Annu. Rev. Ecol. Syst. 2002. 33:49–72
doi: 10.1146/annurev.ecolsys.33.010802.150509

First published online as a Review in Advance on August 6, 2002

Troubleshooting Molecular Phylogenetic Analyses

Michael J. Sanderson and H. Bradley Shaffer
Section of Evolution and Ecology, University of California, Davis, California 95616; email: mjsanderson@ucdavis.edu, hbshaffer@ucdavis.edu

Key Words algorithm, model, data partition, rooting, long-branch attraction

■ **Abstract** The number, size, and scope of phylogenetic analyses using molecular data has increased steadily in recent years. This has simultaneously led to a dramatic improvement in our understanding of phylogenetic relationships and a better appreciation for an array of methodological problems that continue to hinder progress in phylogenetic studies of certain data sets and/or particular parts of the tree of life. This review focuses on several persistent problems, including rooting, conflict among data sets, weak support in trees, strong but evidently incorrect support, and the computational issues arising when methods are applied to the large data sets that are becoming increasingly commonplace. We frame each of these issues as a specific problem to be overcome, review the relevant theoretical and empirical literature, and suggest solutions, or at least strategies, for further investigation of the issues involved.

INTRODUCTION

The explosive growth in the number of molecular phylogenetic studies can be attributed to two advances: the development of rigorous tree-building methods, and the advent of rapid and relatively inexpensive technology for obtaining comparative molecular data. However, the former predates the latter by some time. Parsimony, likelihood, and distance methods for building trees were all introduced and applied to real data in the 1960s (Edwards & Cavalli-Sforza 1967), whereas the explosive growth of GenBank did not begin until the late 1980s. Consequently, the literature on phylogenetic methods is now extensive, with important work appearing not just in journals devoted to systematics and molecular evolution, but also genetics, mathematics, and computer science. In the past decade the mathematical complexity of phylogenetic methodology has taken a quantum leap forward owing to a determined effort to place much of phylogenetic inference in a statistical framework (e.g., Huelsenbeck & Crandall 1997 for review; Felsenstein 2001). A less well advertised but equally significant infusion of ideas from computer science (e.g., Nakleh et al. 2001) has also impacted theoretical phylogenetics. Together these methodological advances have presented investigators who are now finding

0066-4162/02/1215-0049$14.00

it easy to obtain molecular sequence data with an unprecedented and often bewildering set of choices on methodological issues.

Acquiring comparative sequence data is now as simple as accessing GenBank or other databases via the internet, and the number of scientists interested in reconstructing phylogenies as a secondary research interest is growing rapidly. Whereas the inclusion of molecular, developmental, and evolutionary biologists into the world of practical phylogenetics is clearly beneficial for all, a growing knowledge gap between the leading edge of phylogenetic theory and this growing pool of relative novices has developed. Within our own university, and in our interactions with a wide range of colleagues not formally trained in phylogenetic methods, we find ourselves constantly asked how to analyze data sets, what strategies are sufficient, and how to deal with different sorts of comparative information. This review focuses on several issues that repeatedly confront molecular phylogenetic studies, and our view of the recent literature on solutions to these problems. We aim at the relative newcomer to the field who desires advice on troubleshooting rather than an in-depth understanding of all the technical details. The risk of this strategy is that it may give a false impression: Hard problems do not always have simple solutions. Still, the resolution of many problems seems to be taking shape, and solutions to others can be narrowed in scope to a handful of analytical strategies.

Below we list a series of commonly cited problems, and discuss potential strategies for overcoming them. Among the most important of these have been questions about the performance and accuracy of tree-building methods, whether and when to combine disparate sources of data, how to assess levels of support for trees, and how to analyze large data sets with algorithms that often require embarrassingly long running times.

PARTS OF THE TREE SEEM TO BE "WRONG"

Sometimes a phylogenetic analysis will imply relationships that appear dead wrong in the context of previously published results or long-held views. A celebrated example in vertebrates is the placement of bony fish within tetrapods in an analysis of complete mitochondrial genomes (Naylor & Brown 1998). An equally pathological case in plants is the placement of various aquatic or parasitic angiosperms as the solitary sister group to all other angiosperms (*Ceratophyllum*: Chase et al. 1993, *Rafflesia*: Lipscomb et al. 1998). Of course, the possibility always exists that the widely held view is in fact wrong, and that a new phylogenetic paradigm has been uncovered, as appears to be the case for turtles within the amniotes (Rieppel & Reisz 1999), but weight of evidence suggests such cases are probably the exception. Because weak data and sampling error can always lead to apparently incorrect results, the onus is on the investigator to document that a result is both strongly supported and apparently incorrect. Troubleshooting such cases then involves demonstrating that a relationship is questionable and attempting to overcome the cause of the mistaken inferences. This has been an area of active interchange between theoreticians and empirical phylogeneticists and some recommendations have emerged.

Check for Pre-Processing Errors: Alignment, Orthology/Paralogy

The assumption of all phylogenetic inference methods is that the input data consist of a set of homologous characters. For sequence data this means that the sequences are homologs and that their sites have been correctly aligned. As an analytical problem, multiple sequence alignment is of the same order of complexity as phylogenetic inference (Gusfield 1997), yet it rarely receives the same level of attention from phylogeneticists (Lee 2001). Problems can range from fairly trivial to quite serious. Manual sequence alignments, which are commonplace in studies of noncoding DNA data, for example, sometimes can have a misplaced or mistakenly inferred gap that "frameshifts" a large part of a sequence or set of sequences. This can introduce a large number of substitutions into inferred trees, producing false clades that appear strongly supported. Routinely checking alignments with alignment programs (e.g., CLUSTALW; Thompson et al. 1994) can help identify such cases.

Other problems are stickier. Certain genes are notoriously difficult to align either manually or with the assistance of algorithms (e.g., small subunit rDNA), and alignment choices clearly impact phylogenetic inference strongly (Morrison & Ellis 1997). Little consensus is available regarding strategies for alignment in problematic cases (see review in Lee 2001). Three possibilities are (*a*) removing problematic sites, (*b*) concatenating sequences derived from multiple alignment replicates with different parameters (Elision method: Wheeler et al. 1995), and (*c*) extracting common phylogenetic information from trees constructed under different alignments (multiple analysis method: Lee 2001). Of these, removing problematic regions of the alignment is the most conservative but may lead to a loss of resolution.

Phylogenetic error can also arise if paralogs (genes derived via gene duplication) are mistakenly interpreted to be orthologs (genes derived via speciation) in the construction of species trees from sequence data. Paralogs and orthologs are all homologous, but species relationships can only be directly inferred from orthologous sequences (though indirect inferences are possible; see Page & Charleston 1997). Orthology is difficult to discern from sequence data alone, but introns and flanking regions can leave telltale clues. For example, the number and position of introns is a useful diagnostic for distinguishing paralogs from orthologs in the *Adh* gene family of cotton (Small & Wendel 2000). Formal algorithms for inferring species trees from gene family trees (Page & Cotton 2002), are available but are very sensitive to the completeness of sampling of the gene family, a problem with many data sets. The complexities of working with gene families partly explain the popularity of organellar single copy genes in phylogenetics. However, as the data from relatively small organellar genomes is exploited fully, and additional complete nuclear genomes become available, the impetus to identify and use single-copy nuclear genes free of paralogy problems will undoubtedly increase.

Sample Additional Genes with Different Substitution Rates

Yang (1998) and Bininda-Emonds et al. (2001) showed via simulation studies that accuracy—the probability of inferring the correct or nearly correct tree—is maximized at an intermediate nucleotide substitution rate; it drops quickly at lower rates, and more slowly at higher rates. Optimal rates are dependent on the details of the substitution process of the gene sampled and therefore must themselves be estimated for any given data set. However, a useful strategy might be to compare a data set's bootstrap support to that obtained with a faster gene and a slower gene. Given that "too fast" is better than "too slow," a rule of thumb might be to always add a gene with higher rate first. Because rates vary between lineages in most data sets, however, it is necessary to estimate an average rate to permit comparisons between genes. Conventionally, such comparisons are made by reference to a single comparison between two terminal taxa. A more comprehensive estimate of mean rate can be obtained (in PAUP*, for example, Swofford 1999) by rooting a tree, enforcing a molecular clock, obtaining branch lengths, and summing these along one path from root to tip. This value can then be compared to those obtained for the same taxa for other genes. If the age of the root is known it can be put in units of substitutions per site per unit time, but this is not necessary for comparisons between genes. An important caveat is that rates that are too high can lead to long-branch attraction (see next).

Check if Rate Heterogeneity is Sufficient to Cause Long-Branch Attraction

The most widely cited reason for mistaken phylogenetic inferences is long-branch attraction (LBA). Strictly speaking, LBA is an "asymptotic" property: a failure of a method to reconstruct the right tree as the number of characters goes to infinity. This is also known as statistical inconsistency. First noticed for maximum parsimony (MP; Felsenstein 1978), it is now understood to be a problem for parametric methods as well, including maximum likelihood (ML) and neighbor-joining (NJ) using parametric corrections, when model assumptions are violated (Lockhart et al. 1994, Chang 1996). Just as a model of evolution can be constructed in which ML is consistent and MP is not, models containing mixtures of simple substitution models can be constructed in which MP is consistent and ML is not (J. Kim, personal communication). The term long-branch attraction refers to a set of model conditions known to make MP fail. High rates along two branches separated by a branch with low rate will fool MP into joining the long branches together. The LBA problem shows up in many guises, including tree rooting and base compositional bias (see below); it is one of the more pervasive problems in phylogenetics.

However, it is also difficult to document. Huelsenbeck (1997) argued that it is not enough to point to "long" branches in the vicinity of a suspect relationship and suggested two strategies for implicating LBA rigorously. One is to test whether branches are long enough and heterogeneous enough to cause mistaken inferences in a data set given the number of characters observed. The procedure is to choose a model tree, infer reasonable model parameters (including branch lengths) using ML, and then conduct replicate simulations with those parameter values to determine whether or not the tree would be reconstructed with high probability (Huelsenbeck 1997, Maddison et al. 1999, Sanderson et al. 2000, Omilian & Taylor 2001). The Seq-Gen program (Rambaut & Grassly 1997) is an excellent tool for conducting these and other tree-based simulations.

A second strategy is to try different inference methods. In theory, if the model is close to being correct, ML and NJ should be less sensitive to long-branch attraction than MP. However, care should be taken in cases in which MP provides different answers than ML/NJ to make sure the differences are strongly supported—i.e., that the explanation is not just sampling error. The issue of model selection is also critical (Posada & Crandall 2001). The number of substitution models for DNA sequence data is quite large, and minimally one should explore automated model selection methods such as those available in MODELTEST (Posada & Crandall 1998), which balance model complexity with improved likelihood. A nonparametric alternative is provided by Willson's Higher Order Parsimony method (Willson 1999: software available on author's website), which uses characters conventionally viewed as uninformative in standard MP analysis (autapomorphies and constant characters), just as ML does. It is limited to four-taxon trees but may be useful as part of more general quartet methods in larger trees. Simulations suggest it outperforms MP under classical LBA conditions and ML when the model used in ML is incorrect.

Assess the Extent of Base Composition Bias and Heterogeneity

Some genes and genomes are biased in favor of AT or GC base compositions. For example, many animal mitochondrial genomes are AT rich, and this has been proposed as a partial explanation for the occasional reconstruction of incorrect relationships (Steel et al. 1993). Although Conant & Lewis (2001) conducted simulations suggesting that bias must be quite strong before standard methods would be misled, compositional bias has been blamed for many aberrant results, such as the pairing of honeybee and nematode complete mitochondrial protein sequences to the exclusion of other insects (Foster & Hickey 1999), and the removal of *Drosophila willistoni* from its "true" subgenus because its low GC content is similar to more distant outgroups (Tarrio et al. 2001). Inspection of simple base composition statistics provided by many programs (e.g., TREE-PUZZLE, Strimmer & Von Haeseler 1996; PAUP*, Swofford 1999), including χ^2 tests of homogeneity across taxa, should indicate whether compositional bias is a potential issue. More sensitive likelihood ratio tests can then be undertaken to confirm that differences in substitution pattern among extant taxa indicate changes in base composition deeper in the tree (Tarrio et al. 2001). This compares a standard stationary homogeneous model [e.g., the general time-reversible model (GTR); Swofford et al. 1996], to a nonstationary, nonhomogeneous model (Galtier & Gouy 1998).

Possible solutions to the problems induced by composition bias include transforming the data to transversions only (Woese et al. 1991) or translating to amino acid sequences (Hasegawa & Hashimoto 1993). However, Foster & Hickey (1999) showed that compositional bias can propagate to the protein level, producing similar artifacts. Some reconstruction methods are explicitly designed to be robust to composition bias, including distance methods using LogDet distances (Lockhart et al. 1994), modifications of the Tamura (1992) distance that take composition variation into account (Galtier & Gouy 1995), and ML inference with a nonstationary model (Galtier & Gouy 1998). Results from these methods have been mixed, however, with some workers reporting that they helped (Tarrio et al. 2001) and others reporting no improvement (e.g., Foster & Hickey 1999). One problem is that multiple biases may be present and interact, and these problems may be confounded with LBA issues (Whitfield & Cameron 1998, Foster & Hickey 1999). A particularly vexing factor is site-to-site rate variation. Tree reconstruction is exceptionally problematic in the face of both changing base composition through time and different rates between sites (Baake 1998). Until further progress is made with reconstruction methods, simple composition variation can be identified and dealt with, but complex bias remains a difficult obstacle to overcome.

Test for a Covariotide/Covarion Effect

Evidence is mounting that rates of molecular evolution vary not only between sites and between lineages but between both simultaneously, an idea first put forth by Fitch & Markowitz (1970) as the covarion hypothesis (for proteins, covariotides for DNA). An extreme manifestation occurs when some sites are apparently unable to vary in certain taxa but are free to vary in others and vice versa. Several recent studies of deep phylogenies have uncovered such patterns in rDNA (Philippe & Germot 2000) and protein coding genes (Lockhart et al. 1998, 2000; Lopez et al. 1999). Even maximum likelihood inference with complex models can be misled by sequences evolving according to this model (Lockhart et al. 1996). Lockhart and colleagues (Lockhart et al. 1998) provide a relatively simple inequality test for covariotide patterns based on the expected distribution of invariant sites in two sets of sequences. Omilian & Taylor (2001) used this test on large subunit rDNA in the relatively shallow phylogeny of *Daphnia* and found evidence of a covariotide pattern, but they were unable to overcome its effects by removing misleading character classes.

BOOTSTRAP VALUES ARE LOW

Molecular phylogenies sometimes contain clades with low statistical support, even with large amounts of sequence data. Support can be assessed in a variety of ways, but bootstrap resampling (Felsenstein 1985, Efron et al. 1996) is the most commonly used method for molecular data, and its assumptions and interpretation have been scrutinized more intensely than any other (Hillis & Bull 1993, Felsenstein & Kishino 1993, Sanderson 1995, Efron et al. 1996). The general strategy is to resample, with replacement, a data matrix of aligned sequences, where each site (character, column) is a sampling unit. Pseudoreplicate data sets of the same size as the original are constructed and analyzed, and the resulting trees are summarized as a consensus of generally >100 replicates. The bootstrap proportion (BP) has been interpreted in at least two ways: (*a*) as a measure of accuracy, which is the probability that repeated sampling with the same sample size of real data from the universe of characters would recover the true clade; and (*b*) as an estimate of $1 - \alpha$, where α is the conventional type I error (the probability of mistakenly concluding the group is a clade when it is not). Other interpretations of BPs are also possible, including Bayesian ones (see Efron et al. 1996).

Simulation and theory indicate that BP tends to underestimate accuracy when the clade is correct (Zharkikh & Li 1992a,b; Hillis & Bull 1993; Felsenstein & Kishino 1993; Newton 1996; Efron et al. 1996), but that to a first order approximation it does estimate the $1 - \alpha$ significance level (Efron et al. 1996). However, a number of factors can make the latter approximation poor, necessitating second-order corrections. Low BPs may also arise for reasons other than statistical biases. Felsenstein (1985) showed that in the absence of homoplasy BPs are positively correlated with the number of characters changing along the branch subtending a clade, and meta-analyses of real data sets indicate this is true in the presence of homoplasy as well (Sanderson & Donoghue 1996, Bremer et al. 1999). Thus, gathering additional sequence data for the same taxa is one potential cure for low BPs. Slightly less obvious strategies are highlighted below.

Use Bootstrap Correction Factors

Several modified bootstrap procedures have been suggested to obtain BPs closer to true accuracy, including iterated bootstrapping (Rodrigo et al. 1993), the complete-and-partial bootstrap (Zharkikh & Li 1992a) and the approximately unbiased (AU) method (Shimodaira & Hasegawa 2001). Efron et al. (1996) also describe a kind of iterated bootstrap method that explicitly estimates type I error by resampling from data sets constructed to lie on the boundary between regions of tree space that have and do not have the clade of interest. Felsenstein's simple bootstrap method requires some number, K, of iterations, where K is usually >100. These more complex methods often require on the order of K^2 iterations, and thus are computationally very expensive. This is unfortunate given that the bias in BP appears to be exacerbated in intensively sampled clades of large trees that require heavy computation in every replicate (Sanderson & Wojciechowski 2000). Some software is available (e.g., Shimodaira & Hasegawa 2001), but none of these methods are currently implemented in widely used phylogenetics packages.

Check for Rogue Taxa

Because a bootstrap proportion refers to a precise hypothesis about a potentially long list of clade members, it is very sensitive to one or a few "rogue" taxa whose position is unstable. They may be unstable because of missing data, an elevated substitution rate causing homoplasy, or extremely low rates inside and outside the clade, all of which can cause low BPs. Part of the problem is that exact monophyly of a group is a very specific hypothesis, which is all too easily rejected in cases of even a single unstable taxon. Sanderson (1989) suggested relaxing the notion of strict monophyly by putting confidence bounds on *r*, the minimum number of species needed to make a paraphyletic group monophyletic. Wilkinson (1994, 1996) suggested putting (bootstrap) confidence levels on *n*-taxon statements rather than clades. An *n*-taxon statement is a statement of relationship about a subset of taxa on a tree. For example, on the tree (*a*, *b*, *c*, *d*, (*e*, *f*, *g*, *h*)), an example of a 4-taxon statement is (*a*, *b*, (*e*, *f*)). The support for this 4-taxon statement might well be much higher than it is for the monophyly of (*e*, *f*, *g*, *h*), especially if taxon *g*, for example, is problematic. Thorley & Wilkinson (1999) describe a test based on triplets (3-taxon statements) specifically designed to assess the stability of individual taxa (implemented in RadCon: Thorley & Page 2000).

A more general solution is to find maximum agreement subtrees (Finden & Gordon 1985, Steel & Penny 2000) in the collection of bootstrap trees. These are sets of smaller trees constructed by pruning a minimum number of rogue taxa such that the relationships among the remaining taxa agree among all trees. Any clades on the output trees are then supported at the 100% level, although the original tree might not have had any clades supported at this level. Better still would be a fuzzier version of the maximum agreement subtree that returned a majority rule consensus with values higher than that on the original large tree. These approaches extract signal about some relationships at the expense of remaining agnostic about others.

Experiments with taxon sampling from the data available to the investigator can also provide hints about the robustness of BPs. Many studies have found BPs are inversely correlated with the number of taxa (Bremer et al. 1999) or that adding taxa to a study decreases BPs (Sanderson & Wojciechowski 2000, Mitchell et al. 2000), although Omland et al. (1999) found that including more sequences from within each species improved BPs.

Avoid Using "Fast" Bootstrap Methods in Large Trees (if Possible)

In larger data sets, it is common to use heuristic shortcuts in search strategies, such as in PAUP's "fast bootstrap" option, which perform less rigorous searches in the interest of saving computer time. Debry & Olmstead (2000) performed an exhaustive series of simulations, and showed that these strategies lower BPs except in extremely well-supported clades (BP > 90%), and that the deterioration worsened as the number of taxa increased. Mort et al. (2000) found similar results in two real data sets, although those in data clades with BPs of 80% or more were relatively immune. To avoid this deterioration, increasing the stringency of heuristic searches to include at least some minimal branch swapping may be necessary.

When Bootstrapping in Maximum Likelihood Analyses, Experiment with Model Complexity

Bootstrapping in conjunction with maximum likelihood inference is computationally expensive but is a viable option if the number of taxa is not too large (perhaps less than about 25 sequences using commonly available hardware). It is generally assumed that increasing model complexity decreases bias but at the expense of increasing variance (Berry & Gascuel 1996, Burnham & Anderson 1998). Thus, as the number of parameters increases BPs might be expected to decline (Berry & Gascuel 1996). However, Buckley et al. (2001) found little evidence of a consistent pattern in which parameter-rich models had lower BPs than parameter-poor models, although they did find large differences in BPs between models. Computational limitations when bootstrapping ML runs make exploration of the bias-variance tradeoff difficult, and much more work is needed to determine in what direction model choice should be aimed.

SIGNALS FROM DIFFERENT DATA SETS OR PARTITIONS CONFLICT

Since molecular data (including allozymes and DNA sequences) became an important element of phylogenetics, the question of how to best utilize the phylogenetic information in different data partitions has loomed large on the horizon. The problem of conflicts among characters in phylogenetics goes back at least to Hennig (1966), who felt that character conflicts must represent investigator error in making homology statements. However, the current methodological attention to the resolution of conflicts among data partitions dates to Kluge (1989). In that key paper, Kluge made at least two important observations and recommendations. First, to increase the likelihood of character independence, different classes, or kinds of characters should be sampled for a given phylogeny. Second, these different data "partitions" should be combined into a single analysis (often referred to as a "total evidence" approach), rather than testing each partition separately for the presence of different signals. In response to this view, a decade of research has led to new tools for determining whether different partitions have divergent phylogenetic signals, and (to a lesser extent) methods for treating data incongruence. For the practicing phylogeneticist, at least three problems regularly arise. First, how does one determine if data partitions are in conflict? Second, if there is a conflict, how should one obtain a single, best tree from the conflicting data? Third, are there empirical or theoretical guidelines for when one should worry about single-partition trees, given that most practitioners still have data from only a single gene? That is, under some conditions one may feel quite confident that single-gene trees represent true "species trees," whereas in other cases one may feel very uncomfortable with only a single data partition.

Do Different Data Partitions Agree or Disagree in Their Phylogenetic Information?

Over the past two decades, a number of strategies have been proposed to examine the level of disagreement among data partitions. Swofford (1991) provides an excellent review of the first decade. Of the several methods that have stood the test of time, the incongruence length difference test (ILD; also known as the partition homogeneity test), which was originally outlined some 20 years ago (Mickevich & Farris 1981; see also Swofford 1991) is by far the most widely used. The ILD is based on the difference in tree length (that is, the total number of inferred changes under parsimony) between a single tree in which two data partitions are combined, compared to the sum of the tree lengths of each partition on its own maximum parsimony tree. Thus,

$$D_{XY} = L_{(X+Y)} - (L_X + L_Y),$$

where D_{XY} is the length difference, $L_{(X+Y)}$ is the total length of the tree where data partitions X and Y are combined, and L_X, L_Y are the lengths of the most parsimonious trees for each partition separately. Significance is judged by creating random data partitions from the original data set of the same size as X and Y, and asking how frequently a value of D_{XY} as great or greater occurs from this simulated distribution (Farris et al. 1995). When significant incongruence is found, it may be due to differences in the phylogenetic signal of the two data partitions, differences in the level of noise (that is, the randomness of a data set caused by saturation), or a combination of the two (Dolphin et al. 2000). Although the interpretation of the ILD as a test of combinability has received increased scrutiny (Yoder et al. 2001), it remains the test of choice for the moment. It is implemented in PAUP* (Swofford 1999) and other programs.

Several other tests of partition homogeneity have been suggested recently and have considerable appeal. They include an alternative randomization test to the ILD (Rodrigo et al. 1993), the use of congruence among trees rather than data sets (Miyamoto & Fitch 1995), Partitioned Bremer Support (which quantifies the positive or negative contribution of each data partition to the character support for a particular node in a combined data analysis; Baker & Desalle 1997), and a likelihood ratio test that compares differences in likelihood with and without the constraint that the same phylogeny underlies all data partitions (Huelsenbeck & Bull 1996). While each of these recent alternatives has merit, none are as easily implemented as the ILD, and we presume that they will enjoy less widespread use than the ILD test in the near future.

Resolving Conflicts Among Data Partition

As molecular data sets grow to include more genes, conflicts among some or all data partitions appear to be something of a rule. Early comparisons of biochemical and morphological data often found this to be the case (Kluge 1989, reanalyzed in Swofford 1991; Shaffer et al. 1991). However, a small sampling of recent DNA-based analyses with or without morphology on a wide range of plant (Olmstead & Sweere 1994, Nickrent et al. 2000, Sanderson et al. 2000) and metazoan (Baker et al. 1998, Cao et al. 1998, Shaffer et al. 1997, Wiegmann et al. 2000, O'Grady et al. 1998) taxa emphasize that virtually any outcome is possible, with some data sets showing strong conflicts among partitions, and others not. A clear result from these and many other studies is that partitions often conflict in their phylogenetic signal, leaving it to the practitioner to decide how to handle such conflicts.

Fundamentally, there are two strategies for resolving conflicts among data partitions. The first is that advocated by Kluge (1989)—based on first principles: All data should always be included in a combined analysis for any phylogenetic problem. Even when a problematic data partition is successfully identified, some argue that the clearest and most reasonable strategy is to combine (Baker & Desalle 1997). Data conflicts that lead to two well-resolved, but different topologies may result in an unresolved bush when combined (e.g., Shaffer et al. 1991), but even so, advocates feel that this correctly summarizes the current state of knowledge. The alternative view is most clearly laid out by Bull et al. (1993), who advocate a "conditional combinability" strategy. They propose that one should test for homogeneity, and either combine (if the null of homogeneity cannot be rejected), remove the cause of heterogeneity (if the null is rejected and the cause can be identified), or give up (if the null is rejected but the cause cannot be identified). The last option comes down to the "No Obvious Resolution" in Bull et al.'s flow chart. This conditional combinability strategy has been embraced by most systematists (but notable exceptions exist, e.g., Miyamoto & Fitch 1995), and later methodological and simulation work has largely focused on what to do when conflict exists.

Two reasonable solutions appear to be emerging. First, using the ILD (Cunningham 1997), Partitioned Bremer Support (Baker et al. 1998), or visual examination of data conflicts, attempt to identify one gene or data partition within a gene that accounts for much of the conflict (Cao et al. 1998). If a single problematic partition exists, eliminate that partition (Huelsenbeck & Bull 1996). In a similar vein, Johnson (2001) argues for using the ILD to identify problematic taxa and remove them from the analysis. However, as noted by Baker & Desalle (1997) as more gene partitions become available, the likelihood of complex patterns of overlapping conflict becomes ever greater, making it difficult to determine which, if any, partitions should be eliminated (Krzywinski et al. 2001). An emerging theme from several recent analyses is that saturated data (that is, very noisy data suffering from many multiple substitutions per site) is a major source of data conflict (Baker & Desalle 1997, Dolphin et al. 2000). Perhaps because of the problem of long-branch attraction (see above), saturated data partitions are a real source of conflict, and should be eliminated. Unfortunately, these same partitions may be informative in the more recent parts of a tree, but misleading deeper in the tree, making their complete elimination a less desirable choice (Källersjö et al. 1999). In this case, restricting the use of such genes to recently diverged sets of taxa is an attractive analytical solution.

Theoretical Guidelines on Potential Conflicts in New Data Sets

Over the past decade, several guidelines have been proposed regarding the predicted relationship between monophyly in different genome partitions, and how those individual gene partitions reflect the underlying species phylogeny. These guidelines are spread across two divergent literatures, that of phylogenetics/

phylogeography (Ball et al. 1990, Maddison 1997, Moore 1995, Palumbi et al. 2001) and of theoretical population genetics/coalescent modeling (Wu 1991, Hudson 1992). The fundamental issue is the extent to which intra- or interspecific monophyly for a single gene partition predicts monophyly in other partitions, or for the species as a whole. Because single gene mitochondrial (mtDNA) sequences have been used so extensively in metazoan phylogenetics, Moore (1995) used a neutral coalescent framework and the four-fold difference in effective population size of mitochondrial and nuclear genes to calculate the difference in "time to monophyly" (the age of the most recent common ancestor of the sample of sequences) for these two classes of gene partition. Moore concluded that it would take, on average, 16 times as much nuclear DNA sequence information as mitochondrial to achieve equivalent phylogenetic resolution. He thus recommended using mtDNA trees, although some fraction of the time the species trees based on mtDNA will be incorrect. Using very similar logic, Palumbi et al. (2001) formulated their "three times rule" for the predicted concordance of nuclear and mtDNA gene trees. The rule states that, on average, nuclear coalescence (that is, monophyly) is predicted when mtDNA shows a pattern of relatively great divergence between clades compared to the within-clade diversity. In particular, when clades for mtDNA are monophyletic and the ratio of the length of the branch subtending a clade to the depth of the within-clade mtDNA gene tree is greater than about three, nuclear genes may often be monophyletic. Additional theoretical work (R.R. Hudson, M. Turelli, personal communication) has questioned the assumptions leading to the "three times" part of the rule, and emphasizes that guidelines may be lineage-specific with high variances, and therefore have little predictive value. However, the basic premise that low levels of within-clade divergence compared to among-clade divergence in one gene should be associated with similar patterns across loci remains a potentially important result (Hudson 1992, Wu 1991).

The determination of an "x-times rule," including what x may be, is clearly an empirical question, as emphasized by most authors. However, the general framework offers the opportunity to determine, a priori, when a given level of intraclade sampling and a resultant gene tree may predict concordance in other gene partitions. If general rules emerge from empirical results, this could allow molecular systematists to predict when a single gene should suffice, and when many, potentially conflicting, partitions need to be examined to determine organismal phylogenies.

ROOTING THE TREE IS PROBLEMATIC

Rooting phylogenetic trees is a relatively neglected component of many phylogenetic studies. Often considered straightforward, rooting has been identified as "frequently the most precarious step in any phylogenetic analysis" (Swofford et al. 1996, p. 478). Virtually all users of phylogenetic trees want rooted (rather than unrooted) trees, because directionality of evolutionary change and the concept of monophyly itself only have meaning on rooted trees (Smith 1994). By default, software for tree reconstruction typically generates unrooted trees (although they may be displayed as rooted). Rooting, the process of selecting a branch of the unrooted tree into which a root node is inserted, is accomplished afterwards. Although several strategies for rooting trees exist, the one most generally embraced by the phylogenetics community is outgroup analysis, in which some taxa are assigned to a monophyletic ingroup, and one or more taxa to the outgroup. After the analysis is complete, the (unrooted) ingroup tree is rooted at the branch where the outgroup(s) joins the ingroup tree. In many ways, outgroup analysis is nothing more than adding a few additional taxa to an analysis and building a tree, although the initial assumption of monophyly of the ingroup is a strong one. The repercussions of outgroup choice are enormous, however, for they can determine which character states are interpreted as shared derived features and which are ancestral for the entire ingroup (Nixon & Carpenter 1994). In addition, outgroup choice can affect ingroup topology, even for nodes far removed from the presumed root placement (Milinkovitch & Lyons-Weiler 1998, Tarrio et al. 2000).

We discuss two of the more difficult problems with rooting a tree. First, what can be done if reasonable choices exist for potential outgroups, but they are distantly related to the ingroup? Second, what strategies are useful when no obvious potential outgroup(s) can be identified?

Dealing with Distantly Related Outgroups

For previously studied groups, one often knows or can make an educated guess on a set of taxa that are outside of, but closely related to, the ingroup. One or more of these candidates will generally comprise the outgroup (Smith 1994, Nixon & Carpenter 1994). Unfortunately, outgroup sequences are sometimes very divergent from those in the ingroup (Johnson 2001). In the extreme, outgroups may represent little more than randomized, fully-saturated sequences with respect to the ingroup (Wheeler 1990, Maddison et al. 1992). In such cases, outgroup placement reduces to a long-branch attraction problem, and the root will often fall on the longest internal branch of the ingroup tree. As Smith (1994) notes, when a molecular clock is in effect (unusual for most molecular data), this will generally lead to the correct placement of the root. However, the root placement will still result from two long branches (the internal branch and the outgroup branch) attracting, not because of correct phylogenetic signal.

A number of strategies have been suggested that improve the chances of an outgroup correctly rooting a tree. First, attempt to identify and include the sister group to the ingroup for rooting purposes. Generally, this taxon's sequence will be closer to the ingroups than more distantly related outgroups, and will therefore more reliably root the ingroup tree. However, if the sister group has experienced a severe rate speedup, more distantly related—but less divergent—outgroups will provide more reliable evidence on ingroup rooting than the sister group (Lyons-Weiler et al. 1998), although this may be an uncommon scenario.

Second, add additional outgroup taxa (Maddison et al. 1984). Ideally, these additional outgroups should be chosen in such a way that they will break up the branch between the first outgroup and the ingroup, and thus reduce any long-branch attraction problem associated with root placement. The alternative, adding additional outgroups that fall outside of the ingroup-plus-first-outgroup, may simply add long branches to the base of the tree, contributing little to the rooting solution. Unfortunately, choosing multiple outgroups correctly requires prior knowledge of their relationships to each other and to the ingroup, and this knowledge is often not available. Although most recent studies strongly advocate using multiple outgroups, this is not universally accepted (Lyons-Weiler et al. 1998). Sometimes extinct fossil lineages are more relevant outgroups than extant ones, in which case morphological data alone or in combination with molecular data can be used to root the tree (Shaffer et al. 1997).

Finally, if the sister group is only a single species, or a set of very closely related species (Johnson 2001, Maddison et al. 1992), one should test whether the rooting is relatively robust. Strategies include: (*a*) adding different, divergent outgroups to determine if root placement changes, in which case it is suspect (Tarrio et al. 2000, Hutcheon et al. 1998, Dalevi et al. 2001), and (*b*) identifying and removing hypervariable sites from the data matrix, because they probably contribute to any long-branch attraction (Hendy & Penny 1989, Smith 1994). In this latter case, the hypervariable regions may well be informative within the ingroup, in which case the ingroup tree can be built with all sites. This topology can then be fixed, the hypervariable sites removed, and the outgroup(s) added to the analysis. A related problem recently emphasized by Hutcheon et al. (1998) and Tarrio et al. (2000) occurs when codon usage in an outgroup is very different from the ingroup. This is a special case of an outgroup that is so divergent from the ingroup that it has little value in identifying the ingroup root. Midpoint rooting of the ingroup tree (see below) may be preferable to using such a divergent outgroup (Tarrio et al. 2000).

What to Do When No Obvious Outgroup is Known

When studying the phylogeny of very poorly known groups, or groups at the base of the tree of life (Dalevi et al. 2001), there may be no particularly close outgroup with which to root a tree. In this case, exploratory experiments with different strategies may be useful. One possibility is to use midpoint rooting, which assigns the root to the midpoint of the longest path between two terminal taxa in the tree. Under the assumption that the two most divergent lineages in the ingroup tree have evolved at an equal rate, this method correctly roots a tree without reference to an outgroup (Swofford et al. 1996). More generally, reconstructing a tree under the assumption of a molecular clock will also infer the root, although this is a slightly stronger assumption. Alternatively, a variety of outgroups can be tested and compared for their effect on both the ingroup topology and placement of the root. If all give the same result (and even better if it matches the midpoint root), then this lends some support to the root placement being correct. This type of experiment has led to a variety of results ranging from constancy across outgroups (Hutcheon et al. 1998, Dalevi et al. 2001) to considerable heterogeneity (Milinkovitch & Lyons-Weiler 1998). Tarrio et al. (2000) found that, at least in the case of the *willistoni* and *saltans* groups of *Drosophila*, using a simple model of DNA sequence evolution (as opposed to a more complex model) seemed to eliminate outgroup instability, although the generality of this result remains untested.

COMPUTATION TIME IS TOO LONG

Phylogeneticists are spending more time waiting for analyses to finish than ever before. Molecular data sets containing hundreds to thousands of sequences (e.g., Källersjö et al. 1998, Lipscomb et al. 1998, Savolainen et al. 2000, Tehler et al. 2000, Johnson 2001) are increasingly common, pushing the computational ability of phylogenetic algorithms to their limits and beyond. Parametric inference methods such as maximum likelihood are often invoked with complex models of molecular evolution, which also demand time-consuming computation. Assessment of confidence limits in a tree with bootstrapping adds several orders of magnitude to processing time over and above inferring the tree. Besides taking increasingly long coffee breaks, several strategies may help.

Use Fast(er) Algorithms

The dominant factor in determining the running time of phylogenetic algorithms is the number of sequences, N. No known algorithms are guaranteed to find solutions for optimization-based tree-building procedures, such as MP or ML, in running times that scale better than exponential (e^N). Commonly used heuristics, such as those implemented in PAUP* (Swofford 1999) and other programs, have running times that scale polynomially, as N^k, where k is some constant, and this represents a huge improvement in run times. However, these provide no guarantees about how close they will come to finding the optimal tree. Simulation results suggest that the ML heuristics using sequential addition and branch-swapping scale worse than MP, by a factor of about N in standard implementations (Sanderson & Kim 2000). Other heuristics have been proposed more recently (Goloboff 1999, Nixon 1999, Quicke et al. 2001), which may find better scoring trees faster than more widely used heuristics, but these ultimately have the same worst-case running times and likewise lack guarantees on their solutions. Parsimony heuristics using sequential addition of sequences without rearrangement steps scale as N^2, so this may be the method of choice in desperate circumstances. Note that even non-optimization methods such as NJ (Saitou & Nei 1987), which are often regarded as being fast, scale no better than N^3.

A variety of "divide and conquer" methods have been proposed, which break data sets into subsets (often quartets), build the corresponding trees, and then reassemble them. Quartet Puzzling (Strimmer & Von Haeseler 1996) and a recent variant, Weighted Optimization (Ranwez & Gascuel 2001) can be implemented in running times that scale as N^4. They are thus mainly useful for ML searches; they can be faster than standard heuristic search strategies because finding many small trees is much faster than one larger tree. Ota & Li (2000, 2001) took this strategy in a rather different direction with a hybrid between ML and NJ, which identifies well-supported subtrees via NJ and then performs constrained searches across the less-well supported nodes via ML.

One of the main problems in all optimization strategies is the existence of multiple local optima, which are well known in MP (Maddison 1991) and ML (Salter 2001). Stochastic optimization methods such as simulated annealing (Barker 1997, Salter & Pearl 2001) or genetic/evolutionary algorithms (Matsuda 1996; Lewis 1998; Goloboff 1999; Moilanen 1999, 2001; Katoh et al. 2001) strive to avoid being trapped at local solutions by randomly moving away from local optima according to various schemes. Some of these can be significantly faster than traditional heuristic searches in large data sets (Salter & Pearl 2001), but have not been road-tested much on real data.

Thus, a speed-up in search time can generally be obtained by switching algorithms or computer programs. As always, the difficult question concerns whether accuracy is sacrificed. The literature on accuracy of tree-building algorithms is exceptionally large and difficult to distill, mainly because it has relied on simulations done under different conditions of substitution models, parameter values, data set sizes, and measures of accuracy. A new framework for understanding performance of these methods is provided by the notion of fast convergence (Huson et al. 1999, Nakleh et al. 2001). A method is said to be fast converging if the number of characters needed to guarantee a certain level of accuracy grows at worst polynomially with the number of taxa. Unfortunately, the only theoretical results regarding this concept relate to nonstandard methods such as Disk-Covering (Huson et al. 1999), which is fast converging. It is not yet known whether any standard tree-building method is fast converging, but simulation studies hint that NJ and MP are, at least within the class of phylogenies generated by simple branching processes (Bininda-Emonds et al. 2001, Nakleh et al. 2001).

Reduce the Complexity of Parametric Models

Workers using ML methods have developed many shortcuts that decrease running times when using complex models with many parameters, by using approximations for many of the parameter estimates. A typical strategy is to estimate parameters of the substitution model once on a single "seed" tree, such as one obtained by fast NJ or MP analyses, and then fix parameters at the estimates obtained on that one tree during heuristic searches across many trees (e.g., Moncalvo et al. 2000). Various features in PAUP* facilitate these shortcuts. In moderate-sized data sets (<100 sequences), running times can be improved by an order of magnitude or more.

Sequence More Genes for the Same Taxa

A long-term solution for improving running times for data sets may be to acquire more sequence data for the same taxa. Addition of data can dramatically decrease running times of heuristic searches in data sets with many taxa (Hillis 1996, Soltis et al. 1998, Savolainen et al. 2000), presumably by enhancing the signal to noise ratio (Hillis 1996). Thus, although one might assume that running times would be positively correlated with number of characters, because tree-score calculations must be done for all characters independently, this increase is apparently offset by decreases in the volume of tree space that must be examined during heuristic searches. One caveat is that the addition of genes with different evolutionary histories (whether due to recombination, lineage sorting of ancestral polymorphisms, or other processes) may have the opposite effect, because algorithms are forced to try to sort out significantly conflicting partitions in the data. Tests for homogeneity in the phylogenetic signal (see above) between genes should therefore be performed prior to combining.

Estimate the "Consensus Support" Tree Rather Than the Optimal Tree

A fundamentally different strategy is to abandon attempts to obtain the optimal (i.e., best) tree and estimate instead only the best-supported elements of that tree (Goloboff & Farris 2001). Implementations generally involve randomization of some type. The parsimony jackknife (PJ) runs repeated heuristic searches using data matrices randomly subsampled from the original, while discarding clades that appear in fewer than some cutoff frequency of replicates. Bayesian Markov Chain Monte Carlo (MCMC) methods (Mau et al. 1999, Larget & Simon 1999) estimate the marginal posterior probabilities of clades. Both methods are much quicker than optimization per se, but not surprisingly, probably do not provide good estimates of the optimal tree. This is particularly true for MCMC in the case of large data sets (Salter & Pearl 2001), because the Markov chains rarely visit any specific tree, rendering the estimate of its posterior probability subject to large sampling error. However, estimating well-supported groups may often be sufficient or even preferable, because they emphasize areas where the data are informative. Consensus trees obtained by these methods do have limitations if they are to be used in subsequent evolutionary studies. However, reconstructions of character evolution or applications of the comparative method, which generally require resolved trees (Maddison & Maddison 1992), can be weighted by posterior probabilities sampled from an MCMC chain (Huelsenbeck et al. 2001).

Use Faster Hardware

Although computer hardware improvements are unlikely to permit solution of large phylogenetic optimization problems exactly, they can certainly speed up heuristic searches. Processor speed continues to double every 1.5 years (Moore's law), but high performance computing now almost always entails some form of parallel processing. Practical and inexpensive commodity hardware in the form of Beowulf clusters (Sterling et al. 1999) have generated much interest among phylogeneticists, although software to take advantage of these architectures is not yet widely used. Some phylogenetic problems such as bootstrapping are trivially parallelizable, whereas others, such as sequential addition and branch swapping are not (Janies & Wheeler 2001, Stewart et al. 2001). Many stochastic methods requiring simple replication, such as bootstrapping, PJ, genetic algorithms, or MCMC, can be distributed at the operating system level without writing true parallel code. However, truly parallelized optimization software is now available for MP (Gladstein & Wheeler 1996), ML (Ceron et al. 1998, Stewart et al. 2001), Quartet Puzzling (Schmidt et al. 2002) and genetic algorithms (Katoh et al. 2001), but development of additional tools is badly needed.

CONCLUSIONS

The diversity and complexity of methods available for phylogeny reconstruction presents both a steep learning curve for newcomers and a considerable investment of time and energy in exploring and exploiting the information content of a data set. Just as statisticians in the twentieth century were increasingly called upon to provide advice and guidance to biologists, phylogeneticists play a similar role in the twenty-first century as diverse biologists come to terms with the abundance of comparative sequence data. Particularly as data sets become larger and more diverse, creative and user-friendly strategies, computer programs, and trouble-shooting guidelines are necessary for empiricists to continue building the tree of life. Solutions to some of these emerging problems already exist, but many are unresolved challenges worthy of attention. As these problems attract a progressively broader assemblage of workers interested in studying them, we look forward to new solutions to some of the more recalcitrant problems described here.

The *Annual Review of Ecology and Systematics* is online at http://ecolsys.annualreviews.org

LITERATURE CITED

Baake H. 1998. What can and cannot be inferred from pairwise sequence comparisons? *Math. Biosci.* 154:1–21

Baker RH, Desalle R. 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. *Syst. Biol.* 46:654–73

Baker RH, Yu XB, DeSalle R. 1998. Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. *Mol. Phylogenet. Evol.* 9: 427–36

Ball RM Jr, Neigel JE, Avise JC. 1990. Gene genealogies within the organismal pedigrees of random-mating populations. *Evolution* 44:360–70

Barker D. 1997. *LVB 1.0: Reconstructing Evolution with Parsimony and Simulated Annealing*. Edinburgh: Univ. Edinburgh

Berry V, Gascuel O. 1996. On the interpretation of bootstrap trees: appropriate threshold of clade selection and induced gain. *Mol. Biol. Evol.* 13:999–1011

Bininda-Emonds ORP, Brady SG, Kim J, Sanderson MJ. 2001. Scaling of accuracy in extremely large phylogenetic trees.

Presented *Pac. Symp. Biocomput.* 6:547–58

Bremer B, Jansen RK, Oxelman B, Backlund M, Lantz H, Kim K-J. 1999. More characters or more taxa for a robust phylogeny: case study from the coffee family (Rubiaceae). *Syst. Biol.* 48:413–35

Buckley TR, Simon C, Chambers GK. 2001. Exploring among-site rate variation models in a maximum likelihood framework using empirical data: effects of model assumptions on estimates of topology, branch lengths, and bootstrap support. *Syst. Biol.* 50:67–86

Bull JJ, Huelsenbeck JP, Cunningham CW, Swofford DL, Waddell PJ. 1993. Partitioning and combining data in phylogenetic analysis. *Syst. Biol.* 42:384–97

Burnham KP, Anderson DR. 1998. *Model Selection and Inference.* New York: Springer

Cao Y, Janke A, Waddell PJ, Westerman M, Takenaka O, et al. 1998. Conflict among individual mitochondrial proteins in resolving the phylogeny of eutherian orders. *J. Mol. Evol.* 47:307–22

Ceron C, Dopazo J, Zapata EL, Carazo JM, Trelles O. 1998. Parallel implementation of DNAml program on message-passing architectures. *Parallel Comput.* 24:701–16

Chang JT. 1996. Inconsistency of evolutionary tree topology reconstruction methods when substitution rates vary across characters. *Math. Biosci.* 134:189–215

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, et al. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL. *Ann. Mo. Bot. Gard.* 80:528–80

Conant GC, Lewis PO. 2001. Effects of nucleotide composition bias on the success of the parsimony criterion in phylogenetic inference. *Mol. Biol. Evol.* 18:1024–33

Cunningham CW. 1997. Can three incongruence tests predict when data should be combined? *Mol. Biol. Evol.* 14:733–40

Dalevi D, Hugenholtz P, Blackall LL. 2001. A multiple-outgroup approach to resolving division-level phylogenetic relationships using 16S rDNA data. *Int. J. Syst. Evol. Microbiol.* 51:385–91

Debry RW, Olmstead RG. 2000. A simulation study of reduced tree-search effort in bootstrap resampling analysis. *Syst. Biol.* 49:171–79

Dolphin K, Belshaw R, Orme CDL, Quicke DLJ. 2000. Noise and incongruence: Interpreting results of the incongruence length difference test. *Mol. Phylogenet. Evol.* 17:401–6

Edwards AWF, Cavalli-Sforza LL. 1967. Phylogenetic analysis: models and estimation procedures. *Evolution* 21:550–70

Efron B, Halloran E, Holmes S. 1996. Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. USA* 93:13429–34. Erratum. 1996. *Proc. Natl. Acad. Sci. USA* 93(14):7085–96

Farris JS, Källersjö M, Kluge AG, Bult C. 1995. Constructing a significance test for incongruence. *Syst. Biol.* 44:570–72

Felsenstein J. 1978. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. *Biol. J. Linn. Soc.* 16:183–96

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

Felsenstein J. 2001. The troubled growth of statistical phylogenetics. *Syst. Biol.* 50:465–67

Felsenstein J, Kishino H. 1993. Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. *Syst. Biol.* 42:182–92

Finden CR, Gordon AD. 1985. Obtaining common pruned trees. *J. Classif.* 2:255–76

Fitch WM, Markowitz E. 1970. An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. *Biochem. Genet.* 4:579–93

Foster PG, Hickey DA. 1999. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *J. Mol. Evol.* 48:284–90

Galtier N, Gouy M. 1995. Inferring phylogenies from DNA sequences of unequal base compositions. *Proc. Natl. Acad. Sci. USA* 92:11317–21

Galtier N, Gouy M. 1998. Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. *Mol. Biol. Evol.* 15:871–79

Gladstein D, Wheeler W. 1996. POY. Software for direct optimization of DNA and other data. New York: Am. Mus. Nat. Hist.

Goloboff PA. 1999. Analyzing large data sets in reasonable times: solutions for composite optima. *Cladistics* 15:415–28

Goloboff PA, Farris JS. 2001. Methods for quick consensus estimation. *Cladistics* 17: S26–34

Gusfield D. 1997. *Algorithms on Strings, Trees and Sequences*. New York: Cambridge Univ. Press

Hasegawa M, Hashimoto T. 1993. Ribosomal RNA trees misleading? *Nature* 361:23

Hendy MD, Penny D. 1989. A framework for the quantitative study of evolutionary trees. *Syst. Zool.* 38:297–309

Hennig W. 1966. *Phylogenetic Systematics*. Chicago: Univ. Chic. Press. 263 pp.

Hillis DM, Bull JJ. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* 42:182–92

Hudson RR. 1992. Gene trees, species trees and the segregation of ancestral alleles. *Genetics* 131:509–12

Huelsenbeck JP. 1997. Is the Felsenstein zone a fly trap? *Syst. Biol.* 46:69–74

Huelsenbeck JP, Bull JJ. 1996. A likelihood ratio test to detect conflicting phylogenetic signal. *Syst. Biol.* 45:92–98

Huelsenbeck JP, Crandall KA. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. *Annu. Rev. Ecol. Syst.* 28: 437–66

Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294:2310–14

Huson DH, Nettles SM, Warnow TJ. 1999. Disk-covering, a fast-converging method for phylogenetic tree reconstruction. *J. Comp. Biol.* 6:369–86

Hutcheon JM, Kirsch JAW, Pettigrew JD. 1998. Base-compositional biases and the bat problem. III. The question of microchiropteran monophyly. *Philos. Trans. R. Soc. London Ser. B* 353:607–17

Janies DA, Wheeler WC. 2001. Efficiency of parallel direct optimization. *Cladistics* 17: S71–82

Johnson KP. 2001. Taxon sampling and the phylogenetic position of Passeriformes: evidence from 916 avian cytochrome b sequences. *Syst. Biol.* 50:128–36

Källersjö M, Albert VA, Farris JS. 1999. Homoplasy increases phylogenetic structure. *Cladistics* 15:91–93

Källersjö M, Farris JS, Chase MW, Bremer B, Fay MF, et al. 1998. Simultaneous parsimony jackknife analysis of 2538 rbcL DNA sequences reveals support for major clades of green plants, land plants, seed plants and flowering plants. *Plant. Syst. Evol.* 213:259–87

Katoh K, Kuma K-i, Miyata T. 2001. Genetic algorithm-based maximum-likelihood analysis for molecular phylogeny. *J. Mol. Evol.* 53:477–84

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

Krzywinski J, Wilkerson RC, Besansky NJ. 2001. Toward understanding Anophelinae (Diptera, Culicidae) phylogeny: insights from nuclear single-copy genes and the weight of evidence. *Syst. Biol.* 50:540–56

Larget B, Simon DL. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* 16:750–59

Lee MSY. 2001. Unalignable sequences and molecular evolution. *Trends Ecol. Evol.* 16: 681–85

Lewis PO. 1998. A genetic algorithm for maximum-likelihood phylogeny inference using nucleotide sequence data. *Mol. Biol. Evol.* 15:277–83

Lipscomb DL, Farris JS, Källersjö M, Tehler A. 1998. Support, ribosomal sequences and the phylogeny of the eukaryotes. *Cladistics* 14:303–38

Lockhart PJ, Huson D, Maier U, Fraunholz MJ, Van de Peer Y, et al. 2000. How molecules evolve in eubacteria. *Mol. Biol. Evol.* 17:835–38

Lockhart PJ, Larkum AWD, Steel MA, Waddell PJ, Penny D. 1996. Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. *Proc. Natl. Acad. Sci. USA* 93:1930–34

Lockhart PJ, Steel MA, Barbrook AC, Huson DH, Charleston MA, Howe CJ. 1998. A covariotide model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. *Mol. Biol. Evol.* 15:1183–88

Lockhart PJ, Steel MA, Hendy MD, Penny D. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* 11:605–12

Lopez P, Forterre P, Philippe H. 1999. The root of the tree of life in the light of the covarion model. *J. Mol. Evol.* 49:496–508

Lyons-Weiler J, Hoelzer GA, Tausch RJ. 1998. Optimal outgroup analysis. *Biol. J. Linn. Soc.* 64:493–511

Maddison DR. 1991. The discovery and importance of multiple islands of most-parsimonious trees. *Syst. Zool.* 40:315–28

Maddison DR, Baker MD, Ober KA. 1999. Phylogeny of carabid beetles as inferred from 18S ribosomal DNA (Coleoptera: Carabidae). *Syst. Entomol.* 24:103–38

Maddison DR, Ruvolo M, Swofford DL. 1992. Geographic origins of human mitochondrial DNA phylogenetic evidence from control region sequences. *Syst. Biol.* 41:111–24

Maddison WP. 1997. Gene trees in species trees. *Syst. Biol.* 46:523–36

Maddison WP, Donoghue MJ, Maddison DR. 1984. Outgroup analysis and parsimony. *Syst. Zool.* 33:83–103

Maddison WP, Maddison DR. 1992. *MacClade: Analysis of Phylogeny and Character Evolution.* Sunderland, MA: Sinauer

Matsuda H. 1996. Protein phylogenetic inference using maximum likelihood with a genetic algorithm. In *Pacific Symposium on Biocomputing*, ed. L Hunter, TE Klein, pp. 512–13. Singapore: World Sci.

Mau B, Newton MA, Larget B. 1999. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. *Biometrics* 55:1–12

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

Milinkovitch MC, Lyons-Weiler J. 1998. Finding optimal ingroup topologies and convexities when the choice of outgroups is not obvious. *Mol. Phylogenet. Evol.* 9:348–57

Mitchell A, Mitter C, Regier JC. 2000. More taxa or more characters revisited: combining data from nuclear protein-encoding genes for phylogenetic analyses of Noctuoidea (Insecta: Lepidoptera). *Syst. Biol.* 49:202–24

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

Moilanen A. 1999. Searching for most parsimonious trees with simulated evolutionary optimization. *Cladistics* 15:39–50

Moilanen A. 2001. Simulated evolutionary optimization and local search: introduction and application to tree search. *Cladistics* 17:S12–25

Moncalvo J-M, Lutzoni FM, Rehner SA, Johnson J, Vilgalys R. 2000. Phylogenetic relationships of agaric fungi based on nuclear large subunit ribosomal DNA sequences. *Syst. Biol.* 49:278–305

Moore WS. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution* 49:718–26

Morrison DA, Ellis JT. 1997. Effects of nucleotide sequence alignment on phylogeny estimation: a case study of 18S rDNAs of apicomplexa. *Mol. Biol. Evol.* 14:428–41

Mort ME, Soltis PS, Soltis DE, Mabry ML. 2000. Comparison of three methods for estimating internal support on phylogenetic trees. *Syst. Biol.* 49:160–71

Nakleh L, Roshan U, St. John K, Sun J, Warnow

TJ. 2001. Designing fast converging phylogenetic methods. *Bioinformatics* 1:1–9

Naylor GJP, Brown WM. 1998. Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. *Syst. Biol.* 47:61–76

Newton MA. 1996. Bootstrapping phylogenies: large deviations and dispersion effects. *Biometrika* 83:315–28

Nickrent DL, Parkinson CL, Palmer JD, Duff RJ. 2000. Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Mol. Biol. Evol.* 17: 1885–95

Nixon KC. 1999. The Parsimony Ratchet, a new method for rapid parsimony analysis. *Cladistics* 15:407–14

Nixon KC, Carpenter JM. 1994. On outgroups. *Cladistics* 9:413–26

O'Grady PM, Clark JB, Kidwell MG. 1998. Phylogeny of the *Drosophila saltans* species group based on combined analysis of nuclear and mitochondrial DNA sequences. *Mol. Biol. Evol.* 15:656–64

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

Omilian AR, Taylor DJ. 2001. Rate acceleration and long-branch attraction in a conserved gene of cryptic daphniid (Crustacea) species. *Mol. Biol. Evol.* 18:2201–12

Omland KE, Lanyon SM, Fritz SJ. 1999. A molecular phylogeny of the New World orioles (Icterus): the importance of dense taxon sampling. *Mol. Phylogenet. Evol.* 12:224–39

Ota S, Li W-H. 2000. NJML: a hybrid algorithm for the neighbor-joining and maximum-likelihood methods. *Mol. Biol. Evol.* 17:1401–9

Ota S, Li W-H. 2001. NJML+: an extension of the NJML method to handle protein sequence data and computer software implementation. *Mol. Biol. Evol.* 18:1983–92

Page RDM, Charleston MA. 1997. From gene to organismal phylogeny: reconciled trees and the gene tree/species tree problem. *Mol. Phylogenet. Evol.* 7:231–40

Page RDM, Cotton JA. 2002. Vertebrate phylogenomics: reconciled trees and gene duplications. *Pac. Symp. Biocomput.* 7:536–47

Palumbi SR, Cipriano F, Hare MP. 2001. Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. *Evolution* 55:859–68

Philippe H, Germot A. 2000. Phylogeny of eukaryotes based on ribosomal RNA: long-branch attraction and models of sequence evolution. *Mol. Biol. Evol.* 17:830–34

Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817–18

Posada D, Crandall KA. 2001. Selecting the best-fit model of nucleotide substitution. *Syst. Biol.* 50:580–601

Quicke DLJ, Taylor J, Purvis A. 2001. Changing the landscape: a new strategy for estimating large phylogenies. *Syst. Biol.* 50:60–66

Rambaut A, Grassly NC. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Cabios* 13:235–38

Ranwez V, Gascuel O. 2001. Quartet-based phylogenetic inference: improvements and limits. *Mol. Biol. Evol.* 18:1103–16

Rieppel O, Reisz RR. 1999. The origin and early evolution of turtles. *Annu. Rev. Ecol. Syst.* 30:1–22

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

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406–25

Salter LA. 2001. Complexity of the likelihood surface for a large DNA dataset. *Syst. Biol.* 50:970–78

Salter LA, Pearl DK. 2001. Stochastic search strategy for estimation of maximum likelihood phylogenetic trees. *Syst. Biol.* 50:7–17

Sanderson MJ. 1989. Confidence limits on phylogenies the bootstrap revisited. *Cladistics* 5: 113–30

Sanderson MJ. 1995. Objections to bootstrapping phylogenies: a critique. *Syst. Biol.* 44: 299–320

Sanderson MJ, Donoghue MJ. 1996. The relationship between homoplasy and confidence in phylogenetic trees. In *Homoplasy: The Recurrence of Similarity in Evolution*, ed. MJ Sanderson, L Hufford, pp. 67–89. New York: Academic

Sanderson MJ, Kim J. 2000. Parametric phylogenetics? *Syst. Biol.* 49:817–29

Sanderson MJ, Wojciechowski MF. 2000. Improved bootstrap confidence limits in large-scale phylogenies, with an example from Neo-Astragalus (Leguminosae). *Syst. Biol.* 49:671–85

Sanderson MJ, Wojciechowski MF, Hu JM, Khan TS, Brady SG. 2000. Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants. *Mol. Biol. Evol.* 17:782–97

Savolainen V, Chase MW, Hoot SB, Morton CM, Soltis DE, et al. 2000. Phylogenetics of flowering plants based on combined analysis of plastid atpB and rbcL gene sequences. *Syst. Biol.* 49:306–62

Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18:502–4

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

Shaffer HB, Meylan P, McKnight ML. 1997. Tests of turtle phylogeny: molecular, morphological, and paleontological approaches. *Syst. Biol.* 46:235–68

Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17:1246–47

Small RL, Wendel JF. 2000. Copy number lability and evolutionary dynamics of the Adh gene family in diploid and tetraploid cotton (Gossypium). *Genetics* 155:1913–26

Smith AB. 1994. Rooting molecular trees: problems and strategies. *Biol. J. Linn. Soc.* 51:279–92

Soltis DE, Soltis PS, Mort ME, Chase MW, Savolainen V, et al. 1998. Inferring complex phylogenies using parsimony: an empirical approach using three large DNA data sets for angiosperms. *Syst. Biol.* 47:32–42

Steel MA, Lockhart PJ, Penny D. 1993. Confidence in evolutionary trees from biological sequence data. *Nature* 364:440–42

Steel MA, Penny D. 2000. Parsimony, likelihood, and the role of models in molecular phylogenetics. *Mol. Biol. Evol.* 17:839–50

Sterling T, Salmon J, Becker D, Savarese D. 1999. *How To Build A Beowulf: A Guide to the Implementation and Application of PC Clusters.* Cambridge, MA: MIT Press

Stewart CA, Hart D, Berry DK, Olsen GJ, Wernert EA, Fischer W. 2001. *Parallel implementation and performance of fastDNAml-aprogram for maximum likelihood phylogenetic inference.* Presented at SC2001, Denver

Strimmer K, Von Haeseler A. 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* 13:964–69

Swofford DL. 1991. When are phylogeny estimates from molecular and morphological data incongruent? In *Phylogenetic Analysis of DNA Sequences*, ed. MM Miyamoto, J Cracraft, pp. 295–333. New York: Oxford Univ. Press

Swofford DL. 1999. PAUP * 4.0. *Phylogenetic Analysis Using Parsimony and Other Methods.* Sunderland, MA: Sinauer

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM. 1996. Phylogenetic inference. In *Molecular Systematics*, ed. DM Hillis, C Moritz, BK Mable, pp. 407–514. Sunderland, MA: Sinauer

Tamura K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G-plus-C-content biases. *Mol. Biol. Evol.* 9:678–87

Tarrio R, Rodriguez-Trelles F, Ayala FJ. 2000. Tree rooting with outgroups when they differ in their nucleotide composition from the ingroup: the *Drosophila saltans* and *willistoni* groups, a case study. *Mol. Phylogenet. Evol.* 16:344–49

Tarrio R, Rodriguez-Trelles F, Ayala FJ. 2001. Shared nucleotide composition biases among species and their impact on phylogenetic reconstructions of the Drosophilidae. *Mol. Biol. Evol.* 18:1464–73

Tehler A, Farris JS, Lipscomb DL, Källersjö M. 2000. Phylogenetic analyses of the fungi based on large rDNA data sets. *Mycologia* 92:459–74

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–80

Thorley JL, Page RDM. 2000. RadCon: phylogenetic tree comparison and consensus. *Bioinformatics* 16:486–87

Thorley JL, Wilkinson M. 1999. Testing the phylogenetic stability of early tetrapods. *J. Theor. Biol.* 200:343–44

Wheeler WC. 1990. Nucleic acid sequence phylogeny and random outgroups. *Cladistics* 6:363–68

Wheeler WC, Gatesy J, Desalle R. 1995. Elision: a method for accommodating multiple molecular sequence alignments with alignment-ambiguous sites. *Mol. Phylogenet. Evol.* 4:1–9

Whitfield JB, Cameron SA. 1998. Hierarchical analysis of variation in the mitochondrial 16S rRNA gene among hymenoptera. *Mol. Biol. Evol.* 15:1728–43

Wiegmann BM, Mitter C, Regier JC, Friedlander TP, Wagner DM, Nielsen ES. 2000. Nuclear genes resolve Mesozoic-aged divergences in the insect order Lepidoptera. *Mol. Phylogenet. Evol.* 15:242–59

Wilkinson M. 1994. Common cladistic information and its consensus representation: reduced Adams and reduced cladistic consensus trees and profiles. *Syst. Biol.* 43:343–68

Wilkinson M. 1996. Majority-rule reduced consensus trees and their use in bootstrapping. *Mol. Biol. Evol.* 13:437–44

Willson SJ. 1999. A higher order parsimony method to reduce long-branch attraction. *Mol. Biol. Evol.* 16:694–705

Woese CR, Achenbach L, Rouviere P, Mandelco L. 1991. Archaeal phylogeny: reexamination of the phylogenetic position of *Archaeoglobus fulgidus* in light of certain composition-induced artifacts. *Syst. Appl. Microbiol.* 14:364–71

Wu CI. 1991. Inferences of species phylogeny in relation to segregation of ancient polymorphisms. *Genetics* 127:429–36

Yang Z. 1998. On the best evolutionary rate for phylogenetic analysis. *Syst. Biol.* 47:125–33

Yoder AD, Irwin JA, Payseur BA. 2001. Failure of the ILD to determine data combinability for slow loris phylogeny. *Syst. Biol.* 50:408–24

Zharkikh A, Li WH. 1992a. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. I. Four taxa with a molecular clock. *Mol. Biol. Evol.* 9:1119–47

Zharkikh A, Li WH. 1992b. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. II. Four taxa without a molecular clock. *J. Mol. Evol.* 35:356–66