SEPARATE VERSUS COMBINED ANALYSIS OF PHYLOGENETIC EVIDENCE

Alan de Queiroz[1]

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721

Michael J. Donoghue

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138

Junhyong Kim[2]

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721

ABSTRACT

There has been much discussion in the recent systematic literature over whether different data sets bearing on phylogenetic relationships should be analyzed separately or combined and analyzed simultaneously. We review arguments in favor of each of these views. Assuming that the goal is to uncover the true phylogeny of the entities in question, arguments for combining data based on the notions that one should use the "total evidence" available, or that the combined analysis gives the tree with the greatest descriptive and explanatory power, are not compelling. However, combining data sets can enhance detection of real phylogenetic groups. On the other hand, if there is heterogeneity

[1]Present address: University Museum and Department of Environmental, Population and Organismic Biology, University of Colorado, Boulder, Colorado 80309
[2]Present address: Department of Biology, Yale University, New Haven, Connecticut 06511

657

among data sets with respect to some property that affects phylogeny estimation, then combining the data can give misleading results. Thus, there are reasonable arguments on both sides of the debate.

We present a conceptual framework based on the reasons that different data sets may give conflicting estimates of phylogeny. The framework illustrates the point that the precise nature of the difference among data sets is critical in the choice of a method of analysis. In particular, very different approaches are necessary to deal with data sets that differ in processes of character change compared to ones that differ in branching histories. We highlight several recently developed methods designed to deal with these different situations. All of these methods avoid the loss of information that is likely to be associated with summarizing data sets as trees in an intermediate step (an advantage of typical combined analyses), while taking into account heterogeneity among data sets (an advantage of separate analyses). We suggest that the recognition and further development of such methods will help depolarize the debate over combined and separate analysis.

INTRODUCTION

The availability of a variety of sources of evidence on phylogenetic relationships has focused attention on a fundamental question: Should different kinds of data bearing on a given phylogenetic problem be analyzed separately or combined and analyzed simultaneously? Our aim is to review arguments for and against separate and combined analyses and to provide a general conceptual framework within which to explore the basic but sometimes subtle issues associated with the problem. We hope to show that both kinds of analyses are useful in estimating phylogenetic relationships, but that the standard forms of these alternatives by no means exhaust the possible solutions to the problem.

Some additional clarification of our goals is in order, especially to say what we do not intend to provide. First, although we need to refer to particular examples in order to clarify arguments, we do not provide a thorough review of studies in which separate and/or combined analyses have actually been carried out. In addition, although much emphasis has been placed on molecular versus morphological evidence (26, 29, 38, 45, 84, 85, 108), that contrast is not our specific concern here. The partitioning of data into molecular and morphological subsets is only one of many divisions that may be relevant to the problem at hand.

It is also critical to clarify our general perspective from the outset. We assume that the ultimate goals of phylogenetic analysis are to discover the true phylogeny of the entities under investigation, and to understand evolutionary processes. Although this review is written from that perspective, we recognize that certain arguments that we criticize may be valid given different basic goals

(see, in particular, the arguments in favor of combined analysis based on the principle of "total evidence" and on maximizing descriptive and explanatory power). It is also important to appreciate that different sorts of interrelated entities have their own histories, which may or may not coincide with one another. For example, the branching history of a particular gene or organellar genome may not coincide exactly with the branching history of the populations or species of organisms in which it resides (e.g. 3, 29, 109). Consequently, the objects of study must be clearly specified. For the purposes of this paper, we assume that the goal is to estimate relationships among taxa. (Many of our arguments apply equally well to estimation of relationships among other entities—e.g. genes—but may have to be transformed slightly for that purpose.) Finally, we recognize that the "best" method of analysis in a given instance may depend on the relative importance given to resolving power versus avoidance of error (e.g. see 20, 105).

PREVIOUS VIEWS

Much of the discussion in the literature has revolved around a contrast between the "consensus" of trees derived from separately analyzed data sets ("taxonomic congruence" sensu Mickevich, 68; see 50, 57) and what has been called the "total evidence" approach ("character congruence" sensu Mickevich, 68; 56; see below). In general, we consider a method to be a consensus method if the characters in two (or more) data sets are not allowed to interact directly with one another in a single analysis, but instead interact only through the trees derived from them. Given this definition, consensus includes methods such as Brooks parsimony analysis (113, 115) in which two or more trees derived from individual data sets are coded for parsimony analysis as a set of characters that reflect the underlying tree structure (also see 7, 29, 89a, 89b). It is important to recognize that the contrast with which we are primarily concerned—between separate and combined analysis—is not identical to this standard distinction between consensus and the "total evidence" approach. Choosing to analyze data sets separately does not necessitate the use of consensus trees, and proponents of separate analysis have not always condoned the use of consensus techniques (e.g. 10). Instead, separate analyses may be seen as a means of exploring possible disagreements among data sets. Similarly, a combined approach does not necessarily imply incorporation of all character data in a single analysis. Although our discussion of previous views is constrained somewhat by the particular methods that the authors were addressing, ultimately we consider alternatives beyond those typically discussed in the literature (see the "Framework" section).

The idea that one might want to perform separate analyses on subsets of the available data relies on the existence of different classes of evidence with

respect to phylogeny estimation. To qualify as a distinct class of evidence, characters in a data set must, in a statistical sense, be more similar to each other than they are to characters in other data sets with respect to some property that affects phylogeny estimation by the given method (10; see also 19, 55, 96, 105).

Recently, Kluge & Wolf (57; also see 50) have questioned whether such classes of data actually exist. They urged cladists to "question artificial subdivisions of evidence because there is no reason to believe those definitions have discoverable boundaries" (p. 190). This is an important issue to address before proceeding further because, if classes of evidence do not exist, the justification for analyzing data sets separately disappears (72). Several studies suggest that some traditional distinctions (e.g. between molecular, morphological, and behavioral characters; 21, 22, 26, 27, 95) may not be relevant with respect to estimating phylogenies. Nonetheless, molecular studies have made it increasingly clear that distinct, identifiable classes of evidence do exist. The case for the existence of such classes has been articulated most strongly by Miyamoto & Fitch (72; also see 10, 22, 105), who use as an example sequences of the γ^l-globin and 12S rRNA genes, which differ in a number of properties (e.g. substitution patterns, overall rate of evolution, and frequencies of recombination, gene conversion, and gene duplication) that are likely to affect their behavior as indicators of phylogeny. Many other molecular examples could be given. For example, in cases where gene trees may differ from each other and from the overall species tree (see below), data from any pair of unlinked genes may be considered different classes of evidence. As entities with distinct locations in the genome, these different sequences have real, discoverable boundaries. (In stating that the boundaries are real and discoverable we are not suggesting that they are always clean; for example, since linkage varies continuously, classes of characters that are based on linkage relationships may have fuzzy boundaries.) We conclude that the argument of Kluge & Wolf is not universally valid.

Arguments in Favor of Separate Analysis

Bull et al (10; see also 49) presented a persuasive general argument against combining and in favor of separate analysis in some circumstances, which rests on the view that any estimate of phylogeny assumes a model of evolution. If, under the chosen method of estimation, data sets give phylogenetic estimates that are too different to be ascribed to sampling error (due to the limited number of characters and/or taxa sampled), then they must have been governed by different evolutionary rules. The significant difference between the phylogenetic estimates further indicates that the data sets differ in whether or how they violate the assumptions of the method (i.e. either one data set violates the assumptions and the other does not, or both violate the assumptions but in

different ways). Bull et al argued that, if this is the case, the data sets should not be combined unless the method can be changed to account for the difference. In support of this argument they pointed out that assessing heterogeneity prior to combining data is an accepted procedure in science in general, applied, for instance, through analysis of variance and contingency tables.

One argument for separate analysis concerns simply the ability to quickly detect such heterogeneity in the form of areas of agreement and disagreement, which might highlight conflicts caused by natural selection, differential rates of evolution, hybridization, horizontal transfer, or lineage sorting. Comparison of separately analyzed trees has been seen as especially useful in identifying hybrids, wherein, for example, one may see conflicts between uniparentally inherited genomes (most mitochondria and chloroplasts) and nuclear genes and/or morphological characters (e.g. 90, 92). It has been argued that combined analysis may obscure significant patterns of congruence or conflict among characters (10, 19, 105), and it is true that we currently lack efficient methods for keeping track of the behavior of whole suites of characters in combined analyses. However, as noted by Chippindale & Wiens (14), proponents of combined analysis generally also carry out separate analyses to explore such possibilities.

Two arguments that have been made against combining data sets can be viewed as special cases of Bull et al's general argument. One concerns the impact of putting together a "bad" data set with a "good" one, where bad and good refer to the ability to accurately reflect true phylogenetic relationships (10). Within the framework of Bull et al, this can be viewed as an argument against combining a data set that violates the assumptions of the estimation method with one that does not (or at least violates the assumptions less drastically). The idea is that combining "bad" with "good" may actually give a less accurate estimate than using the "good" data by themselves. In the extreme this has led some authors to dismiss large classes of data. For example, some proponents of molecular approaches have written off morphology on the grounds that morphological features are subject to natural selection and therefore may be misleading due to convergent evolution (e.g. 98). As indicated above, there is no compelling evidence that molecular characters are in general better than morphological characters for estimating phylogeny, and the same can be said for some other traditional distinctions (21, 22, 27, 95).

Nonetheless, there do seem to be identifiable "good" and "bad" classes of evidence. For example, certain genes or gene regions may evolve much more rapidly than others; for some levels of divergence, the distribution of nucleotides among taxa for these rapidly evolving characters may be essentially random, whereas genes/gene regions that evolve more slowly may retain phylogenetic information (42, 59, 61, 73, 74). Bull et al (10) explored a similar

case in a series of simulations that show circumstances under which the combined analysis is less likely to recover the true tree. In particular, they examined cases in which characters in one data set evolved at a significantly faster rate than did those in another, and they demonstrated instances in which the best results were obtained from the slowly evolving characters alone. They showed that this result can be obtained even when the estimation method is consistent for both data sets, i.e. converges on the truth as the number of characters is increased (35).

Barrett et al (5) suggested that in such cases characters might be weighted to reflect differences in evolutionary rate and then combined, a possibility also noted by Bull et al (10). Chippindale & Wiens (14) showed that, in the cases examined by Bull et al, such a weighting scheme would indeed render the combined analysis equal or superior to either of the individual analyses in recovering the true tree. However, in this example the true tree was known and various weighting schemes were applied to determine which worked best. Left open is the very real problem of determining an appropriate weighting scheme at the outset of an analysis (see "Framework" section).

The second special case of Bull et al's argument is the concern that one data set may have an inordinately great influence on an analysis, simply by virtue of having a larger number of characters (29, 45, 54). This argument hinges on the possibility that the larger data set might be misleading in some way that the smaller one is not; thus, like the previous argument, it relies on the potential for differential violation of the assumptions of the analysis. Donoghue & Sanderson (26) pointed out that the addition of even a small number of characters can have a significant impact on the outcome, and in practice it often emerges that the smaller of two data sets does have a substantial impact on the resulting trees (e.g. 28). However, the fact that the smaller data set might have an impact on the combined analysis does not in itself indicate lack of any swamping, and it seems certain that swamping must occur in at least some cases.

A number of authors have pointed to independence between data sets (and explicitly or implicitly, nonindependence within data sets) as the basis for arguments in favor of separate analysis (19, 55, 58, 72, 75, 96, 105). In this context, nonindependence within data sets does not necessarily imply functional or physical linkage, but only that characters within a data set are more likely to share some property relevant for phylogeny estimation than are characters in different data sets. We view this as an alternative way of expressing the idea of heterogeneity; if characters within a data set are less independent than characters in different data sets, then there is heterogeneity among data sets. Using the support for conflict among trees from different data sets as a means of assessing such independence (19, 96) can be seen as a test of heterogeneity. In general, we prefer the construction of Bull et al (10) because

the term "nonindependence" conjures up functional or physical links between characters and is thus somewhat misleading.

Focusing on the idea of independence, however, does help to highlight important aspects of the debate. The possibility that different data sets give independent estimates of phylogeny underlies perhaps the most common argument in favor of consensus, namely, that areas of agreement among trees from separate analyses are especially likely to be true and are therefore conservative estimates of phylogeny (16, 19, 45, 55, 68, 72, 75, 86, 87, 96, 105). Because of their lack of independence, characters within a data set might as a whole tend to give misleading results. The same might be true of other data sets, but if there is independence among data sets, they should in general not mislead in the same way. Thus, areas of agreement are likely to represent real groups. [Of course the nature of the data sets must be considered here; data sets that are independent of each other in some ways might still share misleading properties (14).] The same reasoning has been applied in vicariance biogeographic studies (77, 80, 94, 99). Others have wondered whether there really are special advantages derived from assessing confidence through consensus (5, 50, 57), as opposed, for example, to combining the data in one analysis and performing bootstrap and decay analyses with the combined data set. However, for at least some of these latter authors (i.e. 50, 57), this criticism is tied to the idea that "evidence is evidence" (57, p. 190), i.e. that classes of evidence do not exist. The special advantages of consensus derive from the notion that such classes do exist.

The feeling that consensus trees might be safe estimates of phylogeny may have motivated Hillis's (45) suggestion that these be used in formulating classifications, where stability is often a concern (23). Barrett et al (5) challenged this belief by presenting hypothetical data for which a clade supported by strict consensus does not appear in an analysis of the combined data. Although this shows that the consensus result might not be sanctioned by all of the data analyzed together, resolving which approach (if either) is more likely to result in the true tree requires additional arguments (see 6, 14, 49, 76). The frequency of such occurrences with real data sets remains to be examined. It may be that consensus trees—at least those that include only clades found in more than one of the original trees (19)—will tend to contain fewer incorrect clades than do combined trees, at least in part because they generally make fewer claims about relationships.

Kluge & Wolf (57) criticized the argument that consensus analyses may be preferred due to their conservative nature by suggesting that safety in classification is not the goal of cladistics (13, 33). They pointed out that a completely unresolved tree would be maximally conservative; the absurdity of desiring such a tree is apparently meant to imply that conservatism cannot be a compelling criterion in constructing phylogenetic hypotheses. Instead they sug-

gested, citing Popper (88, 89), that completely resolved hypotheses are to be preferred because they are bolder. Counter to the point regarding the completely unresolved tree, it can be argued that conservatism could influence one's choice of a method without leading to a maximally conservative tree. Specifically, one might want a tree in which all clades have received a certain level of support. The idea that bold hypotheses are to be preferred is undoubtedly valid in some contexts, but one can argue that in phylogenetic studies, particularly those in which phylogenies are used as assumptions of an analysis, what one wants are well-supported hypotheses (105).

Chippindale & Wiens (14) questioned the idea that independence between and nonindependence within data sets might favor separate analysis and consensus over combined analysis. They suggested (p. 280) that examples of this sort "do not involve weaknesses unique to data combination; rather they are cases in which the fundamental assumptions of parsimony analysis are violated" (for example, independence of characters and lineages). However, this criticism ignores important differences in the assumptions of consensus versus combined analysis. For example, to estimate relationships among species from several gene sequences (which may have different histories), separate analyses assume that the characters are independent estimators of the gene trees, not of the species tree, and consensus assumes that the different gene trees are not likely to differ from the species history in the same way. A combined analysis of such data, on the other hand, would assume that the characters are independent estimators of the species tree. It is precisely this kind of difference in the assumptions made by a consensus versus a combined analysis that may justify consensus in some circumstances.

Although advocates of separate analysis agree that independence among data sets is important, there is disagreement about the evidence that should compel one not to combine. Some authors suggest that data sets are combinable unless one can show that there is significant conflict among the phylogenies estimated from them (10, 19, 96). This view of combining as the default strategy may be motivated by the potential benefits of combining (see below). Miyamoto & Fitch (72), however, contended that, if there are biological reasons for believing that there is heterogeneity among data sets, then they should not be combined, regardless of the level of disagreement among the phylogenetic estimates. These latter authors placed great emphasis on corroboration of phylogenetic hypotheses by independent data.

A final positive argument for consensus is simply that certain techniques preclude combining some data sets (5, 30, 58, 97). For example, there are no methods for combining DNA hybridization or immunological distance data with a set of morphological characters or molecular sequences. In such cases, consensus is the only option available if one wishes to present a single estimate (5).

Arguments in Favor of Combined Analysis

Arguments for combining data can be divided into five categories:

1. A philosophical argument based on the idea of "total evidence."
2. Objections to arbitrariness in consensus methods.
3. The difficulty of choosing a scheme of partitioning.
4. The greater descriptive and explanatory power of phylogenetic hypotheses generated from the combined data.
5. The greater ability of combined analyses to uncover real phylogenetic groups.

Kluge (56; also see 5, 50, 57) argued in favor of combining data based on the philosophical principle that one should use the total evidence available (12, 41, 43). Conclusions based on all of the relevant evidence are certainly to be preferred. However, general admonitions of this sort are of limited value in choosing among the very particular alternatives in the case of phylogenetic analysis. Probably no current method of analysis takes into account all of the relevant evidence. Consensus methods may lose information in the intermediate step of summarizing individual data sets as trees. However, the standard "total evidence" approach ignores both the problem of data sets being systematically misleading (19), and any trees generated from distance data (58, 89a). The argument for combining based on total evidence stems from the goal of minimizing ad hoc assumptions of homoplasy counted on a character-by-character basis (56, 57). However, if instead the goal of phylogenetic analysis is to uncover the true phylogeny, then evidence beyond what is required to construct the most parsimonious combined tree may argue for separate analysis (19).

The second category of arguments for combining points out that combined analysis of all the data avoids the arbitrariness inherent in consensus analyses. One argument, presented by Kluge (56, also see 57), is that combined analysis circumvents the need to choose among the various methods of consensus, a choice that is characterized as fundamentally arbitrary. A related argument (57) points to the arbitrary nature of deciding how to summarize congruence among data sets when each one may result in two or more best trees. As noted above, one course of action would be to carry out separate analyses without proceeding to a consensus solution (cf 10). However, if the goal is to achieve an estimate of the true phylogeny, rather than simply to explore the conflicts among data sets, then such arguments need to be addressed. It may be possible to defend the choice of a particular consensus method based on the goals of a particular study (as opposed to a more general defense), although such arguments have rarely been articulated in practice. Thus, it might be argued that strict consensus (81, 105) provides the most conservative assessment of

the agreement between trees, or that Adams consensus trees (1) are best at identifying taxa whose position is at odds in two or more trees (39, 45).

A related argument against consensus points out that these methods entail an arbitrary weighting of characters, because the individual trees are accorded equal weight in forming the consensus, regardless of the total number of characters that underlie them, or the number that support particular branches (17, 70; also see 26, 45). That is, the characters in a tree based on more characters will effectively be downweighted in comparison to those in a smaller data set. It might be argued that differential weighting of the characters in different data sets is warranted, but one would then need a further defense of weights that effectively reflect the number of characters in a data set. As Barrett et al (5) argued, weighting decisions should be explicitly defended, rather than being a passive and arbitrary outcome of the method of analysis. An alternative response is possible from those who argue that characters from different sources should not be combined under certain circumstances, namely, that in such instances one wishes to examine trees as bits of evidence rather than characters. For example, if one had confidence in several gene trees, one might wish to use these trees to infer a species phylogeny (83, 114). At this point, one has chosen to ignore individual characters, so their weights become irrelevant.

The third category of arguments is that there are many ways to partition all the data, and it is unclear how a particular scheme of partitioning can be justified (14). For example, DNA sequence data might be partitioned into separate genes or by position in the codon, and morphological data might be partitioned into larval versus adult, cranial versus post-cranial, or soft versus hard anatomy. A general response to this argument is that multiple partitions should be investigated to the extent that this is practical. One can draw an analogy with a multiple regression analysis: Factors should be added to the model if they add significantly to its accuracy. Nonetheless, exactly how one should examine the effects of multiple partitions in phylogenetic analysis remains problematic in many cases. However, with certain methods of analysis, the recognition of distinct classes will not always call for separate analyses, thus simplifying the general problem; for example maximum likelihood, neighbor-joining, and weighted parsimony may account for certain distinctions among classes of characters in a single analysis.

The fourth category of arguments in favor of combined analysis involves criticism of the efficacy of consensus methods as a means of producing phylogenetic hypotheses with descriptive and explanatory power. Miyamoto (70) highlighted the fact that a consensus of trees produced by separate analysis of each data set can be less parsimonious than the tree(s) from a combined analysis of the data. He argued that the consensus approach fails to take into account the underlying evidential support for the fundamental trees (i.e. the trees from

separate analyses) and that consensus trees do not represent the best summary of the character information. He therefore recommended that consensus trees not be used in studies of evolution or as a basis for classification. Instead, the tree from the combined analysis is to be preferred as the most efficient summary of the available evidence. This message appears to have been widely appreciated, and consensus trees are now seldom used to portray character evolution. However, one can argue (105) that polytomies in consensus trees resulting from conflict among the fundamental trees, are "soft" polytomies (62), representing the various possible resolutions of the tree with their attendant character optimizations. Under this view, the only situation in which a consensus tree may be considered less parsimonious than a combined tree is when the consensus actually conflicts with the combined tree (as in the example in 5). Nonetheless, the consensus tree might still be considered a less efficient summary of the evidence if it includes more ambiguity than does the combined tree, as is likely (see below).

In a similar vein, Kluge & Wolf (57) pointed to the greater explanatory power of combined trees. Because consensus trees tend to be less resolved than combined trees, the argument goes, the former are worse at explaining the data than the latter. Here, explanatory power is judged by the ability of the phylogenetic hypothesis to explain shared character states as homologies or, equivalently, to avoid ad hoc assumptions of homoplasy (13, 33, 34; see 101 for objections to this characterization of explanatory power).

From our perspective these arguments based on maximizing descriptive and explanatory power suffer from the same deficiency as the "total evidence" argument. Specifically, they assume that maximizing potential instances of homology counted on a character-by-character basis is the sole criterion of descriptive or explanatory power. The best explanation of the data viewed in this narrow sense may not be the best explanation when all evidence relevant to estimating the phylogeny is considered. For example, consider a case in which separate analyses of several unlinked genes all strongly support the same tree, while one gene gives a different tree. A combined parsimony analysis could give the latter tree (if, for example, the one dissenting gene was larger than the other genes combined); however, although requiring the fewest ad hoc assumptions of character homoplasy, this tree would require ad hoc assumptions to explain why nearly all the genes give the same wrong estimate of phylogeny.

The argument in favor of combining that has perhaps received the most attention recently concerns the ability to uncover real phylogenetic groups. Hillis (45) was concerned with the possibility that two data sets/trees might not be positively at odds, yet standard consensus methods (such as strict consensus) might still yield an unresolved tree. He showed several examples involving differential resolution of various parts of the tree by different data

sets (due perhaps to differences in the rate of evolution). As a means of circumventing this problem, he presented a method of consensus that was subsequently formalized by Bremer (8) as "combinable component consensus" (see 105) and is referred to in PAUP (106) as "semi-strict" consensus. The combinable component consensus is a tree that contains any clade found in any fundamental tree that is not contradicted by another fundamental tree. This consensus is thus always at least as resolved as the strict consensus, and often, more resolved. However, we note that the combinable component consensus lacks a property of both strict and majority-rule consensus that may be desirable in certain circumstances, namely, that any clade in the consensus tree must be found in at least two of the fundamental trees, thus reflecting agreement by (presumably) independent sources of data (19).

In the examples used by Hillis (45), it is assumed that the trees from the separate analyses show real and uncontradicted phylogenetic groups that are then hidden by the use of some consensus methods (e.g. strict consensus). However, it is also possible for a combined analysis to resolve conflicts among trees from separate analyses or even to reveal real groups not present in any of the separate trees. The underlying argument is that with an increasing number of characters the phylogenetic signal is more likely to assert itself over the noise, resulting in a more accurate estimate of the true phylogeny (5, 19). In essence, one is reducing sampling error by increasing the number of data points (see "Framework" section below). Simulation studies have shown that a greater number of characters translates into greater accuracy under a wide variety of circumstances (48).

Chippindale & Wiens (14) summarized a variety of cases in which novel phylogenetic results have been obtained in combined analyses, suggestive of this process. A particularly striking example is provided by an analysis of the angiosperm family Solanaceae by Olmstead & Sweere (79), based on three chloroplast DNA data sets. Each of the data sets results in a tree with some elements not seen in the trees derived from the other two data sets. The combined analysis of any two of the data sets yields a tree that has at least one of the unique elements found in the third data set (Figure 1). This implies that there is indeed signal for this arrangement present in these data, but that the signal is masked in some of the individual data sets and not recovered until they are combined. A similar example using artificial data was presented by Barrett et al (5). Another phenomenon consistent with the notion of enhanced
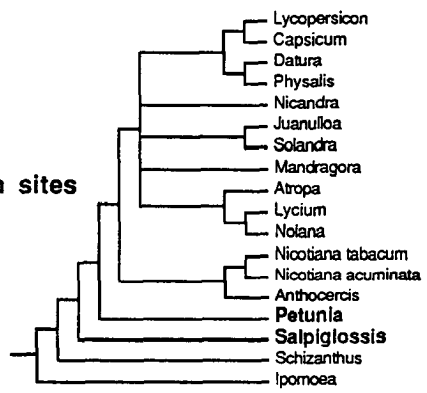
———————————————————————————————→

Figure 1 Strict consensus trees for 17 species of Solanaceae from parsimony analyses of each combination of two data sets from a total of three (*ndh*F and *rbc*L gene sequences, and restriction sites for the entire chloroplast genome). Relationships indicated in boldface were not found through separate analyses of either data set, but were found through analysis of the third data set. Modified from Olmstead & Sweere (79). See text for further explanation.
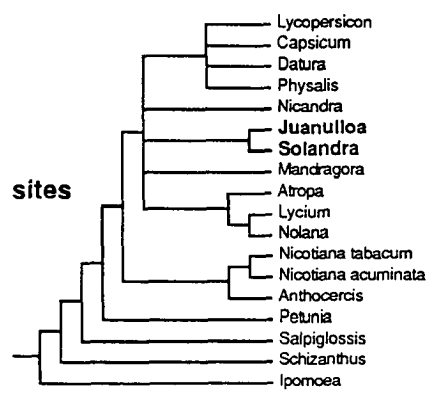
ndhF & rbcL

- Lycopersicon
- Capsicum
- Datura
- Physalis
- Nicandra
- Juanulloa
- Solandra
- Mandragora
- Atropa
- Lycium
- Nolana
- Nicotiana tabacum
- Nicotiana acuminata
- Anthocercis
- Petunia
- Salpiglossis
- Schizanthus
- Ipomoea

ndhF & restriction sites

- Lycopersicon
- Capsicum
- Datura
- Physalis
- Nicandra
- Juanulloa
- Solandra
- Mandragora
- Atropa
- Lycium
- Nolana
- Nicotiana tabacum
- Nicotiana acuminata
- Anthocercis
- Petunia
- Salpiglossis
- Schizanthus
- Ipomoea

rbcL & restriction sites

- Lycopersicon
- Capsicum
- Datura
- Physalis
- Nicandra
- Juanulloa
- Solandra
- Mandragora
- Atropa
- Lycium
- Nolana
- Nicotiana tabacum
- Nicotiana acuminata
- Anthocercis
- Petunia
- Salpiglossis
- Schizanthus
- Ipomoea

signal that is observed frequently is a surprising increase in bootstrap support
for a particular branch in a combined analysis. Such a case was highlighted in
an analysis of seed plant relationships by Doyle et al (28) based on ribosomal
sequences and morphological characters. A "eudicot" clade appeared in the
trees based on morphology, but not in those based on ribosomal sequences.
This clade was found in trees based on the combined data, but with even higher
bootstrap and decay index values than were seen in the morphological trees.
This suggests that signal is present for this clade in the ribosomal data, but
that it is masked until the data sets are combined.

It should be noted that consensus methods that involve recoding trees as
characters (7, 29, 89a, 89b, 113, 115) can also result in such "signal
enhancement." However, because of the probable loss of information asso-
ciated with summarizing a data set as a tree, it seems likely that the effect
will be less frequent than when data sets are combined. Given that our goal
is to uncover phylogenetic relationships, this loss of information may be the
most general argument against consensus and in favor of combining data
sets. To date there has been little attempt to examine the effect of this loss
of information (beyond noting that consensus trees are often less resolved
than combined trees).

Conclusions from Previous Views

Given that our goal in conducting phylogenetic analysis is to discover the true
relationships among the entities in question, arguments against separate analy-
ses based on the principle of total evidence (56, 57), or on preferences for trees
that are bold (13, 57) or efficient descriptions and explanations of data on a
character-by-character basis (57, 70), are not compelling. However, even given
the above goal, legitimate arguments exist on both sides of the debate. The
following points seem especially critical: 1. Combining data sets can give
misleading results if there is heterogeneity among data sets. 2. Combining data
sets can enhance the detection of real groups. In a given case, both combined
and separate analyses can have advantages. Coupled with the fact that inves-
tigators differ in the emphasis they place on high resolution versus avoiding
error, this gives much room for disagreement. Except in extreme cases—where
the benefits of one of the approaches disappear—it is difficult to say what one
should do. This is an important message that seems to have been lost on most
participants in the debate.

Given the potential benefits of both approaches, one "solution" would be
to perform both kinds of analyses (19, 60). However, a better approach may
be to employ nontraditional methods that incorporate the advantages of
combined and separate analyses simultaneously. In the next section, we
highlight a number of such methods that attempt to include "the best of both
worlds."

A CONCEPTUAL FRAMEWORK

The empirical result underlying most arguments in favor of separate analyses is the observation that phylogenetic estimates derived from different sources of data often disagree. As indicated above, Bull et al (10) and Huelsenbeck et al (49) argued that if this disagreement is greater than one would expect from sampling error, then combining the data sets is inappropriate unless the source of the conflict can be identified and accounted for by changing the method of estimation. In this section, we expand upon this framework and argue that the reason for conflict is critical in devising solutions. Although the problem of distinguishing among sources of conflict is an area in need of development (see below), here we assume that one can actually identify the reason(s) for conflict.

Reasons for Conflicting Phylogenies and Possible Solutions

We begin with the assumption that each data set is a sample of the results of a stochastic process that can be specified by a particular tree topology and character change model. The term "character change model" is used broadly here. Examples of differences in the model include overall rate differences, whether different types of character state changes have different probabilities [e.g. Jukes-Cantor (51) vs Kimura (53) models], and differences in branch lengths. Although such models have generally been applied only to molecular data, they could, in theory, apply to other kinds of data as well (e.g. morphological, behavioral, physiological). Under this construction, there are three general, not mutually exclusive, reasons why trees estimated from two or more data sets might differ: 1. sampling error; 2. different stochastic processes acting on the characters; 3. different branching histories.

SAMPLING ERROR   Data sets may be samples from the same tree topology and the same stochastic process and yet give different estimates of phylogeny due purely to sampling error. If sampling error is the only problem, then the data sets should be combined (10). By combining the data, one is increasing the sample size and therefore generally tending to reduce the error around the estimate of phylogeny (48). There seems to be little disagreement on the appropriate course of action in this case (but see 72 for a dissenting view).

This argument assumes that the data are drawn from a distribution such that the estimation method is consistent. If the method is not consistent, then increasing the amount of data may decrease the likelihood of obtaining the correct tree. However, this should not be taken as an argument against combining. The problem here is inconsistency, regardless of whether the data sets are combined or analyzed separately (14), and the solution is to change the method of estimation.

DIFFERENT STOCHASTIC PROCESSES   Data sets may be samples from the same tree topology, but the characters in each set might be affected by distinct stochastic processes (e.g. tend to have different rates of change). It seems clear that if the method of estimation is consistent for one data set and inconsistent for another, then combining the data may lead to problems. Perhaps less obvious is the fact that problems can arise even if the method is consistent for all the data sets. JT Chang (submitted) has shown that with data drawn from different stochastic process distributions, representing what is called a mixture model, even a generally consistent method such as maximum likelihood can be inconsistent. In fact, Steel et al (104) have shown that with general mixture models, for some parameter values every tree will have the same expected data set, which implies that estimation can be impossible regardless of the estimation method. Finally, the sampling error associated with stochastic models over a tree is dependent on the parameter values. Therefore, data drawn from different stochastic process distributions may have different sampling errors. Here the important point is that the sampling error for the combined data may actually be larger than that for some subset of the data (see 10).

A reasonable general strategy in this case is to find the tree (or set of trees) for which the fit combined over all the data sets analyzed separately is the greatest (11). For methods that use a quantitative optimality criterion (e.g. parsimony, maximum likelihood, least squares minimum evolution), the fit of the tree to the data can be assessed using the value of this criterion. The objective function for such methods takes the form f(T,D), where T is the tree topology and D is the data set. The value of this function is the measure of goodness of fit of the tree to the data. For data sets D1, D2,...,Dn the problem is to find the tree such that the combination of the objective functions f(T,D1), f(T,D2),..., f(T,Dn) is optimal.

This general framework leaves open the problem of properly combining the values of this criterion over all of the data sets. For maximum likelihood there is a straightforward solution: The combined optimality criterion should be the sum of the log likelihoods of all the data sets given the tree in question (11). The tree for which this sum is maximized has a straightforward meaning; it is the tree topology that maximizes the joint probability of all data sets.

Cao et al (11) introduced a heuristic use of this approach to estimate phylogenetic relationships among orders of Eutherian mammals. They obtained maximum likelihood trees for each of 13 mitochondrial DNA genes analyzed separately. The model parameter values were allowed to vary among the genes, which is critical because otherwise the analysis does not account for differences in the character change model. For each topology that was the ML estimate for any single gene, the log likelihood for each data set was computed. The tree for which the sum of the 13 log likelihoods was greatest was chosen as the best estimate of the phylogeny. We emphasize that the method is heuristic

because it is possible for some topology other than those that were the ML estimates for single genes to have the greatest summed log likelihood (although not in the specific example analyzed in 11). Ideally one would want to consider all possible topologies.

The approach of Cao et al (see also 112 for a method with some similar properties), with the modification suggested above, has advantages over simply combining the data or using some form of consensus. The method takes into account differences in the character change model, which a simple combined analysis would not do (but see the discussion of weighting below). Differences in the reliability of the data sets should be reflected in their contributions to the overall sum of log likelihoods. The advantage of this approach over consensus is that it does not involve summarizing each data set as a tree, with the probable loss of information that such a summary entails.

A similar approach can be envisioned using parsimony. Under parsimony, the optimality criterion is to minimize tree length, i.e. the number of character steps required. An approach paralleling the maximum likelihood approach above would then be to find the tree(s) for which the length summed over all the data sets is a minimum. Under simple parsimony, length is computed for each character independently and in the same manner for all characters. Thus, the tree that minimizes the sum of lengths of two data sets is also the tree with minimum length for the combined data set. Therefore, it would seem that to achieve the same end as the maximum likelihood approach, we should simply combine the data.

The problem with this parsimony approach is that it does not account for differences in the stochastic processes affecting the characters. When these processes differ it is doubtful whether simple addition of tree lengths is an appropriate procedure. For example, if one data set contains only characters that evolve slowly and show no homoplasy, whereas a second data set contains rapidly evolving, highly homoplasious characters, the meaning of an estimated step in the two data sets is not equivalent. This brings us to what may be a reasonable parsimony solution when stochastic processes differ, namely, differential character weighting (5, 10, 14). Currently, the problem with this procedure under parsimony is that the theoretical justification for any particular scheme of weights has not been well developed (but see, e.g. 32, 36, 40 for some reasonable attempts).

DIFFERENT HISTORIES   Data sets may differ not only in their character change models but in the sequence of branching events they have experienced. In other words, the true tree topology might be different for each data set. Different histories can be the result of differential lineage sorting of ancestral polymorphisms (3, 109; Figure 2a) and/or hybridization/horizontal transfer between taxa (e.g. 52, 66, 100; Figure 2b). Lineage sorting is probably a very
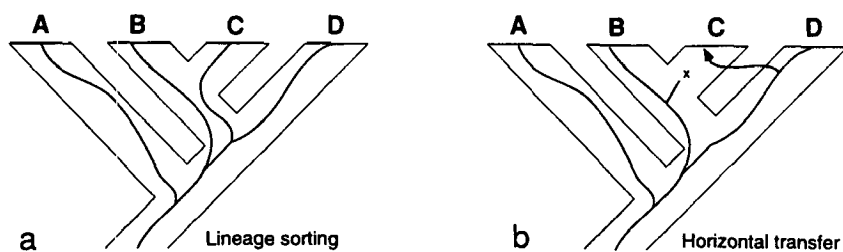
Figure 2 Two processes giving disagreement between gene trees and population trees. The curved lines represent the history of gene lineages and the straight lines represent the history of populations. (a) In the case of lineage sorting, an ancestral polymorphism persists through more than one speciation event, leading to a mismatch between gene and population trees. In this case the gene phylogeny links C and D as sister taxa whereas in the overall population history B and C are sister taxa. (b) In the case of horizontal transfer, discordance results from the transfer of a gene lineage from one population to another. The "x" indicates extinction of a gene lineage. Modified from Maddison (63). See text for further explanation.

common problem for closely related taxa (3, 74, 109), while hybridization/horizontal transfer, although perhaps less common, may be frequent in certain taxa (e.g. many plant groups—103). One of the main points of this review is that conflicts due to different histories present very different analytical problems than those due to different processes of character change. This point has generally been ignored or glossed over in previous considerations of the issue of combined vs separate analysis (e.g. 5, 10, 14, 19, 56).

If data sets have different branching histories, simply combining the data does not solve the problem. That approach is meant to reduce the error around the estimate of a single branching history by compiling as much information as possible from characters that have experienced that history. Allowing for different processes acting on different sets of characters will not solve the problem either; this can account for different expected patterns of character state distribution, but one is still assuming that all characters have experienced the same history. One can imagine a combined analysis of all the data that keeps track of the assignment of each character to its original data set and estimates a tree accounting for the possibility of different histories. However, the development of a generally applicable method of this kind does not appear to be imminent.

When faced with different histories there are at least three obvious ways in which one might represent phylogeny: 1. as the collection of individual histories, 2. as a reticulating tree, 3. as a non-reticulating (i.e. constantly diverging) tree representing the single dominant pattern among data sets. The dominant pattern might be interpreted either as the history of populations/species or just as a description of the central tendency of the distribution of all the histories (64). If there is no correlation among the different histories (i.e.

their topologies are no more similar than randomly chosen topologies), then perhaps only the first option makes sense. However, in most cases data set histories will be largely correlated, and 2. and 3. will be viable options. Here we focus only on the problem of obtaining the single dominant tree. With regard to reticulate trees, some promising methods have been developed (2, 4, 82, 91), but deciding on the number of reticulate connections to allow remains a major difficulty.

To formulate an approach to obtaining the dominant tree, we need to consider the nature of the disagreement among data sets. Specifically, it is important to know whether disagreements are confined to a few taxa or are spread over the tree. The taxon excision and pairwise outlier excision methods described below are designed to deal, respectively, with these two cases. The motivation behind both methods is to allow a single combined analysis while removing data that are misleading with respect to estimating the dominant tree.

Differences that are localized to a small part of the tree can occur, for example, because of lineage sorting associated with a particularly short internal branch near the tips of the tree or through hybridization between a small number of terminal taxa. In such cases, a reasonable procedure may be to excise the taxa involved in the conflict, an idea suggested by Rodrigo et al (93) with a slightly different method than that described here. [Funk (39) and Wagner (110) suggested excising possible hybrids in plants, although not in the context of combining data sets.] A major difficulty is determining which taxa to eliminate. One would begin by testing for significant incongruence among the data sets using a test of heterogeneity (see "Conclusions"). If there were significant incongruence, one would excise the taxon for which a different placement in the estimated trees was most strongly supported, as indicated, for example, by bootstrapping (37) or T-PTP testing (31). One would then cycle back to the heterogeneity test, continuing to excise taxa until there was no significant incongruence among the data sets. The resulting tree from a combined analysis would ideally represent only those taxa for which all data sets had experienced the same history.

Conflicts due to different histories may also be spread widely over the tree. This might occur, for example, if many internal branches were short enough to produce discordances due to lineage sorting, or in situations involving frequent horizontal transfer (e.g. 15). In such cases, excising the "offending" taxa may not be practical since many taxa would have to be eliminated. However, if we have multiple data sets, we can use an approach that relies on the assumption that the relationship between any two taxa will be distorted for only a minority of the data sets. D. Dykhuizen (personal communication to J. Kim) suggested such a method that involves computing distances between pairs of taxa separately for each data set. If a small number of data sets were

involved in a branch rearrangement along the path connecting the two taxa, the distances for those data sets should be outliers in the distribution of distances for all the data sets. The data for such outliers could be eliminated in various ways. If one were using a distance method to estimate the tree, a combined distance for each pair of taxa could be computed excluding the outlier values. Similarly, using a character-based method, one could set the character states in a data set to missing in those taxa for which the data set was an outlier.

Implications

An important implication of this framework is that there is no simple answer to the question of whether to combine data sets for phylogenetic analysis. In some cases, combining all the data, with either equal or unequal weighting of characters, may be appropriate. However, as we have argued above, combining the data implicitly assumes that all data sets are products of the same branching history; when this assumption does not hold, simple combination of all the data may not be the best approach.

Another message of this section is that polarizing the debate into "combining all the data" versus consensus methods obscures some possible solutions. For example, both the taxon excision and pairwise outlier excision methods involve combining information from different data sets, but do not use all the data in the final analysis. Thus they avoid the potential loss of information associated with an intermediate step of summarizing data sets as trees (an advantage of a typical combined analysis), while mitigating the effect of misleading information (an advantage of some consensus methods). The maximum likelihood approach of Cao et al also avoids this loss of information, yet keeps the data sets separate and thus can account for different evolutionary processes acting on them. These methods demonstrate the possibility of incorporating some of the best aspects of traditional combined and separate analyses in one analysis. Nonetheless it is important to note that these methods need to be rigorously tested; it would be premature at this point to claim that they are necessarily superior to more traditional combined or consensus analyses.

A final point concerns the applicability of the framework to nonmolecular data. As suggested above, we believe that, in theory, the framework applies to these kinds of data as well. However, we currently have little idea about what divisions to make in such data. Some traditional divisions may have little relevance in the context of phylogeny estimation (14). An inability to identify relevant classes will clearly limit the kinds of approaches one can take. Our argument, however, is not that one must always partition data sets. Rather the point is that, when faced with biologically defensible divisions, the nature of these partitions should be taken into account in the choice of analytical methods.

CONCLUDING REMARKS

As we have emphasized, there are reasonable arguments in favor of combined analysis of data sets, as well as for separate analysis in some cases. Our goal has been to provide a general conceptual framework within which these options can be better evaluated. This exercise highlights the need for an expanded set of methods to cope with the variety of circumstances that may cause phylogenetically significant heterogeneity among data sets. As indicated above, such methods can incorporate the benefits of traditional combined and separate analyses. Much more effort is also needed to develop and validate specific tests for heterogeneity among data sets (see 60, 69, 93, 105, 107)—tests that are able to pinpoint whether the cause of heterogeneity is, for example, a global difference in evolutionary rate or a highly localized evolutionary event such as hybridization.

In the context of the framework presented above, biogeographic data and data from parasites of the focal organisms may be viewed as peculiar subsets of the problem of diverse data sets. In some cases these data may be products of the exact same branching history as the taxa under study. However, the fact that they may also represent histories somewhat different from the focal taxa should not automatically exclude them from analysis, since methods are available that can mitigate this problem.

A variety of other problems in phylogenetic systematics appear to have a fundamentally similar structure, in that they revolve around the inclusion or exclusion of data from an analysis. For example, there has been controversy over whether (and under what circumstances) fossils should be included in phylogenetic analyses along with information on living organisms (e.g. 24, 47, 78). Likewise there has been disagreement over whether to include the characters being investigated by ecologists and evolutionary biologists using comparative methods (e.g. 9, 18, 65; K de Queiroz, submitted), and whether to include distantly related outgroups (25, 67, 111). Some of the issues raised above are echoed in these controversies. For example, fossils and distant outgroups have been presented as causing more problems than they solve, akin to the idea that a "bad" data set may increase the total sampling error. The exclusion of characters under investigation by comparative biologists has been viewed as conservative, whereas it has been argued that including the characters gives the best estimate of the phylogeny; this distinction mirrors arguments for consensus and combined analyses, respectively. The similarity of these various problems suggests the possibility of a general conceptual framework that would clarify many of the issues involved.

Literature Cited

1. Adams EN. 1972. Consensus techniques and the comparison of taxonomic trees. *Syst. Zool.* 21:390–97
2. Alroy J. 1995. Continuous track analysis: a new phylogenetic and biogeographic method. *Syst. Biol.* 44:152–78
3. Avise JC, Shapira JF, Daniel SW, Aquadro CF, Lansman RA. 1983. Mitochondrial DNA differentiation during the speciation process in *Peromyscus. Mol. Biol. Evol.* 1:38–56
4. Bandelt HJ, Dress AWM. 1989. Weak hierarchies associated with similarity measures—an additive clustering technique. *Bull. Math. Biol.* 51:133–66
5. Barrett M, Donoghue MJ, Sober E. 1991. Against consensus. *Syst. Zool.* 40:486–93
6. Barrett M, Donoghue MJ, Sober E. 1993. Crusade? A response to Nelson. *Syst. Biol.* 42:216–17
7. Baum BR. 1992. Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. *Taxon* 41:3–10
8. Bremer K. 1990. Combinable component consensus. *Cladistics* 6:369–72
9. Brooks DR, McLennan DA. 1991. *Phylogeny, Ecology, and Behavior.* Chicago: Univ. Chicago Press
10. Bull JJ, Huelsenbeck JP, Cunningham CW, Swofford DL, Waddell PJ. 1993. Partitioning and combining data in phylogenetic analysis. *Syst. Biol.* 42:384–97
11. Cao Y, Adachi J, Janke A, Paabo S, Hasegawa M. 1994. Phylogenetic relationships among Eutherian orders estimated from inferred sequences of mitochondrial proteins: instability of a tree based on a single gene. *J. Mol. Evol.* 39:519–27
12. Carnap, R. 1950. *Logical Foundations of Probability.* Chicago: Univ. Chicago Press
13. Carpenter JM. 1988. Choosing among multiple equally parsimonious cladograms. *Cladistics* 4:291–96
14. Chippindale PT, Wiens JJ. 1994. Weighting, partitioning, and combining characters in phylogenetic analysis. *Syst. Biol.* 43:278–87
15. Clark JB, Maddison WP, Kidwell MG. 1994. Phylogenetic analysis supports horizontal transfer of P transposable elements. *Mol. Biol. Evol.* 11:40–50
16. Cracraft J, Helm-Bychowski K. 1991. Parsimony and phylogenetic inference using DNA sequences: some methodological strategies. See Ref. 71, pp. 184–220
17. Cracraft J, Mindell DP. 1989. The early history of modern birds: a comparison of molecular and morphological evidence. See Ref. 38, pp. 389–403
18. Deleporte P. 1993. Characters, attributes and tests of evolutionary scenarios. *Cladistics* 9:427–32
19. de Queiroz A. 1993. For consensus (sometimes). *Syst. Biol.* 42:368–72
20. de Queiroz A, Lawson R. 1994. Phylogenetic relationships of the garter snakes based on DNA sequence and allozyme variation. *Biol. J. Linn. Soc.* 53:209–29
21. de Queiroz A, Wimberger PH. 1993. The usefulness of behavior for phylogeny estimation: levels of homoplasy in

behavioral and morphological charac-
ters. *Evolution.* 47:46–60

22. de Queiroz A, Wimberger PH. 1995.
Comparisons of behavioral and morpho-
logical characters as indicators of phy-
logeny. In *Phylogenies and the
Comparative Method in Animal Behav-
ior,* ed. E Martins. New York: Oxford
Univ. Press. In press

23. de Queiroz K, Gauthier J. 1992. Phylo-
genetic taxonomy. *Annu. Rev. Ecol.
Syst.* 23:449–80

24. Donoghue MJ, Doyle JA, Gauthier J,
Kluge AG, Rowe T. 1989. The impor-
tance of fossils in phylogeny reconstruc-
tion. *Annu. Rev. Ecol. Syst.* 20:431–60

25. Donoghue MJ, Maddison WP. 1986.
Polarity assessment in phylogenetic sys-
tematics: a response to Meacham. *Taxon*
35:534–38

26. Donoghue MJ, Sanderson MJ. 1992.
The suitability of molecular and mor-
phological evidence in reconstructing
plant phylogeny. See Ref. 102, pp. 340–
68

27. Donoghue MJ, Sanderson MJ. 1994.
Complexity and homology in plants. In
*Homology: The Hierarchical Basis of
Comparative Biology,* ed. B Hall, pp.
393–421. San Diego: Academic

28. Doyle JA, Donoghue MJ, Zimmer EA.
1994. Integration of morphological and
rRNA data on the origin of angio-
sperms. *Ann. Mo. Bot. Gard.* 81:419–
50

29. Doyle JJ. 1992. Gene trees and species
trees: molecular systematics as one-
character taxonomy. *Syst. Bot.* 17:144–
63

30. Eernisse DJ, Kluge AG. 1993. Taxo-
nomic congruence versus total evidence,
and amniote phylogeny inferred from
fossils, molecules and morphology. *Mol.
Biol. Evol.* 10:1170–95

31. Faith DP. 1991. Cladistic permutation
tests for monophyly and nonmonophyly.
Syst. Zool. 40:366–75

32. Farris JS. 1977. Phylogenetic analysis
under Dollo's law. *Syst. Zool.* 26:77–88

33. Farris JS. 1979. The information content
of the phylogenetic system. *Syst. Zool.*
28:483–519

34. Farris JS. 1983. The logical basis of
phylogenetic analysis. In *Advances in
Cladistics,* ed. NI Platnick, VA Funk,
2:7–36. New York: Columbia Univ.
Press

35. Felsenstein J. 1978. Cases in which
parsimony or compatibility methods
will be positively misleading. *Syst. Zool.*
27:401–10

36. Felsenstein J. 1981. A likelihood ap-
proach to character weighting and what

it tells us about parsimony and compati-
bility. *Biol. J. Linn. Soc.* 16:183–96

37. Felsenstein J. 1985. Confidence limits
on phylogenies: an approach using the
bootstrap. *Evolution* 39:783–91

38. Fernholm B, Bremer K, Jornvall H.
1989. *The Hierarchy of Life: Molecules
and Morphology in Phylogenetic Analy-
sis.* Amsterdam: Elsevier

39. Funk VA. 1985. Phylogenetic patterns
and hybridization. *Ann. Mo. Bot. Gard.*
72:681–715

40. Goloboff PA. 1993. Estimating charac-
ter weights during tree search. *Cladistics*
9:83–91

41. Good IJ. 1983. *Good Thinking: The
Foundations of Probability and its Ap-
plications.* Minneapolis: Univ. Minn.
Press

42. Graybeal A. 1994. Evaluating the phy-
logenetic utility of genes: a search for
genes informative about deep diver-
gences among vertebrates. *Syst. Biol.*
43:174–93

43. Hempel CG. 1965. *Aspects of Scientific
Explanation and Other Essays in the
Philosophy of Science.* New York: Free
Press

44. Hennig W. 1966. *Phylogenetic Sys-
tematics.* Urbana: Univ. Ill. Press

45. Hillis DM. 1987. Molecular versus mor-
phological approaches to systematics.
Annu. Rev. Ecol. Syst. 18:23–42

46. Hoch PC, Stephenson AG, ed. 1995.
*Experimental and Molecular Ap-
proaches to Plant Biosystematics.* St.
Louis: Monogr. Syst., Mo. Bot. Gard.

47. Huelsenbeck JP. 1991. When are fossils
better than extant taxa in phylogenetic
analysis? *Syst. Zool.* 40:458–69

48. Huelsenbeck JP, Hillis DM. 1993. Suc-
cess of phylogenetic methods in the
four-taxon case. *Syst. Biol.* 42:247–64

49. Huelsenbeck JP, Swofford DL, Cun-
ningham CW, Bull JJ, Waddell PJ. 1994.
Is character weighting a panacea for the
problem of data heterogeneity in phy-
logenetic analysis? *Syst. Biol.* 43:288–
91

50. Jones TR, Kluge AG, Wolf AJ. 1993.
When theories and methodologies clash:
a phylogenetic reanalysis of the North
American ambystomatid salamanders
(Caudata: Ambystomatidae). *Syst. Biol.*
42:92–102

51. Jukes TH, Cantor CR. 1969. Evolution
of protein molecules. In *Mammalian
Protein Metabolism,* ed. HN Munro, pp.
21–132. New York: Academic

52. Kidwell MG. 1993. Lateral transfer in
natural populations of eukaryotes. *Annu.
Rev. Genet.* 27:235–56

53. Kimura M. 1980. A simple method for

estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16: 111–20

54. Kluge AG. 1983. Cladistics and the classification of the great apes. In *New Interpretations of Ape and Human Ancestry,* ed. RL Ciochon, RS Corruccini, pp. 151–77. New York: Plenum

55. Kluge AG. 1984. The relevance of parsimony to phylogenetic inference. In *Cladistics: Perspectives on the Reconstruction of Evolutionary History,* ed. T Duncan, TF Stuessy, pp. 24–38. New York: Columbia Univ. Press

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

57. Kluge AG, Wolf AJ. 1993. Cladistics: What's in a word? *Cladistics* 9:183–99

58. Lanyon SM. 1993. Phylogenetic frameworks: towards a firmer foundation for the comparative approach. *Biol. J. Linn. Soc.* 49:45–61

59. Larson A. 1991. Evolutionary analysis of length variable sequences: divergent domains of ribosomal DNA. See Ref. 71, pp. 221–48

60. Larson A. 1994. The comparison of morphological and molecular data in phylogenetic systematics. In *Molecular Ecology and Evolution: Approaches and Applications,* ed. B Schierwater, B Streit, GP Wagner, R DeSalle, pp. 371–90. Basel, Switzerland: Birkhauser Verlag

61. Li W-H, Graur D. 1991. *Fundamentals of Molecular Evolution.* Sunderland, Mass: Sinauer

62. Maddison WP. 1989. Reconstructing character evolution on polytomous cladograms. *Cladistics* 5:365–77

63. Maddison WP. 1995. The growth of phylogenetic biology. In *Molecular Approaches to Zoology and Evolution,* ed. J Ferraris. New York: Wiley In press

64. Maddison WP. 1995. Phylogenetic histories within and among species. See Ref. 46, pp. 273–87

65. Maddison WP, Maddison DR. 1992. *MacClade: Interactive Analysis of Phylogeny and Character Evolution, Version 3.0.* Sunderland, Mass: Sinauer

66. McDade LA. 1992. Hybrids and phylogenetic systematics. II. The impact of hybrids on cladistic analysis. *Evolution* 46:1329–46

67. Meacham CA. 1986. More about directed characters: a reply to Donoghue and Maddison. *Taxon* 35:538–40

68. Mickevich MF. 1978. Taxonomic congruence. *Syst. Zool.* 27:143–58

69. Mickevich MF, Farris JS. 1981. The implications of congruence in Menidia. *Syst. Zool.* 30:351–70

70. Miyamoto MM. 1985. Consensus cladograms and general classifications. *Cladistics* 1:186–89

71. Miyamoto MM, Cracraft J, ed. 1991. *Phylogenetic Analysis of DNA Sequences.* New York: Oxford Univ. Press

72. Miyamoto MM, Fitch WM. 1995. Testing species phylogenies and phylogenetic methods with congruence. *Syst. Biol.* 44:64–76

73. Moritz C, Dowling TE, Brown WM. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu. Rev. Ecol. Syst.* 18:269–92

74. Nei M. 1987. *Molecular Evolutionary Genetics.* New York: Columbia Univ. Press

75. Nelson G. 1979. Cladistic analysis and synthesis: principles and definitions, with a historical note on Adanson's Familles Des Plantes (1763–1764). *Syst. Zool.* 28:1–21

76. Nelson G. 1993. Why crusade against consensus? A reply to Barrett, Donoghue, and Sober. *Syst. Biol.* 42:215–16

77. Nelson G, Platnick NI. 1981. *Systematics and Biogeography: Cladistics and Vicariance.* New York: Columbia Univ. Press

78. Novacek MJ. 1992. Fossils as critical data for phylogeny. In *Extinction and Phylogeny,* ed. MJ Novacek, QD Wheeler, pp. 46–88. New York: Columbia Univ. Press

79. Olmstead RG, Sweere JA. 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. *Syst. Biol.* 43:467–81

80. Page RDM. 1988. Quantitative cladistic biogeography: constructing and comparing area cladograms. *Syst. Zool.* 37:254–70

81. Page RDM. 1989. Comments on component-compatibility in historical biogeography. *Cladistics* 5:167–82

82. Page RDM. 1994. Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. *Syst. Biol.* 43:58–77

83. Pamilo P, Nei M. 1988. Relationships between gene trees and species trees. *Mol. Biol. Evol.* 5:568–83

84. Patterson C, ed. 1987. *Molecules and Morphology in Evolution: Conflict or Compromise?* Cambridge: Cambridge Univ. Press

85. Patterson C, Williams DM, Humphries CJ. 1993. Congruence between molecu-

lar and morphological phylogenies. *Annu. Rev. Ecol. Syst.* 24:153–88

86. Penny D, Foulds LR, Hendy MD. 1982. Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. *Nature* 297:197–200

87. Penny D, Hendy MD. 1986. Estimating the reliability of evolutionary trees. *Mol. Biol. Evol.* 3:403–17

88. Popper KR. 1965. *Conjectures and Refutations: The Growth of Scientific Knowledge.* New York: Harper & Row

89. Popper KR. 1968. *The Logic of Scientific Discovery.* New York: Harper & Row

89a. Purvis A. 1995. A modification to Baum and Ragan's method for combining phylogenetic trees. *Syst. Biol.* 44:251–55

89b. Ragan MA. 1992. Phylogenetic inference based on matrix representation of trees. *Mol. Phylog. Evol.* 1:53–58

90. Rieseberg LH, Brunsfeld S. 1992. Molecular evidence and plant introgression. See Ref. 102, pp. 151–76

91. Rieseberg LH, Morefield JD. 1995. Character expression, phylogenetic reconstruction, and the detection of reticulate evolution. See Ref. 46, pp. 333–53

92. Rieseberg LH, Soltis DE. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evol. Trends Plants* 5:65–84

93. Rodrigo AG, Kelly-Borges M, Bergquist PR, Bergquist PL. 1993. A randomization test of the null hypothesis that two cladograms are sample estimates of a parametric phylogenetic tree. *NZ J. Bot.* 31:257–68

94. Rosen DE. 1978. Vicariant patterns and historical biogeography. *Syst. Zool.* 27:159–88

95. Sanderson MJ, Donoghue MJ. 1989. Patterns of variation in levels of homoplasy. *Evolution* 43:1781–95

96. Shaffer HB, Clark JM, Kraus F. 1991. When molecules and morphology clash: a phylogenetic analysis of North American ambystomatid salamanders (Caudata: Ambystomatidae). *Syst. Zool.* 40:284–303

97. Sheldon FH, Bledsoe AH. 1993. Avian molecular systematics, 1970s to 1990s. *Annu. Rev. Ecol. Syst.* 24:243–78

98. Sibley CG, Ahlquist JE. 1987. Avian phylogeny reconstructed from comparisons of the genetic material, DNA. See Ref. 84, pp. 95–121

99. Simberloff D. 1987. Calculating probabilities that cladograms match: a method of biogeographic inference. *Syst. Zool.* 36:175–95

100. Smith MW, Feng D-F, Doolittle RF. 1992. Evolution by acquisition: the case for horizontal gene transfers. *Trends Biochem. Sci.* 17:489–93

101. Sober E. 1988. *Reconstructing the Past: Parsimony, Evolution, and Inference.* Cambridge, Mass: MIT Press

102. Soltis PS, Soltis DE, Doyle JJ, ed. 1992. *Molecular Systematics in Plants.* New York: Chapman & Hall

103. Stebbins GL. 1950. *Variation and Evolution in Plants.* New York: Columbia Univ. Press

104. Steel MA, Szekely LA, Hendy MD. 1994. Reconstructing trees when sequence sites evolve at variable rates. *J. Comp. Biol.* 1:153–63.

105. Swofford DL. 1991. When are phylogeny estimates from molecular and morphological data incongruent? See Ref. 71, pp. 295–333

106. Swofford DL. 1993. PAUP: *Phylogenetic Analysis Using Parsimony, Version 3.1.* Champaign: Illinois Nat. Hist. Surv.

107. Swofford DL. 1994. *Inferring phylogenies from combined data. Look before you leap.* Presented at Annu. Meet. Soc. Syst. Biol., Athens, Georgia

108. Sytsma KJ. 1990. DNA and morphology: inference of plant phylogeny. *Trends Ecol. Evol.* 5:104–10

109. Tajima F. 1983. Evolutionary relationships of DNA sequences in finite populations. *Genetics* 105:437–60

110. Wagner WH. 1980. Origin and philosophy of the groundplan-divergence method of cladistics. *Syst. Bot.* 5:173–93

111. Wheeler WC. 1990. Nucleic acid sequence phylogeny and random outgroups. *Cladistics* 6:363–67

112. Wheeler WC. 1991. Congruence among data sets: a Bayesian approach. See Ref. 71, pp. 334–46

113. Wiley EO. 1987. Methods in vicariance biogeography. In *Systematics and Evolution: A Matter of Diversity,* ed. P Hovenkamp, pp. 283–306. Utrecht: Inst. Syst. Bot.

114. Wu C-I. 1991. Inferences of species phylogeny in relation to segregation of ancient polymorphisms. *Genetics* 127:429–35

115. Zandee M, Roos MC. 1987. Component-compatibility in historical biogeography. *Cladistics* 3:305–32

CONTENTS

xii   CONTENTS   (continued)