

Letter to the Editor

Rapid Evaluation of the Phylogenetic Congruence of Sequence Data Using Likelihood Ratio Tests

Peter J. Waddell,* Hirohisa Kishino,† and Rissa Ota‡

*Institute of Molecular BioSciences, Massey University, Palmerston North, New Zealand; †Department of Agriculture and Life Sciences, University of Tokyo, Tokyo, Japan; and ‡Graduate University for Advanced Studies, Tokyo, Japan

Evaluating Congruence Among Gene Trees

Whenever different types of data are used in a scientific analysis, a concern is whether they all point to the same answer except for sampling error. This is no less the case for phylogenetic analysis. If rates of evolution exceed a certain threshold, and edge (branch or internode) lengths and branching pattern meet certain configurations, then maximum parsimony will be inconsistent (Felsenstein 1978). That is, the data will indicate the wrong tree with increasing support as sequence length tends to infinity. Accordingly, it is possible that faster-evolving characters strongly disagree with more slowly evolving characters in recovering the underlying tree, and the combined data set may also indicate an incorrect tree (Bull et al. 1993). Furthermore, inconsistency has been shown to be a problem for all classes of phylogenetic analysis, including distance methods and maximum likelihood (Waddell 1995; Lockhart et al. 1996).

The data set may consist of a mixture of sites, that is, data partitions (a generic term that can mean a gene, a part of a gene, a set of genes, a locus, etc.) which are either consistent or inconsistent with regard to identifying the historical tree. Because of this, it is unwise to automatically combine all data (Bull et al. 1993; Huelssenbeck et al. 1994). Rather, one should make an effort to detect partitions of the data that lead to inferences of statistically distinct trees and then proceed with caution if differences are found (Cunningham 1997). Note that we use the term “congruence” to mean agreement, which does not necessarily mean that trees are the same, but rather that the differences are not statistically significant.

There are other causes of incongruence apart from inconsistency of tree estimators and the attendant systematic error. One is stochastic bias due to finite sample size, i.e., a difference of the mean estimate with finite data from that with infinite data. Accordingly, the total difference between a tree estimate and the historical tree is a function of stochastic variance about the estimator’s mean due to finite data, stochastic bias, and systematic error. However, stochastic bias generally drops away quickly with increasing sequence length and has rarely been shown to be a dominant factor in tree inference.

Key words: phylogenetic congruence likelihood ratio tests, combination of sequence data, phylogenetic analyses, inconsistency.

Address for correspondence and reprints: Peter J. Waddell, E-mail: waddell@onyx.si.edu, pjwaddell@hotmail.com. Communication preferred by E-mail; please advise by E-mail before using other forms of communication.

Mol. Biol. Evol. 17(12):1988–1992, 2000

© 2000 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Another cause of incongruence is ancestral polymorphism. However, branch points of taxa outside of species complexes tend to be separated by reasonably long periods that are typically much longer than the ancestral population coalescence time. Furthermore, given recombination rates similar to substitution rates, finding gene trees that disagree strongly, e.g., with the bootstrap, is even more unlikely (Waddell 1995). A third cause is horizontal transfer, which seems relatively rare in animals but is more common in bacteria and in plants, e.g., via hybridization. Overall, however, there is considerable evidence for inconsistency of tree estimators, e.g., quite different trees from first, second, and third positions, and we feel that this is the a priori most likely explanation in higher-order vertebrate systematics. That is, generally, inconsistency (systematic error) of estimators should be considered first, before looking for other causes. Therefore, associating the tests below with detection of inconsistency is often reasonable a priori, as discussed later.

Nested hypotheses of whether the optimal tree topologies for k data partitions are significantly different can be set up as follows.

H_0 : The trees from data partitions 1, . . . , k are congruent, so we expect that the data sets were generated under the same tree topology (called T_{MED} , the median tree).

H_1 : The trees for data partitions 1, . . . , k are not congruent.

For example, assume there are two genes, A and B, and four taxa. Thus, there are three unrooted resolved trees T_1 , T_2 , and T_3 , with parameters θ_1 , θ_2 , and θ_3 , respectively. Denote the data partitions A and B by \mathbf{X}_A and \mathbf{X}_B , respectively. Then, the maximum likelihood for tree T_i with \mathbf{X}_j is $\ln L(\mathbf{X}_j | \hat{\theta}_i(\mathbf{X}_j), T_i)$, where \ln is the natural log, and $\hat{\theta}_i(\mathbf{X}_j)$ is the maximum-likelihood estimate of θ_i , for T_i , based on data \mathbf{X}_j . Let $\ln L(T_{iA}, T_{jB}) = \ln L(\mathbf{X}_A | \hat{\theta}_i(\mathbf{X}_A), T_i) + \ln L(\mathbf{X}_B | \hat{\theta}_j(\mathbf{X}_B), T_j)$. Furthermore, assume that the set of parameters describing the process of character change is chosen using AIC (e.g., Kishino, Miyata, and Hasegawa 1990). The log likelihood for H_0 , $\ln L(H_0)$, is

$$\max \{ \ln L(T_{1A}, T_{1B}), \ln L(T_{2A}, T_{2B}), \ln L(T_{3A}, T_{3B}) \},$$

and that for H_1 , $\ln L(H_1)$, is

$$\max \{ \ln L(T_{1A}, T_{1B}), \ln L(T_{2A}, T_{2B}), \ln L(T_{3A}, T_{3B}),$$

$$\ln L(T_{1A}, T_{2B}), \ln L(T_{1A}, T_{3B}), \ln L(T_{2A}, T_{1B}),$$

$$\ln L(T_{2A}, T_{3B}), \ln L(T_{3A}, T_{1B}), \ln L(T_{3A}, T_{2B}) \}.$$

The test statistic $cLR = \ln L(H_1) - \ln L(H_0)$. Clearly, the hypotheses are nested, but both consist of multiple

hypotheses (sets of trees), each involving the maximization of a quantity unknown before the data are analyzed. Thus, the usual χ^2 asymptotic distributions of LR statistics that are often so convenient to use when testing simple nested hypotheses do not apply here without invoking special conditions, namely, that the optimal tree for each data partition is fixed a priori. The above formulation follows Huelsenbeck and Bull (1996), except that we do not expect a simple convergence to χ^2 asymptotically.

To test H_0 versus H_1 when a tree is unknown a priori for each gene, brute force Monte Carlo (parametric bootstrap) computer simulation may be used (Huelsenbeck and Bull 1996; Huelsenbeck, Bull, and Cunningham 1996). When the tree that minimizes cLR is searched for (call it $T_{\widehat{\text{MED}}}$; it is also the overall maximum-likelihood tree, as seen later), then n , e.g., 1,000, replicate samples of data A and B are simulated under the optimal parameter values associated with $T_{\widehat{\text{MED}}}$. Each replicate data partition is analyzed in the same way as the original data (including tree search), and the difference in $\ln L$ under H_0 and H_1 is stored. The results of these replicates approximate the true distribution of cLR under H_0 . This test follows the outline of Cox (1962) and is both exact and statistically efficient (for phylogenetic applications, see Reeves 1992; Goldman 1993). This approach is, unfortunately, computationally very expensive, and consequently this test is rarely used.

Another test of congruence (Waddell et al. 1999) takes a tree, e.g., $T_{\widehat{\text{MED}}}$, and asks if the maximum-likelihood trees of each data partition are within sampling error of it. This test measures the distance from $T_{\widehat{\text{MED}}}$ to the nominated tree, T_i , of each data partition in normalized $\Delta \ln L$ units, i.e., $\ln L(T_{\widehat{\text{MED}}}) - \ln L(T_i)$ divided by the standard error of this difference (with the latter estimated using the formula of Kishino and Hasegawa [1989]). These normalized likelihood differences should be close to standard-normally distributed, i.e., $N(0, 1)$. As long as all trees are specified a priori, the expected distribution of the sum of squared normalized deviations approximates χ^2 (for f data partitions, degrees of freedom equals f if the median tree is known a priori, or $f - 1$ when $T_{\widehat{\text{MED}}}$ is chosen to minimize the squared deviations). This test does not take into account the increased uncertainty due to tree searching on the data being analyzed.

We report here a fast way to estimate the distribution of cLR under H_0 . It takes into account the maximization across trees by using centering (e.g., Hall 1992, p. 292) of log likelihood values obtained from bootstrap replicates. Centering was recently used by Shimodaira and Hasegawa (1999) to assist in the selection of a confidence set of trees for a single data partition, and this is a natural application of the centering procedure to another type of phylogenetic test.

A Rapid Likelihood Ratio-Based Congruence Test

Let $\mathbf{X}_1, \dots, \mathbf{X}_k$ be the data partitions $j = 1, \dots, k$, respectively. Next, select an appropriate set of parameters describing the evolution of characters in each data

partition (e.g., using AIC). Select a type I error level, α (e.g., 0.05), and fix n , the number of replicates to be used to infer the distribution under H_0 ($n > 100$ is preferable), then perform the test as follows.

- Step 1. Find the maximum-likelihood tree, i.e., T_1, \dots, T_k , for the k data partitions. If there are other trees to be tested, add in $z - (k + 1)$ other trees, T_{k+1}, \dots, T_z (so that z is the total number of trees included in the test).
- Step 2. The test statistic for the m th tree ($m = 1, \dots, z$) is

$$\text{cLR}_m = \sum_{j=1}^k (\ln L(\mathbf{X}_j | \hat{\theta}_j(\mathbf{X}_j), T_j) - \ln L(\mathbf{X}_j | \hat{\theta}_m(\mathbf{X}_j), T_m)). \quad (1)$$

In this step, all free parameter values, including edge lengths, are reoptimized on each tree. By shifting the summation in equation (1), clearly, the tree that gives the minimum value of cLR_m , and hence the minimum sum of likelihood differences to the maximum-likelihood trees of each data partition, is $T_{\widehat{\text{MED}}}$, with statistic $\text{cLR}_{\widehat{\text{MED}}}$.

- Step 3. (Necessary only if using RELL in step 4). For each data partition j , $j = 1, \dots, k$, store the site pattern log likelihoods for T_m , $m = 1, \dots, z$, in the vector \mathbf{v}_{jm} (giving the set $\mathbf{v}_{11}, \dots, \mathbf{v}_{kz}$).
- Step 4. For each data partition, $j = 1, \dots, k$, generate n bootstrap replicates of the data, storing the h th replicate as \mathbf{I}_{jm}^{*h} . For each h th replicate, evaluate the likelihood of data partition j on the m th tree by reoptimizing all free parameters, and store the likelihood (a single number) as $\ln L_{jm}^{*h}$. The method of resampling estimated site log likelihoods (RELL; see Kishino, Miyata, and Hasegawa 1990) from step 3, which avoids reoptimization of parameters, is a much quicker way to approximate $\ln L_{jm}^{*h}$.
- Step 5. Calculate $1/n \sum_{h=1}^n \ln L_{jm}^{*h}$, which is the average likelihood of data partition j under tree m , for the n replicates. Subtract this number from each likelihood replicate, i.e., $R_{jm}^{*h} = \ln L_{jm}^{*h} - 1/n \sum_{h=1}^n \ln L_{jm}^{*h}$.
- Step 6. For each replicate and each tree, calculate $S_{jm}^{*h} = \max_i \{R_{ji}^{*h}\} - R_{jm}^{*h}$.
- Step 7. For each replicate h and each tree T_m , calculate $E_{hm} = \sum_{j=1}^k S_{jm}^{*h}$, where E_{hm} is the h th replicate of cLR_m .
- Step 8. Calculate $p_m = \#\{E_{hm} > \text{cLR}_m\}/n$, for each $m = 1, \dots, z$. If $p_m < \alpha$, for all $m = 1, \dots, z$, reject H_0 ; otherwise, fail to reject H_0 .

The new test is described in general form. That is, it allows all trees from step 1 to be tested as congruent with the maximum-likelihood trees of the data partitions. If one simply wants to test H_1 versus H_0 , then

rejection of T_{MED} almost certainly means rejection of H_0 (thus, in steps 6–8, only T_{MED} need be considered).

Note that steps 5 and 6 involve centering, which basically acts to conform the sampled log likelihoods of the data under different trees to what they would be if they came from a single tree as H_0 specifies. Unlike the parametric bootstrap, centering forces all trees to have equal expected likelihoods under H_0 , which is typically not the case (i.e., some trees really are better than others). Generally, the test is conservative; i.e., it rejects H_0 too rarely, due to this last point.

The number of replicates, n , can be selected in an adaptive way with minimal impact on the outcome of the test. As a simple example, if $\alpha = 0.05$ and $p_m = 0$ for a trial of 100 replicates, it is sufficient to stop; otherwise, return to step 4 and increase n to 1,000.

It may be assumed that the test is made over all binary trees. However, these trees grow in number exponentially, so it is not feasible for moderate numbers of taxa. Additionally, due to the nature of centering, adding in trees far from T_{MED} only makes the test more conservative. In practice, the test works best when we keep z small and include mostly trees within sampling error of each of the maximum-likelihood trees, including T_{MED} . The test is fast when using REL, so if the site likelihoods of promising trees (e.g., by the KH test) can be stored simultaneously with a tree search, there is minimal additional computational load.

A useful aspect of the new test is that any or all trees, $1, \dots, z$, can be tested to determine if they are acceptable “median” trees, including trees stated a priori (e.g., a morphological tree, a tree from another type of genetic data, or an area cladogram). Note that when testing any T_m as an acceptable median tree, the distribution of the test statistic is evaluated by centering on T_m (steps 6–8). For the parametric bootstrap, this is equivalent to resimulating on T_m (even if it is not an optimal tree). However, a test based on the distribution of cLR from the highest likelihood tree alone will often give the same answer. In testing a completely a priori tree, the new test tends to be more conservative than usual because there is no maximization of likelihood across trees for this tree.

Application to Data Partitions in mtDNA

Consider the question of whether the tree from the concatenated tRNA sequences (22 genes, 1,240 characters) of vertebrate mtDNA is congruent with the tree from the concatenated amino acids (AA; 12 genes, 2,272 noninvariant characters) of the same molecule (Waddell et al. 1999). Each data partition has the same 36 taxa and spans from lamprey to closely related primates (data available from <http://wheat.ab.a.u.-tokyo.ac.jp>). The test of Waddell et al. (1999) did not show a significant difference, although as models became more general, this test came very close to rejecting H_0 with $\alpha = 0.05$. Here, we reevaluate using the new test. First, we analyze the tRNA data with the most general model in PAUP*, version 4.0 (Swofford 1998), that is, the general time-reversible model with eight free parameters other than

the edge lengths, plus a mixture of invariant sites ($\hat{p}_{\text{inv}} = 0.060$) and variable sites following a Γ distribution ($\hat{k} = 0.973$) (using successive updating of the maximum-likelihood estimates of parameters on the optimal tree following a tree bisection-reconnection tree search). The maximum-likelihood tree is slightly different from that found previously, although the differences are not statistically significant. The best tree now puts the unusual fish *Polypterus* with the ray fins, as most morphologists would favor. Also, the platypus joins the marsupials, and the armadillo branches immediately before the elephant, just one short internode from being sister to all the other placentals. tRNA trees of near-optimal likelihood combine local permutations of these parts of the tree, with those of the relative positions of the lamprey, the lungfish, and tetrapods. The amino acids are analyzed using ProtML, version 2.3 (Adachi and Hasegawa 1996), the included mtDNA rate matrix, and removal of invariant sites in proportion to the constant sites (Waddell et al. 1999).

Importantly, we have a priori biological trees from independent data. Traditional methods (paleontology and morphology) favor the tree T_{bio1} , (lamprey, (shark, ((rayfins, *Polypterus*), (coelacanth, (lungfish, (frog, (reptiles, (monotremes, (marsupials, (placentals)))))))))) (e.g., Carrol 1987), while relationships within placentals are often assumed to follow Novacek (1992, p. 64). An additional tree, T_{bio2} , our own estimate of the historical tree, is the same as above except that relationships within placentals follow Waddell, Okada, and Hasegawa (1999). Support for T_{bio2} is largely independent of the mtDNA data. Are these trees congruent with the maximum-likelihood trees from amino acids and tRNA?

In performing the test, $\alpha = 0.05$ and $n = 10,000$. Our set of z trees incorporates 20 trees within sampling range of the maximum-likelihood tree from either tRNA or amino acid data partitions, plus the two biological trees (in this case, trees around the amino acid maximum, T_{AA2} , are also trees around T_{MED}). The results of testing each individual tree are shown in table 1. Clearly, the new test does not reject H_0 for the trees that have near-maximal likelihood based on the amino acid data, but it does reject all the best trees based on the tRNA data. This result implies that the tRNA sequences have low resolving power in parts of the phylogeny where tRNA and amino acid trees disagree. Thus, there is no evidence of the phylogenetic inferences based on tRNA and amino acid sequences being out of harmony with each other, so a joint estimate of the best tree would not seem unreasonable.

As seen from table 1, both a priori biological trees are clearly incongruent with the trees from the data. Explanations include the biological estimates being wrong and/or the mtDNA data not being analyzed adequately. We expect predominantly the latter explanation outside of placentals, while within placentals, nuclear genes tend to support the relationships in T_{bio2} (unpublished data), making it a likely explanation within placentals for T_{bio2} . Regarding mtDNA evolution, the interaction of rate variation across sites and a nonstationary substitution process is very hard to model properly.

Table 1
Testing Trees Under Which the mtDNA Data Might Have Been Generated

Tree	$\ln L$ AA	BP ^a AA	$\ln L$ tRNA	BP ^a tRNA	cLR	P_1^b	P_2^c
AA ₁	-72,736	0.080	-18,871	0.000	39.86	0.668	0.685
AA ₂	-72,739	0.204	-18,864	0.002	35.23	0.746	0.746
AA ₃	-72,766	0.003	-18,870	0.000	68.64	0.355	0.352
AA ₄	-72,753	0.066	-18,860	0.005	45.72	0.630	0.608
AA ₅	-72,741	0.077	-18,872	0.000	46.15	0.593	0.602
AA ₆	-72,739	0.152	-18,878	0.000	49.74	0.554	0.555
AA ₇	-72,740	0.256	-18,885	0.000	57.31	0.477	0.467
AA ₈	-72,744	0.059	-18,876	0.000	52.28	0.525	0.526
AA ₉	-72,742	0.102	-18,872	0.000	46.25	0.609	0.601
AA ₁₀	-72,757	0.001	-18,886	0.000	76.01	0.300	0.290
tRNA ₁	-72,893	0.000	-18,832	0.038	156.96	0.006	0.011
tRNA ₂	-72,861	0.000	-18,832	0.309	124.98	0.029	0.048
tRNA ₃	-72,907	0.000	-18,833	0.042	172.75	0.002	0.005
tRNA ₄	-72,919	0.000	-18,833	0.154	184.51	0.001	0.003
tRNA ₅	-72,867	0.000	-18,833	0.037	132.19	0.018	0.035
tRNA ₆	-72,884	0.000	-18,833	0.044	149.56	0.012	0.016
tRNA ₇	-72,949	0.000	-18,834	0.011	214.59	0.000	0.000
tRNA ₈	-72,936	0.000	-18,834	0.048	202.12	0.000	0.001
tRNA ₉	-72,937	0.000	-18,833	0.177	202.28	0.000	0.001
tRNA ₁₀	-72,949	0.000	-18,832	0.124	213.19	0.001	0.000
bio ₁	-73,208	0.000	-18,871	0.009	511.24	0.000	0.000
bio ₂	-73,009	0.000	-18,873	0.001	314.14	0.000	0.000

NOTE.—AA = amino acid.

^a Bootstrap proportions from RELL with $n = 10,000$.

^b P value of the new test.

^c P value if the distribution of the test statistic is fixed on that of the overall highest total likelihood tree, AA₂.

The estimated distributions of the cLR test statistic when testing T_{AA_2} and T_{bio_1} as possible median trees are shown in figure 1. The long tail with T_{bio_1} illustrates one facet of the conservativeness due to centering. Note that, in contrast to the present test, a failure to record even 1% of bootstrap support for a tree (table 1) does not indicate incongruence, since this type of inspection does not take into account the number and proximity of competing trees.

Discussion

Because the desire to combine conflicting data is strong in real studies (e.g., Liu and Miyamoto 1999),

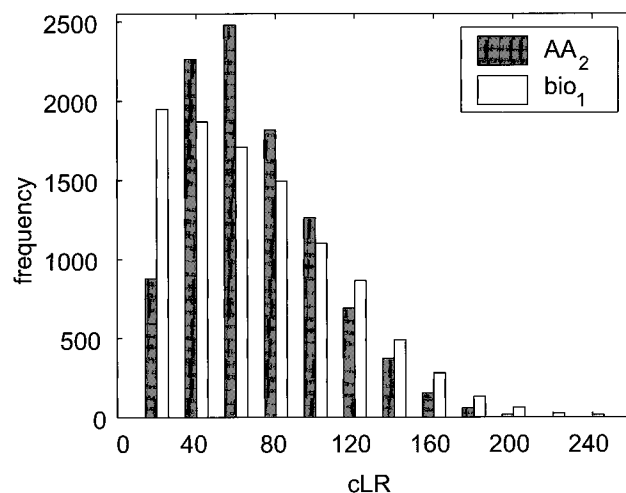


FIG. 1.—The maximally different distributions of cLR generated under different trees (specifically, T_{AA_2} and T_{bio_1}).

the conservativeness of the present test is unlikely to be a major concern for some. Combining data often makes sense if one's major concern is stochastic error. That is, if one expects to get an additional three edges correct by reducing sampling error, but only one additional edge wrong due to systematic error, one has a valid argument for combining.

Conservativeness of a test is also presently desirable, because maximum-likelihood tests of congruence assume that we can identify and generate the data under the true model of character evolution; failure to take this process into account causes tests of homogeneity to reject H_0 too often (Waddell 1995; Huelsenbeck, Bull, and Cunningham 1996). Furthermore, in reality, site evolution is typically correlated, so this breakdown of the independence of sites also leads to the parametric bootstrap rejecting H_0 too often.

If the new test is considered too conservative, a parametric bootstrap can be invoked in those cases where rejection of H_0 is not quite achieved (e.g., $0.05 < P < 0.3$). Initial results of a computationally fast version of Huelsenbeck and Bull's (1996) test are also promising. The test is the same until data sets are evolved according to the model on T_{MED} , except that during the tree search on the original data, trees that are within sampling error of T_{MED} and the maximum-likelihood trees of each partition are also stored. Next, n replicates for j partitions are made. Then, all n replicate sets of characters evolved for partition j are temporarily pooled to allow a single estimate of parameters and site likelihoods on each tree to be made and stored. The site log likelihoods just estimated are substituted in for each pattern in the replicate data sets (like RELL). This al-

lows the likelihood of each tree in each replicate partition to be evaluated by simply summing the site likelihood scores. For each of the n replicates, the sum of log likelihood differences from the tree of the highest total likelihood from each of the best trees for the j partitions, the cLR statistic, is used to build the test distribution. For moderate sequence lengths, this method builds a distribution of cLR similar to that constructed using full reoptimization and tree search, as long as the set of trees stored includes all that are likely to be optimal for each replicate. Because of this, the tree search should be extensive, e.g., based around the TBR function of PAUP*.

Note that the centering test assumes that the data evolved on the same tree topology, but edge lengths do not have to show any relationship across data partitions. The test modifies easily to impose either the same edge lengths for all data partitions or the same edge lengths except for some scalar, such as overall evolutionary rate.

The hypotheses for the new test can be rephrased and used to test for systematic error, which is predominantly due to the inconsistency of tree estimators, when there is a biological reason to expect that the historical trees from all data partitions are identical. For example, in the case with nonrecombining mtDNA, we have H_0 : the estimated trees are mutually consistent; H_1 : at least one of the trees is an inconsistent estimate of the historical tree. Thus, we do not need to know the historical tree to test for inconsistency if we can safely disregard stochastic bias.

The earlier test in Waddell et al. (1999) using squared normalized likelihood differences works well in combination with the present test to diagnose which data partitions most strongly disagree with H_0 . It is also the case that if the test of Waddell et al. (1999) fails to reject H_0 , the present test will most likely do the same. Thus, as a very quick test, it retains utility even when trees are not specified a priori.

Acknowledgments

P.J.W. and R.O. thank the Marsden Fund of New Zealand for support in aid of this work; R.O. thanks Monbusho of Japan; H.K. acknowledges grants JSPS-12554037 and BSAR-497. We thank John Huelsenbeck, Pam Soltis, Jeff Thorne, and the anonymous reviewers for helpful comments.

LITERATURE CITED

- ADACHI, J., and M. HASEGAWA. 1996. MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood. *Comput. Sci. Monogr.* **28**:1–150.
- BULL, J. J., J. P. HUELSENBECK, C. W. CUNNINGHAM, D. L. SWOFFORD, and P. J. WADDELL. 1993. Partitioning and combining data in phylogenetic analysis. *Syst. Biol.* **42**:384–397.
- CARROL, R. L. 1987. *Vertebrate paleontology and evolution*. W. H. Freeman, New York.
- COX, D. R. 1962. Further results on tests of separate families of hypotheses. *J. R. Stat. Soc.* **B34**:406–424.
- CUNNINGHAM, C. W. 1997. Can three incongruence tests predict when data should be combined? *Mol. Biol. Evol.* **14**:733–740.
- FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**:401–410.
- GOLDMAN, N. 1993. Statistical tests of models of DNA substitution. *J. Mol. Evol.* **36**:182–198.
- HALL, P. 1992. *The bootstrap and Edgeworth expansion*. Springer-Verlag, New York.
- HUELSENBECK, J. P., and J. J. BULL. 1996. A likelihood ratio test to detect conflicting phylogenetic signal. *Syst. Biol.* **45**:92–98.
- HUELSENBECK, J., J. J. BULL, and C. W. CUNNINGHAM. 1996. Combining data in phylogenetic analysis. *Trends Ecol. Evol.* **11**:152–158.
- HUELSENBECK, J. P., D. L. SWOFFORD, C. W. CUNNINGHAM, J. J. BULL, and P. J. WADDELL. 1994. Is character weighting a panacea for the problem of data heterogeneity in phylogenetic analysis? *Syst. Biol.* **43**:288–295.
- 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.
- KISHINO, H., T. MIYATA, and M. HASEGAWA. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J. Mol. Evol.* **31**:151–160.
- LIU, F.-G. R., and M. M. MIYAMOTO. 1999. Phylogenetic assessment of molecular and morphological data for eutherian mammals. *Syst. Biol.* **48**:54–64.
- LOCKHART, P. J., A. W. LARKUM, M. A. STEEL, P. J. WADDELL, and D. PENNY. 1996. Evolution of chlorophyll and bacteriochlorophyll: The problem of invariant sites in sequence analysis. *Proc. Natl. Acad. Sci. USA* **93**:1930–1934.
- NOVACEK, M. J. 1992. Fossils, topologies, missing data, and the higher level phylogeny of eutherian mammals. *Syst. Biol.* **41**:58–73.
- REEVES, J. H. 1992. Heterogeneity in the substitution process of amino acid sites of proteins coded for by Mitochondrial DNA. *J. Mol. Evol.* **35**:17–31.
- SHIMODAIRA, H., and M. HASEGAWA. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**:1114–1116.
- SWOFFORD, D. L. 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4.0b1. Sinauer, Sunderland, Mass.
- WADDELL, P. J. 1995. Statistical methods of phylogenetic analysis: including Hadamard conjugations, LogDet transforms, and maximum likelihood. Ph.D. thesis, Massey University, Palmerston North, New Zealand.
- WADDELL, P. J., Y. CAO, J. HAUF, and M. HASEGAWA. 1999. Using novel phylogenetic methods to evaluate mammalian mtDNA, including AA invariant sites—LogDet plus site stripping, to detect internal conflicts in the data, with special reference to the position of hedgehog, armadillo, and elephant. *Syst. Biol.* **48**:31–53.
- WADDELL, P. J., N. OKADA, and HASEGAWA. 1999. Progress in resolving the interordinal relationships of placental mammals. *Syst. Biol.* **48**:1–5.

PAMELA SOLTIS, reviewing editor

Accepted August 25, 2000