

Points of View

Syst. Biol. 50(5):723–729, 2001

Should We Use Model-Based Methods for Phylogenetic Inference When We Know That Assumptions About Among-Site Rate Variation and Nucleotide Substitution Pattern Are Violated?

JACK SULLIVAN¹ AND DAVID L. SWOFFORD²

¹*Department of Biological Sciences, Box 443051, University of Idaho, Moscow, Idaho 83844-3051, USA;
E-mail: jacks@uidaho.edu*

²*Laboratory of Molecular Systematics, Smithsonian Museum Support Center, 4210 Silver Hill Rd., Suitland,
Maryland 27460, USA; E-mail: swofford@lms.si.edu*

An often-expressed concern regarding the use of maximum likelihood and other model-based methods in phylogenetic analysis is that the assumed models are too simple, violating known aspects of sequence evolution in several ways (e.g., Sanderson and Kim, 2000). For example, we know that sites do not truly evolve independently, that the substitution process is not completely homogeneous across sites and through time, and that simple stochastic models fail to adequately represent the complexity of the nucleotide substitution process. Many advances have been made in an effort to model sequence evolution more realistically, including allowance for unequal base composition (Felsenstein, 1981), complex substitution patterns (e.g., Tavaré, 1986; Yang, 1994), among-site rate variation (Yang, 1993), compensatory base substitution (Schöniger and Von Haeseler, 1994), and heterogeneity of the substitution process across the tree (Galtier and Gouy, 1998). Nevertheless, even with these improvements, there is no reason to believe that even the most general and parameter-rich evolutionary models currently available capture all the nuances of the processes that have generated any particular set of sequences. Furthermore, accurate estimation of the parameters of complex evolutionary models can be very difficult, and data sets containing many taxa and sites may be needed for likelihood estimation to consistently yield reliable parameter estimates (Nielsen, 1997; Sullivan et al., 1999).

Fortunately, perfect models are not necessarily a prerequisite for reliable statistical inference (e.g., Burnham and Andersen, 1998). Many statistical methods perform well in the face of violation of common distributional assumptions such as normality (e.g., Boneau, 1960; Donaldson, 1968). However, this observation does not guarantee the robustness of maximum likelihood when applied to the phylogenetic estimation problem. It is therefore important to examine the impact of various model violations on the accuracy of phylogenetic estimation (e.g., Huelsenbeck, 1995). In some cases, biases introduced by violation of a model's assumptions can actually favor recovery of the true tree rather than an incorrect tree (Yang, 1997; Siddall, 1998). For example, Yang (1997) simulated data under a Jukes–Cantor (JC) model with variable rates across sites (following a gamma distribution) and demonstrated that estimation under the JC model with equal rates could be more efficient than estimation in which among-site rate variation was correctly modeled by a gamma distribution. However, the reasons for the improvement in efficiency should be disconcerting (Bruno and Halpern, 1999; Swofford et al., 2001).

These results suggest that application of increasingly general and complex models would sometimes lead to decreased efficiency, despite the fact that the more complex models almost always provide significantly better fit to real data than the simpler models. However, in Yang's study, the improvement

in efficiency was a direct result of a bias (introduced by model violation) that favored the true tree rather than an incorrect tree (Bruno and Halpern, 1999). On the other hand, in cases where the bias favors an incorrect rather than a correct result, accurate modeling of the substitution process becomes more and more important (e.g., Gaut and Lewis, 1995). Practically, in any given empirical study, there is no way to know a priori whether bias will act to increase or decrease accuracy of estimation. We believe most systematists would prefer methods that represent the strength of support for a particular conclusion accurately and that converge to the correct solution as the amount of data increases. This latter property, known as consistency, is often used to justify preference for maximum-likelihood methods in phylogenetics (e.g., Felsenstein, 1978).

Some have argued that consistency is less relevant in the choice of methods for phylogenetic inference (e.g., Farris, 1983, 1999; Siddall and Kluge, 1997; Sanderson and Kim 2000). As emphasized most strongly by Farris (1983, 1999) and reiterated by Sanderson and Kim (2000), proofs of consistency are model dependent. Although one might show that a method is consistent if data are generated precisely according to the assumptions of a model, any guarantee of consistency vanishes if the assumptions of that model are violated. This result is then used to justify a preference for models that are far too general (or an avoidance of models altogether) over the use of more tractable models that are not perfect in every detail. However, a huge gulf separates acceptance of the fact that the consistency of a method usually cannot be assured with mathematical certainty from an outright dismissal of the relevance of consistency. A method may behave consistently over a wide range of conditions despite the violation of some of its assumptions. Although it may seem unsettling to use a wrong model, the cost of using a model known to violate some of its assumptions may be vastly less than that associated with the use of an alternative method that avoids (or seems to avoid) these assumptions altogether. Here, we defend this point, using computer simulations that generate data according to a relatively complex mixed-distribution model of among-site rate variation (the invariable sites plus gamma model of $[I + \Gamma]$ Gu et al. [1995]

and Waddell and Penny [1996]). We analyze our simulated data sets by using maximum likelihood, not only with the correct model but also with several simpler models for which assumptions about the distribution of among-site rate variation and nucleotide-substitution pattern are violated. The results of these analyses reinforce our position that the maximum-likelihood method can behave consistently in the face of violation of its assumptions and show that abandoning model-based methods simply because consistency cannot be proven for real data or because the models are imperfect would be unwise.

SIMULATIONS

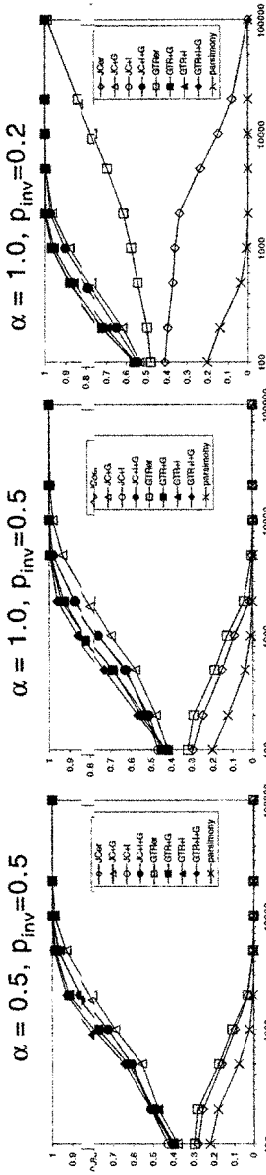
We conducted two sets of simulations. In the first and most extensive, 1,000 replicate data sets were generated for each of several sequence lengths (100, 200, 500, 1,000, 2,000, 5,000, 10,000, 20,000, and 100,000 bp) on each of three four-taxon trees (Fig. 1). These trees represent a "Felsenstein-zone" tree (Fig. 1, top tree), an equal-branch-lengths tree (Fig. 1, middle tree), and an "inverse-Felsenstein-zone" tree (Fig. 1, bottom tree). This last tree shape, with two long branches on the same side of a short internal branch, is equivalent to the "anti-Felsenstein zone" tree of Waddell (1995) and the "Farris zone" of Siddall (1998). The general time-reversible model (GTR; equal to REV of Yang, 1994) was used with a mixed-distribution model ($I + \Gamma$) to generate rate heterogeneity across sites. For all these simulations, the following base-frequency parameters and relative instantaneous rates (see Yang, 1994) were used:

$$\begin{array}{ll} \pi_A = 0.30 & r_{AC} = 5.0 \\ \pi_C = 0.25 & r_{AG} = 10.0 \\ \pi_G = 0.15 & r_{AT} = 3.0 \\ \pi_T = 0.30 & r_{CG} = 1.0 \\ & r_{CT} = 15.0 \\ & r_{GT} = 1.0 \end{array}$$

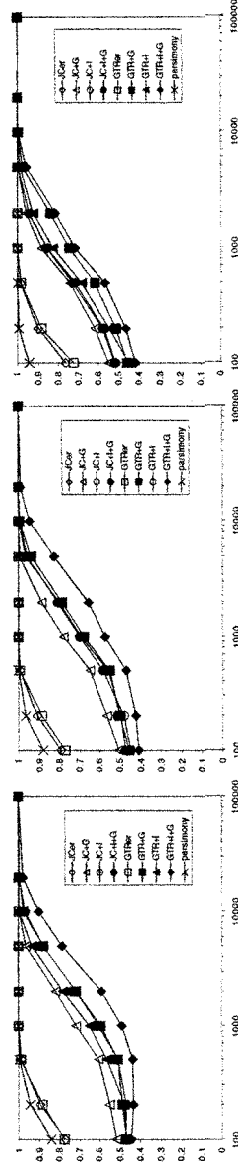
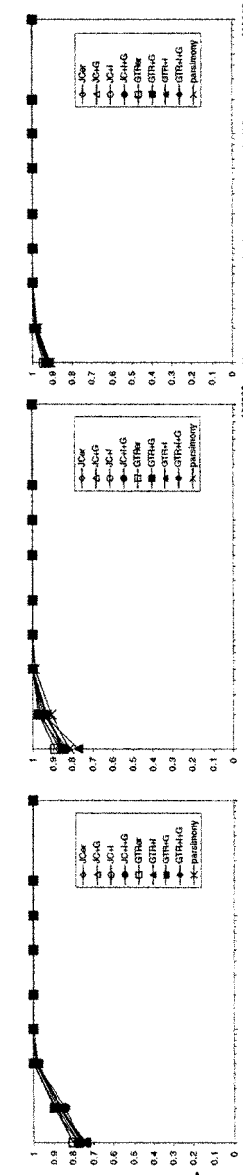
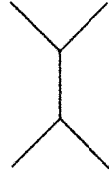
Under this model, some proportion of sites is fixed to be invariable (p_{inv}), and rates at the remaining variable sites are drawn from a gamma distribution, the shape parameter (α) of which determines the strength of the

Rate Heterogeneity

Tree



Proportion Correct



Sequence Length

FIGURE 1. Results of four-taxon simulations. Simulations were conducted on a Felsenstein-zone tree (top row), an equal-branch-lengths tree (middle row), and an inverse-Felsenstein-zone tree (bottom row). In all cases, sequences were generated by using a GTR + I + Γ model of sequence evolution (see text for base frequencies and rate-matrix parameters), with three different rate heterogeneity conditions: extreme (left column), strong (middle column), and weak (right column). In the trees with unequal branches, the long branches were 0.5 expected substitutions/site and short branches were 0.05 expected substitutions/site (top and bottom rows), whereas in the equal branch lengths tree, all branches were set to 0.2 expected substitutions/site.

rate heterogeneity (across variable sites). This mixed-distribution model, which was developed independently by Gu et al. (1995) and Waddell and Penny (1996), was chosen because it is intuitively appealing and has been shown to fit many data sets significantly better than any of its special cases (e.g., Wilgenbusch and de Queiroz, 2000). Furthermore, estimation of the parameters of this model can be particularly difficult for data sets of the size typically used in empirical studies (Sullivan et al., 1999).

Three different rate-heterogeneity conditions were simulated: extreme ($p_{inv} = 0.5$, $\alpha = 0.5$), strong ($p_{inv} = 0.5$, $\alpha = 1.0$), and weak ($p_{inv} = 0.2$, $\alpha = 1.0$). For each combination of tree topology, sequence length, rate heterogeneity, and simulation replicate, phylogeny was estimated by using PAUP* (Swofford, 1999) and the following methods: maximum likelihood under the generating model (GTR + I + Γ), a GTR model with only invariable sites (GTR + I), a GTR model with only gamma-distributed rates (GTR + Γ), an equal-rates GTR model (GTRer), the corresponding JC (Jukes and Cantor, 1969) and HKY (Hasegawa et al., 1985) models, and equally weighted parsimony. Accuracy was calculated as the proportion of replicates in which the true tree was recovered by each method. Consistency was determined by whether the accuracy converged toward 0% or 100% with increasing sequence length, using sequence lengths long enough to overcome small-sample biases that initially can present a misleading picture (see Swofford et al., 2001).

In the second set of simulations, we expanded the number of taxa to eight. The single tree examined (Fig. 2a) was generated from the four-taxon Felsenstein-zone tree (Fig. 1) by splitting each of the terminal branches exactly in half. In this simulation, 100 replicate data sets were generated for five sequence lengths (200, 1,000, 2,000, 5,000, and 10,000 nucleotides) using the HKY + I + Γ model with extreme rate heterogeneity ($p_{inv} = 0.5$, $\alpha = 0.5$). The same set of base frequency parameters were used with κ -parameter of 5 (this defines the transition/transversion rate ratio). Heuristic searches were conducted for each replicate under likelihood with each of the four variants of the HKY model examined above, and with parsimony.

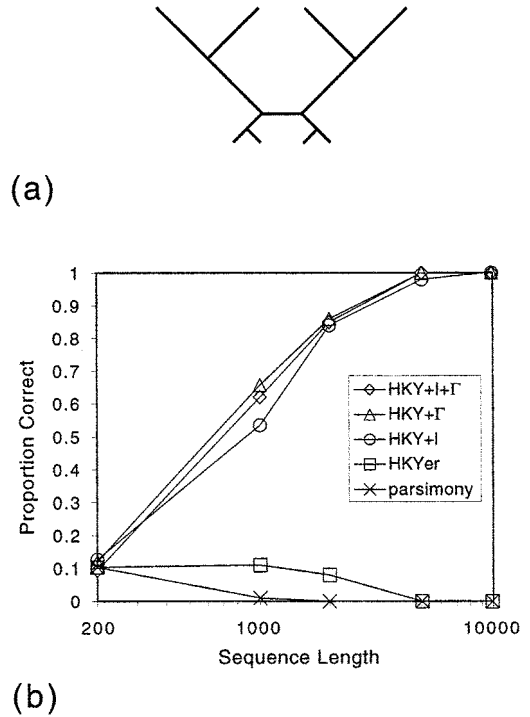


FIGURE 2. Results of eight-taxon simulations. (a) This tree was generated from the four-taxon Felsenstein-zone tree shown in the top left of Figure 1 by splitting all terminal branches in half. (b) Simulations were conducted on the tree under an HKY + I + Γ model with $\kappa = 5$, and extreme rate heterogeneity ($p_{inv} = 0.5$, $\alpha = 0.5$).

RESULTS AND DISCUSSION

The results of the first set of simulations varied across the three tree shapes we examined. In all of the simulations conducted with the Felsenstein-zone tree (Fig. 1, top row), parsimony was inconsistent across all three rate heterogeneity conditions examined. Similarly, as expected, based on the proofs of Chang (1996) and Rogers (1997, 2001), under all three conditions of rate heterogeneity, likelihood estimation under the true (generating) model was consistent. Likelihood analysis under all of the incorrect heterogeneous-rate models was also consistent for all three rate-heterogeneity conditions in the Felsenstein zone. Likelihood estimation of topology under an equal-rates model was inconsistent when the data contained either extreme or strong rate heterogeneity (Fig. 1, top left and top middle panels). However, likelihood estimation under equal rates is apparently consistent when there is weak rate heterogeneity in the data, but only if the substitution matrix is

sufficiently complex (GTRer vs JCe; Fig. 1, top right); however, the proportion correct did not reach 1 for the longest sequences we simulated (100,000 nucleotides). Under this combination of weak rate heterogeneity and a Felsenstein-zone tree, maximum likelihood estimation was inconsistent under an equal-rates JC model. Interestingly, maximum likelihood estimation under an equal-rates HKY model also appeared to be converging on 100% correct (not shown), but did so more slowly than analyses with the GTRer model.

When an equal branch-lengths tree (Fig. 1, middle row) was used as the model tree, estimation using all methods was consistent regardless of the degree of rate heterogeneity in the data. Furthermore, the performance of all methods was essentially indistinguishable.

Contrary to the situation in the Felsenstein zone, results of the simulations conducted with the inverse-Felsenstein-zone tree indicated that all methods are consistent (Fig. 1, bottom row). This result conflicts with the conclusions of Siddall (1998), but is in agreement with both theory (Chang, 1996; Rogers, 1997, 2001) and the simulation results of Swofford et al. (2001). Our results extend those of Chang (1996) and Rogers (1997, 2001) in that their proofs apply only when there is a perfect fit between the model and the data. Similarly the simulation results of Swofford et al. (2001) were also restricted to the case of a perfect fit between model and data. In our study, however, likelihood estimation using the most strongly violated models was dramatically more efficient, as was parsimony analysis (Fig. 1, bottom row). In fact, as observed by both Yang (1997) and Siddall (1998), parsimony analysis was the most efficient of all the methods tested under all rate heterogeneity conditions examined. The reason for the strong performance of parsimony under these conditions is that long-branch attraction, which causes inconsistency in the Felsenstein zone, has the opposite effect in the inverse-Felsenstein zone. Underestimation of the actual number of substitutions occurring along the long branches causes the identical character states that resulted from parallel and convergent events to be misinterpreted as synapomorphies. This phenomenon spuriously increases the apparent strength of support for the true tree (Swofford et al., 2001).

The results of our eight-taxon simulations (Fig. 2) are similar to those for the four-taxon

Felsenstein-zone tree. When there is extreme rate heterogeneity in the data, ignoring it in the phylogeny estimation renders both likelihood and parsimony inconsistent. However, as long as among-site rate heterogeneity is accommodated in some fashion, including use of the incorrect JC + I, JC + Γ , HKY + I, and HKY + Γ models (JC models not shown), likelihood estimation is consistent.

The behavior of likelihood estimation of phylogeny under incorrect models of sequence evolution has received increasing attention recently (Yang, 1997; Bruno and Halpern, 1999). Although the robustness of maximum likelihood estimation of phylogeny from DNA sequences to violation of model assumptions has been demonstrated previously (i.e., Yang, 1997), our simulations indicate one particularly important way in which likelihood methods are robust. Clearly, it is important to deal with the observation that nucleotide sites in commonly utilized genes rarely evolve at a uniform rate (Uzzell and Corbin, 1971; Palumbi, 1989; Yang et al., 1994; Sullivan et al., 1995). However, our results suggest that this rate variation among sites must be incorporated, but need not be modeled precisely, for likelihood estimation to be consistent under conditions for which it would be inconsistent if an equal-rates model were inappropriately assumed. This is a critical finding for at least two reasons. First, Sullivan et al. (1999) demonstrated that the parameters of complex models of nucleotide substitution can be very difficult to estimate when using data sets of the size commonly used in molecular phylogenetics. Our results suggest that the uncertainty surrounding parameter estimates need not adversely affect the performance of maximum likelihood phylogenetic analyses. Second, the distribution of rates among sites is influenced by interactions at several levels, including the structural constraints associated with both intramolecular (secondary and tertiary structure) and intermolecular interactions (e.g., rRNA and ribosomal proteins, and binding between enzymes and both substrate and cofactors). Ideally, likelihood models could be developed incorporating such structural and functional issues (although the effectiveness of these models may be compromised by the difficulty in estimating the potentially large number of parameters required). Until these models are designed, implemented, and tested, we find

it encouraging that likelihood estimation can be consistent even in the face of violated assumptions. The match between the rate-heterogeneity model and the data need not be perfect, and using a model that is not perfect will often be better than using methods that superficially appear to be model free. Of course, we base this conclusion on comparison among a class of simple and admittedly unrealistic models, and strong conclusions about the robustness of maximum likelihood estimation to violation of model assumptions must ultimately survive more rigorous tests (e.g., Sanderson and Kim, 2000). The work reported here is a step in that direction. Furthermore, although the models used here are undoubtedly oversimplifications of biological reality, they appear to provide statistically acceptable explanations of actual data sets in at least some instances (e.g., Goldman, 1993; Sullivan et al., 2000).

In the analyses presented above, we have restricted our consideration of parsimony to equally weighted parsimony. In a sense, this is not completely fair, in that parsimony analyses are commonly conducted by using a variety of weighting schemes (including equal weights). Furthermore, simulations have shown that use of appropriate weighting schemes can decrease the zone of inconsistency for parsimony (e.g., Hillis et al., 1994). However, one limitation of parsimony methods that has yet to be overcome is the inability to directly compare the optimality scores (tree lengths) under different weighting schemes. Perhaps some alternative means for choosing among weighting schemes under the parsimony criterion can be found, but we are not optimistic. Although this should not be taken as a rejection of parsimony methods, the possibility of objective, statistically defensible model comparisons under maximum likelihood remains a decided advantage. Likelihoods provide a common currency for comparing model assumptions, as well as tree topologies. Although the details of the best manner for determining which maximum likelihood model to use are complex, we are unconvinced by the arguments of Sanderson and Kim (2000) that this situation renders maximum likelihood estimation of phylogeny futile.

Finally, suggestions such as those made by Sanderson and Kim (2000:820)—that robustness of maximum likelihood estimation to violations in model assumptions indicates

the presence of strong signal likely to be detected by many methods—are imprecise. Differences in the shape of the underlying true tree influence the relative performance of alternative methods of phylogenetic analysis. Although for some tree shapes all methods perform equally (Fig. 1, middle row), for other tree shapes accurate estimation requires some degree of model complexity (Fig. 1, top row). The relationship between the nature of the topology being estimated and the effect of different types of model violations is complex. As is well known, systematic errors affecting equally weighted parsimony can bias the results away from the true tree, resulting in inconsistency (Felsenstein, 1978). However, if long-branch taxa are indeed closely related on the true tree (the inverse-Felsenstein zone), the bias of equally weighted parsimony may actually favor the true tree; parsimony will be not only a consistent estimator of phylogeny but also the most efficient. This is a consequence of the fact that, in misinterpreting convergence as synapomorphy, parsimony overestimates the amount of phylogenetic signal actually present (Swofford et al., 2001). Although others see this as a potentially useful bias (e.g., Yang, 1997; Siddall, 1998), to us a method that more accurately estimates the strength of support for a particular node is preferable to methods that mistake bias as phylogenetic signal. When there is a reasonable choice, we think it far better to use models that are wrong but nevertheless fairly assess the strength of phylogenetic signal across a variety of tree shapes, than to use methods that relax the assumptions necessary for inference but are prone to misrepresentation or overstatement of the strength of support for a phylogenetic conclusion.

ACKNOWLEDGMENTS

We thank John Demboski, Mark Fishbein, Peter Foster, Jeff Good, John Huelsenbeck, Paul Lewis, Sam Rogers, Peter Waddell, and two anonymous reviewers for discussions of maximum-likelihood estimation of phylogeny and providing helpful comments. We received support from National Science Foundation grant DEB-9974124.

REFERENCES

- BERNHAM, K. P., AND D. R. ANDERSON. 1998. Model selection and inference: A practical information-theoretic approach. Springer, New York.

- BONEAU, C. A. 1960. The effect of violations of assumptions underlying the *t* test. *Psych. Bull.* 57:49–64.
- BRUNO, W. J., AND A. L. HALPERN. 1999. Topological bias and inconsistency of maximum likelihood using wrong models. *Mol. Biol. Evol.* 16:564–566.
- CHANG, J. T. 1996. Full reconstruction of Markov models on evolutionary tree: Identifiability and consistency. *Math. Biosci.* 137:51–73.
- DONALDSON, T. S. 1968. Robustness of the *F*-test to errors of both kinds and the correlation between the numerator and denominator of the *F*-ratio. *J. Am. Stat. Assoc.* 63:660–676.
- FARRIS, J. S. 1983. The logical basis of phylogenetic analysis. Pages 7–36 in *Advances in cladistics*, volume 2 (N. I. Platnick and V. A. Funk, eds.). Columbia Univ. Press, New York.
- FARRIS, J. S. 1999. Likelihood and inconsistency. *Cladistics* 15:199–204.
- FELSENSTEIN, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. *Syst. Zool.* 27:401–410.
- FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* 17:368–376.
- GALTIER, N., AND M. GOUY. 1998. Inferring pattern and process: Maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. *Mol. Biol. Evol.* 15:871–879.
- GAUT, B. S., AND P. O. LEWIS. 1995. Success of maximum likelihood phylogeny inference in the four-taxon case. *Mol. Biol. Evol.* 12:152–162.
- GOLDMAN, N. 1993. Statistical tests of models of DNA substitution. *J. Mol. Evol.* 36:182–198.
- GU, X., Y.-X. FU, AND W.-H. LI. 1995. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Mol. Biol. Evol.* 12:546–557.
- HASEGAWA, M., H. KISHINO, AND T. YANO. 1985. Dating the human–ape split by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22:160–174.
- HILLIS, D. M., J. P. HUELSENBECK, AND D. L. SWOFFORD. 1994. Hobgoblin of phylogenetics? *Nature* 369:363–364.
- HUELSENBECK, J. P. 1995. The robustness of two phylogenetic methods: Four-taxon simulations reveal a slight superiority of maximum likelihood over neighbor joining. *Mol. Biol. Evol.* 12:843–849.
- JUKES, T. H., AND C. R. CANTOR. 1969. Evolution of protein molecules. Pages 21–132 in *Mammalian protein metabolism* (H. N. Munro, ed.). Academic Press, New York.
- LANAVÉ, C., G. PREPARATA, C. SACCONI, AND G. SERIO. 1984. A new method for calculating evolutionary substitution rates. *J. Mol. Evol.* 20:86–93.
- NIELSEN, R. 1997. Site-by-site estimation of the rate of substitution and the correlation of rates in mitochondrial DNA. *Syst. Biol.* 46:346–353.
- PALUMBI, S. J. 1989. Rates of molecular evolution and the fraction of sites free to vary. *J. Mol. Evol.* 29:180–187.
- ROGERS, J. S. 1997. On the consistency of maximum likelihood estimation of phylogenetic trees from nucleotide sequences. *Syst. Biol.* 46:354–357.
- ROGERS, J. S. 2001. Maximum likelihood estimation of phylogenetic trees is consistent when rates vary according to invariable sites plus gamma distribution. *Syst. Biol.*
- SANDERSON, M. J., AND J. KIM. 2000. Parametric phylogenetics? *Syst. Biol.* 49:817–829.
- SCHÖNIGER, M., AND A. VON HAESLER. 1994. A stochastic model for the evolution of autocorrelated DNA sequences. *Mol. Phylogenet Evol.* 3:240–247.
- SIDDALL, M. E. 1998. Success of parsimony in the four-taxon case: Long-branch repulsion by likelihood in the Farris zone. *Cladistics* 14:209–220.
- SIDDALL, M. E., AND A. G. KLUGE. 1997. Probabilism and phylogenetic inference. *Cladistics* 13:313–336.
- SULLIVAN, J., E. ARELLANO, AND D. S. ROGERS. 2000. Comparative phylogeography of Mesoamerican highland rodents: Concerted versus independent response to past climatic fluctuations. *Am. Nat.* 155:754–768.
- SULLIVAN, J., K. E. HOLSINGER, AND C. SIMON. 1995. Among-site rate variation and phylogenetic analysis of 12S rRNA data in sigmodontine rodents. *Mol. Biol. Evol.* 12:988–1001.
- SULLIVAN, J., D. L. SWOFFORD, AND G. J. P. NAYLOR. 1999. The effect of taxon sampling on estimating rate heterogeneity parameters of maximum-likelihood models. *Mol. Biol. Evol.* 16:1347–1356.
- SWOFFORD, D. L. 1999. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Ver. 4.0b3a. Sinauer Associates, Sunderland, Massachusetts.
- SWOFFORD, D. L., P. J. WADDELL, J. P. HUELSENBECK, P. G. FOSTER, P. O. LEWIS, AND J. S. ROGERS. 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Syst. Biol.* in press.
- TAVARÉ, S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect. Math. Life Sci.* 17:57–86.
- UZZELL, T., AND K. W. CORBIN. 1971. Fitting discrete probability distributions to evolutionary events. *Science* 172:1089–1096.
- WADDELL, P. 1995. Statistical methods of phylogenetic analysis, including Hadamard conjugations, LogDet transforms, and maximum likelihood. Ph.D. Dissertation, Massey Univ. Palmerston North, New Zealand.
- WADDELL, P., 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.). Clarendon Press, Oxford, England.
- WILGENBUSCH, J., AND K. DE QUEIROZ. 2000. Phylogenetic relationships among the phrynosomatid sand lizards inferred from mitochondrial DNA sequences generated by heterogeneous evolutionary processes. *Syst. Biol.* 49:529–612.
- 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. 1994. Estimating the pattern of nucleotide substitution. *J. Mol. Evol.* 39:105–111.
- YANG, Z. 1997. How often do wrong models produce better phylogenies? *Mol. Biol. Evol.* 14:105–108.
- YANG, Z., N. GOLDMAN, AND A. FRIDAY. 1994. Comparison of models for nucleotide used in maximum-likelihood phylogenetic estimation. *Mol. Biol. Evol.* 11:316–324.