A LIKELIHOOD-RATIO TEST OF MONOPHYLY

JOHN P. HUELSENBECK,[1] DAVID M. HILLIS,[2] AND RASMUS NIELSEN[1]

[1]*Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA;*
E-mail: johnh@mws4.biol.berkeley.edu (J.P.H.), rasmus@mws4.biol.berkeley.edu (R.N.)
[2]*Department of Zoology, University of Texas, Austin, Texas 78712, USA; E-mail: hillis@bull.zo.utexas.edu*

Abstract.—Many phylogenetic analyses are inspired by or depend upon the monophyly of a group specified a priori. Also, many evolutionary problems for which phylogenies are useful do not require every detail of the phylogeny to be estimated correctly but depend upon the monophyly (or lack thereof) of a particular group. We propose a likelihood-ratio test that compares whether the best trees estimated with and without the constraint of monophyly are significantly different. Simulation suggests that the test is conservative when the null hypothesis (a particular specified constraint) is correct. We applied the likelihood-ratio test of monophyly to the question of the relationship of the presumed-extinct marsupial wolf (*Thylacinus*). Specifically, we examined the null hypotheses that (1) the marsupial wolf does not form a monophyletic group with dasyurids and (2) the marsupial wolf is a basal member of the Australian radiation of mammals. Both hypotheses were rejected using the likelihood-ratio test of monophyly. [Likelihood-ratio test; maximum likelihood; monophyly; phylogenetic methods; *Thylacinus.*]

Systematists are increasingly concerned not only with the accurate estimation of phylogenetic relationships but also with the testing of phylogenetic hypotheses. The method of maximum likelihood provides a powerful framework to address both aspects of a phylogenetic analysis. In terms of the estimation of relationship, maximum likelihood appears to be both efficient and robust (Kuhner and Felsenstein, 1994; Tateno et al., 1994; Huelsenbeck, 1995a, 1995b). However, likelihood also provides a natural way of testing hypotheses through the likelihood-ratio test (Edwards, 1972). The likelihood-ratio test has been used to examine (1) the adequacy of models of DNA substitution (Goldman, 1993), (2) the effect of additional parameters on the fit of a substitution model (Goldman, 1993; Yang et al., 1994; Yang, 1996), (3) the existence of a molecular clock (Felsenstein, 1981), and (4) whether trees estimated from different data partitions are heterogeneous (Huelsenbeck and Bull, 1996).

Commonly, systematists are interested in the monophyly of a group, and this question forms the inspiration for phylogenetic analysis. For example, Pettigrew's (1986, 1991) hypothesis that the "megabats" (e.g., flying foxes) are more closely related to primates than to "microbats" spurred a flurry of DNA sequence analyses with the primary question being, "Are bats monophyletic?" (Bennet et al., 1988; Adkins and Honeycutt, 1991; Mindell et al., 1991; Ammerman and Hillis, 1992; Bailey et al., 1992; Stanhope et al., 1992). Similarly, Grauer et al.'s (1991, 1992) hypothesis that the myomorph rodents (e.g., the rat-like rodents) are more closely related to primates than to the caviomorph rodents (e.g., the guinea pig) motivated several papers examining the monophyly of rodents (Hasegawa et al., 1992; Ma et al., 1993; Cao et al., 1994). In this paper, we propose a likelihood-ratio test to examine the monophyly of a group specified a priori. This test can be extended to more complicated null hypotheses in which the constraint on topology is not simply the monophyly or nonmonophyly of a group. We examined the statistical properties of this method using simulation and provide an example of the method by testing several hypotheses of relationship for the marsupial wolf.

LIKELIHOOD-RATIO TEST OF MONOPHYLY

The method of maximum likelihood depends on the complete specification of a stochastic model of evolution. We will assume aligned DNA sequences for the data. As an example, consider the following $s = 4$ DNA sequences:

Species 1   ACCAGT

Species 2   ACCAGC

Species 3   AGCAGC

Species 4   AGCAGG

Each of the $n = 6$ site patterns represents a different datum in a maximum likelihood analysis. In this case, the observations are $x_1 = (A, A, A, A)^T$, $x_2 = (C, C, G, G)^T$, $x_3 = (C, C, C, C)^T$, $x_4 = (A, A, A, A)^T$, $x_5 = (G, G, G, G)^T$, and $x_6 = (T, C, C, G)^T$, where T is the transpose of the vector.

Assuming independence among sites, the likelihood function is simply the joint probability of observing the data

$$L(\Theta|x_1, \ldots, x_n) = \prod_{i=1}^{n} \Pr[x_i|\Theta],$$

where $\Pr[x_i|\Theta]$ is the probability of observing site pattern i. The vector Θ denotes the parameters of the model of DNA substitution. We assume that the same model of DNA substitution applies to all sites, although this constraint can be relaxed. The probability of observing a given site pattern depends on the topology of the tree, the lengths of the branches of the tree (in terms of expected number of substitutions per site), and the substitution model. We assume a Markov chain model for the process of DNA substitution, as is commonly used (Felsenstein, 1981).

The unknown parameters of the phylogenetic model, Θ, are estimated by maximizing the likelihood function to obtain $\hat{\Theta}$. The topology of the tree is considered a parameter, so the best estimate of phylogeny is simply that tree for which the likelihood is maximized. In principle, the likelihood is maximized over all possible trees (i.e., every possible bifurcating tree is considered and the likelihood calculated), although in practice heuristic search strategies are used instead when the number of sequences is large.

The ratio of the likelihood of the same data under two different hypotheses provides a measure of the support of one hypothesis over the other. For example, if $L(H_1|x_1, \ldots, x_n)/L(H_2|x_1, \ldots, x_i) > 1$, then

H_1 is better supported than H_2. The likelihood ratio provides a framework for the statistical testing of hypotheses.

Our likelihood-ratio test of monophyly compares the likelihood calculated under the search constraint that the group of interest is monophyletic with the likelihood calculated when this search constraint is relaxed. The null hypothesis H_0 is that the group in question is monophyletic. The likelihood is maximized under the null hypothesis to obtain L_0. The alternative hypothesis H_1 is more general in that it makes no assumptions about the monophyly of the group in question. In other words, the best tree that fits the alternative hypothesis is the optimal tree. The likelihood is maximized under the unconstrained hypothesis to obtain L_1. The likelihood-ratio test statistic is

$$\delta = 2(\ln L_1 - \ln L_0).$$

Usually, when the null hypothesis is a subset of the alternative hypothesis, the likelihood ratio can be tested against a χ^2 distribution with $p - q$ degrees of freedom, where p is the number of parameters under the alternative hypothesis and q is the number of parameters under the null hypothesis (Rice, 1995). However, for the phylogeny problem, there are several potential problems with using the χ^2 approximation. First, the number of observations for many of the 4^s possible site patterns is very small or zero (Goldman, 1993). Hence, the general rule of thumb that each cell should contain five observations (or that the number of observations is four or five times the number of site patterns) is violated (McCullagh and Nelder, 1989). However, the χ^2 approximation is known to be robust for comparing two parametric models in the case of sparse data (Haberman, 1977; Agresti and Yang, 1987; Yang, 1996). Another more serious problem is that topology is not a standard statistical parameter (Goldman, 1993). Hence, the usual results from statistics do not necessarily hold (e.g., the χ^2 approximation for nested hypotheses). For the present application, it is not clear how many parameters are represented by topology. In other
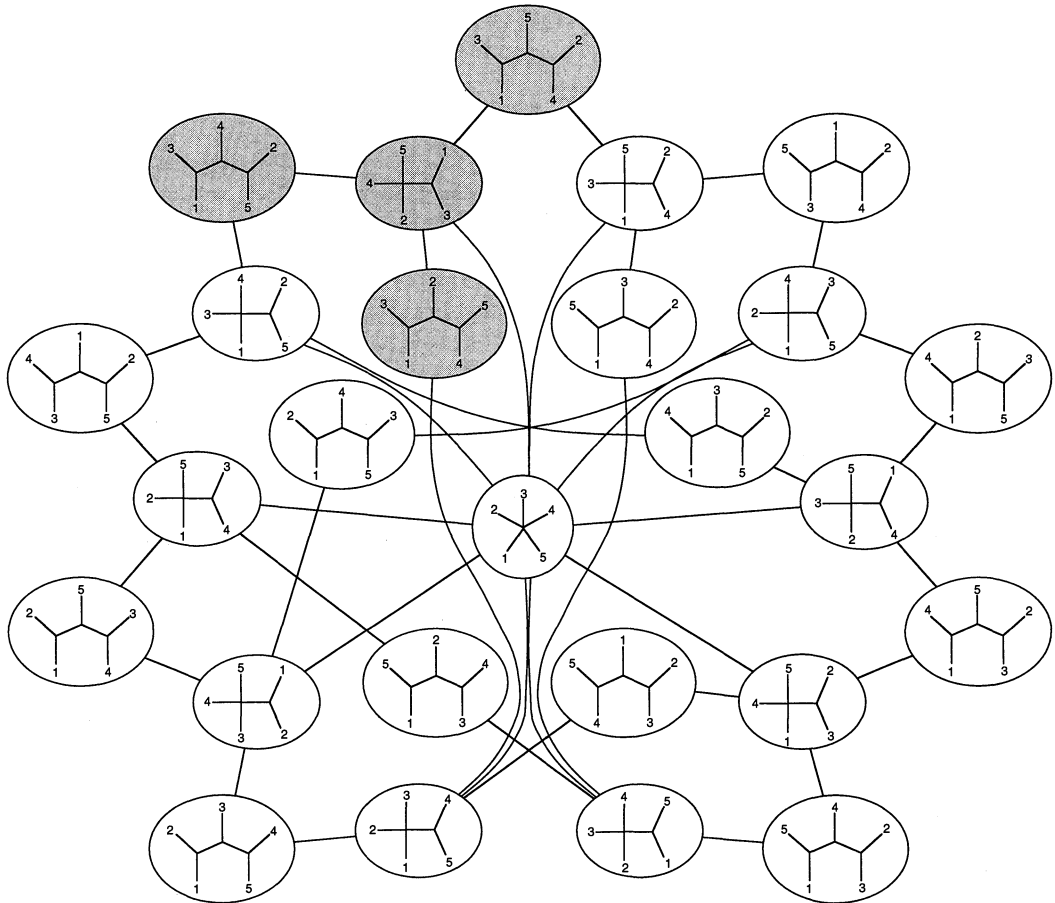
FIGURE 1. The space of all possible unrooted trees for five taxa. The vertices of the graph are represented by the trees. The trees are connected by lines (edges). Trees are adjacent if they are only one perturbation away from each other. In this case, the perturbation is either a contraction or an expansion of an internal branch of the tree. Shaded portions of the tree space represent trees that contain an internal branch separating taxa 1 and 3 from the remaining taxa.

words, the difference in the number of parameters between the general and null hypotheses is unknown. To avoid this problem, we resort to simulation of the null distribution of δ. In the absence of suitable asymptotic results appropriate for all possible values under the null hypothesis, the maximum likelihood values are instead used in the simulations. In this case, the maximum likelihood topology, maximum likelihood estimates of branch lengths, and other parameters (such as the transition: transversion rate ratio) under the null hypothesis are used to simulate replicate data sets of the same size as the original. This

simulation procedure for generating the null distribution is widely used in statistics and is known as parametric bootstrapping (Efron, 1985; Felsenstein, 1988; Goldman, 1993; see also Gouy and Li, 1989; Bull et al., 1993; Huelsenbeck et al., 1996). If the likelihood-ratio statistic δ calculated from the original data is greater than the 95% confidence interval determined through simulation, the null hypothesis that the group in question is monophyletic is rejected.

A representation of the space of all possible trees gives a visual interpretation of our test. Figure 1 shows the space of un-

rooted trees for five taxa. The trees are the vertices (nodes) of a graph and the lines connecting the trees (edges) show the adjacencies of the trees. Trees that are adjacent are only one perturbation away from each other (Charleston, 1995). In Figure 1, the perturbation used is the contraction or expansion of internal branches (Robinson and Foulds, 1981). The tree space would look different for other perturbations (Charleston, 1995). Those trees that are shaded in Figure 1 represent the null hypothesis that taxa 1 + 3 and taxa 2 + 4 + 5 form a bipartition (statements about monophyly of either group require rooted trees, but these shaded trees support the monophyly of taxa 1 + 3 as long as one of the other taxa is the outgroup). The likelihood is maximized over the shaded portion of the tree space to obtain L_0. The whole tree space represents a more general hypothesis. The likelihood is also maximized under the more general hypothesis to obtain L_1. Obviously, the null hypothesis is a subset of the general hypothesis.

This test also can be extended to examine the nonmonophyly of a group. In this case, the null hypothesis, as described above, is reversed and becomes all those trees for which the group of interest is not monophyletic. In terms of the tree space of Figure 1, the null hypothesis could represent all of those trees that are not shaded (i.e., those trees for which 1 + 3 are not monophyletic, again assuming that the outgroup is among taxa 2, 4, and 5). The alternative hypothesis would again be the entire space of trees. The likelihood-ratio test statistic would be calculated as described above, and the distribution of the test statistic δ would be simulated under the null hypothesis.

SIMULATION ANALYSIS

We performed a simulation analysis to examine the behavior of the likelihood-ratio test of monophyly. Specifically, we examined the probability of rejecting the null hypothesis and the behavior of parametric bootstrapping for determining the null distribution of δ. The behavior of the test was examined when the assumptions of
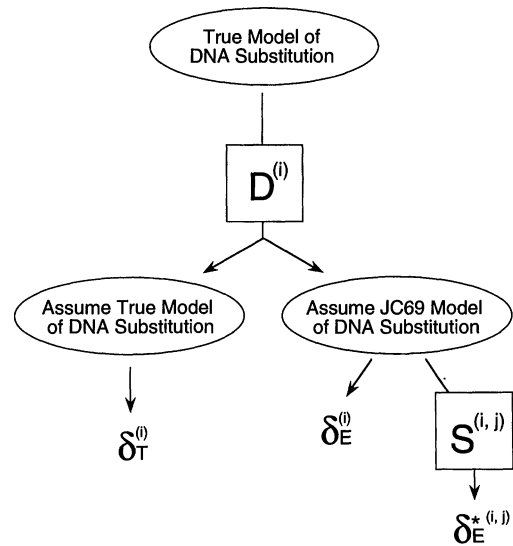

FIGURE 2. The general strategy used to examine the statistical properties of the likelihood-ratio test of monophyly. Data were generated using the "true model of DNA substitution" and then analyzed using maximum likelihood assuming the correct and (possibly) incorrect model of DNA substitution to produce $\delta_T^{(i)}$ and $\delta_E^{(i)}$. Parameters estimated using maximum likelihood under the null hypothesis were used to generate an additional simulated data set ($S^{(i,j)}$) to obtain $\delta_E^{*(i,j)}$. JC69 = Jukes and Cantor (1969).

the test were completely satisfied and also when one of the assumptions was violated. Figure 2 shows the general strategy taken to examine the behavior of the test. We assumed a model of DNA substitution (True Model of DNA Substitution, Fig. 2) and a four- or eight-taxon model tree with specified branch lengths. The Jukes–Cantor (1969) or Kimura (1980) models of DNA substitution were assumed for the true model. Simulation was used to construct n simulated data sets, $D^{(1)}, \ldots, D^{(n)}$. The likelihood-ratio test statistic (twice the difference between the log likelihoods calculated under the null and general hypotheses) for each data set $D^{(i)}$ was then calculated assuming the correct model of DNA substitution (e.g., if the Kimura model was used to generate the data sets $D^{(1)}, \ldots, D^{(n)}$, then the Kimura model was also used in the analysis). This likelihood ratio is denoted $\delta_T^{(i)}$. We also calculated the likelihood ratio statistic assuming the Jukes–Cantor
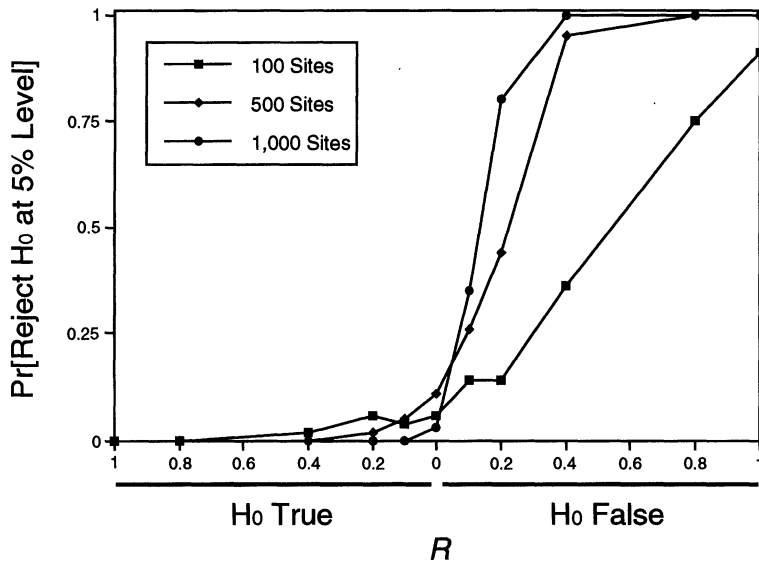
FIGURE 3.   The power of the likelihood-ratio test for 100, 500, and 1,000 simulated nucleotide sites. The rate of change over the entire tree for these simulations was constrained to be 2.0 substitutions per site. R is the ratio of the length of the internal branch to that of one of the four external branches (all external branches are equal in length). As R increases, the internal branch becomes longer.

model, which may or may not represent the model of DNA substitution that was used in generating the simulated data. This likelihood ratio is denoted $\delta_E^{(i)}$. Finally, the tree and branch lengths calculated assuming the Jukes–Cantor model were used to produce m simulated data sets [$S^{(i,1)}, \ldots, S^{(i,m)}$]. This part of the analysis mimics the procedure of parametric bootstrapping used to determine if the null hypothesis of monophyly will be rejected for a particular realization of $D^{(i)}$. For each data set, $S^{(i,j)}$ simulated at this point in the analysis, we calculated the likelihood-ratio test statistic and denote this value as $\delta_E^{*(i,j)}$.

The values $\Delta_T = \delta_T^{(1)}, \ldots, \delta_T^{(n)}$ represent an approximation to the distribution of the likelihood-ratio test statistic if the true model of DNA substitution is assumed in the analysis. Similarly, the values $\Delta_E = \delta_E^{(1)}, \ldots, \delta_E^{(n)}$ are an approximation of the distribution of δ under the Jukes–Cantor model of DNA substitution. If the Jukes–Cantor model of DNA substitution happens to match the process generating the data sets $D^{(1)}, \ldots, D^{(n)}$, then $\Delta_T = \Delta_E$. Finally, the values $\Delta_E^* = \delta_E^{*(1)}, \ldots, \delta_E^{*(n)}$ should

approximate the distribution of Δ_E (and Δ_T as well, if the model of substitution is correct). The match between Δ_E and Δ_E^* was of prime interest in this study because the match or mismatch of these distributions provides an indication of how well the test will perform (i.e., whether the test will reject the null hypothesis more or less often than it should).

The likelihood-ratio test of monophyly appears to be powerful. Figure 3 shows the results of a power analysis in which one of two unrooted four-taxon trees was simulated: tree 1 = ((A, B)C, D) or tree 2 = ((A, C)B, D). Sequences of 100, 500, and 1,000 sites were simulated. For both trees, the lengths of the external branches (those branches leading directly to a tip) were equal, and the expected number of substitutions per site over the entire tree was 2.0. The length of the internal branch, however, differed. Figure 3 shows the ratio of the length of the internal branch to that of the external branch (R) along the x-axis and the probability of rejecting the null hypothesis at the 5% level along the y-axis. When $R = 0$, the internal branch is zero in
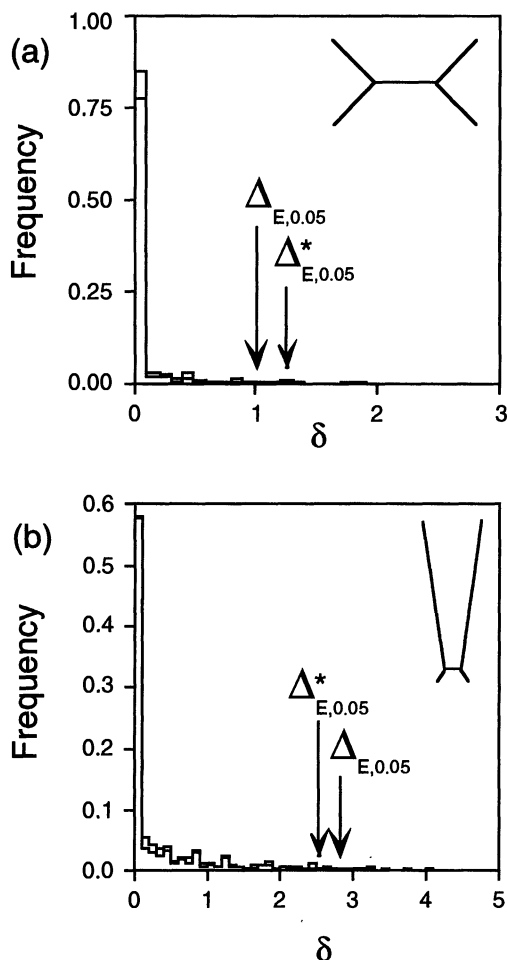
FIGURE 4. The distributions of Δ_E and Δ_E^* for two model trees. The total number of substitutions over the entire tree was constrainted to be 4.0 for both model trees. Using the parametric bootstrap procedure one can closely approximate Δ_E when the assumptions of the method are satisfied. (a) A model tree in which all branches are equal in length. (b) A model tree in which two of the branches are 10 times longer than the other branches on the tree.
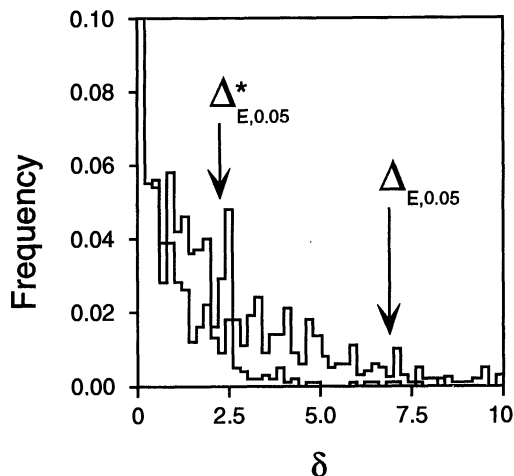

FIGURE 5. The distributions of Δ_E and Δ_E^* when the null hypothesis is false. A four-taxon model tree was assumed with all branches equal in length. The expected number of substitutions per site over the entire tree was set to 4.0. When the null hypothesis is false, the distributions Δ_E and Δ_E^* no longer match. The null hypothesis is rejected with high frequency for the simulations depicted here.

length, and when $R = 1$, the internal branch is equal in length to any one of the four external branches. For this simulation, the null hypothesis was taken to be the constraint that taxa A and B are a group (or the taxon bipartition {A, B}{C, D}). Hence, when tree 1 was simulated, the null hypothesis was correct, and when tree 2 was simulated, the null hypothesis was false. When $R = 0$, the model tree repre-

sents the star phylogeny. Figure 3 shows that the null hypothesis is hardly ever rejected when it is, in fact, true. Similarly, the null hypothesis is rejected with high frequency when false. As expected, the probability that the null hypothesis is rejected depends both on the number of sites in the simulated sequence as well as on the length of the internal branch of the tree.

Parametric bootstrapping closely approximates the distribution of Δ_E and Δ_T when the model of DNA substitution matches the processes generating the sequences. Figure 4 shows the distributions of Δ_E and Δ_E^* for two model trees. In both cases, 100 sites were simulated, with the expected number of substitutions per site over the entire tree constrained to be 4.0. However, for the simulation of Figure 4a, all branches were equal in length, whereas for the simulation of Figure 4b, two of the branches were 10 times longer than the other branches on the tree. For the simulations depicted in Figure 4, the distributions of Δ_E and Δ_E^* closely match. In fact, the 5% critical values for both distributions ($\Delta_{E,0.05}$ and $\Delta_{E,0.05}^*$) are very close. Figure 5 shows the distributions for Δ_E and Δ_E^*

when the null hypothesis is false. In these simulations, all branches were equal in length, and the expected number of substitutions over the entire tree was 4.0. In the simulations of Figure 5, one would not expect the distributions to match because the null hypothesis is false. In fact, not only are the distributions different, but the 5% critical values show that the null hypothesis should, on average, be rejected ($\Delta_{E,0.05}^*$ is much smaller than $\Delta_{E,0.05}$).

The parametric bootstrap procedure does not perform as well for approximating the distribution of Δ_E when the model of DNA substitution assumed in the analysis does not match the processes generating the sequences. Figure 6 shows the distributions of Δ_E and Δ_E^* for two model trees. The conditions of the simulation of Figure 6 were the same as for the simulations of Figure 4 except that transitions had an instantaneous rate of change 10 times that of transversions for the true model of DNA substitution ($\kappa = \alpha/\beta = 10.0$). Maximum likelihood assuming a Jukes–Cantor model of DNA substitution is consistent (will converge to the correct phylogeny given enough sequence data) for the tree of Figure 6a. However, maximum likelihood assuming a Jukes–Cantor model is inconsistent (converges to an incorrect phylogeny) for the tree of Figure 6b. Hence, the simulation results depicted in Figure 6b represent an exceptionally difficult problem for parametric bootstrapping. Not only are the distributions of Δ_E and Δ_E^* different for both simulations, but the expected critical values show that the likelihood-ratio test should reject the null hypothesis more readily than desired ($\Delta_{E,0.05}$ is greater than $\Delta_{E,0.05}^*$). This result suggests that the likelihood-ratio test should be performed with as realistic (parameter rich) models of DNA substitution as possible, otherwise one runs the risk of a high level of Type I error.

Although parametric bootstrapping appears to perform poorly when a wrong substitution model is used for data analysis, the simulations shown in Figure 6 represent an exceptionally difficult problem for the method because the rates of evolu-
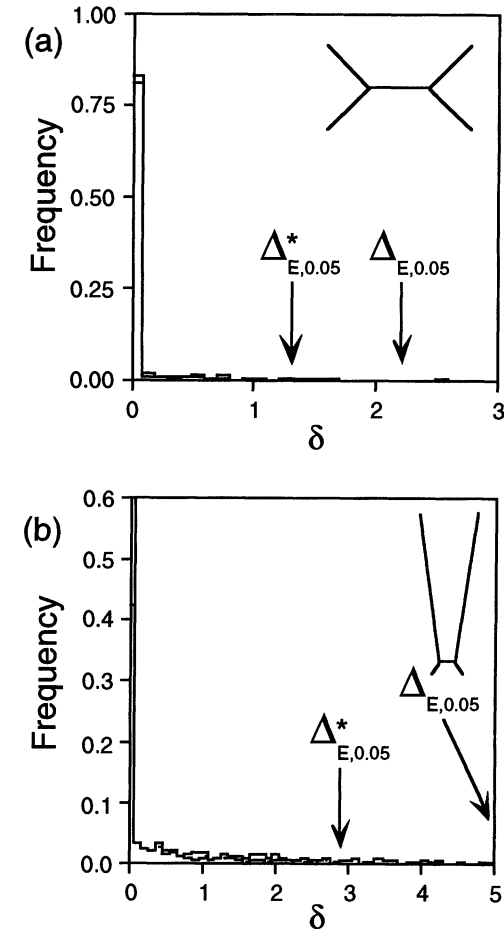

FIGURE 6. The distributions of Δ_E and Δ_E^* when the model of DNA substitution assumed by maximum likelihood is wrong. The true model of substitution followed a Kimura model in which transitions occur at a rate 10 times that of transversions. The expected number of substitutions over the entire tree was constrained to be 4.0 for both model trees. When the model of DNA substitution assumed by maximum likelihood no longer matches the processes of evolution generating the sequence data, parametric bootstrapping performs poorly. The distributions Δ_E and Δ_E^* no longer match. (a) A model tree in which all branches are equal in length. (b) A model tree in which two of the branches are 10 times the length of the remaining three.

tion are very high and, for one case, the method is inconsistent. One would expect the estimates of branch lengths obtained by maximum likelihood to be very poor when rates of evolution are high and the

model is inaccurate. However, when rates of evolution are low, the distribution Δ_E^* obtained by parametric bootstrapping can closely approximate Δ_E even if the assumptions of the method are violated. We also performed simulations in which overall rates of evolution were low (1.0 substitution per site over the entire tree). All other aspects of the simulation were the same as the simulations of Figure 6. In this case, the match between Δ_E and Δ_E^* was close for the simulations in which all branches were equal in length ($\Delta_{E,0.05} = 0.34$ and $\Delta_{E,0.05}^* = 0.82$). For the case in which two of the branches were 10 times longer than the remaining branches, however, the match between Δ_E and Δ_E^* for the low-rate simulations (1.0 substitutions per site) was better than that for the high-rate simulations (4.0 substitutions per site, Fig. 4). However, the critical values show that the null hypothesis would still be rejected more readily than desired ($\Delta_{E,0.05} = 6.37$ and $\Delta_{E,0.05}^* = 2.05$). Phylogenetic problems in which some of the branches are very long represent a difficult estimation problem for most methods of phylogenetic estimation as well as for this test.

We also examined the behavior of the test for a more complicated tree, one with eight taxa. In the case of four taxa, the correct tree under the null hypothesis was always used for simulating the distribution Δ_E^* because only one tree satisfies the null hypothesis (the constraint that taxa A and B form a group). However, in the eight-taxon simulations, there are multiple trees that satisfy the null hypothesis, only one of which can be the true tree. It is possible, then, to simulate the null distribution with an incorrect tree that nonetheless satisfies the null hypothesis. How is parametric bootstrapping affected by inaccurate estimation of topology? The eight-taxon simulations can address this question because often an incorrect tree is used for simulation under the null hypothesis. In these simulations, 1,000 data sets were generated assuming a Jukes–Cantor model of evolution and the tree depicted in Figure 7. All of the branch lengths on the model tree were assumed to be equal, but the total
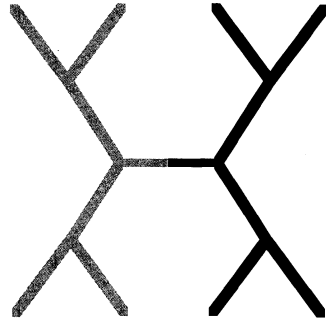

FIGURE 7. The tree assumed in the eight-taxon simulations. All internal and external branches are of equal length. The null hypothesis tested in the simulations constitutes the constraint that the taxon bipartition between the gray and black region of the tree exists.

amount of change on the tree differed. A total of 100 sites were simulated. For each simulated data set, the test statistic δ_E was calculated under the null hypothesis that the taxa in the shaded portion of the tree constitute a monophyletic group. The distribution Δ_E represents the distribution of the test statistic under the null hypothesis. For each of the 1,000 simulated trees, an additional simulation was performed using the most likely tree obtained under the null hypothesis in the previous simulation. Again, δ was calculated for the 1,000 new simulated data sets. The distribution of these δ_E^* values corresponds to the estimated distribution Δ_E^* as obtained by the parametric bootstrap.

The results of the simulations are shown in Figure 8. For both the high-rate (5.0 substitutions per site over the tree) and the low-rate (3.0 substitutions per site over the tree) cases, the true distribution of the test statistic is very similar or identical to the estimated distribution of the test statistic. For the low-rate simulations, $\Delta_{E,0.05} = 0.0$, whereas $\Delta_{E,0.05}^*$ as determined using parametric bootstrapping is 0.67. Similarly, for high-rate simulations, $\Delta_{E,0.05} = 2.14$, whereas $\Delta_{E,0.05}^* = 2.72$. The parametric bootstrap procedure appears conservative with respect to rejection of the null hypothesis, which confirms that parametric bootstrapping in this case may be an appropriate method to obtain the null distribution of
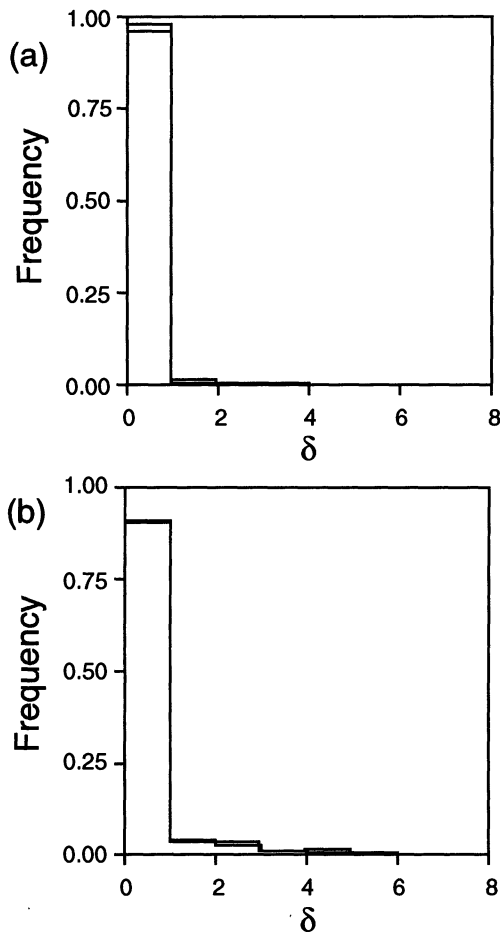
(a)

(b)

FIGURE 8. The distributions of Δ_E and Δ_E^* for the eight-taxon simulations. The parametric bootstrap procedure for generating the null distribution of δ performs well. The match between Δ_E and Δ_E^* is very close for both (a) the low-rate case in which there is a total of 3.0 substitutions per site over the entire tree and (b) the high-rate case in which there is a total of 5.0 substitutions per site.

δ_E, even for a moderately large number of taxa, when the model of DNA substitution is accurate.

These simulations show that the likelihood-ratio test of monophyly appears to be powerful and that the level of false significance (Type I error) is low when the assumptions of the analysis are satisfied. Like most tests, the likelihood-ratio test proposed here appears sensitive to violation of its assumptions. However, whether

violation of assumptions will cause spurious results should depend on the rate of evolution, the severity of the violation, and the difficulty of the phylogenetic problem.

AN APPLICATION TO MARSUPIAL RELATIONSHIPS

Thomas et al. (1989) examined the relationship of the presumed-extinct marsupial wolf (*Thylacinus cynocephalus*) using 12S ribosomal RNA (rRNA) data. The tree estimated using the parsimony criterion placed the marsupial wolf as the sister taxon to the Dasyuridae (*Sarcophilus harrisii* [Tasmanian devil] and *Dasyurus maculatus* [tiger cat]), supporting the Australian origin for the thylacine (Thomas et al., 1989). Some morphological evidence, however, supports a close relationship of the marsupial wolf with the South American borhyaenids (Archer, 1982). Faith (1991) used a randomization procedure to show that the degree of support for a *Thylacinus* + Dasyuridae clade was significant (as determined with the T-PTP test; Faith, 1991), even though he could not reject the null hypothesis of random signal for the data set as a whole (as determined with the PTP test; Faith, 1990).

We reexamined the marsupial wolf relationship using the likelihood-ratio test of monophyly. The 12S rRNA and cytochrome *b* sequences of Thomas et al. (1989) were used in this analysis. We examined two null hypotheses: $H_0' = $ *Thylacinus* + Dasyuridae are not monophyletic, and $H_0'' = $ *Thylacinus* is a basal member of the Australian radiation of mammals. The first hypothesis (H_0') is equivalent to the hypothesis tested by Faith (1991). However, this hypothesis cannot be used to examine whether the marsupial wolf is basal to the other Australian and New Guinean marsupials. To test the second hypothesis (H_0''), we calculated the likelihood under the constraint that *Bos*, *Philander* (opossum), and *Thylacinus* form a group (when the tree estimated under this constraint is rooted using the cow sequence, the resulting tree forces the marsupial wolf to a basal position).

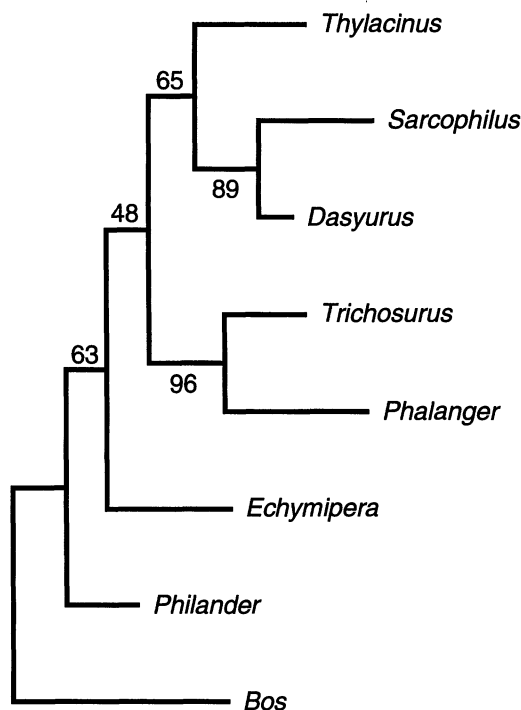Figure 9 shows the tree estimated using

FIGURE 9. The maximum likelihood estimate of phylogeny for the analysis of the 12S rRNA and cytochrome *b* sequences of cow (*Bos*), opossum (*Philander*), bandicoot (*Echymipera*), phalangers (*Tricosurus, Phalanger*), the marsupial wolf (*Thylacinus*), and the Dasyuridae (*Sarcophilus, Dasyurus*). The HKY85 model of DNA substitution was assumed in the analysis with the transition:transversion bias estimated using maximum likelihood. The best estimate of phylogeny had a log likelihood score of −1005.937. The numbers at the nodes represent nonparametric bootstrap values.
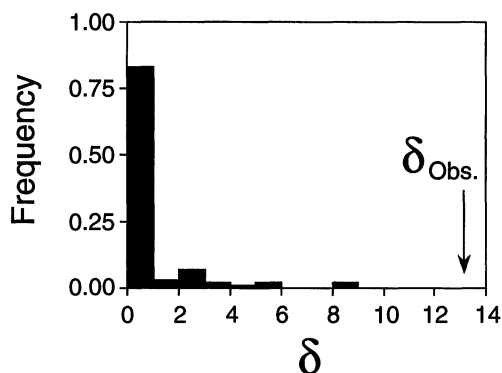

FIGURE 10. The simulated distribution of δ under the null hypothesis that the marsupial wolf is basal to the Australian and New Guinean marsupials (H_0''). The distribution is based on 100 replicates. The observed δ is much greater than the 5% critical value determined through parametric bootstrapping.

maximum likelihood implemented with the HKY85 model of DNA substitution (Hasegawa et al., 1985) when the 12S rRNA and cytochrome *b* sequences were analyzed (a total of 212 sites with gaps omitted). The HKY85 model allows for a different rate for transitions and transversions and for unequal base frequencies. This tree places *Thylacinus* as the sister taxon to the Dasyuridae, as did Thomas et al.'s (1989) original parsimony analysis of the 12S rRNA data. The log likelihoods calculated for the hypotheses examined here were −1005.937 for the unconstrained hypothesis, −1011.095 for the *Thylacinus* + Dasyuridae nonmonophyly hypothesis (H_0'), and −1012.519 for the basal marsu-

pial wolf hypothesis (H_0''). Likelihoods were calculated using the program PAUP* 4.0d52 (Swofford, 1996). Both null hypotheses are rejected using our likelihood-ratio test (H_0': δ = 10.315, P < 0.01; H_0'': δ = 13.163, P < 0.01; see Fig. 10). The test is able to reject the null hypotheses despite the small number of sites analyzed. Hence, useful hypotheses can be examined using the likelihood-ratio test even when some groups on the maximum likelihood tree have little support (as evidenced by nonparametric bootstrapping). The T-PTP test also rejects H_0' at P < 0.01 but could not reject H_0'' (P = 0.54). However, Swofford et al. (1996) showed that the null hypothesis of random covariation among characters used by the T-PTP test is inappropriate. We argue that the null hypothesis assumed by the likelihood-ratio test of monophyly, however, is appropriate. An appropriate null hypothesis consists of a tree (or a group of trees) as well as a model of DNA substitution.

DISCUSSION

The likelihood-ratio test of monophyly proposed here can be used to test a wide variety of hypotheses of relationship. Besides constructing hypotheses in which groups are constrained to be monophyletic or nonmonophyletic, more complex restric-

tions of the space of all trees can be envisioned. At this time, however, no publicly available phylogenetic program allows constraints more complex than monophyly or nonmonophyly of particular groups to be imposed on the search. Such general constraints could be used to examine questions such as the African origin of modern humans; the null hypothesis would be those trees that are consistent with an origination of modern humans in Africa. This hypothesis cannot be accommodated by a simple constraint of monophyly.

Numerous other tests are currently available to test one tree against another. Some of these tests do not make strong assumptions about the probability density distribution underlying the data (e.g., nonparametric tests such as Templeton's test; Templeton, 1983). Typically, however, nonparametric tests are not as powerful as parametric tests (Sokal and Rohlf, 1981). Kishino and Hasegawa (1989) proposed an alternative parametric test that compares the likelihood of two trees. Their test assumes that the variance of the likelihood of the two trees is asymptotically normally distributed. Faith (1991) proposed a permutation test that can be used with the parsimony criterion to examine the monophyly or nonmonophyly of a group. However, the distribution of the test statistic used by Faith (1991) is not related to the null hypothesis for the phylogeny problem (Swofford et al., 1996). A thorough study exploring the behavior of the available tests in terms of Type I and Type II errors under a variety of evolutionary scenarios might establish which methods generally have acceptable statistical properties.

The likelihood-ratio test proposed here should be implemented with as realistic a model of DNA substitution as possible, otherwise the null hypothesis may be rejected too easily. Currently, the most parameter-rich models of DNA substitution allow different rates of change from one nucleotide to another, different numbers of substitutions along each branch of the tree, and rate heterogeneity among sites (as determined by a gamma distribution or by allowing a certain proportion of the sites

to be invariant; Yang, 1993; Waddell and Penny, 1996). Also, the models of DNA substitution can be made to more closely fit the observed sequence data by estimating parameters separately for different data partitions (e.g., first, second, and third codon positions; Yang, 1996). Using more realistic models of DNA substitution may alleviate the problem of an elevated Type I error.

Phylogenetic analysis cannot proceed without making specific assumptions about the process of evolution. These assumptions often evoke skepticism about a method. For example, maximum likelihood and distance methods have often been criticized because they make explicit assumptions about the process of DNA substitution (Carpenter, 1994). However, evolutionary models should motivate an examination of their adequacy in a hypothesis-testing framework and a search for more realistic models rather than their complete dismissal as useless. This approach has led to a demonstrable improvement in the models of DNA substitution used in phylogenetic analysis over the past 5 years (Yang et al., 1994). Furthermore, by making assumptions explicit, our models can be improved and other hypotheses (such as the molecular clock) can be tested. Here, we have shown how a wide variety of hypotheses of phylogenetic relationship can be tested in a likelihood framework.

PROGRAM AVAILABILITY

The likelihood-ratio test of monophyly can be implemented if PAUP* 4.0 is used in conjunction with a program to simulate data under the appropriate model of DNA substitution. A program written by J.P.H. for simulating DNA sequence data for the HKY85+Γ model of DNA substitution (Hasegawa et al., 1985; Yang, 1993) is available from http://mw511.biol.berkeley.edu/homepage.html/.

REFERENCES

ADKINS, R. M., AND R. L. HONEYCUTT. 1991. Molecular phylogeny of the superorder Archonta. Proc. Natl. Acad. Sci. USA 88:10317–10321.

AGRESTI, A., AND M. C. YANG. 1987. An empirical investigation of some effects of sparseness in contingency tables. Comput. Stat. Data Anal. 5:9–21.

AMMERMAN, L. K., AND D. M. HILLIS. 1992. A molecular test of bat relationships: Monophyly or diphyly? Syst. Biol. 41:222–232.

ARCHER, M. 1982. A review of Miocene thylacinids (Thylacinidae, Marsupialia), the phylogenetic position of the Thylacinidae and the problem of apriorisms in character analysis. Pages 445–476 in Carnivorous marsupials (M. Archer, ed.). Royal Zoological Society of New South Wales, Sydney.

BAILEY, W. J., J. L. SLIGHTOM, AND M. GOODMAN. 1992. Rejection of the "flying primate" hypothesis by phylogenetic evidence from the e-globin gene. Science 256:86–89.

BENNET, S., L. J. ALEXANDER, R. H. CROZIER, AND A. G. MACKINLAY. 1988. Are megabats flying primates? Contrary evidence from a mitochondrial DNA sequence. Aust. J. Biol. Sci. 41:327–332.

BULL, J. J., C. W. CUNNINGHAM, I. J. MOLINEUX, M. R. BADGETT, AND D. M. HILLIS. 1993. Experimental molecular evolution of bacteriophage T7. Evolution 47:993–1007.

CAO, Y., J. ADACHI, T.-A. YANO, AND M. HASEGAWA. 1994. Phylogenetic place of guinea pigs: No support of the rodent-polyphyly hypothesis from maximum-likelihood analyses of multiple protein sequences. Mol. Biol. Evol. 11:593–604.

CARPENTER, J. 1994. Successive weighting, reliability and evidence. Cladistics 10:215–220.

CHARLESTON, M. 1995. Toward a characterisation of landscapes of combinatorial optimisation problems, with special reference to the phylogeny problem. Comput. Biol. 2:439–450.

EDWARDS, A. W. F. 1972. Likelihood. Cambridge Univ. Press, Cambridge, England.

EFRON, B. 1985. Bootstrap confidence intervals for a class of parametric problems. Biometrika 72:45–58.

FAITH, D. P. 1990. Chance marsupial relationships. Nature 345:393–394.

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

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Mol. Evol. 17:368–376.

FELSENSTEIN, J. 1988. Phylogenies from molecular sequences: Inference and reliability. Annu. Rev. Genet. 22:521–565.

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

GOUY, M., AND W.-H. LI. 1989. Phylogenetic analysis based on rRNA sequences supports the archaebacterial rather than the eocyte tree. Science 339:145–147.

GRAUER, D., W. A. HIDE, AND W.-H. LI. 1991. Is the guinea-pig a rodent? Nature 351:649–652.

GRAUER, D., W. A. HIDE, A. ZHARKIKH, AND W.-H. LI. 1992. The biochemical phylogeny of guinea-pigs and gundis, and the paraphyly of the order Rodentia. Comp. Biochem. Physiol. B 101:495–498.

HABERMAN, S. J. 1977. Log-linear models and frequency tables with small expected cell counts. Ann. Stat. 5:1148–1169.

HASEGAWA, M., Y. CAO, J. ADACHI, AND T. YANO. 1992. Rodent polyphyly? Nature 355:595.

HASEGAWA, M., H. KISHINO, AND T. YANO. 1985. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 21:160–174.

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

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

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

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

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

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

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

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

MA, D.-P., A. ZHARKIKH, D. GRAUER, J. L. VANDEBERG, AND W.-H. LI. 1993. Structure and evolution of opossum, guinea pig, and porcupine cytochrome b genes. J. Mol. Evol. 36:327–334.

MCCULLAGH, P., AND J. A. NELDER. 1989. Generalized linear models, 2nd edition. Chapman and Hall, London.

MINDELL, D. P., C. W. DICK, AND R. J. BAKER. 1991. Phylogenetic relationships among megabats, microbats and primates. Proc. Natl. Acad. Sci. USA 88:10322–10326.

PETTIGREW, J. D. 1986. Flying primates? Megabats have the advanced pathway from eye to midbrain. Science 231:1304–1306.

PETTIGREW, J. D. 1991. Wings or brain? Convergent

evolution in the origins of bats. Syst. Zool. 40:199–216.

RICE, J. A. 1995. Mathematical statistics and data analysis. Duxbury Press, Belmont, California.

ROBINSON, D. F., AND L. R. FOULDS. 1981. Comparison of phylogenetic trees. Math. Biosci. 53:131–147.

SOKAL, R. R., AND F. J. ROHLF. 1981. Biometry, 2nd edition. W. H. Freeman, New York.

STANHOPE, M. J., J. CZELUSNIAK, J.-S. SI, J. NICKERSON, AND M. GOODMAN. 1992. A molecular perspective on mammalian evolution from the gene encoding interphotoreceptor retinoid binding protein, with convincing evidence for bat monophyly. Mol. Phylogenet. Evol. 1:148–160.

SWOFFORD, D. L. 1996. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0. Sinauer, Sunderland, Massachusetts.

SWOFFORD, D. L., J. L. THORNE, J. FELSENSTEIN, AND B. M. WIEGMANN. 1996. The topology-dependent permutation test for monophyly does not test for monophyly. Syst. Biol. 45:575–579.

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

TEMPLETON, A. R. 1983. Phylogenetic inference from restriction site endonuclease cleavage site maps with particular reference to the humans and apes. Evolution 37:221–244.

THOMAS, R. H., W. SCHAFFNER, A. C. WILSON, AND S. PÄÄBO. 1989. DNA phylogeny of the extinct marsupial wolf. Nature 340:465–467.

WADDELL, P. J., AND D. PENNY. 1996. Evolutionary trees of apes and humans from DNA sequences. Pages 53–73 in Handbook of symbolic evolution (A. J. Lock and C. R. Peters, eds.). Oxford Univ. Press, Oxford, England

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

YANG, Z. 1996. Maximum likelihood models for combined analyses of multiple sequence data. J. Mol. Evol. 42:587–596.

YANG, Z., N. GOLDMAN, AND A. FRIDAY. 1994. Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. Mol. Biol. Evol. 11:316–324.