Phylogenetics, a term often used as a shorthand synonym for phylogenetic inference or phylogeny reconstruction, is the attempt to assess and portray how organisms are related historically in terms of their evolutionary paths of descent and divergence. The basic steps of phylogeny reconstruction would seem to be straightforward: (1) defining the primary biological units of study, called taxa or OTU (operational taxonomic units), which are often species but may be units at a lower (e.g., subspecies, populations) or higher (e.g., genera, families, orders, classes, phyla) level of organization; (2) characterizing taxa by the states of their characters, which might be any attributes (genetic, morphological, physiological, etc.) that vary in their expression among taxa; and (3) arranging the taxa by grouping them according to shared character states or degree of similarity. The arrangement is usually hierarchical, with more inclusive groups being characterized by more and more shared characters. The hypothesized hierarchical arrangement (a phylogeny) is usually portrayed graphically as a tree and may be used, for example, as the basis for a taxonomic classification or as an interpretation of temporal evolutionary relationships.

Despite the apparent simplicity of these steps, phylogenetic inference is not straightforward; on the contrary, it is a highly controversial discipline. This is because different characters do not necessarily define the same groups. In fact, different characters may be completely incompatible in their information about hierarchical relationships. For example, possession of a dorsal hollow nerve chord defines the chordates, which are almost universally considered to be an evolutionarily monophyletic group (i.e., derived from a single ancestral taxon). Possession of a complex eye defines a quite different group, on the other hand, partly overlapping with the chordates, that is not generally considered to be monophyletic. The incompatibilities among characters, and the consequent necessity for systematists to choose (i.e., weight) one kind of
character over another, provide both the fundamental sources of controversy in phylogenetics and the reason why there is more to systematics than the simple definition of groups.

**Model-Free versus Model-Laden Phylogenetics**

Characters can be classified in different ways, such as absolute (green) versus relative (forelimb longer than hindlimb); discrete (five toes) versus continuous (wing 8 mm long); developmentally early versus late; evolutionarily stable versus labile; evolutionarily ancestral versus derived. The first three of these contrasts can usually be evaluated objectively using the organisms at hand, while the latter two depend on prior hypotheses about evolutionary patterns and processes.

There is a spectrum of opinion among systematists about the desirability of incorporating process-level evolutionary assumptions into phylogenetic inference, depending to some extent on the intended use of the resulting phylogeny. If the object of a study is to obtain the best (i.e., most realistic) phylogeny possible from the data, then it may be desirable to use a tree-building method that incorporates as much as is known or surmised about evolutionarily processes. The resulting phylogeny will then reflect the biological assumptions built into the methodology. Molecular systematists characteristically take this approach, attempting to incorporate specific models of nucleotide-sequence evolution (Jukes and Cantor 1969; Kimura 1980; Lake 1987; Cheeseman and Kanefsky 1990). However, if the resulting phylogeny is to be used to test independent hypotheses about biological phenomena such as sequences of character evolution, rates of evolution, or biogeographic distributions, then the tree-building method should be as independent as possible of prior biological assumptions about the relevant phenomena. For example, if the objective of a study is to test for heterogeneity of rates of diversification among different lineages, then the phylogeny on which the test is based should have been constructed without invoking an assumption of rate homogeneity. The most extreme position in assumption-independence is that of the so-called transformed or pattern cladists, who have attempted to avoid the evolutionary framework entirely in inferring phylogenetic relationships among organisms (Platnick 1979; Patterson 1980; Ridley 1986). Most systematists consider the complete exclusion of evolutionary assumptions to be untenable (Sober 1989).

**Trees, Cladograms, and Phylogenies**

Putative historical relationships, and their implied time components, are conventionally portrayed as branching phylogenetic trees. However, because in-
formation about nested sets of character states (from which historical relationships among taxa are inferred) is also portrayed in the form of branching trees, several important distinctions must be made (Eldredge and Cracraft 1980).

Although not all systematists agree on terminology, the terms cladogram, tree, and phylogeny are often used to imply different kinds of information about character states and taxa. A cladogram is a branching diagram of species, clustered by their shared derived character states (see Marcus, this volume, chap. 2). The connections among the species indicate only the joint possession of derived characters, and there is no implicit suggestion of ancestral or temporal relationships. A tree, on the other hand, is usually considered to be a branching diagram depicting patterns of character-state distributions and transitions among taxa at various hierarchical levels. Hypothetical ancestral species are placed at the branching nodes, rather than at the tips of branches as in cladograms. The branching relationships among taxa then imply or represent evolutionary lineages.

A tree may be either rooted or unrooted; the root of a tree is the position of the putative ancestral taxon. An unrooted tree is often called a network, although this term is discouraged by some theorists because it conflicts with the standard usage in topology. Many of the quantitative procedures described below produce unrooted trees, for which the position of the root must be judged from other information. Typically one or more related taxa that are assumed to lie outside the group of interest (the in-group) can be included in the analysis. The position at which these so-called out-group taxa join the tree then defines the root with respect to the in-group (Maddison et al. 1984).

The term phylogeny or phylogenetic tree often implies the addition of a time dimension, particularly when stratigraphic, molecular-clock, or other time-referential data are available to indicate relative or absolute rates of evolution.

**Phylogenetic Inference under Uncertainty**

The expression phylogeny reconstruction implies a deterministic process of reconstructing the past, but phylogenetics is best viewed as an estimation procedure. The objective of the procedure is to produce the best possible estimate of historical relationships based on incomplete and often contradictory information about the past, primarily utilizing contemporary morphologies and molecules (although fossils can sometimes serve as a valuable source of phylogenetic information). Putative historical relationships are then portrayed as branching cladograms or trees, but many such trees are possible, and systematists need a basis for selecting one or more preferred trees from the spectrum of possibilities.

Contemporary phylogenetic methods determine preferred trees in one of two ways (Swoford and Olsen 1990): (1) by defining a specific sequence of steps (an algorithm) by which the preferred tree is produced; or (2) by defining
a quantitative criterion by which to compare alternatives and then deciding which of a series of trees is best by this criterion.

The first class of methods combines the definition of the preferred tree and the process of tree building into a single step. All of the various forms of hierarchical pair-group cluster analysis are of this type, as are the additive-tree methods described below. The primary advantage to these methods is their efficiency; because a single tree is generated algorithmically according to some optimality criterion, the current implementations are quite fast and require relatively little computer memory. Their primary disadvantage is that, although the resulting tree can be compared with the original data to assess the degree of fit, it is generally not possible to construct and evaluate alternative suboptimal trees. A second problem is that these methods are heuristic and often fail to address, or even make explicit, the underlying evolutionary assumptions.

The second class of methods divides tree building and tree assessment into two distinct logical steps. (1) An optimality criterion (also called an objective or merit function) is first defined, by which to evaluate individual trees. Optimality criteria are often based on explicit evolutionary assumptions and allow scores to be assigned to trees indicating their relative suitability. (2) Various algorithms are then used to search for the one or more trees having the best value (maximum or minimum) of the criterion. The evolutionary assumptions invoked in the first step, presumably based upon scientific criteria, are thus decoupled from the searching algorithms of the second step, which are based on computer science research and are improving rapidly with time. Because every tree can be assigned a score, the primary advantage to this class of methods is that alternative phylogenies can be ranked in order of preference and assessed in terms of their biological properties. The cost of this is reduced speed, a consequence of having to search for the best tree among an often enormous number of possibilities.

Kinds of Data

There are two broad categories of data used in phylogenetic inference: character data, which provide information about individual taxa, and distance or similarity data, which describe pairwise relationships among taxa.

Character Data

Character data are specified as single numeric values codifying the state of each character for each taxon. There is some disagreement among systematists about the terms character and character state, but those who employ numerical methods generally use the term character to be a synonym for
variable and consider a character state to be one of the set of possible observed values of the character. A character matrix \( X_{(n \times p)} \) for \( n \) taxa and \( p \) characters is a rectangular matrix of character-state values \( x_{ij} \). Characters can be either qualitative or quantitative, and qualitative characters can be further classified as binary (having two states, often indicating the presence or absence of traits) or multistate (having three or more states). For tree-building methods requiring binary characters, multistate characters can be recoded as a series of binary characters by additive binary coding.

Multistate characters are often considered to be unordered (nominal), such that any state can transform (over evolutionary time) to any other state. This is the usual assumption for DNA-sequence data, for which the four nucleotides (A, C, G, T) are the possible states at a particular sequence position. Alternatively, multistate characters can be ordered into a transformation series reflecting a set of assumptions about sequences of character-state evolution. For example, the relative development of a particular bony crest might be coded as an ordered sequence 0–3, for which the states represent crest absent (0), crest present but small (1), crest present and elongated (2), and crest present and bifurcated (3). The corresponding assumption would be that evolution from no crest to a bifurcated crest must proceed through the intermediate states, thus ruling out a punctuated evolution of the bifurcated crest. Whether this assumption is tenable is a biological problem, not a numerical one. Because ordering represents a constraint on character evolution, the tree length for an ordered character must be greater than or equal to that for the corresponding unordered character.

Character polarity is a separate but related concept. Whereas ordering refers to the permitted character-state transformations, polarity refers to hypothesized directions of character evolution through time. Polarity assessment thus presumes the determination of the ancestral state, that of the most recent common ancestor of the group in question. Putative character polarities can be assessed in one of two ways: (1) one character at a time at the beginning of a study (prior polarities), usually by comparison with an out-group condition; or (2) from the results of the study (posterior polarities), after the preferred tree has been rooted. A prior polarity represents a constraint on character evolution beyond the assumption of ordering and thus increases the lengths of preferred trees.

Quantitative character data differ from discrete characters in that they vary incrementally on an interval scale or continuously on a ratio scale with a fixed zero point. Thus they are inherently ordered and have a set of possible states limited only by the resolution of the measuring device. However, the problem of polarity, the assessment of the ancestral state, still holds.

Regardless of whether characters are discrete or continuous, ordered or unordered, there are two assumptions that are fundamental to all character-based phylogenetic studies. The first is that of character independence. The
assumption that evolutionary character-state transitions are uncorrelated across characters permits phylogenetic problems to be partitioned into simpler, independent subproblems. For example, in parsimony methods corresponding branch lengths can be summed across characters to give a total branch length, and corresponding probabilities can be multiplied across characters in maximum likelihood methods. The assumption also allows us to ignore character covariances, which greatly simplifies computational methods. It should be noted that the assumption of character independence, though necessary, is biologically unrealistic and very unsatisfactory to many systematists (Donoghue 1989).

The second fundamental assumption is that of homology (Patterson 1982b; Roth 1988; Smith 1990), usually defined to be the similarity of structure due to descent from a common ancestor. Characters are defined by systematists such that the states among observed taxa are assumed to have been derived via modification from the state of a common ancestor. The critical assumption of homology implies that states among taxa are scientifically as well as numerically comparable and is the methodological basis for studies of character-state transformations and the diagnosis of monophyletic groups. This is not to say that the recognition of homology, or even its operational definition, is a trivial problem, and several different kinds (e.g., operational, taxic) have been proposed and implemented by systematists of various philosophical persuasions (Smith 1990). Note, however, that the evolutionary definitions of homology are not synonymous with mere similarity, as in the term sequence homology used by some molecular biologists.

Pairwise Similarity and Distance Measures

Whereas character data describe individual taxa, similarity and distance data describe the relationship between pairs of taxa. Some procedures, such as immunology and nucleic acid hybridization, provide distance data directly; other kinds of character data (e.g., allozymes, nucleotide sequences) are routinely converted to distances for analysis (but see Farris 1981, 1986). A distance matrix is a symmetrical matrix of which the elements are the pairwise distances among taxa.

Phylogenetic methods based on distance data usually assume the distances to have particular mathematical properties, of which two are of special interest: the additivity and ultrametric properties (Swofford and Olsen 1990).

Additive Distances. A set of additive pairwise distances are mutually compatible to the extent that all distances can be fitted exactly by an unrooted tree, such that, for all pairs of taxa, the branch lengths along the paths connecting the taxa sum exactly to the observed distances between them. A number of methods described below (the additive-tree methods) assume that the ob-
served data are additive; if so, then all should provide the same tree. Because real data are seldom exactly additive due to systematic and random variation, the various procedures often produce different results and can be assessed by their performance when the assumption is violated.

*Ultrametric Distances.* Ultrametric distances are much more highly constrained. They must be additive, but in addition can be fitted exactly by a tree such that all terminal taxa connected to a particular node are equidistant from that node. This in turn implies that the tree must be rooted such that all of the terminal taxa are equidistant from the root. The ultrametric criterion corresponds to the strict evolutionary assumption that all lineages have diverged to equal extent with equal rates of change (the molecular-clock assumption).

It is very unlikely that ultrametric data can be sampled from nature; even if amounts of evolutionary divergence are identical for two or more lineages, statistical sampling fluctuations will lead to deviations from the ultrametric condition. The criterion is important in systematics primarily because cluster analysis techniques (discussed below), which have been widely used to construct phylogenetic trees for more than twenty years, implicitly assume that the distances being clustered are ultrametric.

### Kinds of Methods

The methodologies currently used in phylogenetics can be classified into three broad categories (Felsenstein 1982; Swofford and Olsen 1990): distance methods, parsimony methods, and methods invoking explicit models of evolutionary change over time. The best known of the latter methods are the maximum-likelihood procedures, although the category could include a number of specialized methods geared toward particular kinds of data (such as DNA sequence data).

#### Pairwise-Distance Methods

*Cluster Analysis Methods.* The cluster analysis algorithms typically used in phylogenetic applications are a family of algorithms known as *hierarchical agglomerative procedures* (Rohlf 1970; Williams 1971), because they produce hierarchical trees by beginning with individual taxa and iteratively fusing them into more and more inclusive clusters. Such algorithms are conceptually straightforward. Given a matrix of pairwise distances among taxa, all of these methods first link together the most similar (i.e., least distant) pair, followed by a series of successive linkages involving more and more distant pairs. At each step the two taxa or groups of taxa being linked are merged into a single cluster. The result is usually called a *dendrogram* because it is a
rooted dichotomous treelike structure that portrays clusters of taxa based on their relative similarities.

All hierarchical agglomerative algorithms involve four basic steps: (1) finding the current taxa or clusters \( C_i \) and \( C_j \) having the minimum distance value \( d_{ij} \); (2) defining the depth of branching between the clusters (the position of the node) to be half the distance \( d_{ij} \), according to the ultrametric criterion; (3) merging the two clusters to form a new cluster \( C_h \); and (4) defining (and thus computing) the distances between the new cluster \( C_h \) and all of the other existing taxa or clusters \( (C_k) \) to be a linear function of \( d_{ik} \) and \( d_{jk} \). For \( N \) terminal taxa, the steps are repeated \( N - 1 \) times, each time with one less taxon or cluster, until all taxa are merged into a single cluster.

Out of a large number of possible clustering algorithms, only a few are typically used in phylogenetics, particularly single linkage or nearest neighbor; complete linkage or furthest neighbor; UPGMA (unweighted pair-group method using arithmetic averages); and WPGMA (weighted pair-group method using arithmetic averages). These differ only in the definition of the distances between new and existing clusters (step 4 above). Indeed, Lance and Williams (1967) showed that these and many other algorithms are special cases of a four-parameter general linear model (the so-called flexible strategy) that is applicable to any measure of intertaxon distance. For a given distance matrix, a series of dendrograms ranging from balanced (nearly symmetrical) to chained (completely asymmetrical) can be produced by altering the parameters of the model. The agreement between the resulting dendrogram and the original data matrix can be assessed by the so-called cophenetic correlation between the elements of the matrix and the pairwise distances implied by the minimal nodes of the dendrogram (Rohlf 1970). Of the commonly used procedures, empirical studies have shown UPGMA to yield rather good fits.

Clustering procedures have historically been very popular, primarily because they are fast and easy to implement and because they invoke few assumptions about the data, other than that the intertaxon distances are ultrametric (Colless 1970). The primarily disadvantages are that there seems to be no objective definition of what constitutes the best tree when the data are not ideal. Furthermore, alternative methods are available that are suitable for data that are merely additive, which is a more moderate assumption than that of ultrametric relationships.

**Additive Tree Methods.** Additive tree methods comprise a broad class of methods based on the assumption that the lengths of the branches of a tree can be summed to provide a quantity that is proportional to the amount of evolution that has taken place along a lineage. There are three basic kinds of additive tree methods, all of which produce unrooted trees.

1. The Fitch-Margoliash methods (Fitch and Margoliash 1967; Cavalli-Sforza and Edwards 1967) differ conceptually from cluster analysis in that
they involve the minimization of a merit function that measures the summed deviations between the original pairwise intertaxon distances and those predicted or accounted for by the branch lengths along the tree. The deviations to be minimized may be specified in various ways, for example, as absolute differences or squared differences (the least-squares criterion), and can be weighted by various criteria. As with other minimum-length tree methods, finding the tree with the minimum value of the merit function generally requires two considerations: optimizing the branch lengths of a given tree topology to find the best correspondence between observed and predicted and identifying the particular tree topology having the minimum function value of all possible tree topologies. For a particular topology, the merit function can be minimized iteratively by successive refinement or analytically by linear or quadratic programming. The more general problem of finding the shortest of all possible trees is discussed below ("Searching for Globally Optimum Trees").

2. The so-called distance-Wagner methods (Farris 1972; Swofford 1981; Tateno et al. 1982) are similar conceptually to the Fitch-Margoliash methods in that they find a tree having a minimum value for the merit function: in this case, the total of all absolute branch lengths of the tree. Corresponding to this computational difference, however, is an important conceptual distinction. The conceptual basis of the Fitch-Margoliash methods is that the observed intertaxon distances are estimates of the true evolutionary distances, with some estimates being greater than the true evolutionary values and some being smaller. Thus the estimated or predicted distance, as determined from the branch lengths of the resulting tree, may be either greater or less than the observed distances. In contrast, because a significant amount of evolution may have taken place that is not reflected in the single observed distance value for two taxa, the distance-Wagner method presumes that the observed intertaxon distances are actually lower bounds for the true values, uncorrected for superimposed changes. Thus the observed distances are taken to be lower bounds for the fitted branch lengths, so that the length of the path connecting any two taxa is constrained to equal or exceed the corresponding observed distance. A number of algorithms for finding such minimum-length trees have been described. If the observed distances are exactly additive then the optimal solution can always be found, but if this is not the case then the effects on the resultant tree are unpredictable.

3. The neighbor-joining method (Saitou and Nei 1987; Studier and Keppler 1988) is conceptually related to cluster analysis, in that it is iterative and involves a sequential recomputing of linear functions of distances, but it relaxes the assumption that the intertaxon distances are ultrametric. The procedure begins with the standard distance matrix, but keeps track of the positions of nodes rather than clusters. It computes a modified distance matrix in which the amounts of separation between pairs of nodes are adjusted on the basis of their average deviations from all other nodes. The tree is then con-
structured by joining the least distant pair of adjusted nodes into a new node, removing the branches originating from the new node, and successively repeating this process until only two nodes remain, separated by a single branch. The entire tree is reconstructed from the set of accumulated nodal values. As with the distance-Wagner methods, if the observed intertaxon distances are exactly additive then the optimal solution will always be found. The procedure is relatively new and has few applications in the literature.

Parsimony Methods

Parsimony methods, which invoke the criterion of maximum parsimony, have been the most widely used tree-building procedures in systematics. The principle of maximum parsimony as it applies to phylogenetic inference equates simplicity of explanation with the explanation of shared attributes of two or more taxa being due to inheritance from a common ancestor (Sober 1989). When character conflicts occur this simplest explanation cannot hold, and additional assumptions of homoplasy (convergence, parallelism, or reversal of character states) must be invoked to account for the observed data. The parsimony methods summarized here are all based directly on character data, as coded in a character matrix $X_{(n \times p)}$. The result of parsimony analysis is a cladogram depicting nested sets of derived character states, from which inferences about the historical relationships are then made. Such cladograms are often called most-parsimonious reconstructions (MPR).

Parsimony algorithms operate by finding, for any particular tree topology, the distribution of character states that minimizes the total tree length, where each branch length is taken to be proportional to the number of required character-state changes (evolutionary steps) along it (Fitch 1975, 1977). Such trees are known in graph theory as Steiner minimum trees (Gilbert and Pollak 1968). The tree-construction methods differ primarily in the underlying evolutionary assumptions or constraints that they invoke about the nature of character-state changes along branches, assumptions that must be made to determine the total tree length. However, they share the difficult necessity of searching for the particular tree topology that results in the shortest total length. Thus, parsimony methods in general are distinguished by their optimality criteria rather than by the particular algorithms used to search for optimal trees.

By far the most widely used parsimony methods used have been the Wagner and Fitch procedures. The so-called Wagner method, formulated by Kluge and Farris (1969) and Farris (1970) on the basis of Wagner's (1961, 1969) tree construction protocol, assumes only that each character is measured on an interval scale, which is an appropriate assumption for continuous characters and for binary and ordered multistate characters. Fitch (1971) generalized the method to allow the use of unordered multistate characters, as are common
with DNA (nucleotide) and protein (amino acid) data. The Wagner and Fitch procedures both permit the free reversibility of character states along the tree; that is, they assume that the probabilities of changes among contiguous character states are symmetrical. The Wagner and Fitch algorithms as originally formulated find only one tree (MPR), even though others may exist (Swoford and Maddison 1987), but they can be modified to handle multifurcations.

While the Wagner and Fitch methods assume probabilities of character-state changes to be symmetrical along branches, the Dollo parsimony model (Farris 1977) imposes an asymmetry on possible directions of change: a derived character state may be independently evolved any number of times, but once lost, it cannot be regained. Such a model seems to be appropriate for molecular restriction-site data, for which the "gaining" of the character state can be interpreted as the loss of the cleavage site, and vice versa (Debry and Slade 1985); however, if a particular restriction site does originate independently in two lineages, the Dollo model can significantly overestimate the actual number of evolutionary changes. The original parsimony method of Camin and Sokal (1965), which was the first discrete-character parsimony procedure to be formalized, makes an asymmetry assumption that is far more strict than the Dollo assumption: namely, that any derived condition cannot reverse to the ancestral character state. This is the strongest assumption about character-state evolution made by any of the parsimony methods.

These various assumptions about ordered versus unordered character states and the symmetries of character-state transitions can be subsumed into a generalized parsimony method (Sankoff 1975; Swoford and Olsen 1990) that assigns a weight or "cost" for the transformation of each character state to every other state. The transformation weights are tabulated in a cost matrix, special cases of which can be determined for the Wagner, Fitch, and Dollo parsimony models. An exact dynamic-programming algorithm can be used both to determine the minimum tree length required on a given tree topology for any particular assignments of costs and to obtain one or more of the MPR corresponding to this length (Sankoff and Cedergren 1983). Such an algorithm is very expensive computationally, which is a current if not future limitation. A more general problem is that it may not be clear in any particular case how to assign the costs between character states (Williams and Fitch 1989). Details about conceptual and computational aspects of these parsimony algorithms and particular examples of each are provided by Swoford and Olsen (1990).

Maximum-Likelihood Methods

A final (if arbitrary) category of tree-building methods includes those that invoke explicit probabilistic models of evolutionary change. The most familiar of this category are the maximum-likelihood methods, originally proposed
by Edwards and Cavalli-Sforza (1964; Cavalli-Sforza and Edwards 1967) but further developed and described in detail by Felsenstein (1973; 1981a,b; 1982; 1984b; 1985; 1988). Maximum-likelihood methods are common in hypothesis-testing applications in which a clear pair of alternative explanations of the data is presented. They are more difficult to apply when numerous alternative and complex hypotheses are evaluated, which is the case in phylogenetic inference.

However, the rationale behind maximum-likelihood procedures becomes clearer if we view phylogenetic inference as a means of finding the phylogenetic hypothesis (as summarized by a tree having a specific topology and set of branch lengths) that is most consistent with the data. Given an explicit model of the evolutionary process that converts one character state into another, a maximum-likelihood approach would estimate the likelihood that the given evolutionary model will yield the observed set of character-state distributions among taxa. If we could evaluate this likelihood for a collection of possible phylogenies, we could then select the one (or more) that provides the greatest likelihood of "explaining" the data.

The sources of controversy about maximum-likelihood procedures center on the necessity of defining particular models of evolution. Several models have been used, most of which are very simplistic. For morphological data the model typically assumes that all characters are evolving independently, each following a random sequence (Brownian motion) with a mean displacement of zero and a constant variance in displacement per unit time. This model is tractable analytically because, after t units of time, the resulting phenotypes can be considered to change by an amount drawn from a normal distribution with a mean of zero and a variance proportional to t. It is then possible to obtain maximum-likelihood estimates of the best-fitting evolutionary trees. Such trees can be estimated from continuous or discrete data, given such a random-walk model of independent evolution.

For molecular sequence data the evolutionary models can be made more explicit in terms of patterns of nucleotide substitution; however, the procedures in current use are still based on very simple models of sequence evolution. For example, the Jukes and Cantor (1969) model invokes the assumptions that the four nucleotides are equally frequent and that all substitutions are equally likely. The Kimura (1980) model is somewhat less restrictive in that it assumes independent substitution rates for transitions versus transversions. Both models make explicit quantitative assumptions about the nature of the relationship between substitution rate and sequence similarity. Felsenstein (1988) has generalized the Kimura model, replacing the assumption that the nucleotides are equally frequent with a more relaxed assumption that the relative nucleotide frequencies are maintained at equilibrium.

Regardless of the assumptions by which the branch lengths of a maximum-likelihood tree are optimized, the result is a likelihood value for the particular
tree topology. Several different methods have been proposed for assessing the statistical significance of the likelihood values and for comparing the likelihoods of two or more alternative trees (Felsenstein 1988; Kishino and Hasegawa 1989).

Searching for Globally Optimum Trees

As noted above, all of the tree-building methods that are associated with specific optimality criteria distinguish the problem of finding the particular optimal combination of branch lengths for a given tree topology from that of finding the one or more trees (from the extremely large number of possibilities) that provide the globally optimum solution. For data sets of up to approximately twenty taxa the exact solutions can usually be found, either by exhaustive search (systematically evaluating all possible trees) or using the so-called branch-and-bound methods (Hendy and Penny 1982; Swofford 1990). The latter are conceptually related to the combinatorial exhaustive-search algorithms but instead systematically search through the possible branch rearrangements in such a way as to localize the subset of possible trees of minimum length.

For data sets of more than about twenty-two taxa (given the current technology) even the most efficient branch-and-bound methods would require a prohibitive amount of computer time, and consequently other methods must be used, methods that are approximate and that cannot be guaranteed to find the globally optimum trees. These heuristic approaches generally operate by hill-climbing methods, seeking rearrangements of the tree that result in improved values of the optimality criteria and continuing until no further improvements can be found (Swofford 1990). Like hill-climbing methods in general, such approaches will find local optima that might or might not correspond to global optima, and various heuristic fixes can be applied to perturb the search or to begin it from a different starting position. One particularly useful type of perturbation is branch-swapping, a systematic predefined interchanging of branches designed to move the search off of local optima in the hope of finding a better hill to climb.

Other Issues

There are a large number of other important issues that follow from these basic considerations but that cannot be elaborated here. These include such things as: (1) decisions about weighting or not weighting different characters and about how to manage missing data; (2) methods for detecting internal inconsistencies and systematic biases in data sets; (3) problems associated with estimating ancestral character states and identifying appropriate and
inappropriate character-state transitions; (4) evaluation of the reliability and robustness of cladograms and trees, based either on asymptotic statistical assumptions or on resampling techniques (bootstrapping and jackknifing); (5) comparisons of trees derived from different kinds of data sets; (6) choice of the most appropriate optimality criterion for continuous characters; (7) justification of the use of parsimony as a universal optimality criterion; (8) mapping of trees onto so-called character spaces and other kinds of geometric manifolds; (9) potential ramifications of character correlations, either among terminal taxa or along evolutionary lineages; (10) methods of dealing with multiple solutions (e.g., consensus trees); and (11) the role of fossils as hypothetical ancestral taxa. Some of these problems are inherently biological and must be tackled by evolutionary biologists, while others are numerical and involve the development or refinement of computational algorithms and more realistic probabilistic evolutionary models. The discipline of phylogenetic inference lies at the heart of contemporary evolutionary biology and hence is likely to remain a dynamic and controversial enterprise for some time to come.
LITERATURE CITED


