Syst. Biol. 51(5):754–760, 2002
DOI: 10.1080/10635150290102429

Identifiability of Parameters in MCMC Bayesian Inference of Phylogeny

BRUCE RANNALA

Department of Medical Genetics, University of Alberta, Edmonton, Alberta, T6G 2H7, Canada

Abstract.—Methods for Bayesian inference of phylogeny using DNA sequences based on Markov chain Monte Carlo (MCMC) techniques allow the incorporation of arbitrarily complex models of the DNA substitution process, and other aspects of evolution. This has increased the realism of models, potentially improving the accuracy of the methods, and is largely responsible for their recent popularity. Another consequence of the increased complexity of models in Bayesian phylogenetics is that these models have, in several cases, become overparameterized. In such cases, some parameters of the model are not identifiable; different combinations of nonidentifiable parameters lead to the same likelihood, making it impossible to decide among the potential parameter values based on the data. Overparameterized models can also slow the rate of convergence of MCMC algorithms due to large negative correlations among parameters in the posterior probability distribution. Functions of parameters can sometimes be found, in overparameterized models, that are identifiable, and inferences based on these functions are legitimate. Examples are presented of overparameterized models that have been proposed in the context of several Bayesian methods for inferring the relative ages of nodes in a phylogeny when the substitution rate evolves over time. [Bayesian phylogenetic inference; Markov chain Monte Carlo; overparameterization; parameter identifiability.]

An important factor that has delayed the adoption of Bayesian methods in biology and other fields is the mathematical difficulty of many Bayesian calculations. The recent increase in popularity of Bayesian statistical methods in genetics (Shoemaker et al., 1999) and other areas of biology is largely due to advances in computing power that have allowed numerical techniques such as Monte Carlo simulation to be implemented to perform Bayesian analysis using complex models. In the past, Bayesian inference was largely limited to simple models for which analytical results were available; the choice of a model was too often based on mathematical convenience, and biologists have been justified in their scepticism of such methods. With the advent of new numerical techniques for evaluating Bayesian equations, powerful computers to carry out the calculations, and flexible programming languages, model choice in Bayesian analysis has become less arbitrary, and Bayesian techniques for statistical inference are becoming increasingly accepted among biologists.

The development of numerical techniques for generating posterior distributions for models of arbitrary complexity, beginning in the 1950s, presented the prospect that Bayesian analysis could be carried out for scientific problems in which prior distributions and likelihoods were chosen as best suited to the problem at hand. The most popular approaches, Markov chain Monte Carlo (MCMC) methods, have used the Metropolis–Hastings algorithm (Metropolis et al., 1953; Hastings, 1970) and its many variants (reviewed by Gilks et al., 1996). The basic principle underlying MCMC methods is that a Markov chain can be constructed with a stationary distribution that is the joint posterior probability distribution of the parameters of the model. The parameters are assigned arbitrary initial values, and the chain is simulated until it appears to converge to the stationary distribution. Observations from the chain at stationarity are used to estimate the joint posterior probabilities of the parameters.

The posterior probability distribution of parameters, obtained by applying the Bayes theorem, forms the foundation for Bayesian inference. Let $\mathbf{X} \in \Omega$ be a vector of observed random variables (the data) taking values on a state space Ω, which defines the set of possible data configurations, and let $\theta \in \Theta$ be a vector of one or more parameters that completely specify the form of the probability distribution of $\mathbf{X}$ on Ω. The posterior probability distribution of the parameters θ, given $\mathbf{X}$, is

$$f(\theta \mid \mathbf{X}) = \frac{f(\mathbf{X} \mid \theta) g(\theta)}{\int_{\theta \in \Theta} f(\mathbf{X} \mid \theta) g(\theta)\, d\theta}, \qquad (1)$$

where $g(\theta)$ is the prior probability distribution of θ (i.e., the probability distribution before examination of the data) and $f(\boldsymbol{X} \mid \theta)$ is the probability of the data given the parameters (the likelihood). Point estimates and confidence intervals for θ can be obtained from the posterior distribution in various ways. The mode, or mean, of the posterior distribution is often used as a point estimate of θ, and the α percent credible set is used as a confidence interval, containing the "true" parameter value with probability α.

An attractive feature of MCMC methods is that each iteration of the Markov chain requires only that ratios of the likelihood function (and possibly the priors) be calculated, eliminating the need to evaluate the denominator of Equation 1, which is often a higher dimensional integral, or sum, for complex models. Moreover, marginal distributions of parameters are easily obtained by monitoring the values of particular parameters in the chain at stationarity, again avoiding the need to evaluate integrals or sums. The ease with which the number of parameters may be expanded in a Bayesian model (with no apparent cost) carries some risk. It is possible to overparameterize a model such that it is not identifiable, meaning that the model leads to sample configuration probabilities identical to those of a simpler model with fewer parameters. Overparameterization will increase the importance of the prior; even with an infinite amount of data, the prior will continue to influence the posterior distribution. Overparameterization may also lead to improper posterior distributions (i.e., the posterior distribution does not satisfy the laws of probability) in cases where an improper prior is used and possibly to poor convergence of the MCMC algorithm (see Carlin and Louis, 1996).

Several authors have recently proposed MCMC methods for Bayesian phylogenetic inference (Yang and Rannala, 1997; Mau et al., 1999). Problems of identifiability and overparameterization can arise in the context of MCMC Bayesian phylogenetic analysis and are the focus here. To set the context, I provide a brief overview of standard theory relating to identifiability of parameters. A more extensive discussion of these concepts was presented by Casella and Berger (1990). I then present a simple model that is not identifiable to study the effect of nonidentifiability on the posterior density of parameters and the rate of convergence of a MCMC algorithm in this case. To illustrate the potential for identifiability problems in Bayesian MCMC phylogenetic inference, I discuss recent approaches that are aimed at modeling molecular evolution when the rate of nucleotide substitution varies over time (Thorne et al., 1998; Huelsenbeck et al., 2000; Kishino et al., 2001).

The examples presented, i.e., the variable clock models of Thorne et al. (1998), Huelsenbeck et al. (2000), and Kishino et al. (2001), were chosen merely to illustrate how the complex models used in Bayesian phylogenetic inference can become overparameterized. There is no indication that overparameterization causes any problems for phylogenetic inference in the cases presented, and my observation that the models are overparameterized should not be perceived as a criticism. Overparameterization is a property of the likelihood, and so similar problems may arise in maximum likelihood analysis. Nonidentifiable parameters are more likely to be recognized in a likelihood analysis, however, because they will result in multiple maxima for the likelihood of the nonidentifiable parameters and will therefore present problems for maximization algorithms. One solution for an overparameterized model (under either a likelihood or a Bayesian approach) is to find a function of the parameters that is identifiable and to estimate the value of this function instead. A potential advantage of Bayesian analysis over likelihood is that if an informative prior is available proper inferences can be obtained despite the fact that a model is overparameterized.

Identifiability of Parameters

Here, I define the concept of identifiability and give an example of a model with parameters that are not identifiable. Following earlier notation, let $\mathbf{X}$ be a vector of observed random variables, where $\mathbf{X} \in \Omega$. Define f to be a probability distribution function for a model completely specified by parameters θ. If there exists some $\theta_1 \neq \theta_2$ satisfying

$$f(\mathbf{X} \mid \theta_1) = f(\mathbf{X} \mid \theta_2),$$

for all $\mathbf{X} \in \Omega$, then the parameters of the model are not identifiable, i.e., all possible sets of observations have identical

probabilities for two different sets of parameters. Identifiability is a problem of model specification rather than one of inference, but inference problems can arise because of misspecified models. From a Bayesian perspective, nonidentifiability of parameters may also be manifest as a strong correlation among parameters in the posterior density, despite the fact that the parameters are independent under the prior density.

If an informative prior is used for one or more of the nonidentifiable parameters, legitimate Bayesian inference may still be possible. For example, if $\theta_1 = \{\alpha_1^{(1)}, \alpha_1^{(2)}\}$ and $\theta_2 = \{\alpha_2^{(1)}, \alpha_2^{(2)}\}$ and the likelihood is a function of $\alpha^{(1)} + \alpha^{(2)}$ only, then the α parameters are not separately identifiable. However, if the prior for $\alpha^{(2)}$ specifies that $\alpha^{(2)} = x$ with probability 1, then $\theta_1 = \theta_2$ if and only if $\alpha_1^{(1)} = \alpha_2^{(1)}$, and the model becomes identifiable. If an informative prior is available, then Bayesian inference is possible, even in cases where the model is not identifiable (from the perspective of the likelihood).

NONIDENTIFIABILITY: A SIMPLE EXAMPLE

In the following example, there are two simple models, one that is overparameterized and another that is not. The intention is to illustrate the concept of nonidentifiability in a simple yet concrete example and to study the effect on the posterior density of the parameters and the convergence of an MCMC algorithm. Let $\mathbf{X} = \{X_j\}$, where the X_j are independent identically distributed (i.i.d.) random variables, each exponentially distributed with parameter λ. To simplify notation:

$$\Delta = \sum_{j=1}^{n} X_j.$$

The probability density of the data given λ (the likelihood of λ) is

$$f_1(\mathbf{X} \mid \lambda) = \lambda^n e^{-\lambda\Delta}.$$

In model 1, let the prior density of the parameter λ be a gamma density with parameters k and β:

$$g_1(\lambda \mid \beta, k) = \frac{\beta^k \lambda^{k-1} e^{-\lambda\beta}}{\Gamma(k)}.$$

The marginal probability of the data is

$$f_1(\mathbf{X} \mid \beta, k) = \int_0^\infty \frac{\beta^k \lambda^{n+k-1}}{\Gamma(k)} e^{-\lambda(\Delta+\beta)}\, d\lambda = \frac{\Gamma(n+k)\beta^k}{\Gamma(k)(\Delta+\beta)^{n+k}},$$

and the posterior density of λ given $\mathbf{X}$ is

$$f_1(\lambda \mid \mathbf{X}, \beta, k) = \frac{(\Delta+\beta)^{n-k}\, \lambda^{n+k-1} e^{-\lambda(\Delta+\beta)}}{\Gamma(n+k)},$$

which is a gamma density with parameters $n + k$ and $\Delta + \beta$.

In model 2, the prior density of the rate parameter for the n exponential random variables is determined by a sum, $\Lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_k$, of k i.i.d. exponential random variables with parameter β. The $k - 1$ additional parameters in this model are not identifiable because an uncountably infinite number of combinations of $\lambda_1, \lambda_2, \ldots, \lambda_k$ will result in the same value of Λ and thus the same probability density of the data (likelihood). The probability of the data $\mathbf{X}$, given Λ, is

$$f_2(\mathbf{X} \mid \Lambda) = \Lambda^n e^{-\Lambda\Delta},$$

and the prior probability of λ_i is

$$g_2(\lambda_i \mid \beta) = \beta e^{-\beta\lambda_i}.$$

The marginal probability of the data is

$$\begin{aligned} f_2(\mathbf{X} \mid \beta, k) &= \int_0^\infty \cdots \int_0^\infty \left(\sum_{i=1}^{k} \lambda_i\right)^n \beta^k \\ &\quad \times \exp\left[-\sum_{i=1}^{k} \lambda_i (\Delta+\beta)\right] d\lambda_1 \cdots d\lambda_k \\ &= \frac{\Gamma(n+k)\beta^k}{\Gamma(k)(\Delta+\beta)^{n+k}}, \end{aligned}$$

which is identical to the marginal probability of the data under model 1. The sum of k i.i.d. exponential random variables with parameter β is distributed as a gamma (k, β) density, and this is the prior chosen for model 1. The joint posterior density of the parameters

under model 2 is

$$f_2(\lambda_1, \lambda_2, \ldots, \lambda_k \mid \mathbf{X}, \beta, k) = \frac{\Gamma(k)}{\Gamma(n+k)} \Lambda^n (\Delta + \beta)^{n+k} e^{-\Lambda(\Delta+\beta)}.$$

The two models can be used to study the effect of overparameterization on the posterior probability density of Λ and the rate of convergence of MCMC. Model 2 has $k-1$ additional parameters by comparison with model 1, and only the sum of the parameters is identifiable. The effect of arbitrary degrees of overparameterization can be studied for model 2 by simply modifying k.

In the simple case that there is one additional parameter under the second model ($k = 2$), the expected correlation, ρ_{12}, of the parameters λ_1 and λ_2 in the posterior density is

$$\rho_{12} = -\frac{n}{n+6}.$$

The parameters are negatively correlated in the posterior density, and this correlation increases with increasing sample size, ultimately tending to -1. This result is intuitive because with few observations there is little information available about the parameters; the prior, which models λ_1 and λ_2 as i.i.d., then dominates. With increased data, the likelihood dominates and the nonidentifiability is manifest as a strong negative correlation between the variables. For complex models, it may be impossible to recognize overparameterized models a priori by analytical analysis. An alternative approach to detect overparameterization might be to empirically estimate the correlation among parameters by jointly sampling parameter values from the chain in an MCMC analysis and calculating a correlation coefficent based on these samples. Although a very high degree of correlation among parameters in the posterior density (especially when they are independent under the prior) may provide a useful indicator of overparameterization, a low degree of correlation will be less informative because it could also occur simply because of uninformative data and consequentially a dominance of the posterior by the prior. Often, parameters are correlated in the posterior distribution, although still identifiable, and this criterion should be used only as a rough guide for detecting potential cases of overparameterization and will be most useful when very large sample sizes are available.

Overparameterization and MCMC Convergence

Overparameterization may retard convergence in MCMC because of the resulting strong negative correlations among parameters in the posterior probability density. The effect of overparameterization on the rate of convergence of MCMC can be studied using the two models developed above. For model 1, the log-likelihood to be evaluated in the MCMC algorithm is

$$\log L_1(\lambda) = n \log(\lambda) - \lambda\Delta.$$

For model 2, the log-likelihood to be evaluated in the MCMC algorithm is

$$\log L_2(\Lambda) = n \log(\Lambda) - \Lambda\Delta.$$

At each iteration of the chain, new values are proposed for either λ (model 1) or successively $\lambda_1, \lambda_2, \ldots, \lambda_k$ (model 2). For both models, potential parameter values are proposed by adding a uniform random variable, δ, chosen on an appropriate interval $(-D, D)$, to the current value of the parameter (reflecting negative parameter values back onto the positive axis). Under model 1, a proposed change will then be $\lambda' = \lambda + \delta$, and under model 2 (at step i) a proposed change will be $\lambda_i' = \lambda_i + \delta$. But the proposed change under model 2 is equivalent to $\Lambda' = \lambda_1 + \cdots + \lambda_i + \delta + \cdots + \lambda_k = \Lambda + \delta$. Ignoring the fact that the proposed values may need to be reflected back more often under model 2, the two MCMC algorithms are essentially identical regardless of the number of additional parameters in model 2, and the rate of convergence is not affected by overparameterization in this example (MCMC programs were also written to generate the posterior density under the two models, and no difference in the rate of convergence was observed).

Models of a Variable Molecular Clock

Parametric maximum likelihood and Bayesian methods of phylogenetic inference typically model the process of DNA substitution as a continuous-time Markov process.

If the rate of substitution is a constant, μ, substitutions occur at the mth nucleotide site in the lineage separating descendent species i from its ancestor, according to a homogeneous Poisson process. If v_i is the time separating the ancestral and descendent species, the expected number of substitutions is μv_i. Gillespie (1984) and others have considered a process in which the rate of DNA substitution may change over time. Define the rate of substitution in lineage i at time t to be $\mu_i(t)$. Substitutions now occur according to a nonhomogenous Poisson process, and the expected number of substitutions is

$$\int_0^{v_i} \mu_i(t)\, dt. \qquad (2)$$

In both the constant and variable rate substitution models, the total number of substitutions at a site in a descendent will follow a Poisson distribution. The parameter of the probability distribution is either μv_i (constant rate model) or the result obtained by evaluating the integral in Equation 2.

The form of the distribution of substitutions is Poisson under either a constant or variable substitution model, and the models can therefore not be distinguished because one can always choose a constant rate model (with rate μ_0) that will give an identical probability distribution to the variable rate model if μ_0 is chosen to satisfy Equation 2. Among-lineage and among-site rate variation can be identified. Among-lineage rate variation would be implied by the observation that $\mu_i v_i \neq \mu_j v_j$ when $v_i = v_j$, suggesting that $\mu_i \neq \mu_j$ (define μ_i as the rate of substitution in the ith lineage, etc). This is the basis for a test of the molecular clock. Among-site rate variation would be implied by the observation $\mu_{il} v_i \neq \mu_{im} v_i$, where μ_{il} is the rate of substitution at site l in lineage i, etc.

Several authors have recently proposed methods for carrying out phylogenetic inference that allow the rate of substitution to evolve over time (Sanderson, 1997; Thorne et al., 1998; Huelsenbeck et al., 2000; Kishino et al., 2001). In these models, the rate of evolution of the substitution rate determines the amount of information available for inferring the ages of nodes in the tree. With a rapidly evolving substitution rate, little information is preserved about node ages, and there are few constraints among branches in terms of their average substitution rates. If rates evolve very slowly the constraints of rates among branches are closer to those assumed under a strict molecular clock. These methods are very appealing because they can allow information to be extracted about node ages even though rates are not perfectly constant. Here, two Bayesian methods are studied that have recently been proposed (Huelsenbeck et al., 2000; Kishino et al., 2001) for inferring the relative ages of nodes in the presence of an evolving substitution rate. These methods, which both use MCMC methods to estimate node ages, provide instructive examples of overparameterized models.

The only parameters that are identifiable in evolving rate models are the average rates of substitution on branches. An evolving rate of substitution potentially generates a greater correlation between the mean substitution rates on branches that are closer to one another on a phylogeny. Kishino et al. (2001) considered a very simple model in which the average substitution rate on a branch is the average of the rates at the ancestral and descendent nodes. The logarithm of the substitution rate of a descendent node is normally distributed with a mean chosen such that the expectation of the substitution rate in the descendent is equal to the rate of its ancestor (i.e., the process of rate evolution is unbiased). The variance of the (normal) distribution of the substitution rate in descendents determines the rate of evolution of the substitution rate; greater variance results in greater changes in rates across the phylogeny and less correlation in substitution rates among branches. Because there are $2s - 1$ ancestral nodes that are assigned rates in this model but only $2s - 2$ identifiable substitution rate parameters (e.g., the mean substitution rate for each of the $2s - 2$ branches), the model is overparameterized. This problem was recognized by Kishino et al. and was dealt with by constraining one branch to have a descendent rate equal to the ancestral rate, reducing the number of rates at nodes to $2s - 2$.

A more complex model of rate variation was used by Huelsenbeck et al. (2000), who considered a model in which substitution rates change at discrete times over a phylogeny. For a given tree, they defined ξ as the number of rate change events, $\mathbf{z} = \{z_1, z_2, \ldots, z_\xi\}$ as the positions of rate change

events on the tree, and $\mathbf{r} = \{r_1, r_2, \ldots, r_\xi\}$ as a set of rate multipliers. At the ith rate change event, the current rate is multiplied by r_i to obtain a new rate. Various priors are imposed for ξ, $\mathbf{z}$, and $\mathbf{r}$, but in this case no additional information is available for constructing the prior densities on the rate parameters (it seems unlikely that any such information will be available in future either), so the likelihood is the relevant term in assessing whether parameters are identifiable. For simplicity, I focus on a single branch of the phylogeny (of length T), letting ξ be the number of rate changes on the branch, $\mathbf{z}$ the points in time at which rate changes occur on the branch, etc. Let μ_0 be the substitution rate in the immediate ancestor of this lineage. The substitution rate on interval (z_i, z_{i+1}) is

$$\mu_0 \prod_{j=0}^{i} r_j,$$

and the mean rate on the branch is

$$M(\mathbf{r}, \mathbf{z}) = \mu_0 \sum_{i=0}^{\xi} \left[(z_{i+1} - z_i) \prod_{j=0}^{i} r_j \right],$$

where $r_0 = 1$, $z_0 = 0$, and $z_{\xi+1} = T$. The parameters $\mathbf{r}$ and $\mathbf{z}$ are not identifiable. Assume that $\xi = 2$, then

$$M(\mathbf{r}, \mathbf{z}) = \mu_0 \left[(T - z_2) r_2 r_1 + (z_2 - z_1) r_1 + z_1 \right].$$

To prove nonidentifiability of these parameters, we need only show that some $\mathbf{r}' \neq \mathbf{r}$ and $\mathbf{z}' \neq \mathbf{z}$ exist satisfying

$$M(\mathbf{r}', \mathbf{z}') = M(\mathbf{r}, \mathbf{z}).$$

The average substitution rate M on a branch is the only factor influencing the likelihood; parameter changes that do not alter M do not alter the likelihood. If

$$r_2' = r_2 + \frac{\delta}{(T - z_2) r_1}$$

and

$$z_1' = z_1 - \frac{\delta}{1 - r_1}$$

for any δ satisfying $r_2' > 0$ and $z_2 > z_1' > 0$, a new combination $\mathbf{z}'$ and $\mathbf{r}'$ results that corresponds to the same value of M. An uncountably infinite set of parameter values exist that yield the same value of the likelihood. If inference is focused on the posterior density of M, the mean substitution rate for each branch, rather than $\mathbf{z}$ or $\mathbf{r}$, the fact that the model is overparameterized in this case need not be a concern unless overparameterization leads to convergence problems. In making inferences about $\mathbf{z}$ or $\mathbf{r}$, however, only a specific function of these parameters, M, is influenced by the data.

DISCUSSION

The ease with which complex models can be incorporated into a phylogenetic analysis using Bayesian MCMC methods can lead to overparameterization. Parameters that are nonidentifiable are influenced by the data only through certain functions of the parameters that are identifiable; certain aspects of their posterior distributions are therefore insensitive to the data and can unduly increase the influence of the prior. In such cases, the posterior density will be nontrivial, even with an infinite amount of data, because the prior continues to influence the posterior density.

In simple cases, identifiable functions of the parameters of an overparameterized model can be determined and can be effectively studied by MCMC through use of an overparameterized model. Overly complex models with many nonidentifiable parameters may lead to large correlations among parameters in the posterior density, possibly retarding convergence of an MCMC algorithm, although this is not always the case as was found for the simple example presented here. Determining the robustness of the posterior density to the prior can be helpful in identifying overparameterized models. In such cases, the posterior density will remain sensitive to the form of the prior, despite an increase in the amount of data. Careful analysis and monitoring of correlations among parameters in the posterior density when carrying out MCMC can also provide a tool for identifying inference problems arising because of overparameterization, although correlations among parameters in the posterior distribution also commonly occur in models that are not overparameterized. If simpler models are available, model fitting may help by reducing the need for overly complex models.

ACKNOWLEDGMENTS

I thank Paul Lewis and two anonymous reviewers for helpful comments. This research was supported by NIH grant HG01988.

REFERENCES

CARLIN, B. P., AND T. A. LOUIS. 1996. Bayes and empirical Bayes methods for data analysis. Chapman and Hall, New York.

CASELLA, G., AND R. BERGER. 1990. Statistical inference. Duxbury Press, Belmont, California.

GILKS, W. R., S. RICHARDSON, AND D. J. SPIEGELHALTER . 1996. Introducing Markov chain Monte Carlo. Pages 1–16 *in* Markov chain Monte Carlo in practice (W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, eds.). Chapman and Hall, New York.

GILLESPIE, J. H. 1984. The molecular clock may be an episodic clock. Proc. Natl. Acad. Sci. USA 81:8009–8013.

HASTINGS, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–101.

HUELSENBECK, J. P., B. LARGET, AND D. SWOFFORD. 2000. A compound Poisson process for relaxing the molecular clock. Genetics 154:1879–1892.

KISHINO, H., J. L. THORNE, AND W. J. BRUNO. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Mol. Biol. Evol. 18:352–361.

MAU, B., M. A. NEWTON, AND B. LARGET. 1999. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. Biometrics 55:1–12.

METROPOLIS, N., A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER, AND E. TELLER. 1953. Equations of state calculation by fast computing machine. J. Chem. Phys. 21:1087–1091.

SANDERSON, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14:1218–1232.

SHOEMAKER, J. S., I. S. PAINTER, AND B. S. WEIR. 1999. Bayesian statistics in genetics: A guide for the uninitiated. Trends Genet. 15:354–358.

THORNE, J. L., H. KISHINO, AND I. S. PAINTER. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15:1647–1657.

YANG, Z., AND B. RANNALA. 1997. Bayesian phylogenetic inference using DNA sequences: A Markov chain Monte Carlo method. Mol. Biol. Evol. 14:717–724.

First submitted 29 September 2001; reviews returned 11 December 2001; final acceptance 5 April 2002
Associate Editor: Rasmus Neilsen