Annu. Rev. Ecol. Syst. 1997. 28:437–66
Copyright © 1997 by Annual Reviews Inc. All rights reserved

PHYLOGENY ESTIMATION AND HYPOTHESIS TESTING USING MAXIMUM LIKELIHOOD

John P. Huelsenbeck
Department of Biology, University of Rochester, Rochester, NY 14627;
e-mail: johnh@onyx.si.edu

Keith A. Crandall
Department of Zoology and M. L. Bean Museum, Brigham Young University, Provo, UT 84602

KEY WORDS: maximum likelihood, phylogeny, hypothesis test, evolutionary model, molecular evolution

ABSTRACT

One of the strengths of the maximum likelihood method of phylogenetic estimation is the ease with which hypotheses can be formulated and tested. Maximum likelihood analysis of DNA and amino acid sequence data has been made practical with recent advances in models of DNA substitution, computer programs, and computational speed. Here, we describe the maximum likelihood method and the recent improvements in models of substitution. We also describe how likelihood ratio tests of a variety of biological hypotheses can be formulated and tested using computer simulation to generate the null distribution of the likelihood ratio test statistic.

INTRODUCTION

Only recently has phylogenetics been recognized as a field that has basic relevance to many questions in biology. Phylogenies have proven to be important tools of research in fields such as human epidemiology (42, 86), ecology (7), and evolutionary biology (43). In fact, for any question in which history may be a confounding factor, phylogenies have a central role (25). To the outsider interested in using a phylogeny, one of the most frustrating aspects of the field

0066-4162/97/1120-0437$08.00

of systematics is the lack of agreement as to which of the many methods of analysis to use (54). Which method of analysis is best suited for a particular data set is a question of current, and often vehement, debate. The evolutionary biologist is often left asking not only, "Which method of analysis do I use?", but, if different methods lead to conflicting genealogies, "Which tree do I believe?". As researchers become drawn to the opinion that phylogeny reconstruction is a problem of statistical inference, it is important to examine phylogenetic methods for their statistical properties and assumptions.

Several criteria can be used as a basis to choose among methods. One criterion is accuracy; that is, how well do different methods estimate the correct tree? This criterion has captured most of the attention of systematists, and there is a veritable scientific cottage industry producing papers that examine the performance of different phylogenetic methods for simulated data sets (29, 44, 49, 50, 54, 62, 97, 108), well-supported phylogenies (2, 17, 52), and experimental phylogenies (16, 48). However, criteria besides accuracy are also important. For example, a phylogenetic method should provide some means of falsifying the assumptions made during the analysis (88). All phylogenetic methods, by necessity, must make specific assumptions about the evolutionary process. Typical assumptions include a bifurcating tree as a model to describe the genealogy of a group and a model of character change. Yet, many methods do not provide a means of testing these assumptions. Furthermore, some provision for choosing among different models of evolution should be available. In choosing between a simple model of character change and a more complex model, for example, how can one justify using the complex model in a phylogenetic analysis? Finally, the methods should be able to estimate the confidence in the phylogeny and provide a framework for testing phylogenetic hypotheses.

Inasmuch as phylogenies are important for many evolutionary questions, the criteria posed above are important. In this review, we concentrate on one method of phylogenetic estimation—maximum likelihood. Recent advances have made maximum likelihood practical for analysis of DNA and amino acid sequence data. Many of the advances consist of improvements in the models of DNA substitution implemented by maximum likelihood. However, increased computer speed and improved computer programs have also played an important role. In this paper, we review the recent advances made in maximum likelihood estimation of phylogenetic trees. Specifically, we examine how maximum likelihood has been used for phylogeny estimation and hypothesis testing.

THE LIKELIHOOD PRINCIPLE

The method of maximum likelihood is usually credited to the English statistician RA Fisher, who described the method in 1922 and first investigated its

properties (28). The method of maximum likelihood depends on the complete specification of the data and a probability model to describe the data. The probability of observing the data under the assumed model will change depending on the parameter values of the model. The maximum likelihood method chooses the value of a parameter that maximizes the probability of observing the data.

A Coin Tossing Experiment

Consider the simple experiment of tossing a coin with the goal of estimating the probability of heads for the coin. The probability of heads for a fair coin is 0.5. For this example, however, we assume that the probability of heads is unknown (perhaps the fairness of the coin is being tested) and must be estimated. The three main components of the maximum likelihood approach are (*a*) the data, (*b*) a model describing the probability of observing the data, and (*c*) the maximum likelihood criterion.

Assume that we performed the coin flip experiment, tossing a coin n times. An appropriate model that describes the probability of observing h heads out of n tosses of a coin is the binomial distribution. The binomial distribution has the following form

$$\Pr[h \mid p, n] = \binom{n}{h} p^h (1-p)^{n-h}, \qquad 1.$$

where p is the probability of heads, the binomial coefficient $\binom{n}{h}$ gives the number of ways to order h successes out of n trials, and the vertical line means "given."

Assuming independence of the individual and discrete outcomes, the likelihood function is simply the joint probability of observing the data under the model. For the coin toss experiment in which a binomial distribution is assumed, the likelihood function becomes

$$L(p \mid h, n) = \binom{n}{h} p^h (1-p)^{n-h}. \qquad 2.$$

Often the log likelihood is used instead of the likelihood for strictly computational purposes. Taking the natural log of the function does not change the value of p that maximizes the likelihood.

Figure 1 shows a plot of likelihood, L, as a function of p for one possible outcome of $n = 10$ tosses of a coin (six heads and four tails). The likelihood appears to be maximized when p is the proportion of the time that heads appeared in our experiment. This illustrates a computational way to find the maximum likelihood estimate of p (change p in small increments until a maximum is found). Alternatively, the maximum likelihood estimate of p can be found analytically by taking the derivative of the likelihood function with respect to p and finding where the slope is zero. If this is done, we find that the estimate

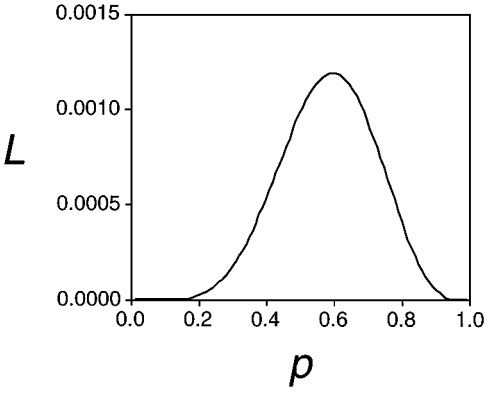


Figure 1 The likelihood surface for one possible realization of the coin tossing experiment. Here, six heads and four tails are observed. The likelihood appears to be maximized when $p = 0.6$.

of p is $\hat{p} = {}^{h}/_{n}$. The estimate of p is simply the proportion of heads that we observed in our experiment.

Maximum likelihood estimates perform well according to several criteria. Statistical estimators are evaluated according to their consistency, efficiency, and bias. These criteria provide an idea of how concentrated an estimate of a parameter is around the true value of that parameter. A method is consistent if the estimate converges to the true value of the parameter as the number of data increases. An efficient method provides a close estimate of the true value of the parameter (i.e., the variance of the estimate is small), and an unbiased estimator does not consistently under- or over-estimate the true value of the parameter. Mathematically, an estimate $\hat{\theta}_n$ of a parameter θ based on a sample of size n is consistent if

$$P(|\hat{\theta}_n - \theta| > \varepsilon) \rightarrow 0, \quad \text{as } n \rightarrow \infty \tag{3.}$$

for any $\varepsilon > 0$ where θ is the true value of the parameter. The mean square error (MSE) measures the variance and bias of an estimate. The MSE of an estimate $\hat{\theta}$ is

$$\begin{aligned} MSE(\hat{\theta}) &= E(\hat{\theta} - \theta)^2 \\ &= \text{Var}(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2. \end{aligned} \tag{4.}$$

The first term of the MSE is the variance of the estimate, and the second term is the bias. A method is unbiased if the expectation of the estimate equals the true value of the parameter $[E(\hat{\theta}) = \theta]$. When two competing estimators of a parameter are considered, the estimate with the smaller MSE is said to be more

efficient. Maximum likelihood estimates are typically consistent under the model. Furthermore, they are asymptotically efficient, meaning that the variance of a maximum likelihood estimate is equal to the variance of any unbiased estimate as the sample size increases. However, maximum likelihood estimates are often biased (e.g., the maximum likelihood estimate of the parameter σ^2 of a normal distribution is biased).

MAXIMUM LIKELIHOOD IN PHYLOGENETICS

The application of maximum likelihood estimation to the phylogeny problem was first suggested by Edwards & Cavalli-Sforza (20). However, they found the problem too computationally difficult at the time and attempted approximate methods instead (thereby introducing the minimum evolution—later to be known as the parsimony—and the least squares methods of phylogeny estimation). The subsequent history of maximum likelihood can be read as a steady progress in which computational barriers were broken and the models used were made more biologically realistic. For example, Neyman (83) applied maximum likelihood estimation to molecular sequences (amino acids or nucleotides) using a simple model of symmetric change that assumed substitutions were random and independent among sites. It was not until Felsenstein's implementation (24), however, that a general maximum likelihood approach was fully developed for nucleotide sequence data. Below, we outline the basic strategy for obtaining maximum likelihood estimates of phylogeny, given a set of aligned nucleotide or amino acid sequences. We then show how the basic strategy can be complicated by biological reality.

Conceptually, maximum likelihood in phylogenetics is as simple as the example given above for estimating the probability of heads in a coin toss experiment. The data for molecular phylogenetic problems are the individual site patterns. We assume that the sequences have been aligned, though the alignment procedure can be explicitly incorporated into the estimation of phylogeny (111, 112). For example, for the following aligned DNA sequences of $s = 4$ taxa,

Taxon 1 ACCAGC
Taxon 2 AACAGC
Taxon 3 AACATT
Taxon 4 AACATC,

the observations are $\mathbf{x}_1 = \{\text{A, A, A, A}\}'$, $\mathbf{x}_2 = \{\text{C, A, A, A}\}'$, $\mathbf{x}_3 = \{\text{C, C, C, C}\}'$, $\mathbf{x}_4 = \{\text{A, A, A, A}\}'$, $\mathbf{x}_5 = \{\text{G, G, T, T}\}'$, and $\mathbf{x}_6 = \{\text{C, C, T, C}\}'$. If one were interested in coding the data as amino acids, the above sequences, if in frame, would be represented as $\mathbf{x}_1 = \{\text{Thr, Asn, Asn, Asn}\}'$ and $\mathbf{x}_2 = \{\text{Ser, Ser, Ile, Ile}\}'$. The sample consists of n vectors (as many vectors as there

are sites in the sequence, with the elements of each vector denoting the nucleotide state for each taxon for site *i*.

Note that two of the sites exhibit the same site pattern ($\mathbf{x}_1$ and $\mathbf{x}_4$). There are a total of $r = 4^s$ site patterns possible for *s* species. The number of sites exhibiting different site patterns can also be considered as the data in a phylogenetic analysis. For example, the above data matrix can also be described as

Taxon 1	AAAAAAAAA...C...C...C...G...TTT
Taxon 2	AAAAAAAAA...A...C...C...G...TTT
Taxon 3	AAAACCCCG...A...C...T...T...TTT
Taxon 4	ACGTACGTA...A...C...C...T...CGT
Number	2 0 0 0 0 0 0 0 0...1...1...1...1...0 0 0

where the matrix is now a 4×256 matrix of all $r = 4^4 = 256$ site patterns possible for four species. The site patterns are labeled $1, 2, \ldots, r$. Most of the possible site patterns for our sample data set are not observed. However, 5 site patterns are observed (now labeled $\mathbf{y}_1 = \{\text{A, A, A, A}\}'$, $\mathbf{y}_{65} = \{\text{C, A, A, A}\}'$, $\mathbf{y}_{86} = \{\text{C, C, C, C}\}'$, $\mathbf{y}_{94} = \{\text{C, C, T, C}\}'$, and $\mathbf{y}_{176} = \{\text{G, G, T, T}\}'$). The numbers of sites exhibiting each site pattern are contained in a vector $\mathbf{n}$ ($n_1 = 2$, $n_{65} = 1$, $n_{86} = 1$, $n_{94} = 1$, $n_{176} = 1$, with all other $n_i = 0$ for the example data matrix).

Maximum likelihood assumes an explicit model for the data. Just as with the coin tossing experiment, the data are considered as random variables. However, instead of two possible outcomes, there are $r = 4^s$ possible outcomes for DNA sequences. Hence, the data can be described using a multinomial distribution. The multinomial distribution is a generalization of the binomial distribution and has the following form:

$$\Pr[n_1, n_2, \ldots, n_r \mid p_1, p_2, \ldots, p_r] = \binom{n}{n_1, n_2, \ldots, n_r} \prod_{i=1}^{r} p_i^{n_i}, \qquad 5.$$

where $\binom{n}{n_1, n_2, \ldots, n_r}$ is the number of ways that *n* objects can be grouped into *r* classes, n_i is the number of observations of the *i*th site pattern, and p_i is the probability that site pattern *i* occurs. A maximum likelihood estimate of p_i is $\hat{p}_i = {}^{n_i}/_{n}$ (that is, the probability of the *i*th class is the proportion of the time it was observed). The likelihood, then, can be calculated assuming a multinomial distribution by setting the likelihood equal to Equation 5. However, by using a multinomial distribution, one cannot estimate topology or other biologically interesting parameters. Hence, models that incorporate phylogeny are assumed. The difference in the log likelihood of the data under multinomial and phylogenetic models, however, represents the cost associated with assuming

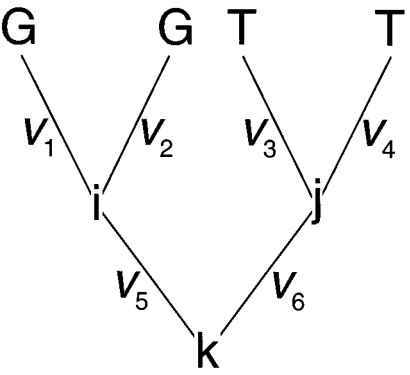


Figure 2 The likelihood method assumes that observed sequences (here, the nucleotides for site pattern 176 from the example in the text; $\mathbf{y}_{176} = \{\text{G, G, T, T}\}'$) are related by a phylogenetic tree (τ) with branch lengths ($v_1, \ldots, v_6$) specified in terms of expected number of substitutions per site. The probability of observing the data ($\mathbf{y}_{176}$) is a sum over the possible assignments of nucleotides to the internal nodes, *i*, *j*, and *k*.

a phylogenetic tree and substitution model (32, 80, 93) and has also been used to show that maximum likelihood in phylogenetics is consistent (117).

Just like the multinomial probability model, phylogenetic models specify the probability of observing different site patterns. At a minimum, a phylogenetic model for molecular data includes a tree (τ) relating the sequences with branch lengths of the tree ($\mathbf{v}$) specified in terms of expected number of changes per site and a model of sequence change. Consider just one of the nucleotide site patterns, above ($\mathbf{y}_{176}$), for the tree of Figure 2. Because the identities of the nucleotides at the internal nodes *i*, *j*, and *k* are unknown, the probability of observing site pattern 176 is a sum of 64 terms (the $4^3 = 64$ possible assignments of nucleotides to nodes *i*, *j*, and *k*),

$$\begin{aligned}\Pr[\mathbf{y}_{176} = \{\text{G, G, T, T}\} \mid \tau, v_1, \ldots, v_6, \Theta] \\ = \sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{4} \pi_k p_{Gi}(v_1, \Theta) p_{Gi}(v_2, \Theta) p_{Tj}(v_3, \Theta) \\ \times\ p_{Tj}(v_4, \Theta) p_{ik}(v_5, \Theta) p_{jk}(v_6, \Theta),\end{aligned} \qquad 6.$$

where nucleotides, A, C, G, or T are assumed at the internal nodes if *i*, *j*, or *k* are equal to 1, 2, 3 or 4, respectively, π_i is the equilibrium frequency of nucleotide *i*, and $p_{xy}(v_i, \Theta)$ is the probability of observing nucleotides *x* and *y* at the tips of a branch given the branch length and other parameters, Θ, of the substitution model. Felsenstein (24) pointed out that the above calculation can be performed

much more quickly by taking advantage of the tree structure when performing the summations over nucleotides at interior nodes. If instead of DNA sequences, amino acid sequences are used, the above equation involves a summation over $20^3 = 8000$ possible assignments of amino acids to the internal nodes, and $p_{xy}(v_i, \Theta)$ represents the probability of observing amino acids x and y at the tips of a branch given the branch length and other parameters of the substitution model.

Assuming independence among sites, the likelihood of a tree (τ) is

$$L(\tau, \mathbf{v}, \Theta \mid \mathbf{y}_1, \ldots, \mathbf{y}_r) = \binom{n}{n_1, n_2, \ldots, n_r} \prod_{i=1}^{r} \Pr[\mathbf{y}_i \mid \tau, \mathbf{v}, \Theta]^{n_i}, \qquad 7.$$

where $\mathbf{v}$ is a vector containing the lengths of the branches and is either $\mathbf{v} = (v_1, \ldots, v_{2s-2})$ for rooted trees or $\mathbf{v} = (v_1, \ldots, v_{2s-3})$ for unrooted trees (s is the number of sequences), and r is the total number of site patterns possible for s sequences. The multinomial coefficient $\binom{n}{n_1, n_2, \ldots, n_r}$ is a constant and is usually disregarded when calculating the likelihood of a tree. Also, to speed computation of the likelihood, the product is taken only over observed site patterns. The likelihood as formulated in Equation 7 does not consider the order of the site patterns. However, for several models of DNA substitution, the order of the sites is of interest and cannot be disregarded (27, 120). For such problems, the likelihood function is the product over all sites

$$L(\tau, \mathbf{v}, \Theta \mid \mathbf{x}_1, \ldots, \mathbf{x}_n) = \prod_{i=1}^{n} \Pr[\mathbf{x}_i \mid \tau, \mathbf{v}, \Theta]. \qquad 8.$$

The method of Felsenstein (24) is to choose the tree that maximizes the likelihood as the best estimate of phylogeny. This application of likelihood is unusual because the likelihood function changes depending on the tree (81, 124). In principle, to find the maximum likelihood tree, one must visit each of the $\frac{(2s-5)!}{2^{s-3}(s-3)!}$ possible unrooted bifurcating trees in turn (23). For each tree, one finds the combination of branch lengths and other parameters that maximizes the likelihood of the tree (that maximizes the likelihood function, above). The maximum likelihood estimate of phylogeny is the tree with the greatest likelihood.

This procedure of visiting all possible trees and calculating the likelihood for each is computationally expensive. Fortunately, there are many short cuts that can substantially speed up the procedure. As mentioned above, Felsenstein described an efficient method to calculate the likelihood by taking advantage of the tree topology when summing over all possible assignments of nucleotides to internal nodes (24). There are also efficient ways of optimizing branch lengths that involve taking the first and second derivatives of the likelihood function with respect to the branch of interest (see 67). Finally, rather than visiting each

possible tree, search algorithms concentrate on only those trees that have a good chance of maximizing the likelihood function (see 105).

POISSON PROCESS MODELS

To calculate the probability of observing a given site pattern, the transition probabilities $[P_{xy}(v_i, \Theta)]$ need to be specified. All current implementations of likelihood assume a time-homogeneous Poisson process to describe DNA or amino acid substitutions.

An Example with Two-Character States

As an example of how transition probabilities are calculated, consider a very simple case for which only two character states exist (0 or 1). The rate of change from 0 to 1 or from 1 to 0 in an infinitesimal amount of time, δt, is specified by the rate matrix, **Q**

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} -\lambda\pi_1 & \lambda\pi_1 \\ \lambda\pi_0 & -\lambda\pi_0 \end{pmatrix}. \quad 9.$$

The states are ordered 0, 1 along the rows and diagonals; λ is the rate of change from 0 to 1 or from 1 to 0; and π_0 and π_1 are the equilibrium frequencies of states 0 and 1, respectively. The diagonals of the rate matrix are negative to satisfy the mathematical requirement that the row sums are zero. The matrix may be multiplied by a constant such that the average rate of substitution is one, and time (t) is then measured by the expected number of substitutions per site (v). The matrix **Q**, above, is reversible because it satisfies the requirement that $\pi_i q_{ij} = \pi_j q_{ji}$.

To calculate the probability of observing a change over an arbitrary interval of time, t, the following matrix calculation is performed: $\mathbf{P}(t, \Theta) = p_{ij}(t, \Theta) = e^{\mathbf{Q}t}$ (15). The vector Θ contains the parameters of the substitution model (in this case $\Theta = \{\pi_0, \pi_1\}$). For many substitution models, the transition probability matrix $\mathbf{P}(t, \Theta)$ can be calculated analytically. For example, the transition probabilities for the two-state case are

$$\mathbf{P}(t, \Theta) = \{p_{ij}(t, \Theta)\} = \begin{pmatrix} \pi_0 + (1-\pi_0)e^{-\lambda t} & \pi_1 - \pi_1 e^{-\lambda t} \\ \pi_0 - \pi_0 e^{-\lambda t} & \pi_1 + (1-\pi_1)e^{-\lambda t} \end{pmatrix}. \quad 10.$$

However, for complicated rate matrices, the probability matrix can be calculated numerically (e.g., 118).

Models of DNA Substitution

Many of the advances in maximum likelihood analysis in phylogenetics have come through improvements in the models of substitution assumed. One of

the simplest models of DNA substitution—the Jukes-Cantor (JC69) model—assumes that the base frequencies are equal ($\pi_A = \pi_C = \pi_G = \pi_T$) and that the rate of change from one nucleotide to another is the same for all possible changes (58). However, the JC69 model, like several other models, is simply a special case of a general model of DNA substitution for which the instantaneous rate matrix **Q** has the following form:

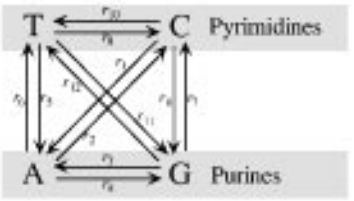
$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} \cdot & r_2\pi_C & r_4\pi_G & r_6\pi_T \\ r_1\pi_A & \cdot & r_8\pi_G & r_{10}\pi_T \\ r_3\pi_A & r_7\pi_C & \cdot & r_{12}\pi_T \\ r_5\pi_A & r_9\pi_C & r_{11}\pi_G & \cdot \end{pmatrix} \qquad 11.$$

(3, 94, 109, 118). The rows and columns are ordered A, C, G, and T. The matrix gives the rate of change from nucleotide i (arranged along the rows) to nucleotide j (along the columns). For example, $r_2\pi_C$ gives the rate of change from "A" to "C". Let $\mathbf{P}(v, \Theta) = \{p_{ij}(v, \Theta)\}$ be the transition probability matrix, where $p_{ij}(v, \Theta)$ is the probability that nucleotide i changes into j over branch length v. The vector Θ contains the parameters of the substitution model (e.g., π_A, π_C, π_G, π_T, r_1, r_2, r_3, ...). As for the two-state case, to calculate the probability of observing a change over a branch of length v, the following matrix calculation is performed: $\mathbf{P}(v, \Theta) = e^{\mathbf{Q}v}$.

A model based on the matrix **Q** (Equation 11) represents the most general 4×4 model of DNA substitution currently available (Figure 3). The model is nonreversible, meaning that the rate matrix **Q** does not satisfy the reversibility condition ($\pi_i q_{ij} = \pi_j q_{ji}$). Because the model is nonreversible, the likelihood of an unrooted tree changes depending on the root position. Therefore, the likelihood must be maximized over the $\frac{(2s-3)!}{2^{s-2}(s-2)!}$ rooted trees. For reversible models of DNA substitution, on the other hand, the likelihood is maximized over the $\frac{(2s-5)!}{2^{s-3}(s-3)!}$ unrooted trees because, for a given unrooted topology, the likelihood is the same regardless of where the tree is rooted. Many commonly used models of DNA substitution are subsets of this general model. Table 1 shows the parameter settings of the general model that give a variety of models of DNA substitution. Maximum likelihood explicitly incorporates a model of substitution into the estimation procedure, as do distance methods; parsimony methods, on the other hand, incorporate variations of these models implicitly. Because other models of DNA substitution are subsets of this general model, and because they are often subsets of one another as well, likelihood ratio tests of the model of DNA substitution can be easily performed testing whether a particular parameter provides a significant improvement in the likelihood (as is discussed later).

The four-state character models describe the substitution process at a single site. The assumption of independence among sites is necessary in phylogenetic

A



B

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} \cdot & r_2\pi_C & r_4\pi_G & r_5\pi_T \\ r_1\pi_A & \cdot & r_8\pi_G & r_{10}\pi_T \\ r_3\pi_A & r_7\pi_C & \cdot & r_{12}\pi_T \\ r_6\pi_A & r_9\pi_C & r_{11}\pi_G & \cdot \end{pmatrix}$$

Figure 3 The likelihood method assumes that substitutions follow a Poisson process, with the rate of nucleotide substitution specified by a rate matrix, **Q**. The rates of change from one nucleotide to another are specified by the parameters $r_1, \ldots, r_{12}$ (A). The rate matrix, **Q**, corresponding to the diagram (*A*) is shown in *B*. Different models of substitution are just special cases of the general matrix shown here. For example, to obtain the Jukes-Cantor (JC69; 58) model, the rates of change among all nucleotides are equal ($r_1 = r_2 = \cdots = r_{12}$) and base frequencies are equal ($\pi_A = \pi_C = \pi_G = \pi_T$).

analyses and is assumed when calculating the likelihood of a tree (the joint probability of observing all site patterns is the product of the individual site patterns). However, biologists know that the substitution processes at different sites in a sequence often are not independent. For example, hydrogen-bonded sites in the stem regions of ribosomal genes are not independent because a substitution in one nucleotide changes the probability that a compensatory substitution will occur in its partner.

Several authors have examined the effect of non-independent substitution among pair-bonded stem nucleotides. For example, Dixon & Hillis (18) examined the appropriate weights that stem sites should receive to correctly accommodate non-independence in a parsimony analysis. Others have devised time-homogeneous Poisson process models to describe substitutions in stem regions. Instead of a 4×4 matrix, these models assume a 16×16 rate matrix (**Q**) of all possible nucleotide doublets possible in a stem (76, 95, 98, 99). The instantaneous rate of change is set to zero if more than one substitution is required to change from doublet *i* to doublet *j* (e.g. $q_{ij} = 0$ for AC $\rightarrow$ CG). The other rate parameters of the **Q** matrix are specified in different ways depending

Table 1 Parameter settings for a variety of evolutionary models employed in maximum likelihood analyses. Parameters of the substitution matrix, **Q**, are shown in Figure 3.

Model	Nucleotide frequencies	Rates of change	Reference
JC69	$\pi_A = \pi_C = \pi_G = \pi_T$	$r_1 = r_2 = r_3 = r_4 = r_5 = r_6 = r_7 = r_8 = r_9 = r_{10} = r_{11} = r_{12}$	58
K80	$\pi_A = \pi_C = \pi_G = \pi_T$	$r_3 = r_4 = r_9 = r_{10}; r_1 = r_2 = r_5 = r_6 = r_7 = r_8 = r_{11} = r_{12}$	60
K3ST	$\pi_A = \pi_C = \pi_G = \pi_T$	$r_3 = r_4 = r_9 = r_{10}; r_5 = r_6 = r_7 = r_8; r_1 = r_2 = r_{11} = r_{12}$	61
F81	$\pi_A; \pi_C; \pi_G; \pi_T$	$r_1 = r_2 = r_3 = r_4 = r_5 = r_6 = r_7 = r_8 = r_9 = r_{10} = r_{11} = r_{12}$	24
HKY85	$\pi_A; \pi_C; \pi_G; \pi_T$	$r_3 = r_4 = r_9 = r_{10}; r_1 = r_2 = r_5 = r_6 = r_7 = r_8 = r_{11} = r_{12}$	45
TrN	$\pi_A; \pi_C; \pi_G; \pi_T$	$r_3 = r_4; r_9 = r_{10}; r_1 = r_2 = r_5 = r_6 = r_7 = r_8 = r_{11} = r_{12}$	107
SYM	$\pi_A = \pi_C = \pi_G = \pi_T$	$r_1 = r_2; r_3 = r_4; r_5 = r_6; r_7 = r_8; r_9 = r_{10}; r_{11} = r_{12}$	126
GTR	$\pi_A; \pi_C; \pi_G; \pi_T$	$r_1 = r_2; r_3 = r_4; r_5 = r_6; r_7 = r_8; r_9 = r_{10}; r_{11} = r_{12}$	64

on the assumptions the biologist is willing to make. For example, Schöniger & von Haeseler (98) considered the simplest case by setting $q_{ij} = \pi_j$ if doublets i and j differ at one nucleotide (e.g., $q_{AC,AG} = \pi_{AG}$).

Other models consider the substitution process at the level of the codon. Fundamental to interpreting changes in substitution rates is an accurate assessment of how these changes influence the resulting protein; that is, does a substitution produce a change in the amino acid (a nonsynonymous change), or does the substitution not alter the protein (a synonymous change) (100). An increase in the relative amount of nonsynonymous change can be strong evidence for adaptive evolution (72). Many methods exist for estimating synonymous and nonsynonymous substitution rates (68–70, 75, 82, 87, 89). Recently, maximum likelihood has been used to estimate synonymous and nonsynonymous rates of change (33, 77, 78). These authors modeled the substitution process at the level of the codon. They used a 61×61 matrix to describe the instantaneous rate of change from one codon to another (the three stop codons are not considered). For the model of Muse & Gaut (78) the instantaneous rate of change from codon i to j is zero ($q_{ij} = 0$) if the change requires more than one substitution, $q_{ij} = \beta\pi_{n_{ij}}$ if the substitution causes a change in the amino acid, and $q_{ij} = \alpha\pi_{n_{ij}}$ if the substitution is synonymous (where n_{ij} is the equilibrium frequency of the substituted nucleotide). The model of Goldman & Yang (33) is similar except that they allow a transition/transversion rate bias and consider the physicochemical

properties of the 20 amino acids by using Grantham's distances (34). The parameters of both models can be estimated using maximum likelihood. Unlike many of the standard methods, these approaches do not rely on the assumption that the number and location of silent/replacement sites do not change over time. The codon-based approach has been used successfully to explain the rate heterogeneity found in the chloroplast genome as being due primarily to differences in nonsynonymous substitution rates (78). As an alternative to modeling the substitution process at the level of the codon, the substitution process has also been modeled at the amino acid level (1, 9, 110).

The assumption of equal rates at different sites can also be relaxed. Several models of among-site rate heterogeneity have been developed (e.g., assume that the sites are log normally distributed—39, 85; assume an invariant rate class—12, 45, 91; estimate the rates in different data partitions separately, see 105; or use a combination of different rate distributions—38, 114). Probably the most widely used model is one that assumes that rates are gamma distributed (57, 116). The gamma distribution is a continuous probability density function that has wide application in probability and statistics (the well-known χ^2 and exponential distributions are special cases of the gamma distribution). Systematists have co-opted this distribution for their own use because the shape of the gamma distribution changes dramatically depending on the value of the shape parameter, α, and the scale parameter, λ. Systematists set $\alpha = 1/\lambda = a$ so that the mean of the gamma distribution is 1.0 and the variance is $1/a$. Rates at different sites, then, are thought of as random variables drawn from a gamma distribution with shape parameter a. When a is equal to infinity, the gamma model of among-site rate heterogeneity collapses to the equal rates case. However, most empirical estimates of the shape parameter a fall in the range of 0.1 to 0.5 (121), indicating substantial rate variation among sites. Yang (116, 119) provides details on how to calculate the likelihood under a gamma model of rate heterogeneity.

An advantage of the likelihood approach is that the models can be made complicated to incorporate other biologically important processes. For example, the models of substitution can be modified to account for insertion and deletion events (5, 111, 112), secondary structure of proteins (76a, 110), and correlated rates among sites (11, 27, 120). In the course of estimating phylogeny, the maximum likelihood method provides estimates of model parameters that may be of interest to the biologist. If the biologist is only interested in phylogeny, then these additional parameters are considered nuisance parameters (i.e., parameters not of direct interest to the biologist but which must be accommodated in the analysis by either integrated likelihood or maximum relative likelihood methods; see 31). However, maximum likelihood estimates of parameters such as the variance in the rate of substitution among sites or the bias in the substitution

process are of interest to students of molecular evolution. Interestingly, many of the models that are currently implemented in likelihood were algebraically intractable and therefore unusable for other methods of phylogenetic inference. However, now that computer speed is affordable, these models and estimates of their parameters are feasible. Furthermore, Monte Carlo methods promise to make tractable models that are currently difficult or impossible to implement (37, 63, 84, 125).

LIKELIHOOD RATIO TESTS IN PHYLOGENETICS

Currently, one of the most debated subjects in the field of phylogenetics concerns the role that assumptions play in a phylogenetic analysis (8, 21, 74). All phylogenetic methods make assumptions about the process of evolution. An assumption common to many phylogenetic methods, for example, is a bifurcating tree to describe the phylogeny of species. However, additional assumptions are made in a phylogenetic analysis. For example, the assumptions of a maximum likelihood analysis are mathematically explicit and, besides the assumption of independence among sites, include parameters that describe the substitution process, the lengths of the branches on a phylogenetic tree, and among-site rate heterogeneity. The assumptions made in a parsimony analysis include independence and a specific model of character transformation (often called a step-matrix or weighting scheme; a commonly used weighting scheme is to give every character transformation equal weight). Phylogenetic methods can estimate the correct tree with high probability despite the fact that many of the assumptions made in any given analysis are incorrect. In fact, the maximum likelihood, parsimony, and several distance methods appear to be robust to violation of many assumptions, including making incorrect assumptions about the substitution process, among-site rate variation, and independence among sites (50, 55, 99). The advantage of making explicit assumptions about the evolutionary process is that one can compare alternative models of evolution in a statistical context. Instead of being viewed as a disadvantage, the use of explicit models of evolution in a phylogenetic analysis allows the systematist not only to estimate phylogeny, but to learn about processes of evolution through hypothesis testing.

One measure of the relative tenability of two competing hypotheses is the ratio of their likelihoods. Consider the case in which L_0 specifies the likelihood under the null hypothesis, whereas L_1 specifies the likelihood of the same data under the alternative hypothesis. The likelihood ratio is

$$\Lambda = \frac{\max[L_0(\text{Null Model} \mid \text{Data})]}{\max[L_1(\text{Alternative Model} \mid \text{Data})]} \qquad 12.$$

(19, 59, 71, 101). Here, the maximum likelihood calculated under the null hypothesis (H_0) is in the numerator, and the maximum likelihood calculated under the alternative hypothesis (H_1) is in the denominator. When Λ is less than one, H_0 is discredited and when Λ is greater than one, H_1 is discredited. Λ greater than one is only possible for non-nested models. When nested models are considered (i.e., the null hypothesis is a subset or special case of the alternative hypothesis), $\Lambda < 1$ and $-2 \log \Lambda$ is asymptotically χ^2 distributed under the null hypothesis with q degrees of freedom, where q is the difference in the number of free parameters between the general and restricted hypotheses.

An alternative means of generating the null distribution of $-2 \log \Lambda$ is through Monte Carlo simulation (parametric bootstrapping; 13, 14). Felsenstein (26) first suggested the use of the parametric bootstrap procedure in phylogenetics. Goldman (32) was among the first to apply the method in phylogenetics and to demonstrate that the usual χ^2 approximation of the null distribution is not appropriate for some tests involving phylogeny. In parametric bootstrapping, replicate data sets are generated using simulation under the assumption that the null hypothesis is correct. The maximum likelihood estimates of the model parameters under the null hypothesis are used to parameterize the simulations. For the phylogeny problem, these parameters would include the tree topology, branch lengths, and substitution parameters (e.g., transition:transversion rate ratio or the shape parameter of the gamma distribution). For each simulated data set, $-2 \log \Lambda$ is calculated anew by maximizing the likelihood under the null and alternative hypotheses. The proportion of the time that the observed value of $-2 \log \Lambda$ exceeds the values observed in the simulations represents the significance level of the test. Typically, the rejection level is set to 5%; if the observed value for the likelihood ratio test statistic is exceeded in less than 5% of the simulations, then the null hypothesis is rejected. Although there are good statistical reasons for test statistics based on the probability density of the data, the parametric bootstrap procedure may also be used to determine the null distribution of other test statistics (47).

Maximum likelihood allows the easy formulation and testing of phylogenetic hypotheses through the use of likelihood ratio tests (though also see 96). Furthermore, likelihood ratio tests are known to have desirable statistical properties. For example, they are known to be uniformly most powerful when simple hypotheses are considered and often outperform other hypothesis tests for composite hypotheses (92). Over the past two decades, numerous likelihood ratio tests have been suggested. These include tests of the null hypotheses that (*a*) a model of DNA substitution adequately explains the data (32, 80, 93), (*b*) rates of nucleotide substitution are biased (32, 80, 93), (*c*) rates of substitution are constant among lineages (24, 65, 79, 115), (*d*) rates are equal among sites (123), (*e*) rates of substitution are the same in different data partitions (30, 66, 122),

(*f*) substitution parameters are the same among data partitions (122), (*g*) the same topology underlies different data partitions (51), (*h*) a prespecified group is monophyletic (53), (*i*) hosts and associated parasites have corresponding phylogenies (56), (*j*) hosts and parasites have identical speciation times (56), and (*k*) rates of synonymous and nonsynonymous substitution are the same (77). Here, we describe a few of these tests. Our goal is not to provide an exhaustive list and description of all the likelihood ratio tests that have been proposed in phylogenetics but rather to illustrate how biological questions can be addressed in a simple way using likelihood.

Testing the Model of DNA Substitution

Current models implemented by maximum likelihood and distance methods assume that DNA substitutions follow a time-homogeneous Poisson process. As mentioned above, these models have been made complex to incorporate biological reality by the addition of parameters that allow for biased substitution and among-site rate variation. Likelihood provides the systematist with a rationale for choosing among different possible models through the use of likelihood ratio tests.

Typically, the question asked is "Does the addition of a substitution parameter provide a significant increase in the likelihood"? For example, one possible null hypothesis to consider is that transitions and transversions occur at the same rate. The Felsenstein (24; designated F81) model assumes equal rates for all substitutions and could be used to calculate the likelihood under the null hypothesis. The likelihood under the null hypothesis is compared to the likelihood under an alternative hypothesis that assumes a model of DNA substitution that allows a different rate of substitution for transitions and transversions. In this case, the HKY85 (45) model would be an appropriate model of DNA substitution to assume for the alternative hypothesis; the F81 and HKY85 models are identical except that the HKY85 model includes a parameter that allows for a different rate for transitions and transversions. The likelihood ratio test statistic ($-2 \log \Lambda$) is calculated, and the significance level is approximated by comparing $-2 \log \Lambda$ to χ^2 distribution with 1 degree of freedom.

As an example, consider data from five species of vertebrates. The data consist of aligned DNA sequences of 1383 sites from the albumin gene (Fish: *Salmo salar*, X5297; Frog: *Xenopus laevis*, M18350; Bird: *Gallus gallus*, X60688; Rat: *Rattus norvegicus*, J00698; Human: *Homo sapiens*, L00132). For this set of taxa, the phylogeny is almost certainly (Fish,(Frog,(Bird,(Rat,Human)))) (52). As mentioned above, these data could be analyzed using any one of several models. Here, we consider a hierarchy of hypotheses (Figure 4). Ideally, the hierarchy of hypotheses to be considered should be formulated before analysis begins. The null hypotheses considered are (*a*) base frequencies are equal,

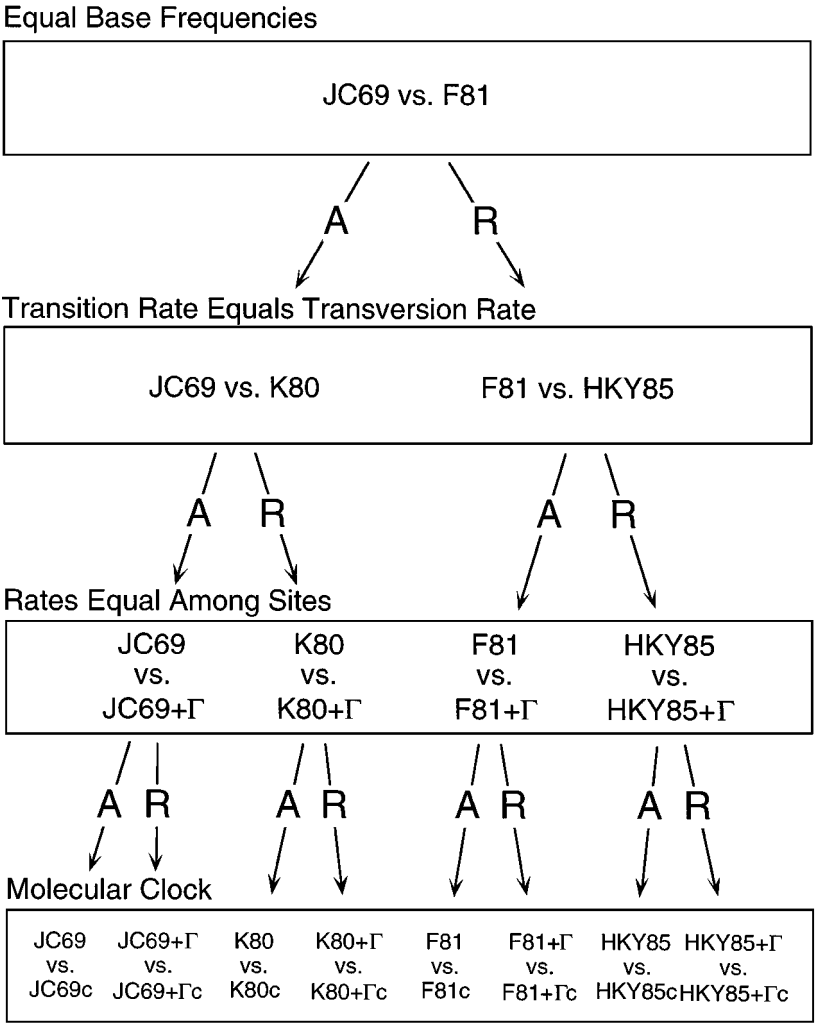


Figure 4 The hierarchy of hypotheses examined for the albumin data from five vertebrates. The parameters of the models are explained in Table 1. At each level, the null hypothesis is either accepted, "A," or rejected, "R."

(*b*) transitions and transversions occur at the same rate, (*c*) the rate of substitution is equal among sites, and (*d*) rates among lineages are constant through time (i.e., the molecular clock holds).

Figure 5 shows the trees and log likelihoods of the albumin data under several different models of DNA substitution. For the first null hypothesis considered—that base frequencies are equal—the likelihood under the JC69 model is compared to the likelihood under the alternative hypothesis calculated assuming the F81 model. These two models are identical except that the F81 model allows for unequal base frequencies and the JC69 model assumes equal base frequencies. Furthermore, the models are nested because the JC69 model is simply a special case of the F81 model. Therefore, the likelihood ratio test statistic, $-2 \log \Lambda$, can be compared to a χ^2 distribution with 3 degrees of freedom. For the test of equal base frequencies, $-2 \log \Lambda = 17.56$, a value much greater than 7.82, which represents the 95% critical value from a χ^2 distribution with 3 degrees of freedom. Therefore, the null hypothesis that base frequencies are equal is rejected, and the F81 model is preferred to the JC69 model of DNA substitution. The other three null hypotheses considered can also be tested using likelihood ratio tests. Table 2 shows the results of these tests. For the albumin data set, the most parameter-rich model considered (HKY85+Γ) is the best fitting model (i.e., provides a statistically significant increase in likelihood over the other models considered). Although the HKY85+Γ model was found to best fit the data, a general test of model adequacy indicates that the model is inadequate to explain the data (H_0: HKY85+Γ, H_1: unconstrained model, H_0 rejected at $P < 0.01$; 32). As expected, our models of DNA substitution do not fully describe the process of evolution leading to the observed sequences.

Although this conclusion sounds dire, it should be taken with a grain of salt because we know a priori that our models are inadequate to explain all the features of the evolutionary process. However, although the model is in some sense false, this does not detract from the utility of the model for estimating parameters such as topology and branch lengths, especially given the observation that phylogenetic methods in general, and maximum likelihood in particular, can be robust to violation of assumptions (50). Note that although a hierarchy of hypotheses was considered in this example, an alternative means of specifying the tests would be to treat the most general model as the alternative against which the other models are compared. Also note that the same tree was estimated for each of these models and that this tree is the one generally acknowledged as the best based on other sources of evidence (e.g., fossil and morphological data; 4). This is true even though the assumptions of all of the models are violated to some degree and some of the models considered (e.g., the JC69 model) poorly describe the data. Hence, the contention that methods using wrong models are

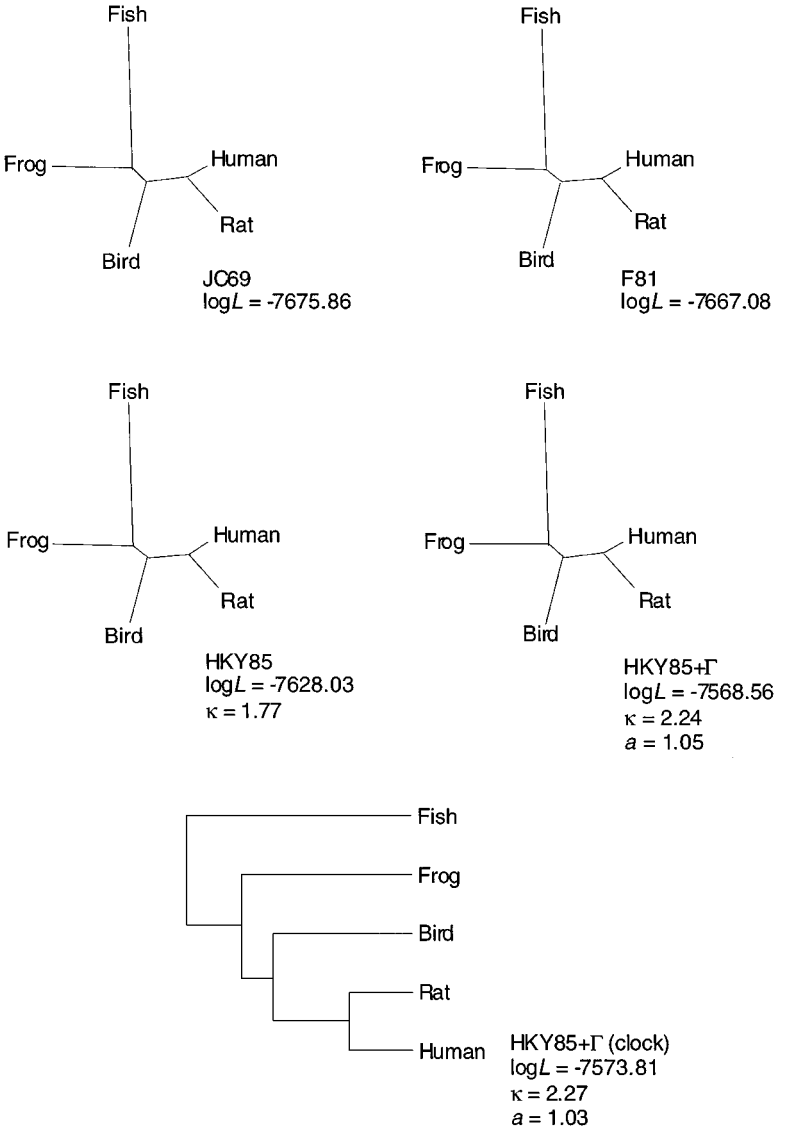


Figure 5 The phylogenies estimated for the albumin data under five different models of DNA substitution. The same phylogeny is estimated in each case, and this phylogeny is consistent with the traditional phylogeny of vertebrates [(Fish,(Frog,(Bird,(Rat,Human))))]. Note that the estimate of the transition:transversion rate ratio (κ) changes depending on whether or not among-site rate variation is accounted for; κ is underestimated when rate variation is not accounted for. Models that assume gamma distributed rates are denoted with a "+Γ."

Table 2 The results of likelihood ratio tests performed on the albumin DNA data from five vertebrates

Null hypothesis	Models compared	log L_0	log L_1	$-2 \log \Lambda$	d.f.	P
Equal base frequencies	H_0: JC69 H_1: F81	−7675.86	−7667.08	17.56	3	2.78×10^{-5}
Transition rate equals transversion rate	H_0: F81 H_1: HKY85	−7667.08	−7628.03	78.10	1	9.75×10^{-19}
Equal rates among sites	H_0: HKY85 H_1: HKY85+Γ	−7628.03	−7568.56	118.94	1	0
Molecular clock	H_0: HKY85+Γc H_1: HKY85+Γ	−7573.81	−7568.56	10.5	3	1.47×10^{-2}

L_0 and L_1 denote the likelihoods under the null (H_0) and alternative (H_1) hypotheses, respectively. P represents the probability of obtaining the observed value of the likelihood ratio test statistic ($-2 \log \Lambda$) if the null hypothesis were true. Because multiple tests are performed, the significance value for rejection of the null hypothesis should be adjusted using a Bonferroni correction (hence, the significance level for rejection of the null hypothesis is set to 1.25×10^{-2}).

"poor estimators of hierarchical pattern when the assumptions of the models are violated" (8, p. 426) appears overstated. Just as with other methods of phylogenetic estimation, the maximum likelihood method can be robust to a variety of model violations (50).

Tests of Topology

IS A PRESPECIFIED GROUP MONOPHYLETIC? A taxonomic group is monophyletic if all its members share a most recent common ancestor. Many phylogenetic studies are aimed at determining the monophyly, or lack thereof, of some group of organisms. The most controversial of these studies have questioned the monophyly of groups, such as the rodents, bats, and toothed whales (35, 36, 73, 90), long defined as having a common evolutionary history based on morphological similarities. Often, the monophyly of a group has important evolutionary implications, particularly with respect to selection and adaptation. Pettigrew, for example, argued on the basis of neurological characters that megachiropterans (flying foxes) are more closely related to primates than they are to microchiropterans (90). This hypothesis of relationship implies that bats are not a monophyletic group and that either flight evolved twice (independently) in mammals or that flight evolved once in mammals but was subsequently lost in the primates.

How can the Pettigrew hypothesis that bats do not form a monophyletic group be tested? An analysis of the interphotoreceptor retinoid binding protein (IRBP) gene, that has been sequenced in primates, bats, and other mammals, provides an example of a likelihood ratio test that can be generalized to other questions

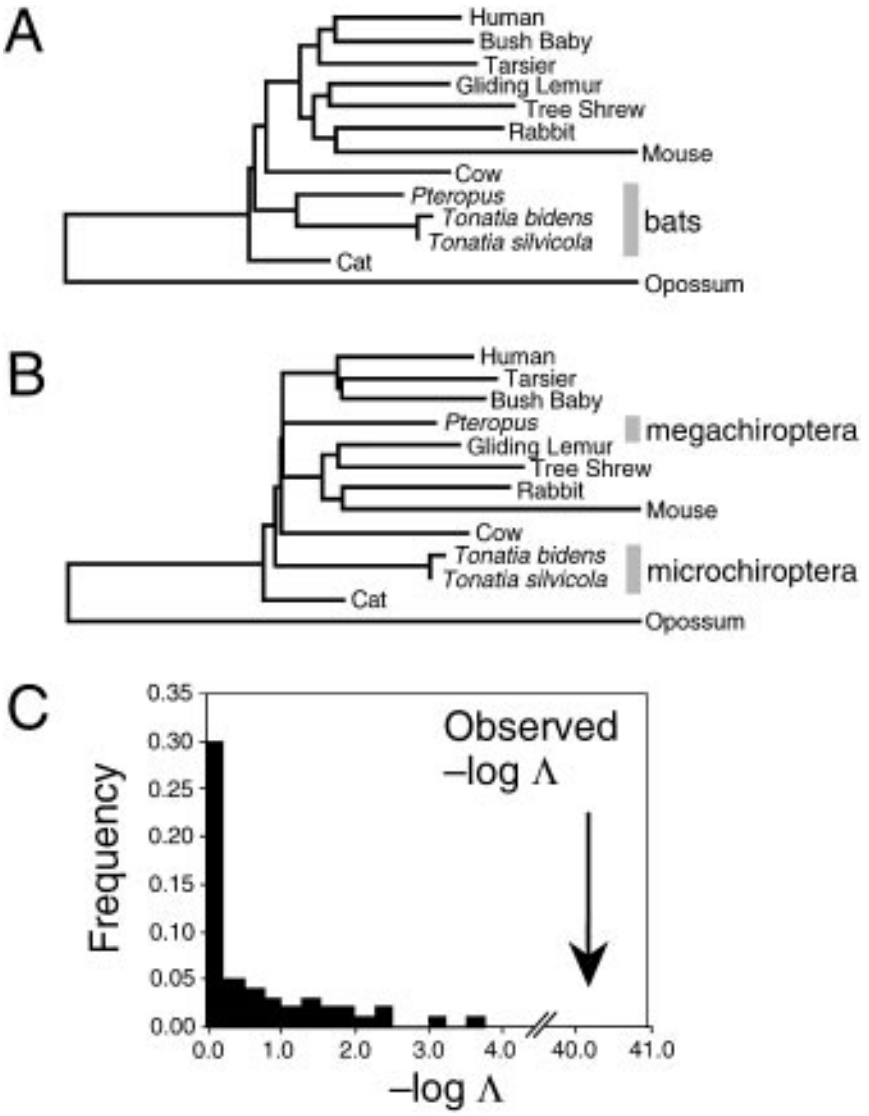
of relationship (53, 103, 113). The maximum likelihood tree for the IRBP gene is shown in Figure 6*A* (113). The tree is consistent with the monophyly of bats. The best tree based on the assumption that bats are not monophyletic has a log likelihood 40.16 less than the best tree (Figure 6*B*). There are two ways to explain this result: (*a*) the Pettigrew hypothesis that bats are not monophyletic is correct, but estimation error has resulted in a tree consistent with bat monophyly; or (*b*) bats are a monophyletic group.

The ratio of the likelihoods calculated under the null model (bats constrained to be nonmonophyletic) and under the alternative hypothesis (no constraints placed on relationships) provides a measure of the relative support of the two hypotheses. In this case, the ratio of the likelihoods is $-\log \Lambda = 40.16$ (113). How damaging is this likelihood ratio to the Pettigrew hypothesis? Figure 6*C* shows the distribution of $-\log \Lambda$ that would be expected if the null hypothesis that bats are not monophyletic is true. The observed value of $-\log \Lambda$ is much greater than would be expected under the null hypothesis. Hence, the Pettigrew hypothesis can be rejected for the IRBP gene with a significance level of less than 1% (113).

What are the statistical properties of the likelihood ratio test of monophyly? Simulation study suggests that the method can be powerful (i.e., frequently rejects the null hypothesis when, in fact, the null hypothesis is false). Figure 7 shows the results of a study in which one of two trees was simulated (53). Simulating data for Tree 1 generates the distribution under the null hypothesis, whereas simulating data for Tree 2 generates data under the alternative hypothesis. The graph shows that the likelihood ratio test of monophyly can be powerful; the power of the test increases, as expected, when the number of sites in the analysis is increased. Although promising, the simulations presented in Figure 7 represent an ideal situation for the likelihood ratio test of monophyly. Additional simulations suggest that the test also performs well when the overall rate of substitution is low and an incorrect model is implemented in the likelihood analysis. However, when an incorrect model is used and the overall rate of substitution is high, the test rejects the null hypothesis too often. Hence, the likelihood ratio test of monophyly should be implemented with as biologically realistic a model as possible to prevent false rejection of the null hypothesis.

DO DIFFERENT DATA SETS CONVERGE TO SIGNIFICANTLY DIFFERENT TREES?

The comparison of phylogenetic trees estimated from different data partitions has been used to address a variety of biological questions. Here, a data partition is defined as a division of characters into two or more subsets. The characters in each subset are either suspected to or have been demonstrated to have evolved according to different processes (e.g., different rates of substitution, different levels of selection, or different underlying phylogenies). Examples of potential


A
Human
Bush Baby
Tarsier
Gliding Lemur
Tree Shrew
Rabbit
Mouse
Cow
Pteropus
Tonatia bidens
Tonatia silvicola
bats
Cat
Opossum
B
Human
Tarsier
Bush Baby
Pteropus
Gliding Lemur
Tree Shrew
Rabbit
Mouse
Cow
Tonatia bidens
Tonatia silvicola
megachiroptera
microchiroptera
Cat
Opossum
C
Observed
–log Λ
Frequency
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.0
1.0
2.0
3.0
4.0
40.0
41.0
–log Λ

partitions of DNA characters include first- and second- versus third-codon positions, different genes, different coding regions within genes, or different genomic regions (e.g., nuclear versus mitochondrial, or different viral segments). The incongruence of the trees estimated from different genes in bacteria has been used to demonstrate horizontal gene transfer and recombination in bacteria (40). Similarly, the incongruence of trees estimated from different viral DNA segments has been used to show that reassortment of the segments has frequently occurred in the hanta virus (46). Such an approach can also be used to indicate method failure; if the partitioned data have evolved on the same underlying phylogenetic tree, but different trees have been estimated from each data partition, then either sampling error or systematic bias by the phylogenetic method are at fault (10). Finally, the congruence of trees estimated from host and associated parasites can be used to infer cospeciation (6, 41, 102).

How can incongruence of phylogenetic trees estimated from different data partitions be tested? A simple likelihood ratio test can be used to test whether the same phylogeny underlies all data partitions (51). The likelihood under the null hypothesis (L_0) is calculated by assuming that the same phylogenetic tree underlies all of the data partitions. However, branch lengths and other parameters of the substitution model are estimated independently in each. The likelihood under the alternative hypothesis (L_1) is calculated by relaxing the constraint that the same tree underlies each data partition. The alternative hypothesis allows the possibility that the histories of all data partitions are different. The likelihood ratio test statistic [$-2 \log \Lambda = -2 (\log L_0 - \log L_1)$] is compared to a null distribution generated using parametric bootstrapping.

Rejection of the null hypothesis of homogeneity (i.e., the same phylogeny for all data partitions) can indicate one of several different processes. One possibility is that a different history underlies different data partitions; incongruence of this sort could be caused by recombination, horizontal gene transfer, or ancestral polymorphism. Another possibility is that the phylogenetic methods have failed for one or more data partitions. All phylogenetic methods can

←

Figure 6 The phylogenetic relationship of bats and other mammals. Trees were estimated using maximum likelihood under a model that allows an unequal transition:transversion rate, unequal base frequencies, and among-site rate heterogeneity (the HKY85+Γ model). Maximum likelihood trees were obtained using a tester version of the program PAUP* 4.0 (104). Bats form a monophyletic group in the maximum likelihood tree ($\log L = -5936.52$) (*A*). The best tree under the Pettigrew hypothesis (that megachiroptera are constrained to be a sister group with the primates) ($\log L = -5976.69$) (*B*). The distribution of the likelihood ratio test statistic under the assumption that the null hypothesis (the Pettigrew hypothesis) of relationship is correct (*C*). The observed likelihood ratio test statistic ($-\log \Lambda = 40.16$) is significant at $P < 0.01$. Hence, the Pettigrew hypothesis is rejected.

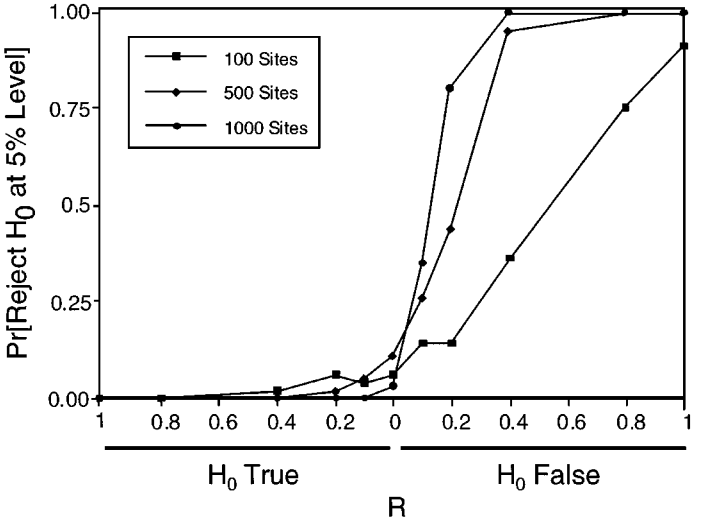


Figure 7 The power of the likelihood ratio test of monophyly. The graph represents the results of a computer simulation of two unrooted four taxon trees. The null hypothesis is true when the tree ((A, B), C, D) is simulated and false when ((A, C), B, D) is simulated. The external branches were constrained to be equal in length. *R* represents the ratio of the length of the internal branch to the external branches; when $R = 1.0$, all branches of the tree are equal in length. The null hypothesis is that the taxa A and B are a group (that the taxon bipartition {A, B} {C, D} exists). The null hypothesis is rarely rejected when it is true and frequently rejected when false.

produce biased estimates of phyolgeny when their assumptions are severely violated. For example, the parsimony method is known to produce inconsistent estimates of phylogeny for some very simple four-species trees (22). A method is inconsistent when it converges to an incorrect phylogenetic tree as more nucleotide or amino acid sites are included in the analysis. Regardless of the cause of the incongruence, combining the data is unwarranted because one runs the risk of either obscuring an interesting evolutionary phenomenon (e.g., different histories for different data partitions) or providing a poor estimate of phylogeny when method failure for one or more data partitions is at fault (10).

Null Distributions: χ^2 or Parametric Bootstrapping?

For many of the tests discussed here, nested hypotheses are considered (i.e., the null hypothesis is a special case of the alternative hypothesis). For most statistical problems involving nested hypotheses, the χ^2 distribution with q degrees of freedom (where q is the difference in the number of free parameters between the alternative and null hypotheses) can be used to test the significance of the likelihood ratio test statistic ($-2 \log \Lambda$). However, the phylogeny problem is

an unusual statistical problem, and for many nested phylogenetic hypotheses, it is known that the χ^2 distribution is not appropriate (32). For example, for the tests of monophyly and topological incongruence among different data partitions, the χ^2 distribution cannot be used to determine the significance level of the likelihood ratio test statistic. On the other hand, the χ^2 distribution appears to be appropriate when testing whether the addition of a substitution parameter provides a significant improvement in the likelihood. The reason that the χ^2 distribution is appropriate for some phylogenetic probelms but not for others appears to be related to the fact that a tree topology is not a standard statistical parameter (124). In fact, for tests involving maximization over trees, it is difficult to determine the appropriate degrees of freedom because it is not clear how many parameters a topology represents. Probably the safest course for many likelihood ratio tests is to generate the null distribution using computer simulation (parametric bootstrapping). For null hypotheses that are composite, the maximum likelihood values under the null hypothesis can be used to parameterize the simulation (13, 14). This procedure has the advantage that it does not rely on asymptotic results and can be applied to non-nested as well as to nested hypotheses. The sensitivity of the paramertic bootstrap procedure to incorrect assumptions, however, has not been widely tested (though see 53).

CONCLUSIONS

Phylogenies are being applied to a wider variety of biological questions than ever before. One of the challenges for systematists is to develop appropriate tests to address the questions posed by evolutionary biologists. Statistical tests of phylogenetic questions can be formulated in many different ways. However, for many of the tests posed to date, the underlying assumptions are not clear. In fact, for some tests the null hypothesis is unclear (106). The testing of phylogenetic hypotheses in a likelihood framework should prove useful in the future. Likelihood provides a unified framework for the evaluation of alternative hypotheses. With the use of parametric bootstrapping, likelihood ratio tests can be applied to questions for which the null distribution is difficult to determine analytically.

The application of likelihood ratio tests in phylogenetics is a recent phenomenon, with most of the research activity occurring in the past five years. However, in that time the approach has proven powerful. Likelihood ratio tests have already provided information on the pattern of DNA substitution. Furthermore, the approach has been applied to questions involving topology and has even allowed the examination of whether hosts and parasites cospeciated (56). Future research can investigate the statistical properties of likelihood ratio tests with the objective of determining the power and robustness of the tests.

Another avenue of research involves the development of likelihood ratio tests to address additional null hypotheses. The likelihood approach should be particularly useful because the development of a likelihood ratio test is straightforward as long as the null and alternative hypotheses can be precisely described.

ACKNOWLEDGMENTS

We thank Rasmus Nielsen, Jack Sites, and Spencer Muse for comments on earlier versions of this manuscript. Spencer Muse, especially, made many helpful comments pointing out inadequacies of an earlier manuscript. This work was supported by a Miller postdoctoral fellowship awarded to JPH and an Alfred P. Sloan Young Investigator Award to KAC.

Visit the *Annual Reviews home page* at http://www.annurev.org.

Literature Cited

1. Adachi J, Hasegawa M. 1992. Amino acid substitution of proteins coded for in mitochondrial DNA during mammalian evolution. *Jpn. J. Genet.* 67:187–97
2. Allard MW, Miyamoto MM. 1992. Perspective: testing phylogenetic approaches with empirical data, as illustrated with the parsimony method. *Mol. Biol. Evol.* 9:778–86
3. Barry D, Hartigan JA. 1987. Statistical analysis of hominoid molecular evolution. *Stat. Sci.* 2:191–210
4. Benton MJ. 1990. Phylogeny of the major tetrapod groups: morphological data and divergence dates. *J. Mol. Biol.* 30:409–24
5. Bishop MJ, Thompson EA. 1986. Maximum likelihood alignment of DNA sequences. *J. Mol. Biol.* 190:159–65
6. Brooks DR. 1981. Hennig's parasitological method: a proposed solution. *Syst. Zool.* 30:229–49
7. Brooks DR, McLennan DA. 1991. *Phylogeny, Ecology, and Behavior.* Chicago, IL: Univ. Chicago Press. 434 pp.
8. Brower AVZ, DeSalle R, Vogler A. 1996. Gene trees, species trees, and systematics: a cladistic perspective. *Annu. Rev. Ecol. Syst.* 27:423–50
9. Bruno WJ. 1996. Modeling residue usage in aligned protein sequences via maximum likelihood. *Mol. Biol. Evol.* 13:1368–74
10. Bull JJ, Huelsenbeck JP, Cunningham CW, Swofford DL, Waddell PJ. 1993. Partitioning and combining data in phylogenetic analysis. *Syst. Biol.* 42:384–97
11. Churchill GA. 1989. Stochastic models for heterogeous DNA sequences. *Bull. Math. Biol.* 51:79–94
12. Churchill GA, von Haeseler A, Navidi WC. 1992. Sample size for a phylogenetic inference. *Mol. Biol. Evol.* 9:753–69
13. Cox DR. 1961. Tests of separate families of hypotheses. *Proc. 4th Berkeley Symp. Math. Stat. Prob.* 1:105–23
14. Cox DR. 1962. Further results on tests of separate families of hypotheses. *J. R. Stat. Soc. B* 24:406–24
15. Cox DR, Miller HD. 1977. *The Theory of Stochastic Processes.* London: Chapman & Hall
16. Crandall KA. 1994. Intraspecific cladogram estimation: accuracy at higher levels of divergence. *Syst. Biol.* 43:222–35
17. Cummings MP, Otto SP, Wakeley J. 1995. Sampling properties of DNA sequence data in phylogenetic analysis. *Mol. Biol. Evol.* 12:814–22
18. Dixon MT, Hillis DM. 1993. Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. *Mol. Biol. Evol.* 10:256–67
19. Edwards AWF. 1972. *Likelihood.* Cambridge: Cambridge Univ. Press
20. Edwards AWF, Cavalli-Sforza LL. 1964. Reconstruction of evolutionary trees. In *Phenetic and Phylogenetic Classification*, ed. J McNeill, pp. 67–76. London: Syst. Assoc.
21. Farris JS. 1983. The logical basis of phylogenetic analysis. In *Advances in Cladistics*, ed. NI Platnick, VA Funk, 2:7–36.

New York: Columbia Univ. Press
22. Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27:401–10
23. Felsenstein J. 1978. The number of evolutionary trees. *Syst. Zool.* 27:27–33
24. Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17:368–76
25. Felsenstein J. 1985. Phylogenies and the comparative method. *Am. Nat.* 125:1–15
26. Felsenstein J. 1988. Phylogenies from molecular sequences. *Annu. Rev. Genet.* 22:521–65
27. Felsenstein J, Churchill GA. 1996. A hidden Markov model approach to variation among sites in rate of evolution. *Mol. Biol. Evol.* 13:93–104
28. Fisher RA. 1922. On the mathematical foundations of theoretical statistics. *Philos. Trans. R. Soc. London Ser. A* 222:309–68
29. Gaut BS, Lewis PO. 1995. Success of maximum likelihood phylogeny inference in the four-taxon case. *Mol. Biol. Evol.* 12:152–62
30. Gaut BS, Weir BS. 1994. Detecting substitution-rate heterogeneity among regions of a nucleotide sequence. *Mol. Biol. Evol.* 11:620–29
31. Goldman N. 1990. Maximum likelihood inference of phylogenetic trees with special reference to a Poisson process model of DNA substitution and to parsimony analyses. *Syst. Zool.* 39:345–61
32. Goldman N. 1993. Statistical tests of models of DNA substitution. *J. Mol. Evol.* 36:182–98
33. Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* 11:725–36
34. Grantham R. 1974. Amino acid difference formula to help explain protein evolution. *Science* 185:862–64
35. Grauer D, Hide WA, Li W-H. 1991. Is the guinea-pig a rodent? *Nature* 351:649–52
36. Grauer D, Hide WA, Zharkikh A, Li W-H. 1992. The biochemical phylogeny of guinea-pigs and gundis, and the paraphyly of the order rodentia. *Comp. Biochem. Physiol. B* 101:495–98
37. Griffiths RC, Tavaré S. 1994. Simulating probability distributions. *Theor. Popul. Biol.* 46:131–59
38. Gu X, Fu Y-X, Li W-H. 1995. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Mol. Biol. Evol.* 12:546–57
39. Guoy M, Li W-H. 1989. Phylogenetic analysis based on rRNA sequences supports the archaebacterial rather than the eocyte tree. *Nature* 339:145–47
40. Guttman DS. 1997. Recombination and clonality in natural populations of *Escherichia coli*. *Trends Ecol. Evol.* 12:16–22
41. Hafner MS, Nadler SA. 1988. Phylogenetic trees support the coevolution of parasites and their hosts. *Nature* 332:258–59
42. Harvey PH, Nee S. 1994. Phylogenetic epidemiology lives. *Trends Ecol. Evol.* 9:361–63
43. Harvey PH, Pagel MD. 1991. *The Comparative Method in Evolutionary Biology.* Oxford: Oxford Univ. Press. 239 pp.
44. Hasegawa M, Kishino H, Saitou N. 1991. On the maximum likelihood method in molecular phylogenetics. *J. Mol. Evol.* 32:443–45
45. Hasegawa M, Kishino K, Yano T. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22:160–74
46. Henderson WW, Monroe MC, St. Jeor SC, Thayer WP, Rowe JE, et al. 1995. Naturally occurring sin nombre virus genetic reassortants. *Virology* 214:602–10
47. Hillis DM. 1997. Biology recapitulates phylogeny. *Science* 276:218–19
48. Hillis DM, Bull JJ, White ME, Badgett MR, Molineux IJ. 1992. Experimental phylogenetics: generation of a known phylogeny. *Science* 255:589–91
49. Huelsenbeck JP. 1995. Performance of phylogenetic methods in simulation. *Syst. Biol.* 44:17–48
50. Huelsenbeck JP. 1995. The robustness of two phylogenetic methods: four-taxon simulations reveal a slight superiority of maximum likelihood over neighbor joining. *Mol. Biol. Evol.* 12:843–49
51. Huelsenbeck JP, Bull JJ. 1996. A likelihood ratio test to detect conflicting phylogenetic signal. *Syst. Biol.* 45:92–98
52. Huelsenbeck JP, Cunningham CW, Graybeal A. 1997. The performance of phylogenetic methods for a well supported phylogeny. *Syst. Biol.* In press
53. Huelsenbeck JP, Hillis DM, Nielsen R. 1996. A likelihood-ratio test of monophyly. *Syst. Biol.* 45:546–58
54. Huelsenbeck JP, Hillis DM. 1993. Success of phylogenetic methods in the four-taxon case. *Syst. Biol.* 42:247–64
55. Huelsenbeck JP, Nielsen R. 1997. The effect of non-independent substitution on phylogenetic accuracy. *Syst. Biol.* Submitted
56. Huelsenbeck JP, Rannala B, Yang Z. 1997. Statistical tests of host parasite

cospeciation. *Evolution* 51:410–19
57. Jin L, Nei M. 1990. Limitations of the evolutionary parsimony method of phylogenetic inference. *Mol. Biol. Evol.* 7:82–102
58. Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In *Mammalian Protein Metabolism,* ed. HM Munro, pp. 21–132. New York: Academic
59. Kendall M, Stuart A. 1979. *The Advanced Theory of Statistics,* 2:240–80. London: Charles Griffin
60. Kimura M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16:111–20
61. Kimura M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl. Acad. Sci. USA* 78:454–58
62. Kuhner MK, Felsenstein J. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.* 11:459–68
63. Kuhner MK, Yamato J, Felsenstein J. 1995. Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics* 140:1421–30
64. Lanave C, Preparata G, Saccone C, Serio G. 1984. A new method for calculating evolutionary substitution rates. *J. Mol. Evol.* 20:86–93
65. Langley CH, Fitch WM. 1974. An examination of the constancy of the rate of molecular evolution. *J. Mol. Evol.* 3:161–77
66. Learn GH, Shore JS, Furnier GR, Zurawski G, Clegg MT. 1992. Constraints on the evolution of plastid introns: the group II intron in the gene encoding tRNA-Val(UAC). *Mol. Biol. Evol.* 9:856–71
67. Lewis PO, Huelsenbeck JP, Swofford DL. 1996. Maximum likelihood. In *PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods,) Version 4.0.* Sunderland, MA: Sinauer
68. Lewontin RC. 1989. Inferring the number of evolutionary events from DNA coding sequence differences. *Mol. Biol. Evol.* 6:15–32
69. Li W-H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.* 36:96–99
70. Li W-H, Wu C-I, Luo C-C. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* 2:150–74
71. Lindgren BW. 1976. *Statistical Theory.* New York: Macmillan. 3rd ed.
72. Messier W, Stewart C-B. 1997. Episodic adaptive evolution of primate lysozymes. *Nature* 385:151–54
73. Milinkovitch MC, Orti G, Meyer A. 1993. Revised phylogeny of whales suggested by mitochondrial DNA. *Nature* 361:346
74. Mindell DP, Thacker CE. 1996. Rates of molecular evolution: phylogenetic issues and applications. *Annu. Rev. Ecol. Syst.* 27:279–303
75. Miyata T, Yasunaga T. 1980. Molecular evolution of mRNA: a method for estimating rates of synonymous and amino acid substitution from homologous sequences and its application. *J. Mol. Evol.* 16:23–26
76. Muse SV. 1995. Evolutionary analyses when nucleotides do not evolve independently. In *Current Topics on Molecular Evolution,* ed. M Nei, N Takahata, pp. 115–24. University Park, PA: Penn. State Univ., Inst. Mol. Evol. Genet.
76a. Muse SV. 1995. Evolutionary analysis of DNA sequences subject to constraints on secondary structure. *Genetics* 139:1429–39
77. Muse SV. 1996. Estimating synonymous and nonsynonymous substitution rates. *Mol. Biol. Evol.* 13:105–14
78. Muse SV, Gaut BS. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates with application to the chloroplast genome. *Mol. Biol. Evol.* 11:715–24
79. Muse SV, Weir BS. 1992. Testing for equality of evolutionary rates. *Genetics* 132:269–76
80. Navidi WC, Churchill GA, von Haeseler A. 1991. Methods for inferring phylogenies from nucleic acid sequence data by using maximum likelihood and linear invariants. *Mol. Biol. Evol.* 8:128–43
81. Nei M. 1987. *Molecular Evolutionary Genetics.* New York: Columbia Univ. Press
82. Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3:418–26
83. Neyman J. 1971. Molecular studies of evolution: a source of novel staistical problems. In *Statistical Decision Theory and Related Topics,* ed. SS Gupta, J Yackel, pp. 1–27. New York: Academic
84. Nielsen R. 1997. A likelihood approach to populations samples of microsatellite alleles. *Genetics* 146:711–16
85. Olsen GJ. 1987. Earliest phylogenetic branchings: comparing rRNA-based evolutionary trees inferred with various

techniques. *Cold Spring Harbor Symp. Quant. Biol.* 52:825–37
86. Ou C-Y, Ciesielski CA, Myers G, Bandea CI, Luo C-C, et al. 1992. Molecular epidemiology of HIV transmission in a dental practice. *Science* 256:1165–71
87. Pamilo P, Bianchi NO. 1993. Evolution of the *Zfx* and *Zfy* genes: rates and interdependence between the genes. *Mol. Biol. Evol.* 10:271–81
88. Penny D, Hendy MD, Steel MA. 1992. Progress with methods for constructing evolutionary trees. *Trends Ecol. Evol.* 7:73–79
89. Perler R, Efstratiadis A, Lomedico P, Gilbert W, Klodner R, Dodgson J. 1980. The evolution of genes: the chicken preproinsulin gene. *Cell* 20:555–66
90. Pettigrew JD. 1986. Flying primates? Megabats have the advanced pathway from eye to midbrain. *Science* 231:1304–6
91. Reeves JH. 1992. Heterogeneity in the substitution process of amino acid sites of proteins coded for by mitochondrial DNA. *J. Mol. Evol.* 35:17–31
92. Rice JA. 1995. *Mathematical Statistics and Data Analysis.* Belmont, CA: Duxbury
93. Ritland K, Clegg MT. 1987. Evolutionary analysis of plant DNA sequences. *Am. Nat.* 130:S74–100
94. Rodriguez F, Oliver JF, Marin A, Medina JR. 1990. The general stochastic model of nucleotide substitutions. *J. Theor. Biol.* 142:485–501
95. Rzhetsky A. 1995. Estimating substitution rates in ribosomal RNA genes. *Genetics* 141:771–83
96. Rzhetsky A, Nei M. 1995. Tests of applicability of several substitution models for DNA sequence data. *Mol. Biol. Evol.* 12:131–51
97. Saitou N, Imanishi T. 1989. Relative efficiencies of the Fitch-Margoliash, maximum-parsimony, maximum-likelihood, minimum-evolution, and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. *Mol. Biol. Evol.* 6:514–25
98. Schöniger M, von Haeseler A. 1994. A stochastic model for the evolution of autocorrelated DNA sequences. *Mol. Phylogeny Evol.* 3:240–47
99. Schöniger M, von Haeseler A. 1995. Performance of the maximum likelihood, neighbor joining, and maximum parsimony methods when sequence sites are not independent. *Syst. Biol.* 44:533–47
100. Sharp PM. 1997. In search of molecular darwinism. *Nature* 385:111–12
101. Silvey SD. 1975. *Statistical Inference.* London: Chapman & Hall
102. Simberloff D. 1987. Calculating probabilities that cladograms match: a method of biogeographic inference. *Syst. Zool.* 36:175–95
103. Stanhope MJ, Czelusniak J, Si J-S, Nickerson J, Goodman M. 1992. A molecular perspective on mammalian evolution from the gene encoding Interphotoreceptor Retinoid Binding Protein, with convincing evidence for bat monophyly. *Mol. Phylogeny Evol.* 1:148–60
104. Swofford DL. 1996. *PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods), Version 4.0.* Sunderland, MA: Sinauer Assoc.
105. 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. 2nd ed.
106. Swofford DL, Thorne JS, Felsenstein J, Wiegmann BM. 1996. The topology-dependent permutation test for monophyly does not test for monophyly. *Syst. Biol.* 45:575–79
107. Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10:512–26
108. Tateno Y, Takezaki N, Nei M. 1994. Relative efficiencies of the maximum-likelihood, neighbor-joining, and maximum-parsimony methods when substitution rate varies with site. *Mol. Biol. Evol.* 11:261–77
109. Tavaré S. 1986. Some probabilistic and statistical aspects of the primary structure of nucleotide sequences. In *Lectures on Mathematics in the Life Sciences,* ed. RM Miura, pp. 57–86. Providence, RI: Am. Math. Soc.
110. Thorne JL, Goldman N, Jones DT. 1996. Combining protein evolution and secondary structure. *Mol. Biol. Evol.* 13:666–73
111. Thorne JL, Kishino H, Felsenstein J. 1991. An evolutionary model for maximum likelihood alignment of DNA sequences. *J. Mol. Evol.* 33:114–24
112. Thorne JL, Kishino H, Felsenstein J. 1992. Inching toward reality: an improved likelihood model of sequence evolution. *J. Mol. Evol.* 34:3–16
113. Van Den Bussche RA, Baker RJ, Huelsenbeck JP, Hillis DM. 1997. Base compositional bias and phylogenetic analysis: a test of the "flying DNA" hypothesis. *Mol. Phyl. Evol.* Submitted

114. Waddell PJ, Penny D. 1996. Extending hadamard conjugations to model sequence evolution with variable rates across sites. Available by anonymous ftp from onyx.si.edu.
115. Weir BS. 1990. *Genetic Data Analysis.* Sunderland, MA: Sinauer
116. Yang Z. 1993. Maximum likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol. Biol. Evol.* 10:1396–401
117. Yang Z. 1994. Statistical properties of the maximum likelihood method of phylogenetic estimation and comparison with distance methods. *Syst. Biol.* 43:329–42
118. Yang Z. 1994. Estimating the pattern of nucleotide substitution. *J. Mol. Evol.* 39:105–11
119. Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* 39:306–14
120. Yang Z. 1995. A space-time process model for the evolution of DNA sequences. *Genetics* 139:993–1005
121. Yang Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol. Evol.* 11:367–72
122. Yang Z. 1996. Maximum likelihood models for combined analyses of multiple sequence data. *J. Mol. Evol.* 42:587–96
123. Yang Z, Goldman N, Friday AE. 1994. Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. *Mol. Biol. Evol.* 11:316–24
124. Yang Z, Goldman N, Friday AE. 1995. Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. *Syst. Biol.* 44:384–99
125. Yang Z, Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. *Mol. Biol. Evol.* 14:In press
126. Zharkikh A. 1994. Estimation of evolutionary distances between nucleotide sequences. *J. Mol. Evol.* 39:315–29