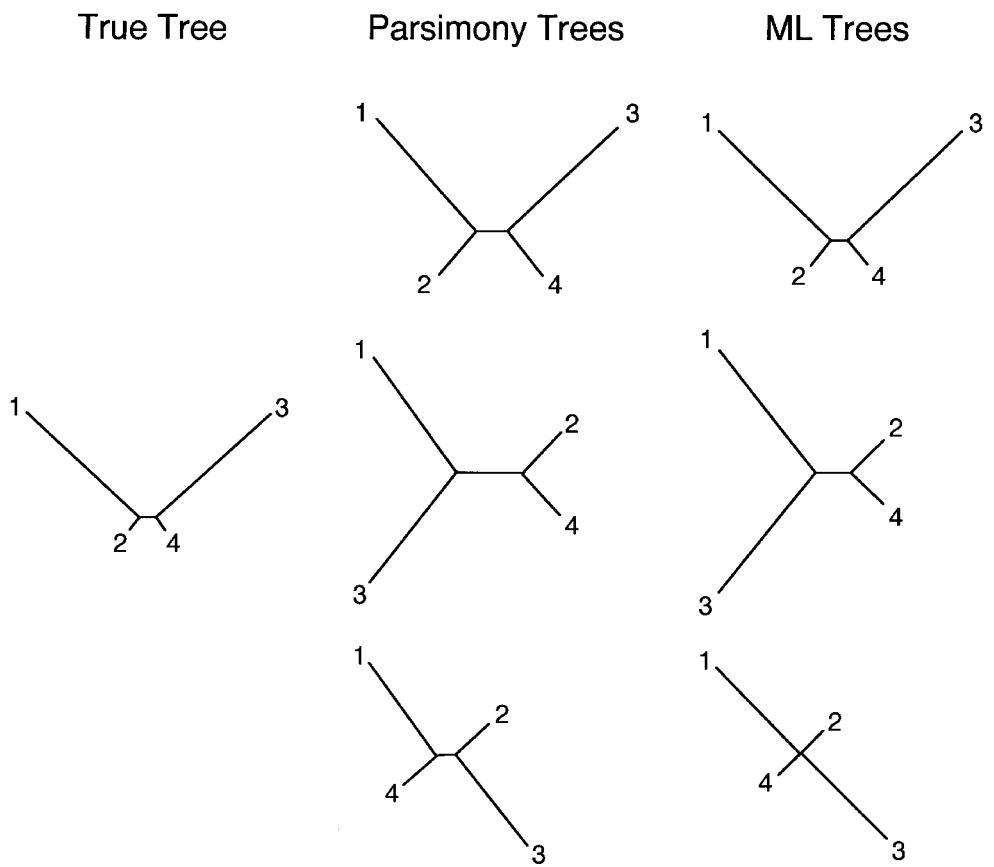


FIGURE 2. When long branches are separated by short internal branches on the true tree, parsimony and maximum likelihood (assuming the wrong model) will converge to the wrong estimate of phylogeny. Here, the actual process of substitution obeys a Jukes-Cantor (1969) model with gamma-distributed rate variation among sites (shape parameter, $\alpha = 0.5$). The two long opposing peripheral branches are 10 times longer than the remaining branches ($R = 10$), and the expected number of substitutions per site over the entire tree is 2 ($S = 2$). Parsimony assumed Fitch (1971) character optimization, and maximum likelihood assumed a Jukes-Cantor model of DNA substitution. Both maximum likelihood and parsimony converge to a tree that incorrectly places the long branches together: $((n_1, n_3), n_2, n_4)$. The estimated branches for both parsimony and maximum likelihood have long branches leading to n_1 and n_3. However, parsimony more severely underestimates the branch lengths.

in the rate of substitution that occurred before the speciation event leading to the long branches. Figure 3 illustrates this point. For the tree of Figure 3 (a four-taxon tree with the long branches grouped together), all three possible trees have long branches leading to species n_1 and n_2. For the tree of Figure 3, both maximum likelihood and parsimony correctly estimate phylogeny. The pattern of branch lengths on the estimated trees are very similar when the true phylogeny has the long branches together or separate.

How well can parsimony and maximum likelihood distinguish between the situations illustrated in Figures 2 and 3? The conditions under which phylogenetic methods become inconsistent have been extensively explored for the four-taxon case with the restriction that two of the opposing peripheral branches are equal in length ($\nu_1 = \nu_3$) and that the remaining three branches are equal in length ($\nu_2 = \nu_4 = \nu_5$) (Felsenstein, 1978; Jin and Nei, 1990; Nei, 1991; Huelsenbeck and Hillis, 1993; Tateno et al., 1994; Gaut and Lewis, 1995; Huelsenbeck, 1995a, 1995b; Rzhetsky and Sitnikova, 1995). However, only a few

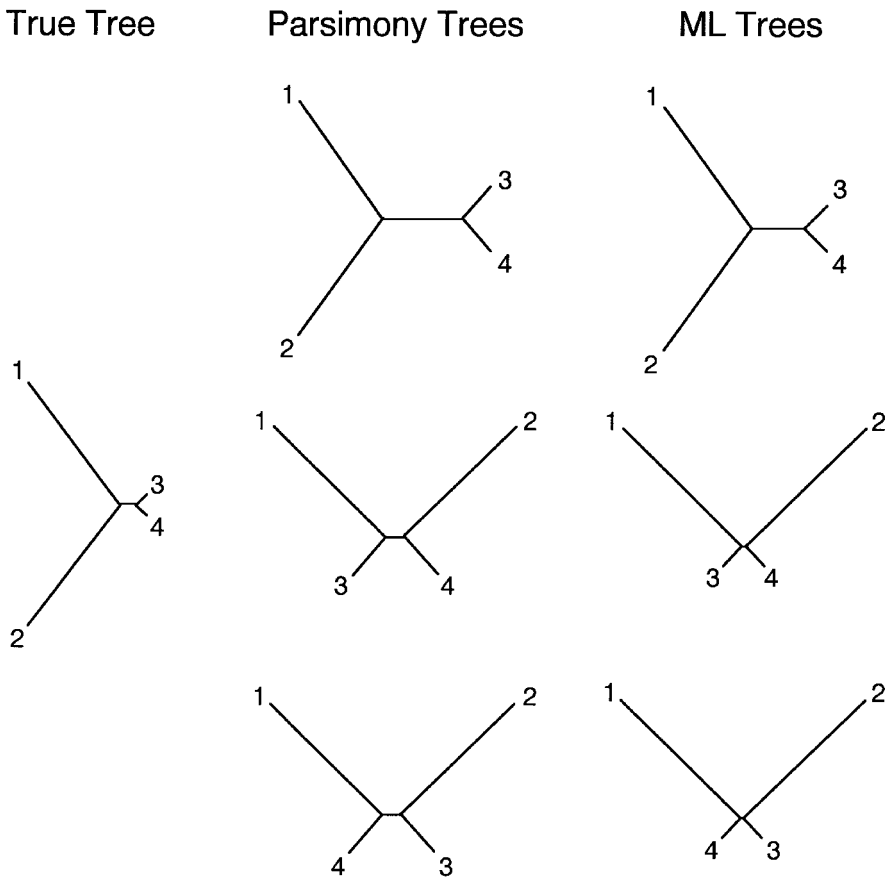


FIGURE 3. When the long branches are adjacent on the true tree both parsimony and maximum likelihood are consistent. The process of DNA substitution followed a Jukes-Cantor (1969) model with gamma distributed rate variation among sites ($\alpha = 0.5$). Two of the branches were 10 times longer than the other branches ($Q = 10$), and the expected number of substitutions over the entire tree per site was 2 ($S = 2$). Both maximum likelihood and parsimony estimate the true tree with infinite number of sites.

authors have explored the statistical properties of methods when the two long branches do, in fact, belong together (Waddell, 1995; Yang, 1996; D. Swofford, pers. comm.).

To illustrate the results from simulations of four species, I explore two types of model trees. One set of model trees (**R**) has the constraint that two opposing peripheral branches are equal in length and the remaining branches are equal in length ($\nu_1 = \nu_3$; $\nu_2 = \nu_4 = \nu_5$). The overall substitution rate (S; in terms of expected number of substitutions per site over all branches) and the ratio of the long branches to the short branches ($R = \nu_1/\nu_2$) were varied. The other set of model trees (**Q**) has the constraint that two adjacent peripheral branches are equal in length and the remaining branches are equal in length ($\nu_1 = \nu_2$; $v_3 = \nu_4 = \nu_5$). Here, the long branches are closest relatives in the true phylogeny (Fig. 4). Again, the overall substitution rate (S) and the ratio of the long branches to the short branches ($Q = \nu_1/\nu_3$) were varied. The parameter space explored is shown in Figure 4. The parameter space can be visualized as two planes that meet where all branches are equal in length ($R = Q = 1$; $\nu_1 = \nu_2 = \nu_3 = \nu_4 = \nu_5$). The parameter space was explored in the region $0.0 < S \leq 5$ and $1 \leq Q, R \leq 25$.

The results of the simulations for the two parameter spaces (**Q** and **R**) are summarized in Appendix 1. Figure 5 shows the results for

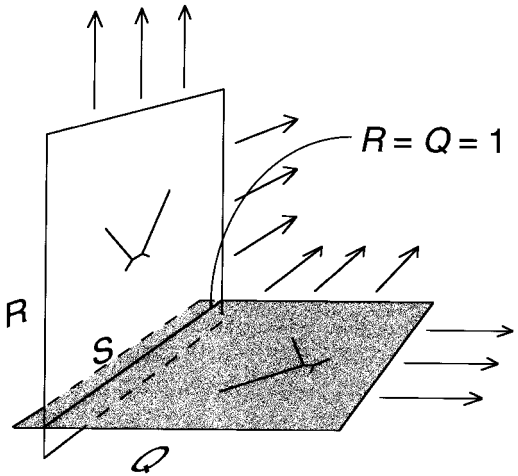


FIGURE 4. Four-taxon simulations were performed that had the long branches together ($Q > 1$) and separate ($R > 1$). The two sets of simulations can be visualized as two planes that meet where $R = Q = 1$ (i.e., all branches equal in length). In this diagram, Q is plotted on the x-axis, R is plotted on the y-axis, and the expected number of substitutions per site over the entire tree (S) is plotted on the z-axis.

the consistency of maximum parsimony and maximum likelihood in the **R** parameter space. The expected number of substitutions per site over the entire tree (S) is plotted along the abscissa and the ratio of long to short branches (R) is plotted along the ordinate. Parsimony was implemented with Fitch (1971) optimization of characters and maximum likelihood assumed a Jukes-Cantor model of DNA substitution. Fitch (1971) character optimization and the Jukes-Cantor substitution model are similar in spirit because they both assume equal weights for different character transformations. However, the Jukes-Cantor model assumes that substitutions follow a Poisson process and takes into account multiple substitutions at the same site. Parsimony implemented with Fitch (1971) character optimization does not account for multiple substitutions at the same site (but parsimony can be made to assume a specific model of DNA substitution; Steel et al., 1993). The assumptions of maximum likelihood are violated in the analyses

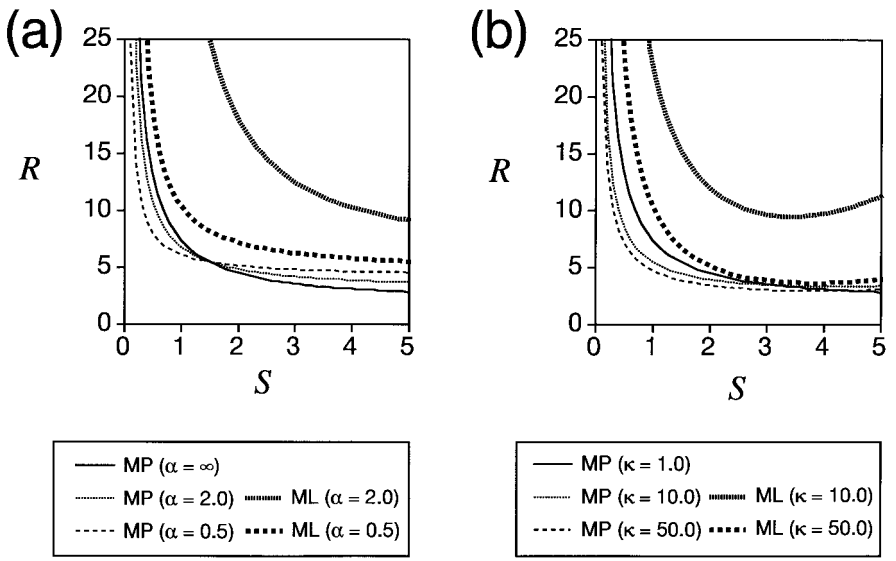


FIGURE 5. The parameter conditions under which maximum parsimony (MP) and maximum likelihood (ML) are inconsistent in the **R** space. A method is inconsistent above its respective line. Maximum parsimony assumed Fitch (1971) optimization of characters, and maximum likelihood assumed a Jukes-Cantor (1969) model of DNA substitution. The process of substitution generating the sequences was varied. (a) The process generating the sequences included gamma-distributed rate variation among sites (with gamma shape parameter, α). (b) Transitions occurred at a higher rate than transversions (with transition/transversion rate parameter, κ). When $\kappa = 1$ and $\alpha = \infty$, the assumptions of maximum likelihood are satisfied.

depicted in Figure 5 either by ignoring gamma-distributed rate variation among sites (Fig. 5a, with gamma shape parameter α; Jin and Nei, 1990; Yang, 1993) or by allowing transitions to occur at a higher rate than transversions (Fig. 5b, with transition/transversion rate parameter κ; Kimura, 1980). In Figure 5, a method is inconsistent above the line. The set of conditions under which a method becomes inconsistent has been termed the Felsenstein zone (Huelsenbeck and Hillis, 1993). Note that both parsimony and maximum likelihood are inconsistent when R becomes large and that the problem is more severe when the substitution rate is high (S is large). Also note that the parameter space for which maximum likelihood is inconsistent is smaller than the set of conditions under which parsimony is inconsistent. Similar results for the relative performance of maximum likelihood and parsimony hold when the efficiency of the methods is examined in the **R** space (Appendix 1) for a limited number of sites; maximum likelihood is generally more efficient when the model tree has two long branches separated by a short internal branch.

Both maximum likelihood and parsimony are consistent over the set of conditions examined here for the **Q** space. However, parsimony is much more efficient than maximum likelihood when adjacent branches are long. This result is similar to those of Yang (1996) and Waddell (1995). An intuitive explanation for this result is that maximum likelihood is less biased toward any one tree because the difference between trees that have long branches separate versus adjacent is very small (Fig. 6); to move from tree A to tree B in Figure 6, one only has to move the branches a small distance (it is necessary to contract and expand the small internal branch to move from tree A to tree B; Kuhner and Felsenstein, 1994). Parsimony, in effect, counts any parallel changes to the same state that occur along the long branches as synapomorphies (i.e., assumes that they changed along the branch leading to the two long branches). Also, the performance of maximum likelihood increases when the model of DNA substitution fits poorly (Yang, 1996).

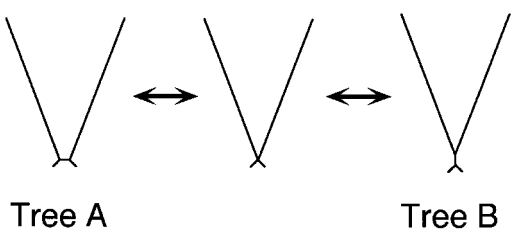


FIGURE 6. Trees with long branches separate (tree A) or together (tree B) are very similar. To move from tree A to tree B, one only has to contract and expand the small internal branch (moving through the intermediate star phylogeny).

The simulations in the **R** and **Q** spaces suggest that maximum likelihood is better able to distinguish between trees with the long branches separate versus adjacent. However, it may take very long sequences before the two types of trees can be statistically distinguished from each other. Figure 7 shows the results of simulations showing the power of the Kishino-Hasegawa (1989) test for trees with long branches separate and together. The Kishino-Hasegawa test uses the asymptotic variance of the likelihoods of different trees to test for significance between the trees. This test is implemented in the programs DNAML (Felsenstein, 1995) and PAUP* 4.0 (provided by David Swofford). The three possible trees are labelled tree 1 $[((n_1, n_2), n_3, n_4)]$, tree 2 $[((n_1, n_3), n_2, n_4)]$, and tree 3 $[((n_1, n_4), n_2, n_3)]$; tree 1 was used to generate the sequences. When the branch lengths are equal in length ($R = Q = 1$), the Kishino-Hasegawa test easily distinguishes the true tree from the other possible trees (tree 1 from tree 2 and tree 1 from tree 3); only a few hundred sites are needed before the P value drops below the 5% level for a comparison of trees 1 and 2 or trees 1 and 3. However, when there are long branches on the tree, thousands of sites are needed before tree 1 becomes significantly different from trees 2 and 3. Although maximum likelihood can correctly discriminate between phylogenies that have the long branches together and separate, the method may require thousands of sites before any confidence can be placed in the estimate.

Although the common view is that long branches are a problem only when, in reality, the long branches occur in disparate parts of a phylogenetic tree, long branches are also a problem when they are adjacent on the true tree. Regardless of whether the long branches are separate or together on the true tree, the maximum parsimony method is more likely to place

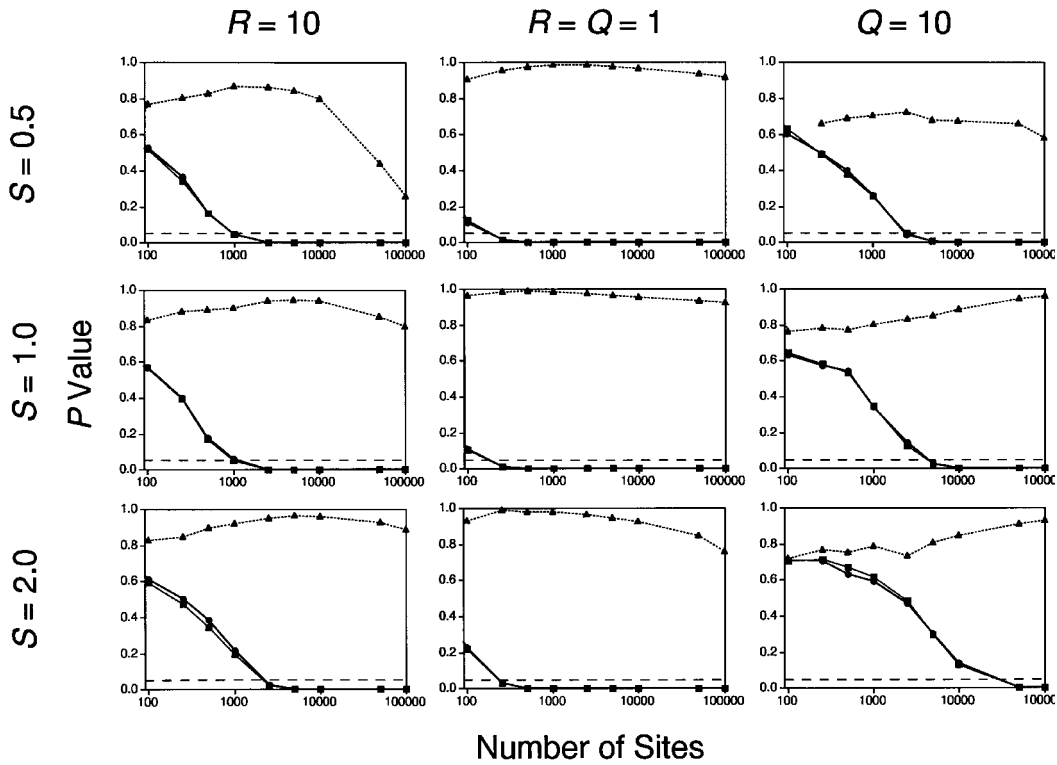


FIGURE 7. Graphs showing the power of the Kishino-Hasegawa (1989) test for distinguishing between the three possible trees for four species. On each graph the length of the sequences is plotted along the abscissa and the average *P* value for 100 simulations is plotted along the ordinate. When the *P* value drops below 5% (the dashed lines) the test is considered significant. Trees 1 and 2 place the long branches together or separate, respectively. Three tests were performed: between trees 1 and 2 (●), between trees 1 and 3 (■), and between trees 2 and 3 (▲). Trees 1 and 2 and trees 1 and 3 are easily distinguished when R= $Q = 1$. However, many more sites are needed to statistically distinguish between trees 1 and 2 when $R = 10$ or $Q = 10$. *S* is the expected number of substitutions over the entire tree. The Jukes-Cantor (1969) model of DNA substitution was used to generate and analyze the sequences.

the long branches together on the estimated tree. Maximum likelihood, on the other hand, is better able to distinguish between the two cases. If the long branches are separated on the true tree (as with the simulations in the **R** space), maximum likelihood has a better chance than parsimony of correctly estimating the phylogeny even if the assumptions of the analysis are violated. The performance advantage of maximum likelihood in the **R** space apparently comes with a cost in the **Q** space where its performance is lower than that of parsimony. However, it appears that maximum likelihood more faithfully depicts the relative support of trees with long branches separated or together. As noted above, the move from a tree with the long branches separate to a tree with the long branches together involves contracting and expanding the branches a small amount.

How can the systematist avoid long-branch attraction? One possible solution to the problem is to sample more species from within the clade of interest. Hillis (1996), for example, found that parsimony and minimum evolution were able to accurately estimate the phylogeny of a large number (228) of species with surprisingly few sites. The hope is that by sampling more species, the long branches are subdivided and the problem is converted from one that has several long branches to one that has no long branches and for which phylogenetic analysis using parsimony (or other phylogenetic methods) is not problematic. Another possible resolution of the problem is to use phylogenetic

methods that are less sensitive to the problem of long-branch attraction. As shown in Figure 5, by using a method of phylogeny estimation that corrects for the multiple substitutions that occur along long branches, the problem of long-branch attraction can be reduced. The problem of long-branch attraction can be reduced even if the correction for multiple substitutions is imperfect, suggesting that some correction is better than none at all.

Is Long-Branch Attraction a Problem for Real Data?

Although long-branch attraction has been recognized as a potential problem in phylogenetics for almost 20 years, few empirical demonstrations of the phenomenon exist. One possible explanation for the paucity of empirical examples is that for real phylogenetic problems, systematic bias is rarely a problem. Another possible explanation concerns the difficulty of detecting phylogenetic problems for which method inconsistency misleads the systematist. Detecting phylogenetic problems that are in the Felsenstein zone is a lot like detecting black holes; such phenomena can only be inferred indirectly.

Several criteria have been used to detect cases of possible long-branch attraction. One criterion involves looking for long branches that are placed together in a phylogenetic tree. If two long branches are adjacent on an estimated tree, then long branch attraction may be suspected. Another criterion involves asking whether the long branches are long enough to attract in a parsimony analysis. Huelsenbeck et al. (1996; also see Huelsenbeck, 1997) suggested a way of addressing this question using Monte Carlo simulation. The method involves estimating parameters of a parameter-rich model of DNA substitution on a tree that has the long branches separated. Many data sets are simulated on this tree and analyzed using parsimony (or any other method). If the long branches are placed together a high proportion of the time in the parsimony analyses, then the branches are long enough to attract under the assumed substitution model. Finally, long-branch attraction may be suspected if a method that is less sensitive to long branches provides an estimate of phylogeny that places the long branches in separate parts of the tree.

Long-Branch Attraction in Amniotes

One possible case of long-branch attraction has been identified for amniotes (Hedges, 1994). For 18S rDNA sequences collected from four species, parsimony, distance, and maximum likelihood methods estimate a phylogeny that has long branches leading to birds and to mammals (Hedges, 1994; Huelsenbeck and Bull, 1996; Huelsenbeck et al., 1996). The best estimate of phylogeny places the birds and mammals together despite paleontological (Gauthier et al., 1988), morphological (Gauthier et al., 1988), and molecular data (Hedges, 1994) that suggest that the closest relatives of birds are crocodylians. Monte Carlo simulation of a subset of the amniote species shows that the long branches are long enough to attract in a parsimony analysis (Huelsenbeck et al., 1996).

Long-Branch Attraction in Rodents

Several recent analyses have challenged the monophyly of Rodentia by arguing that the guinea pig is not a rodent (Graur et al., 1991; Li et al., 1992; Ma et al., 1993; D'Erchia et al., 1996). The analysis of D'Erchia et al. was particularly persuasive because it was based on very long sequences (complete mitochondrial DNA sequences from 16 species). D'Erchia et al.'s original analysis of these data using parsimony, distance, and maximum likelihood methods produced an estimated phylogeny for which Rodentia is not monophyletic. Sullivan and Swofford (1997), however, argued that the mitochondrial DNA results do not provide conclusive evidence against rodent monophyly because D'Erchia et al. assumed substitution models that did not account for among-site rate variation (their analyses assumed that all sites evolve at the same rate). When Sullivan and Swofford analyzed the mitochondrial data using maximum likelihood implemented with a substitution model that allows for among-site rate variation, the ingroup topology of the estimated tree was consistent with rodent monophyly and they found very little difference in the likelihoods for the best tree with rodents monophyletic versus the best tree with rodents nonmonophyletic. Furthermore, in the Sullivan

and Swofford analysis the branches leading to the opossum and hedgehog were very long and placed in separate parts of the tree (whereas with the D'Erchia analysis, the two long branches were placed together on the tree). Finally, when they simulated sequence data sets on the best tree for which rodents were monophyletic, parsimony estimated trees with rodents nonmonophyletic a high proportion (68%) of the time.

Long-Branch Attraction in Insects

Another potential instance of long-branch attraction was identified by Carmean and Crespi (1995). They noted that in an analysis of 18S rDNA sequence data from insects, that the branches leading to the flies (Diptera) and to Strepsiptera (a small group of parasitic insects; Kathirithamby, 1989) were long and placed as sister taxa on the maximum parsimony tree. This estimate of phylogeny was incongruent with some hypotheses of insect phylogeny based on morphology that place the Strepsiptera with the beetles (Coleoptera) (Kristensen, 1991). They argued that this pattern is suggestive of long-branch attraction.

I applied the criteria for detecting possible branch attraction, outlined above, to a data set consisting of 13 rDNA sequences from insects collected by Carmean and Crespi (1995) (Huelsenbeck, 1997). I extended the analysis of Carmean and Crespi by asking whether the branches leading to Strepsiptera and Diptera were long enough to attract in a parsimony analysis and by analyzing the data using maximum likelihood and neighbor joining, methods that are less sensitive to long-branch attraction (Huelsenbeck, 1997). The Monte Carlo simulation study showed that the branches leading to Strepsiptera and Diptera were long enough to attract and reanalysis of the data using maximum likelihood provided an estimated tree that placed the long branches on separate parts of the tree (the Strepsiptera were placed with the Coleoptera with low nonparametric bootstrap support). These results were suggestive of long-branch attraction misleading parsimony.

Whiting et al. (1997) collected 18S and 28S rDNA sequences and scored morphological features for 85 insects. They analyzed these data using the parsimony criterion (and no other methods) and the most parsimonious trees placed Strepsiptera and Diptera as sister taxa. They argued that long-branch attraction was not a problem for several reasons (Whiting et al., 1997): (1) the phylogenies among all three data sets were congruent, (2) the branches leading to Strepsiptera and Diptera appeared about as long as the branches leading to other insect groups, (3) the greatest sequence difference was not between Strepsiptera and Diptera (using raw sequence similarity and no corrections for multiple substitutions), and (4) their analysis had more thorough taxon sampling. In the following section, I reanalyze the sequence data collected by Whiting et al. and will show that the problem of long-branch attraction may exist for both the 18S and 28S rDNA sequences despite the more thorough taxon sampling. I maintain that the support for Strepsiptera + Diptera monophyly is over-stated.

A Reanalysis of the Whiting et al. DNA Data

Whiting et al. (1997) analyzed approximately 1.0 kb of 18S rDNA sequence from 85 insect taxa and approximately 0.4 kb of 28S rDNA sequence from 51 insect taxa. The 18S rDNA sequences used by Whiting et al. (1997) overlapped the 18S rDNA sequences used by Carmean et al. (1992) by about 0.75 kb. Primer sites were excluded from the analysis by Whiting et al. and are also excluded here. Whiting et al. aligned the sequences using the program MALIGN (version 1.93; Wheeler and Gladstein, 1994). This program implements the strategy of Sankoff et al. (1973) by finding the alignment that minimizes the overall number of insertion, deletion, and substitution events as reconstructed by the parsimony criterion. Simultaneous sequence alignment and tree estimation can, in principle, also be performed using maximum likelihood (Bishop and Thompson, 1986; Thorne et al., 1991, 1992). However, maximum likelihood alignment of sequences is not feasible for large numbers of species at this time and the alignment of Whiting et al. was used even though this alignment may be sensitive to the same conditions that can cause

the maximum parsimony criterion to fail (e.g., long-branch attraction). The aligned sequences have 24 and 14 insertions for the 18S and 28S rDNA sequences, respectively (Whiting et al., 1997). These inserts were coded as missing data in this analysis. Because many of the likelihood analyses performed in this study are computationally intensive, only the holometabolous insects (a well-supported clade) were included. Three data sets were analyzed: (1) 18S rDNA sequences from 65 holometabolous insects, (2) 28S rDNA sequences from 39 holometabolous insects, and (3) combined 18S and 28S rDNA sequences from 65 holometabolous insects (28S rDNA sequences were not available for all holometabolous insects examined by Whiting et al. and were coded as missing in the combined analysis).

In all likelihood and distance analyses, the Hasegawa-Kishino-Yano (1985) substitution model was assumed (HKY85). This model allows for different base frequencies and for a transition/transversion rate ratio ($0 < \kappa < \infty$). Among-site rate variation was modeled using a discrete gamma distribution (Yang, 1994). The shape parameter of the gamma distribution ($0 < \alpha < \infty$) is related to the variance in the substitution rate (r) among sites by $\mathrm{Var}[r] = 1/\alpha$. Five rate categories were used to approximate the continuous gamma density and the shape parameter, α, was estimated using maximum likelihood. For the parsimony analyses, Fitch (1971) optimization of characters was assumed.

Searching the space of trees for the maximum likelihood tree proved difficult, especially for the complex models of DNA substitution (e.g., for substitution models that include a transition-transversion rate bias and allow for among site rate variation). Heuristic searches were performed as follows: (1) a starting tree was obtained using parsimony or neighbor joining, (2) on this starting tree, κ and α were estimated using maximum likelihood, (3) κ and α were fixed to the maximum likelihood values, (4) the starting tree was perturbed using nearest neighbor interchange (NNI) or taxon bisection and reconnection (TBR) and for each neighboring tree visited, the likelihood was maximized, and (5) the tree was perturbed until a neighboring tree under the perturbation with a better likelihood could not be found. Where possible, steps 2–5 were repeated until a tree on which κ and α were estimated survived a full round of perturbation under NNI and TBR. The approach used here avoids estimating κ and α for each tree visited but, obviously, does not guarantee that a global maximum likelihood tree has been found. For the parsimony analysis, a heuristic search starting from 10 random-addition sequences was used to find minimum length trees. A test version of the program PAUP* 4.0 was used for all analyses.

The maximum likelihood trees under the constraints that Strepsiptera + Diptera are/are not monophyletic are shown in Figure 8 for the 18S, 28S, and combined analyses, respectively. The scores of the best trees for the parsimony and maximum likelihood criteria are summarized in Table 1. Because only unrooted trees are considered, I will use the term Strepsiptera + Diptera monophyly to indicate that Strepsiptera + Diptera form a taxon bipartition on a tree and Strepsiptera + Diptera nonmonophyly to indicate that they do not form a taxon bipartition.

The best estimate of phylogeny under the parsimony criterion is consistent with Strepsiptera + Diptera monophyly for all three analyses. The best trees for which Strepsiptera + Diptera are not monophyletic are 8, 5, and 15 steps longer than the best trees for the 18S, 28S, and combined analyses, respectively. These results are consistent with those of Whiting et al. (1997). Interestingly, the maximum-likelihood estimate of phylogeny for the 18S rDNA data is consistent with Strepsiptera + Diptera monophyly but the maximum-likelihood estimate of phylogeny for the 28S rDNA data is not. However, the difference between the best tree for which Strepsiptera + Diptera are and are not monophyletic is not significant for any of the analyses performed using maximum likelihood (Kishino and Hasegawa, 1989; 18S analysis, $|\log L_M - \log L_{NM}| = 2.234$, $SD = 13.801$, $T = 0.162$, $P = 0.872$; 28S analysis, $|\log L_M - \log L_{NM}| = 0.252$, $SD = 4.303$, $T = 0.059$, $P = 0.953$; 18S + 28S analysis, $|\log L_M - \log L_{NM}| = 1.644$, $SD = 7.115$, $T = 0.231$, $P = 0.817$). The Kishino-Hasegawa test results suggest that there is very little support for Strepsiptera + Diptera monophyly or nonmonophyly.

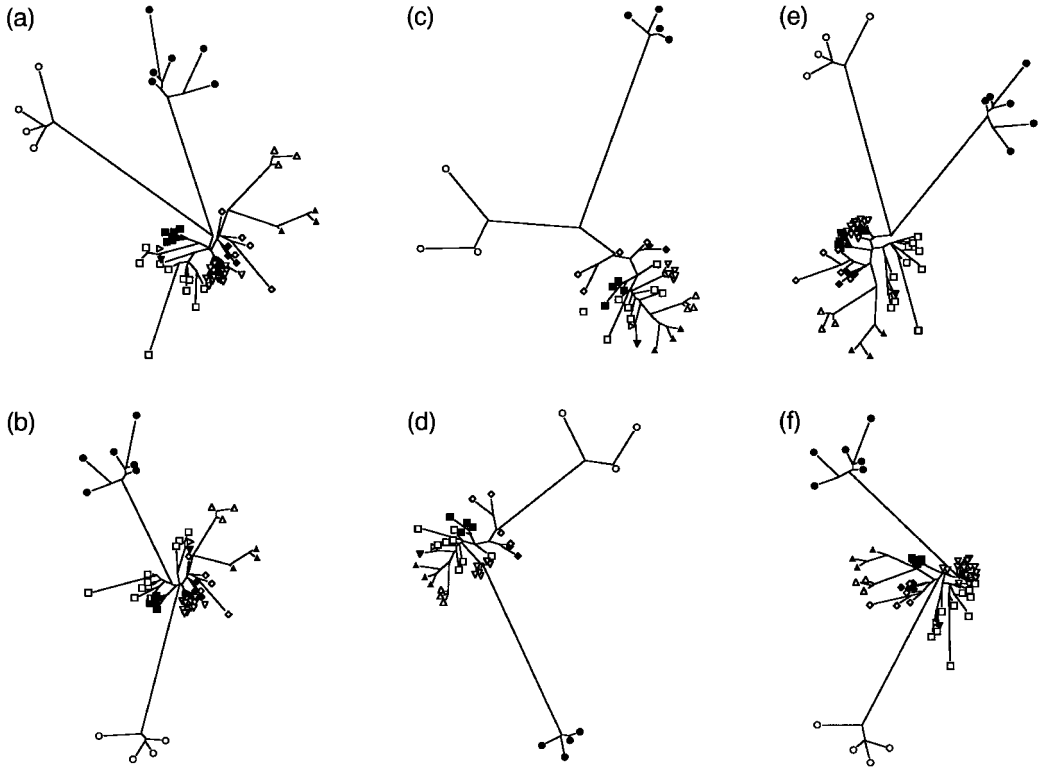


FIGURE 8. The maximum likelihood estimates of phylogeny under the constraints that Strepsiptera + Diptera are monophyletic (a) and nonmonophyletic (b) for the 18S rDNA data set, monophyletic (c) and nonmonophyletic (d) for the 28S rDNA data set, and monophyletic (e) and nonmonophyletic (f) for the 18S + 28S rDNA data set. Coleoptera = □ , Neuroptera = ■ , Hymenoptera = ▽ , Megaloptera = ▼ , Mecoptera = ◇ , Siphonaptera = ◆ , Lepidoptera = △ , Trichoptera = ▲ , Raphidioptera = ▷ , Diptera = ○ , Strepsiptera = ● .

Figure 9 summarizes the results of the neighbor joining analyses (Saitou and Nei, 1987). These graphs show Strepsiptera + Diptera monophyly as a function of κ and α. Black areas denote parameter conditions that lead to a nonmonophyletic Strepsiptera + Diptera whereas white areas denote parameter values that lead to Strepsiptera + Diptera monophyly. The white dots denote the maximum likelihood values of κ and α. Obviously, the results of the neighbor joining analysis are sensitive to the assumptions of the model. Neither Strepsiptera + Diptera monophyly nor nonmonophyly are strongly supported by neighbor joining.

TABLE 1. The scores for the best trees found in the three analyses performed in this study. Monopoly, the best tree under the constraint of Strepsiptera + Diptera monophyly; Nonmonophyly, the best under the constraint of Strepsiptera + Diptera nonmonophyly; κ transition/transversion rate ratio (estimated via maximum likelihood); α, shape parameter of gamma distribution for among-site rate variation (estimated via maximum likelihood).

	Parsimony (tree length)		Maximum likelihood					
			Monophyly			Nonmonophyly		
Analysis	M	NM	ln L	κ	α	ln L	κ	α
18S	1180	1188	−7427.052	3.311	0.297	−7429.286	3.309	0.292
28S	618	623	−3194.192	3.147	0.227	−3193.940	3.160	0.230
18 + 28S	1831	1846	−10876.713	3.187	0.259	−10878.232	3.204	0.258

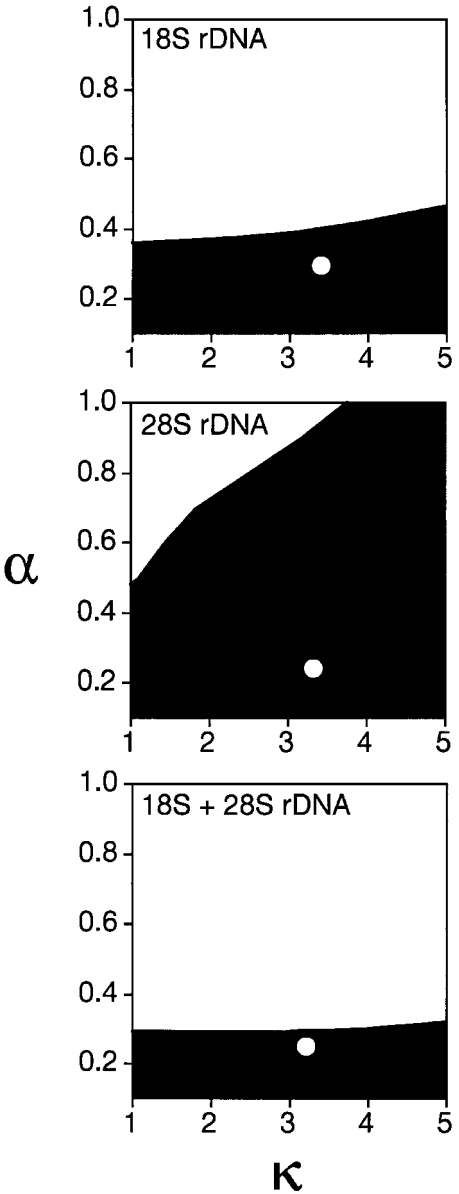


FIGURE 9. The results of the neighbor-joining analysis are sensitive to the assumptions used to generate the distances. Maximum likelihood was used to estimate the distances between all pairs of species under the Hasegawa-Kishino-Yano (1985) model of DNA substitution. The transition/tranversion rate parameter (κ) and the shape parameter of the gamma distribution for among-site rate variation (α) were varied. The results were sensitive to the values that κ and α took in the analysis because either Strepsiptera + Diptera monophyly (white areas) or nonmonophyly (black areas) was obtained for different combinations of κ and α. The white dot indicates the maximum likelihood values for κ and α.

Are the branches leading to Strepsiptera and Diptera the longest?—The first criterion for long branch attraction—that there are long branches that may be problematic in the analysis—appears to be satisfied. In all analyses, the branches leading to Strepsiptera and to Diptera were very long. This assertion can be verified by examining the unrooted phylograms depicted in Figure 8. Figure 10 shows frequency histograms of the lengths of all branches on the trees. The branches leading to Diptera and to Strepsiptera are denoted D and S, respectively. Branch lengths for all methods are in the same units (expected number of substitutions per site). As expected, the parsimony method severely underestimates the lengths of the branches leading to Strepsiptera and Diptera (branch length estimates were obtained by averaging the lengths of the branches over all equally parsimonious reconstructions of ancestral characters using MacClade 3.0; Maddison and Maddison, 1992). However, even for the parsimony method, the branches leading to the Strepsiptera are among the longest on the tree (ranks for branches leading to Strepsiptera [S] and Diptera [D], rank 1 being longest: 18S monophyly, $S = 3$, $D = 1$; 18S nonmonophyly, $S = 3$, $D = 1$; 28S monophyly, $S = 1$, $D = 12$; 28S nonmonophyly, $S = 1$, $D = 5$). The lengths of the branches leading to Strepsiptera and to Diptera are the longest two branches in the maximum likelihood analyses. In fact, for the 28S rDNA sequences, the length of the Strepsiptera and Diptera branches are among the longest ever observed (approximately 1.0 substitution per site).

The extreme lengths of the branches leading to Strepsiptera and to Diptera compared to the other branches of the phylogeny are virtually unparalleled in phylogenetic analysis. Interestingly, the more thorough sampling of species in the Whiting et al. (1997) study did very little to break up the branches leading to Strepsiptera and Diptera; all of the Strepsipteran and Dipteran species join to the long branches leading to the clades near the very tips. Although thorough taxon sampling is important, sampling by itself cannot be relied upon to solve the long-branch problem. Also, the fact that long branches occur for both 18S

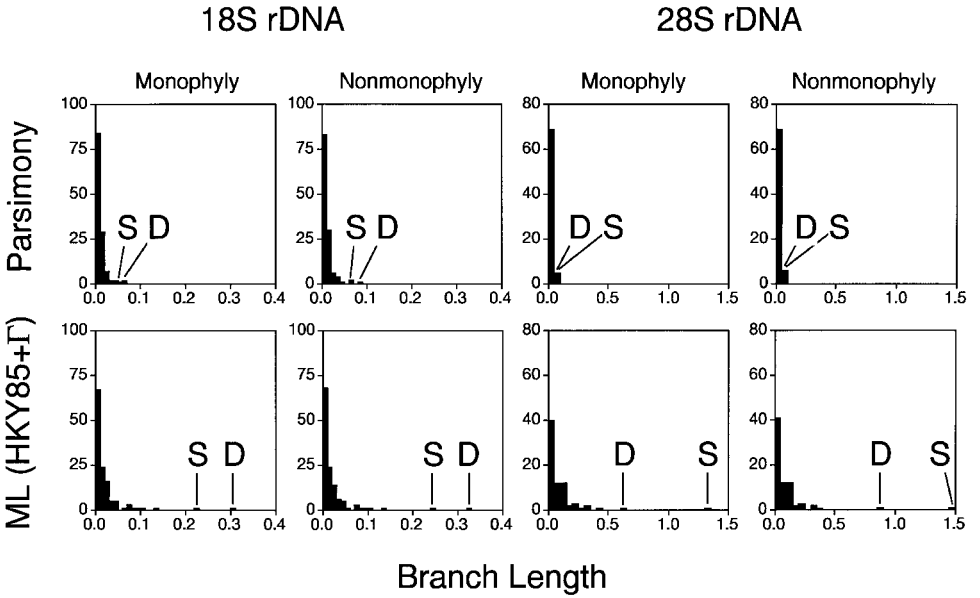


FIGURE 10. Frequency histograms of the branch length estimates under the parsimony and maximum likelihood (ML) criteria with constraints of monophyly (M) and nonmonophyly (N). The parsimony method grossly underestimates the lengths of the branches, especially the branches leading to Strepsiptera and to Diptera (D).

and 28S rDNA is not surprising because these genes are closely linked.

Are the branches leading to Strepsiptera and Diptera long enough to attract?— I used Monte Carlo simulation to determine whether the long branches leading to Strepsiptera and to Diptera are long enough to attract one another in a parsimony analysis even if the branches are separate on the true tree (Huelsenbeck et al., 1996; Huelsenbeck, 1997). Data were simulated on the maximum likelihood estimates of phylogeny that were consistent with Strepsiptera + Diptera nonmonophyly for 18S and 28S rDNA (trees B and D, Fig. 8). The maximum likelihood estimates of the branch lengths, transition/transversion rate ratio, and shape parameter of the gamma distribution for rate variation among sites were used in the simulations. The observed base frequencies were also used in the simulations. Figure 11 shows the results of the simulations. The number of sites simulated is plotted on the abscissa and the probability that the parsimony estimate was consistent

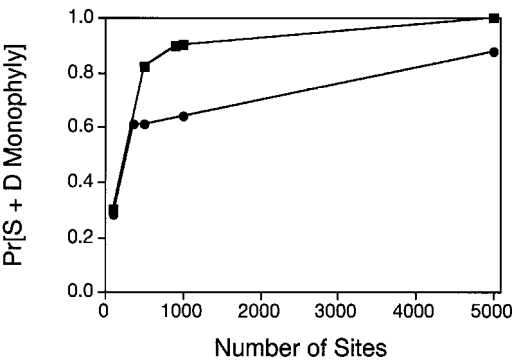


FIGURE 11. The branches leading to Strepsiptera and to Diptera (D) are long enough to attract one another in a parsimony analysis. The figure shows the probability of the branches leading to Strepsiptera and to Diptera being estimated as sister taxa as a function of the sequence length. The model trees used in the simulations placed Strepsiptera and Diptera apart (Figs. 8b and 8d for the 18S [■] and 28S [●] simulations, respectively). The maximum likelihood values for the branch lengths, κ, and α on the model trees and the observed base frequencies were used in the simulations. Maximum parsimony assumed Fitch (1971) character optimization. No branch swapping was performed in the heuristic searches. The observed lengths of the 18S and 28S rDNA sequences were 902 and 358 bp, respectively.

with Strepsiptera + Diptera monophyly (i.e., the proportion of the time that the long branches attracted one another in a parsimony analysis) is plotted on the ordinate. The probability that the long branches will attract in a parsimony analysis is high for the simulations that matched the observed number of sites for the 18S and 28S analyses (18S analysis, 902 sites, Pr[long branch attraction] = 0.90; 28S analysis, 358 sites, Pr[long branch attraction] = 0.61). Moreover, as the simulated sequences become longer, the probability that long branch attraction will occur in a parsimony analysis increases, suggesting that parsimony is inconsistent for the model tree that places Strepsiptera and Diptera in disparate parts of the tree.

DISCUSSION

Traditionally, long branches are considered to be problematic in phylogenetic analysis only when they occur in separate parts of the true tree. It is well known that many phylogenetic methods will converge to the wrong phylogeny if the branches and sequences are long enough. Yet, long branches are also problematic when they occur together on the true tree. They are not problematic because such phylogenies are difficult to estimate, but rather because it is difficult to distinguish between a phylogenetic estimate that has the long branches together versus the long branches separate.

Which phylogenetic methods are best when long branches occur in a phylogeny? I maintain that methods that attempt to correct for the multiple substitutions that occur along long branches are better in such cases. In this article, only the performance of maximum likelihood was explored in any detail. However, many of the results should also apply to additive distance methods using corrected distances. Methods that correct for multiple substitutions, for one, are better able to estimate phylogeny when the long branches are separate and their performance is adequate when the long branches are together on the true tree. Moreover, they more accurately depict the uncertainty in the phylogenetic estimate. Trees with long branches separate versus together are very similar when one considers how little branch length must be contracted and expanded to convert one type of tree into the other. I favor methods that depict this uncertainty more faithfully. Maximum parsimony, although very efficient when the long branches are adjacent on the true phylogeny, is very inefficient or inconsistent when the long branches are apart on the true phylogeny.

Although systematists are often concerned with the fit of the models of DNA substitution assumed in maximum likelihood and distance analyses, it appears that even an imperfect attempt to correct for multiple substitutions that occur on long branches is better than failing to correct for multiple substitutions at all. In this study and others (Tateno et al., 1994; Huelsenbeck, 1995a, 1995b), when maximum likelihood assumed a false model, its performance was better than that of parsimony when the long branches are separate (i.e., the method was consistent and more efficient over a larger range of parameter conditions).

Instances of possible long branch attraction have been identified in amniotes (Huelsenbeck et al., 1996) and in rodents (Sullivan and Swofford, 1997). I argue that the phylogeny of Strepsiptera and Diptera is also problematic because of the long branches leading to each of these two clades (also see Carmean and Crespi, 1995; Huelsenbeck, 1997). The branches leading to the Strepsiptera and to Diptera are very long regardless of whether parsimony (a method that will severely underestimate the lengths of the long branches) or maximum likelihood is used to estimate the branch lengths. Moreover, the simulation results presented in this note demonstrate that the branches are long enough to attract in a parsimony analysis; that is, even if the long branches leading to Strepsiptera and to Diptera occur in separate parts of the true phylogeny, parsimony will very likely produce an estimate of phylogeny that places these two taxa together on the tree. Hence, the results of the Whiting et al. (1997) analysis are not surprising.

Neither Strepsiptera + Diptera monophyly nor Strepsiptera + Diptera nonmonophyly is strongly supported by the 18S and 28S rDNA. Analysis of the 18S and 28S rDNA sequences using maximum likelihood (1) produced conflicting trees for 18S and 28S data (the 18S tree was consistent with Strep-

sipter + Diptera monophyly whereas the 28S tree was consistent with the nonmonophyly of Strepsiptera + Diptera) and (2) the best trees for which Strepsiptera + Diptera are and are not monophyletic were not significantly different using the Kishino-Hasegawa (1989) test for the 18S, 28S, and combined analyses.

The Strepsiptera + Diptera problem can be visualized as lying in either the **R** space or the **Q** space for the 18S and 28S rDNA data (Fig. 12). What can be done to distinguish between these two possible scenarios? One possible solution is to add more Strepsiptera and Diptera species with the hope of breaking up the long branches. This solution, however, is not guaranteed to work because the additional species may join at the tips of the very long branches. Another solution is to sequence additional nonlinked genes. If the phylogeny from an additional gene is consistent with Strepsiptera + Diptera monophyly and the long branches do not appear to be a problem, then the monophyly of these two taxa is supported.

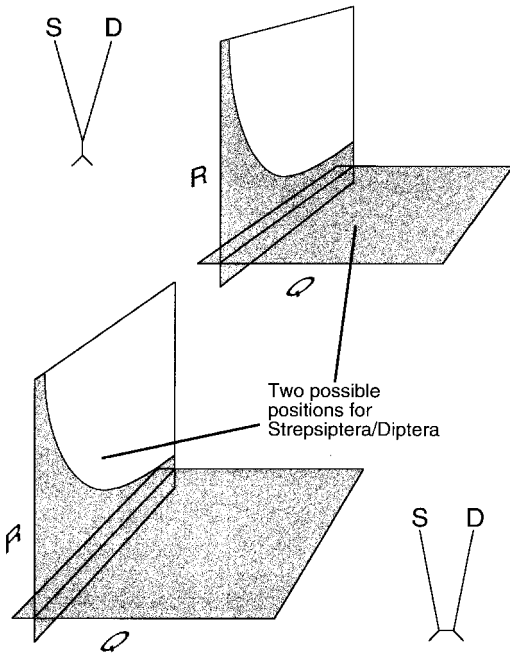


FIGURE 12. Strepsiptera and Diptera either are or are not monophyletic. If they are monophyletic, then the true tree can be thought of as lying in the **Q** space whereas if they are not, the true tree lies in the **R** space. In either case, maximum parsimony will estimate a phylogeny with the long branches adjacent on the tree so that the true tree will be inferred as being in the **Q** space. The shaded areas of the planes denote conditions under which parsimony is consistent.

Regardless of whether or not the Strepsiptera and Diptera are monophyletic, the data suggest that something interesting is occurring in the 18S and 28S rDNA sequences. The rates of substitution, for one, are much higher in the Strepsiptera and Diptera. Also, the 18S rDNA sequences of Strepsiptera are much longer than the 18S sequences in other insects (Chalwatzis et al., 1995) with lower G + C content due to A + T-rich expansion segments (e.g., the Strepsiptera species *Xenos vesparum* is 3,316 bp in length with 28.1% G + C content whereas *Polistes dominulus* and *Meloe proscarabaeus* are 1,919 bp and 1,934 bp in length, respectively, and have 49.3% and 48.2% G + C content, respectively). Although interesting from a molecular evolutionary point of view, these features of the 18S and 28S rDNA in Strepsiptera and Diptera make phylogenetic analysis difficult. In addition to the long branch attraction problem discussed here, alignment of ribosomal sequences in general, and Strepsiptera sequences in particular, is difficult. It would not be surprising if the best estimate of phylogeny changed depending on the alignment used (although it would be unusual if the alignment changed the significance of the difference between the best Strepsiptera + Diptera monophyly and nonmonophyly trees). Also, the stem regions of ribosomal genes present problems because the substitutions in stems are not independent (Dixon and Hillis, 1993; Wheeler and Honeycutt, 1988), a basic assumption of all phylogenetic methods. One possible solution to the nonindependence of the substitutions in the stem regions is to model the stems separately. Several authors have proposed Poisson process models that account for the dependence of Watson-Crick paired stem bases (e.g., Schöniger and von Haeseler, 1994).

Phylogenetic analysis is a complex problem in inference. It is not surprising, then, that systematists take widely divergent positions on the different methods of analysis and the principles upon which they are based. No one method of phylogenetic analysis can be expected to perform best for all possible data sets. It is important, then, to keep in mind the limitations of any method of analysis and pro-

ceed cautiously when the data appear to indicate that the method may be providing misleading results.

ACKNOWLEDGMENTS

I thank M. Whiting for sending me the sequence data used in his analysis. This paper benefited from discussions with J. Felsenstein, R. Nielsen, B. Rannala, D. Swofford, and Z. Yang. This research was supported by a Miller Postdoctoral Fellowship and a NSF Sloan Postdoctoral Fellowship. I thank the NOW project at the University of California at Berkeley, and especially E. Brewer, for making available the many SUN Ultrasparc workstations used to perform the simulation and maximum likelihood analyses.

REFERENCES

BISHOP, M. J., AND E. A. THOMPSON. 1986. Maximum likelihood alignment of DNA sequences. J. Mol. Biol. 190:159–165.

CARMEAN, D., AND B. CRESPI. 1995. Do long branches attract flies? Nature 373:666.

CARMEAN, D., L. S. KIMSEY, AND M. L. BERBEE. 1992. 18S rDNA sequences and holometabolous insects. Mol. Phylogenet. Evol. 1:270–278.

CHALWATZIS, N., A. BAUR, E. STETZER, R. KINZELBACH, AND F. K. ZIMMERMAN. 1995. Strongly expanded 18S rRNA genes correlated with a peculiar morphology in the insect order Strepsiptera. Zoology 98:115–126.

DEBRY, R. W. 1992. The consistency of several phylogeny-inference methods under varying evolutionary rates. Mol. Biol. Evol. 9:537–551.

D'ERCHIA, A. M., C. GISSI, G. PESOLE, C. SACCONE, AND U. ARNASON. 1996. The guinea pig is not a rodent. Nature 381:597–600.

DIXON, M. T., AND D. M. HILLIS. 1993. Ribosomal RNA secondary structure: Compensatory mutations and implications for phylogenetic analysis. Mol. Biol. Evol. 10:256–267.

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

FELSENSTEIN, J. 1995. PHYLIP: Phylogeny inference package, version 3.57c. Department of Genetics, Univ. Washington, Seattle.

FITCH, W. M. 1971. Toward defining the course of evolution: Minimum change for a specific tree topology. Syst. Zool. 20:406–416.

GAUT, B. S., AND P. O. LEWIS. 1995. Success of maximum likelihood phylogeny inference in the four-taxon case. Mol. Biol. Evol. 12:152–162.

GAUTHIER, J., A. G. KLUGE, AND T. ROWE. 1988. Amniote phylogeny and the importance of fossils. Cladistics 4:105–209.

GOLDMAN, N. 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. 36:182–198.

GRAUR, D, W. A. HIDE, AND W.-H. LI. 1991. Is the guinea-pig a rodent? Nature 315:649–652.

HASEGAWA, M., H. KISHINO, AND T. YANO. 1985. Dating of the human-ape splitting by the molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.

HEDGES, S. B. 1994. Molecular evidence for the origin of birds. Proc. Nat. Acad. Sci., USA 91:2621–2624.

HENDY, M. D., AND D. PENNY. 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38:297–309.

HILLIS, D. M. 1996. Inferring complex phylogenies. Nature 383:130–131.

HUELSENBECK, J. P. 1995a. Performance of phylogenetic methods in simulation. Syst. Biol. 44:17–48.

HUELSENBECK, J. P. 1995b. The robustness of two phylogenetic methods: Four-taxon simulations reveal a slight superiority of maximum likelihood over neighbor joining. Mol. Biol. Evol. 12:843–849.

HUELSENBECK, J. P. 1997. Is the Felsenstein zone a fly trap? Syst. Biol. 46:69–74.

HUELSENBECK, J. P, AND J. J. BULL. 1996. A likelihood ratio test to detect conflicting phylogenetic signal. Syst. Biol. 45:92–98.

HUELSENBECK, J. P., AND D. M. HILLIS. 1993. Success of phylogenetic methods in the four-taxon case. Syst. Biol. 42:247–264.

HUELSENBECK, J. P., D. M. HILLIS, AND R. JONES. 1996. Parametric bootstrapping in molecular phylogenetics: Applications and performance. Pages 19–45 *in* Molecular zoology: Advances, strategies, and protocols (J. D. Ferraris and S. R. Palumbi, eds.). Wiley-Liss, New York.

JIN, L., AND M. NEI. 1990. Limitations of the evolutionary parsimony method of phylogenetic analysis. Mol. Biol. Evol. 7:82–102.

JUKES, T. H., AND C. R. CANTOR. 1969. Evolution of protein molecules. Pages 21–132 *in* Mammalian protein metabolism (H. Munro, ed.). Academic Press, New York.

KATHIRITHAMBY, J. 1989. Review of the order Strepsiptera. Syst. Entomol. 14:41–92.

KIMURA, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16:111–120.

KISHINO, H., AND M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29:170–179.

KRISTENSEN, N. P. 1991. Phylogeny of extant hexapods. Pages 125–140 *in* The insects of Australia, 2nd edition (I. D. Naumann, P. B. Carne, J. F. Lawrence, E. S. Nielsen, J. P. Spradberry, R. W. Taylor, M. J. Whitten, and M. J. Littlejohn, eds.). CSIRO, Melbourne Univ. Press, Melbourne.

KUHNER, M. K., AND J. FELSENSTEIN. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Mol. Biol. Evol. 11:459–468.

LI, W.-H., W. A. HIDE, A. ZHARKIKH, D.-P. MA, AND D. GRAUR. 1992. The molecular taxonomy and evolution of the guinea pig. J. Hered. 83:174–181.

MA, D.-P., A. ZHARKIKH, D. GRAUR, J. L. VANDEBERG, and W.-H. LI. 1993. Structure and evolution of opossum,

guinea pig, and porcupine cytochrome *b* genes. J. Mol. Evol. 36:327–334.

MADDISON, W. P., AND D. R. MADDISON. 1992. MacClade: Analysis of phylogeny and character evolution, version 3.0. Sinauer, Sunderland, Massachusetts.

NEI, M. 1991. Relative efficiencies of different tree-making methods for molecular data. Pages 90–128 *in* Phylogenetic analysis of DNA sequences (M. M. Miyamoto and J. Cracraft, eds.). Oxford Univ. Press, Oxford, England.

RZHETSKY, A., AND T. SITNIKOVA. 1995. Using a wrong substitution model in tree-making. Pages 125–135 *in* Current topics on molecular evolution (M. Nei and N. Takahata, eds.). Pennsylvania State Univ., University Park.

SAITOU, N., AND M. NEI. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406–425.

SANKOFF, D., C. MOREL, AND R. J. CEDERGREN. 1973. Evolution of 5S RNA and the non-randomness of base replacement. Nature 245:232–234.

SCHÖNIGER, M., AND A. VON HAESELER. 1994. A stochastic model and the evolution of autocorrelated DNA sequences. Mol. Phylogenet. Evol. 3:240–247.

STEEL, M. A., M. D. HENDY, AND D. PENNY. 1993. Parsimony can be consistent! Syst. Biol. 42:581–587.

SULLIVAN, J., AND D. L. SWOFFORD. 1997. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. J. Mammal. Evol. 4:77–86.

SWOFFORD, D. L., G. P. OLSEN, P. J. WADDELL, AND D. M. HILLIS. 1996. Phylogenetic inference. Pages 407–514 *in* Molecular systematics, 2nd edition (D. M. Hillis, C. Moritz, and B. K. Mable, eds.). Sinauer, Sunderland, Massachusetts.

TATENO, Y., N. TAKEZAKI, AND M. NEI. 1994. Relative efficiencies of the maximum-likelihood, neighbor-joining, and maximum-parsimony methods when substitution rate varies with site. Mol. Biol. Evol. 11:261–277.

THORNE, J. L., H. KISHINO, AND J. FELSENSTEIN. 1991. An evolutionary model for maximum likelihood alignment of DNA sequences. J. Mol. Evol. 33:114–124.

THORNE, J. L., H. KISHINO, AND J. FELSENSTEIN. 1992. Inching toward reality: An improved likelihood model of sequence evolution. J. Mol. Evol. 34:3–16.

WADDELL, P. 1995. Statistical methods of phylogenetic analysis, including Hadamard conjugations, LogDet transforms, and maximum likelihood. Ph.D. Dissertation, Massey Univ., Palmerston North, New Zealand.

WHEELER, W. C., AND D. L. GLADSTEIN. 1994. MALIGN, version 1.93. American Museum of Natural History, New York.

WHEELER, W. C., AND R. L. HONEYCUTT. 1988. Paired sequence difference in ribosomal RNAs: Evolutionary and phylogenetic implication. Mol. Biol. Evol. 5:90–96.

WHITING, M. F., J. C. CARPENTER, Q. D. WHEELER, AND W. C. WHEELER. 1997. The Strepsiptera problem: Phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal sequences and morphology. Syst. Biol. 46:1–68.

YANG, Z. 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol. Biol. Evol. 10:1396–1401.

YANG, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. J. Mol. Evol. 39:306–314.

YANG, Z. 1996. Phylogenetic analysis using parsimony and likelihood methods. J. Mol. Evol. 42:294–307.

Received 7 August 1997; accepted 10 February 1998
Associate Editor: D. Cannatella

APPENDIX

The accuracy of maximum parsimony (MP) and maximum likelihood (ML) for simulated sequences of 100 and 1000 sites. The process of substitution that generated the sequences assumed equal base frequencies but allowed for a transition/transversion rate ratio (κ) and for gamma distributed rate variation among sites (with gamma shape parameter α). Parsimony assumed Fitch (1971) optimization of characters and maximum likelihood assumed a Jukes-Cantor (1969) model of DNA substitution. When $\kappa = 1$ and $\alpha = \infty$ the assumptions of maximum likelihood were satisfied. *R*, *Q*, and *S* are defined in the text.