Syst. Biol. 51(5):703–714, 2002
DOI: 10.1080/10635150290102375

Effects of Models of Rate Evolution on Estimation of Divergence Dates with Special Reference to the Metazoan 18S Ribosomal RNA Phylogeny

STÉPHANE ARIS-BROSOU AND ZIHENG YANG

Department of Biology, University College London, Darwin Building, Gower Street, London WC1E 6BT, England

Abstract.—The molecular clock, i.e., constancy of the rate of evolution over time, is commonly assumed in estimating divergence dates. However, this assumption is often violated and has drastic effects on date estimation. Recently, a number of attempts have been made to relax the clock assumption. One approach is to use maximum likelihood, which assigns rates to branches and allows the estimation of both rates and times. An alternative is the Bayes approach, which models the change of the rate over time. A number of models of rate change have been proposed. We have extended and evaluated models of rate evolution, i.e., the lognormal and its recent variant, along with the gamma, the exponential, and the Ornstein–Uhlenbeck processes. These models were first applied to a small hominoid data set, where an empirical Bayes approach was used to estimate the hyperparameters that measure the amount of rate variation. Estimation of divergence times was sensitive to these hyperparameters, especially when the assumed model is close to the clock assumption. The rate and date estimates varied little from model to model, although the posterior Bayes factor indicated the Ornstein–Uhlenbeck process outperformed the other models. To demonstrate the importance of allowing for rate change across lineages, this general approach was used to analyze a larger data set consisting of the 18S ribosomal RNA gene of 39 metazoan species. We obtained date estimates consistent with paleontological records, the deepest split within the group being about 560 million years ago. Estimates of the rates were in accordance with the Cambrian explosion hypothesis and suggested some more recent lineage-specific bursts of evolution. [18S rRNA; local molecular clocks; Markov chain Monte Carlo; Metazoa; Metropolis–Hastings algorithm; molecular clock; Ornstein–Uhlenbeck process; phylogeny; posterior Bayes factor; rate of evolution.]

Since it was proposed by Zuckerkandl and Pauling (1965) almost four decades ago, the molecular clock hypothesis, i.e., the constancy of evolutionary rate over time, has been a matter of debate (e.g., Gillespie, 1991). A number of tests have been developed to examine its validity, such as the relative rate test (Sarich and Wilson, 1967; Wu and Li, 1985) and the likelihood ratio test (Felsenstein, 1988). These tests often reject the molecular clock in real data sets (see Nei and Kumar, 2000:188). When the evolutionary rate is not constant across lineages, it is interesting to know whether it fluctuates at random or evolves following some specific trends.

Violation of the molecular clock assumption has also caused seriously biased estimates of divergence dates (Yoder and Yang, 2000; Soltis et al., 2002). Recently, several investigators have attempted to relax the molecular clock assumption when estimating divergence times. One approach is to construct local molecular clock models in the likelihood framework (Rambaut and Bromham, 1998; Yoder and Yang, 2000), where independent evolutionary rates are assigned to some lineages while all the other branches evolve at the same rate. Dates and rates are then parameters in the model and are estimated by maximum likelihood (ML). This approach is straightforward to apply if the branches with different rates can easily be identified a priori. However, when such information is unavailable, date estimates might be sensitive to the assumptions about the rates.

Another approach is to use a stochastic process to describe evolutionary rate change over lineages, relying on the observation that closely related lineages tend to have similar rates (Sanderson, 1997). A Bayesian approach is then used to derive the posterior distributions of rates and dates. One such model assumes that rates are autocorrelated across speciation events; the rate of a branch is sampled from a lognormal distribution centered on the rate of the ancestral branch (Thorne et al., 1998). Other models of rate evolution have also been suggested. Following Gillespie (1991), Bickel (2000) and Cutler (2000) proposed a model based on a doubly stochastic Poisson process (Cox process), which extends the constant-rate

Poisson process first described to model the accumulation of substitutions since divergence (Zuckerkandl and Pauling, 1965). Huelsenbeck et al. (2000) modeled the evolutionary rate as a point process, assuming that the rate of evolution changes according to a Poisson process along the tree, and the rate parameter of the Poisson has a prior distribution.

There seems to be some arbitrariness in the choice of the model of rate evolution. Thus, it is important to know how sensitive the estimates of divergence times are to the choice of the model of rate change. In this study, we implemented and compared different models of autocorrelated rate change over time, focussing on two points: the effect of the model of rate change and the effect of the parameterization of each model to relax the clock. We also compared these Bayesian methods with the likelihood-based local clock analysis. We used the hominoid tRNA gene (Horai et al., 1992) and the metazoan 18S ribosomal RNA (rRNA) gene (Bromham et al., 1998) as test data sets.

The Bayesian Approach

In the framework of ML, the most general model assumes that the substitution rate r_i for branch i is allowed to vary among branches. The branch length is given by the product of the rate and the time duration for that branch, $b_i = r_i t_i$. The likelihood, i.e., the probability of observing the data X, depends on the vector of branch lengths B and is denoted $p(X \mid B)$. Branch lengths can be estimated using classical hill-climbing algorithms to maximize the likelihood (see Gill et al., 1981). Because rate and time are confounded, it is not possible to estimate one without making assumptions regarding the other. For instance, the molecular clock hypothesis assumes that all rates are equal; branch lengths are then proportional to divergence times, and the problem reduces to ML estimation. Models of local clocks are similar; some prespecified branches are assigned independent rate parameters while all other branches have the same rate.

To relax the molecular clock in a Bayesian framework, we assign a prior distribution $p(R, T)$ for rates of evolution R and divergence times T. The Bayes theorem is then used to derive the (posterior) probability of times and rates:

$$p(R, T \mid X) = \frac{p(X \mid R, T)p(R, T)}{p(X)}. \quad (1)$$

A sensible way to factorize the joint prior distribution is $p(R, T) = p(R \mid T)p(T)$. We therefore assume that speciation events are generated by a random process independent of the rates of molecular evolution and that the rate for a given branch is dependent on the time duration of that branch. Moreover, if the prior for the rates is independent of divergence times, we have $p(R \mid T)p(T) = p(R)p(T)$.

The probability $p(X \mid R, T)$ is the traditional likelihood, and its calculation requires a nucleotide substitution model. Here, we use the HKY85 model (Hasegawa et al., 1985) incorporating among-site rate variation modeled by a gamma distribution (Yang, 1994). The parameters in the substitution model are $\psi = \{\kappa, \alpha, \pi\}$, where κ is the transition:transversion rate ratio, α is the shape parameter of the gamma distribution, and π is the vector of the base frequencies. We have also extended this model to take into account heterogeneous site partitions in the sequence (e.g., the three codon positions of a gene). Usually ψ is assumed to follow a uniform prior distribution with its components mutually independent and independent of R and T. The prior distribution of the complete model is then $p(R \mid T)p(T)p(\kappa)p(\alpha)p(\pi)$. In this paper, we will concentrate on prior models for times and rates, with ψ set to its ML estimates (MLEs) obtained without the clock.

Prior Distribution for Divergence Times

The prior distribution for divergence times is generated by a process of cladogenesis, the generalized birth and death process (BDP) with species sampling, as described by Yang and Rannala (1997) (see also Kendall, 1948; Thompson, 1975; Nee et al., 1994). The model assumes a constant speciation rate λ and an extinction rate μ per lineage. Node times are conditioned on the time of the root, arbitrarily set to 1. Species sampling is modeled as a mass extinction event occurring at the sampling time with a probability ρ. This process is flexible and can accommodate more shapes of trees than the Yule process as used by Thorne et al. (1998) but seems comparable

with the generalized Dirichlet distribution of Kishino et al. (2001). To accommodate the uncertainty in the hyperparameters (λ, μ, and ρ), they are integrated out of the model by a standard Bayes averaging method. Independent uniform distributions were used as priors for λ, μ, and ρ.

Lower and upper bounds on node times can be discerned from fossil dates. This approach is expected to improve convergence of the algorithm, because the times for the constrained nodes do not have to explore the whole sample space. This process has been implemented recently by Kishino et al. (2001), but as pointed out by those authors, no closed-form prior under such a constraint has been suggested. As a result, we have not incorporated this feature in our implementation.

Prior Distributions for Rates of Evolution

Each branch in the tree has a rate, which is the mean rate over the time period covered by the branch. If the time period between two speciation events is short, we may not expect the rate of evolution to change dramatically. However, the longer this period, the more likely the rate changes. Thus it is natural to model the rate for a descendent branch as if it were drawn from a distribution with the mean as the ancestor's rate and the variance increasing as the time along the branch increases. We use σ^2 as a measure of how the variance in the rate increases as a function of time. The model tends to the clock for small σ^2 and represents highly variable rates when σ^2 is large. The prior model does not assume a systematic trend in the rate, either upward or downward. Several different distributions are implemented, as described below.

Lognormal Prior Distribution and Its Stationary Variant

We first briefly review two models of rate change developed by Thorne et al. (1998) and Kishino et al. (2001). The consideration that rates are positive motivated the choice of the lognormal distribution instead of the normal distribution (Thorne et al., 1998). Let r_i be the rate of branch i, r_A the rate of the ancestral branch of i, and $\varphi(r_i, r_A, s^2)$ the Gaussian density $\exp[-(r_i - r_A)^2/(2s^2)]/\sqrt{2\pi s^2}$. In the former implementation (Thorne et al., 1998), the rate r_i follows the lognormal distribution $\varphi(r_i, r_A, s^2)/r_i$ and has two parameters: r_A is the rate of the ancestor, and s^2 is a variance parameter that controls how much the model is constrained by the clock. This lognormal model is hereafter referred to as LND. If the time period between two speciation events is short, it is natural to think that the rate of evolution of a given gene may not change dramatically. However, the longer this period, the more likely the rate changes. Therefore, s^2 was assumed to be proportional to this time period, Δt, with $s^2 = \sigma^2 \Delta t$. Parameter σ^2 measures the departure from the strict clock assumption; the model tends to the molecular clock for small σ^2 and represents highly variable rates when σ^2 is large. The time duration Δt was measured by the difference between the two midpoints of the current and ancestral branches by Thorne et al. (1998) and by the time duration of the current branch by Kishino et al. (2001). Here, we adopt the implementation of Thorne et al. (1998).

The mean of the lognormal distribution is not the ancestral rate r_A but $r_A e^{s^2/2}$. The rate of evolution therefore exhibits an upward trend, so the process is time dependent. A remedy to this problem, proposed by Kishino et al. (2001), is to subtract $s^2/2$ from the logarithm of the ancestral rate so that the probability density function (pdf) becomes $\varphi(r_i, r_A e^{s^2/2}, s^2)/r_i$. We refer to this modified distribution as the stationary lognormal distribution (SLD).

To reduce the computational demand, Thorne et al. (1998) and Kishino et al. (2001) used MLEs of branch lengths $\hat{B}$ as pseudodata, approximating the likelihood function by a multivariate normal distribution centered on $\hat{B}$. Our implementation of the LND and SLD models is similar to those of Thorne et al. (1998) and Kishino et al. (2001), but we adopted an exact and more expensive likelihood computation using the sequence alignment.

Gamma and Exponential Distributions

We also implemented two simple models of rate change: the gamma and exponential distributions (GD and ED, respectively). The rate of a branch is assumed to be drawn from a GD or an ED, with the mean rate equal to the rate of the ancestral branch. As with the models discussed above, the variance of the GD

is set proportional to Δt, the time duration of the considered branch, whereas with the ED, the variance is a function of the mean only. Therefore, these two models are not nested, and ED implicitly assumes that the larger the rate, the more variable it is.

Ornstein–Uhlenbeck Process

Another model we implemented allows the rate to evolve according to the Ornstein–Uhlenbeck process (OUP), a continuous-time Gaussian Markov process. OUP was originally designed to model the speed of a particle (not just its position, as in the Brownian process) as a function of time. The speed of the particle is reduced by frictional resistance from the medium and altered by random collisions with neighboring particles. According to the process, the pdf of rate r_i is $\varphi[r_i, r_A e^{-\beta\Delta t}, \sigma^2(1 - e^{-2\beta\Delta t})/(2\beta)]$ (e.g., Karlin and Taylor, 1981:170–173). The mean of the distribution is now $r_A e^{-\beta\Delta t}$, which tends to the ancestral rate r_A as β or Δt goes to zero. As before, σ^2 is the parameter measuring departure from the molecular clock; the variance of the distribution, $\sigma^2(1 - e^{-2\beta\Delta t})/(2\beta)$, tends to $\sigma^2\Delta t$ for small β or Δt.

The simplest model of rate change is when all the branches of the tree have the same rate. This model is essentially the Bayesian version of the molecular clock hypothesis, the only difference with the traditional clock being the prior distribution for the speciation times.

Posterior Distribution and Its Approximation

In a Bayesian framework, the marginal posterior distribution of a variable is obtained by integrating out other variables. For example, the marginal posterior distribution of the times T is derived from Equation 1 by integrating $p(R, T \mid X)$ over the rates, and the hyperparameters:

$$p(T \mid X) = \int p(X \mid B)\, p(R \mid T, \sigma^2)\, p(T \mid \lambda, \mu, \rho) \times p(\lambda)\, p(\mu)\, p(\rho)\, dR\, d\lambda\, d\mu\, d\rho / p(X). \quad (2)$$

Equation 2 can be further simplified for the ED prior for rates because $p(R \mid T) = p(R)$. Because one of the objectives of this study is to examine the effect of the molecular clock assumption on time estimates, σ^2 is not integrated out, but is estimated from the data.

In general, it is very expensive to calculate the normalizing constant $p(X)$ or the integral (Eq. 2) Markov chain Monte Carlo (MCMC) (e.g., Gilks et al., 1996) is employed to approximate the (marginal) posterior distributions. At each step of the chain, a new state $\theta^* = \{R^*, T^*, \psi^*\}$ is proposed to change parameters from a proposal distribution Q, which is assumed to be a normal distribution with the mean centered at the current state θ. The variance of the normal distribution is a tuning parameter. The new state θ^* is accepted with probability h (Metropolis et al., 1953; Hastings, 1970):

$$\begin{aligned} h &= \min\left[1, \frac{p(\theta^* \mid X) \cdot Q(\theta^* \to \theta)}{p(\theta \mid X) \cdot Q(\theta \to \theta^*)}\right] \\ &= \min\left[1, \frac{p(\theta^* \mid X)}{p(\theta \mid X)}\right] \\ &= \min\left[1, \frac{p(X \mid R^*, T^*)}{p(X \mid R, T)} \times \frac{p(R^* \mid T^*)\, p(T^*)\, p(\kappa^*)\, p(\alpha^*)\, p(\pi^*)}{p(R \mid T)\, p(T)\, p(\kappa)\, p(\alpha)\, p(\pi)}\right]. \end{aligned} \quad (3)$$

The simplification is because the proposal distribution Q is symmetrical. The $p(X)$ term in $p(\theta^* \mid X)$ and $p(\theta \mid X)$ cancels.

The updating scheme cycles through two steps. In step 1, a divergence time is chosen at random to be updated together with parameters λ, μ, and ρ of the BDP. In step 2, the rate of a randomly selected branch is updated. When the proposed value of the parameter is out of the parameter space, it is reflected back into the correct interval. The chain moves to the proposed state when it is accepted; otherwise, it remains in the current state. The tuning parameters for rates and times were adjusted by running preliminary chains to attain a balance between acceptance rate and mixing. Depending on the parameterization, the acceptance rate was between 60% and 5% (for extreme parameterizations).

Sampling from the posterior distribution can start when the chain has reached stationarity. Although there exist heuristic tests to determine when the MCMC has converged,

none of them seem infallible (Gilks et al., 1996). We have monitored convergence by plotting time series of the studied variables (times and rates). Multiple chains were run from very different starting points. Linear regressions were performed on time series of each variable, testing the significance of the slope: The P value should be large, indicating a slope not significantly different from zero, and the autocorrelation functions should not detect any structure in the samples. Sampling starts after a burn-in period defined as the time the chain takes to forget the initial state and reach stationarity. The chain is sampled every 100 accepted states, hereby "thinning" the chain (Raftery and Lewis, 1996) and reducing autocorrelation between successive samples. The likelihood $p(X \mid R, T)$ reached stationarity quickly, whereas times and rates typically converged more slowly, especially for large data sets. We have used the median of the estimated posterior distribution as the best point estimate of that parameter. For most of the analyses reported here, the MCMC was run multiple times, and the results were very close.

Prior Model Selection

Our primary interest is to estimate divergence dates. However, different models of rate change can lead to different date estimates, and choosing the model that best fits the data can be important. In a Bayesian framework, inference proceeds usually from the posterior distribution $p(\theta \mid X)$, where θ stands for parameters R, T, and the hyperparameters of the BDP. However, $p(\theta \mid X)$ does not allow us to evaluate the goodness of fit of the model nor does it permit comparison between models, which have different sets of parameters.

The marginal probability $p(X)$ under a given model M_k, also denoted $p(X \mid M_k)$, contains information for assessing model performance. One approach is to use the Bayes factor to compare models M_1 and M_2:

$$BF_{12} = p(X \mid M_1)/p(X \mid M_2). \quad (4)$$

The $p(X)$ for each model is obtained by averaging (and not maximizing, as for the likelihood ratio test) over the parameter space, with respect to the prior distribution. The so-called prior mean is defined as

$$p(X \mid M_k) = \int p(X \mid R, T)p(R, T \mid \eta)p(\eta) \times dR\, dT\, d\eta, \quad (5)$$

where η includes the hyperparameters from the birth–death process $\eta = (\lambda, \mu, \rho)$. Computing the right side of Equation 5 is difficult (Raftery, 1996). Instead of computing the prior mean, Aitkin (1991) proposed using the posterior mean under each model, which we use here because it can be calculated easily by sampling from the MCMC:

$$L_k^{post} = E_{post}[p(X \mid R, T)p(R, T \mid \eta)]. \quad (6)$$

Thus, the posterior Bayes factor comparing models M_1 and M_2 is L_1^{post}/L_2^{post}. The posterior Bayes factor is controversial (see Aitkin, 1991), but because it is directly estimable from the MCMC outputs, we have used it here.

For the LND, SLD, GD, and OUP models of rate change, an empirical Bayes approach is used to estimate the hyperparameter σ^2. Under each model, L_k^{post} was evaluated for different values of σ^2. The value that best fits the data is chosen as the estimate. The same approach was used for the hyperparameters σ^2 and β under OUP.

Performance of the Different Models on a Small Data Set

Comparison of the Different Bayesian Models of Rate Change

We analyzed a small data set that consists of the tRNA-coding genes of the mitochondrial genome of six hominoid species: common chimpanzee, pygmy chimpanzee, human, gorilla, orangutan, and siamang (Horai et al., 1992). Alignment gaps were removed, leaving 762 nucleotides in the sequences. The phylogenetic relationship of these species seems well established, and the tree shown in Figure 1 will be assumed throughout. The data set was analyzed under the HKY85 + Γ model of nucleotide substitution (Hasegawa et al., 1985; Yang, 1994). The orangutan divergence was set at 13 million years ago (MYA) and was used as a calibration point

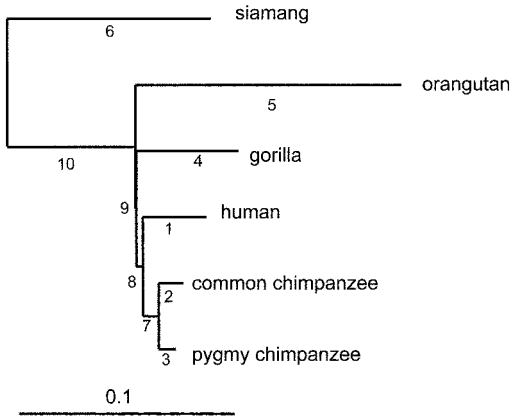

FIGURE 1. ML tree for six species of hominoids. The branch lengths of the unrooted tree were estimated under the HKY85 + Γ model of nucleotide substitution. The root of the tree is placed on the siamang branch.

(Horai et al., 1992). The molecular clock assumption was not rejected by the likelihood ratio test (LRT); the LRT statistic is $2\Delta\ell = 2[-1785.96 - (-1789.65)] = 7.38$, with $P = 0.12$ and $df = 4$. MLEs of substitution parameters without the clock are $\hat{\kappa} = 45.20$ and $\hat{\alpha} = 0.187$. These values were used in the MCMC runs in the Bayes analysis. The hyperparameters of the BDP prior for times are from the uniform distributions $\lambda \sim U(0, 15)$, $\mu \sim U(0, 5)$, and $\rho \sim U(0, 0.01)$. Each chain was run with a burn-in of 10^4 steps, after which 10^4 samples were collected every 100 steps.

When the variance for the rate (σ^2) is very small, all the models essentially make the clock assumption and produced similar estimates for the divergence times (Table 1). SLD has the largest L_k^{post}, but the posterior Bayes factor is always <0.53 when this model is compared with any other model, so that the differences are not significant (see Kass and Raftery, 1995:777).

The exponential model does not have any hyperparameter to control its variance. In all other models, increasing the variance for the rate (σ^2) relaxes the clock assumption. The same σ^2 in the different models means different levels of rate variation. Figure 2 shows the influence of σ^2 on the estimates of two rates: r_5 for the branch ancestral to orangutan and r_7 for the branch ancestral to the two chimpanzee species (see Fig. 1). Rates and divergence times are sensitive to the hyperparameter σ^2 for all models when σ^2 is small. When σ^2 is large, LND (not shown) and GD reach a plateau, and the estimates are not sensitive to σ^2.

We maximize L_k^{post} to estimate the hyperparameter σ^2 in the LND, SLD, and GD models and β and σ^2 in the OUP model. Figure 3a shows that the models behave differently. Under LND and GD, the posterior mean L_k^{post} reaches a plateau for large values of σ^2 and does not decrease until σ^2 is very large; date and rate estimates are insensitive to σ^2 when σ^2 is large in these two distributions (Fig. 2). Under SLD and OUP, L_k^{post} is sensitive to σ^2. The optimum value under SLD is about $\sigma^2 = 1$ (Fig. 3a), whereas the optimum values for OUP are about $\beta = 100$ and $\sigma^2 = 1$ (Fig. 3b); estimates of dates and rates are somewhat sensitive to these hyperparameters under SLD and OUP. When optimum parameters are used, the date estimates are similar among the different models.

TABLE 1. Bayes estimates (posterior medians ± SE) of the divergence times in clocklike and nonclocklike analyses.

Analysis	Chimpanzees	Human	Gorilla	Siamang	$\log L_k^{post}$
Clocklike					
Clock	2.14 ± 0.70	4.79 ± 1.09	7.03 ± 1.38	18.33 ± 2.01	−1792.34
LND ($\sigma^2 = 10^{-4}$)	2.09 ± 0.65	4.69 ± 1.01	6.89 ± 1.29	19.03 ± 1.98	−1791.95
SLD ($\sigma^2 = 10^{-4}$)	2.08 ± 0.65	4.71 ± 1.02	6.92 ± 1.26	18.87 ± 1.95	−1791.90
GD ($\sigma^2 = 10^{-4}$)	2.11 ± 0.81	4.76 ± 1.21	7.00 ± 1.46	19.40 ± 2.12	−1792.43
OUP ($\beta = 100$, $\sigma^2 = 10^{-4}$)	2.17 ± 0.87	4.79 ± 1.28	7.02 ± 1.54	19.22 ± 2.16	−1792.29
Nonclocklike[a]					
LND ($\sigma^2 = 10$)	5.66 ± 2.63	8.82 ± 2.63	10.96 ± 2.60	17.07 ± 2.50	−1790.17
SLD ($\sigma^2 = 1$)	5.51 ± 1.92	8.64 ± 2.04	10.49 ± 2.10	16.57 ± 2.14	−1790.73
GD ($\sigma^2 = 9$)	5.02 ± 2.84	7.99 ± 2.82	10.38 ± 2.77	16.92 ± 2.58	−1790.08
OUP ($\beta = 100$, $\sigma^2 = 1$)	4.54 ± 1.95	7.89 ± 2.08	10.21 ± 2.09	15.09 ± 1.79	−1788.75
ED	5.89 ± 2.21	9.11 ± 2.20	11.02 ± 2.13	15.01 ± 1.91	−1789.72

[a] Hyperparameters β and σ^2 are chosen to maximize the posterior mean L_k^{post}.

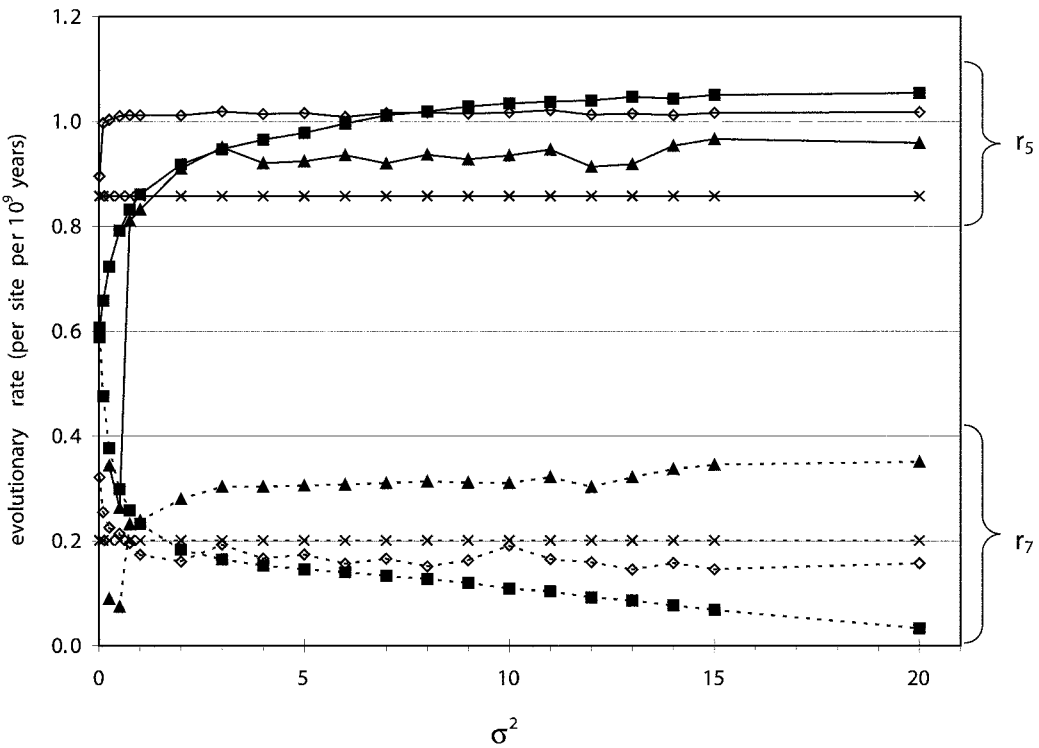

FIGURE 2. Posterior medians of evolutionary rates for branches 5 and 7 in Figure 1 under different models of rate change: SLD (■), OUP (▲), GD (◇), and ED (×). Rates are measured by the expected number of substitutions per site per 10^9 years. The hyperparameter β of OUP is set to 100.

We also used L_k^{post} to compare models of rate change (Fig. 3a). For OUP, the hyperparameter β has been set to 100, which is close to the optimal value. OUP outperformed the other models, probably because it has more hyperparameters. The posterior Bayes factor, computed from those probabilities, ranges from 1.0 to 2.0 on the log scale for comparison between OUP and the other models (Table 1), indicating a strong preference for OUP (Kass and Raftery, 1995).

Comparison with ML Analysis under Local Clock Models

Comparison of the Bayesian approach (Table 1) with ML (Table 2) is a good means of testing the MCMC implementation. Under the molecular clock assumption, both approaches should give similar estimates, with larger SEs from the Bayes models. The ML date estimates for nodes younger than the calibration point are slightly younger (say 4.3 MYA for the human–chimpanzee divergence) than the Bayes estimates (4.7–4.8 MYA). For nodes older than the calibration point, the difference is also small but in the opposite direction. The observed discrepancy appears to be due to the BDP prior for divergence times used in the Bayes approach. The use of a small sampling fraction, $\rho \sim$ U(0, 0.01), has the effect of shortening the internal branches (Yang and Rannala, 1997).

When the molecular clock is relaxed, the MLEs of dates are very different from and much older than those under the clock (Table 2). For example, the date for human–chimpanzee divergence changed from 4.3 MYA under the clock to 8.8 MYA under a two-rate local-clock model, although both estimates involve large sampling errors. Thus, the date estimates are sensitive to the clock assumption, although the molecular clock was not rejected by the LRT. In the Bayes approach, relaxing the clock assumption also had considerable effect on date estimates, but the effect is somewhat different. The chimpanzee, human, and gorilla divergences become much older than those under the clock, and the siamang divergence becomes slightly younger. Although the Bayes

TABLE 2. Maximum likelihood[a] estimates of divergence times (±SE) under the clock and local-clock[b] models.

	Clock (one rate)	Two rates	Three rates	Four rates
Chimpanzees	1.77 ± 0.54	3.74 ± 1.13	3.76 ± 1.16	5.88 ± 1.82
Human	4.28 ± 0.91	8.85 ± 1.86	7.59 ± 2.61	10.68 ± 3.80
Gorilla	6.50 ± 1.18	13.00 ± 2.32	13.00 ± 2.54	12.64 ± 6.13
Orangutan	13	13	13	13
Siamang	19.56 ± 3.49	35.86 ± 6.35	37.61 ± 7.08	57.57 ± 11.51
ℓ	−1773.21	−1770.90	−1770.52	−1769.42
$\hat{r}_1$	1	3.41	3.60	6.23
$\hat{r}_2$	1	1	1.51	1.84
$\hat{r}_3$	1	1	1	2.45

[a]ML analyses were performed under the HKY85 + Γ model ($\kappa = 45.20$, $\alpha = 0.187$). The calibration point was set at 13 MYA for the orangutan.

[b]Local clock settings: r_1 for orangutan; r_2 for human; r_3 for gorilla; $r_0 = 1$ for all other branches.

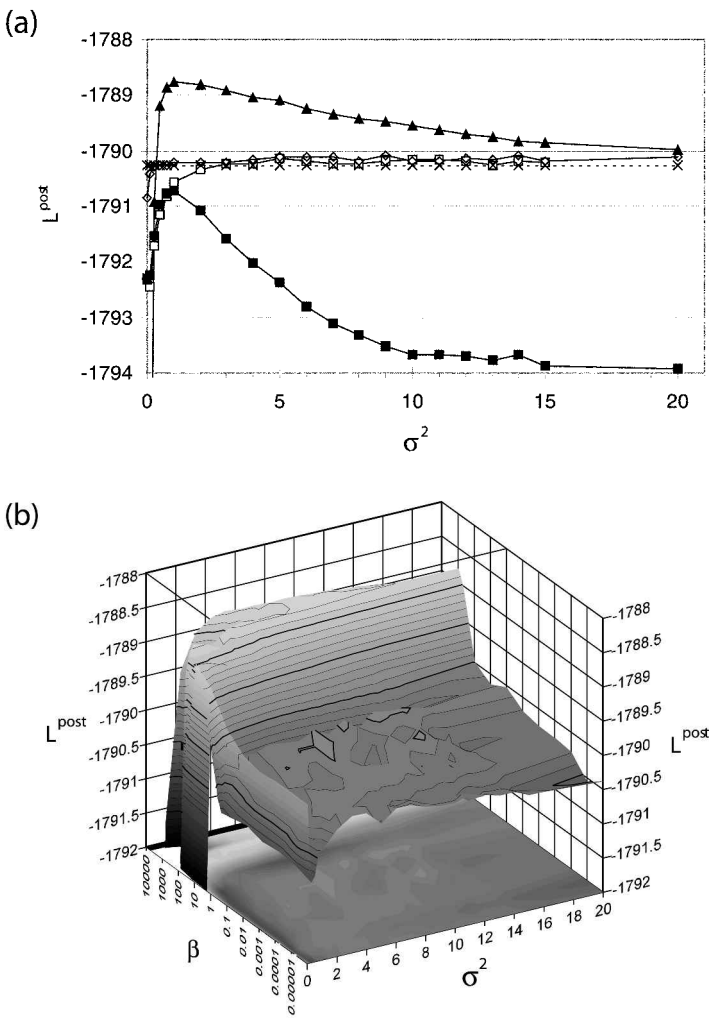

FIGURE 3. (a) Approximate posterior means L_k^{post} as a function of the hyperparameter σ^2 under different models of rate change: LND (□), SLD (■), OUP (▲), GD (◇), and ED (×). Under OUP, $\beta = 100$ is fixed. (b) Approximate surface of the posterior mean L_{OUP}^{post} as a function of the hyperparameters β and σ^2.

date estimates for the recent nodes are similar to the MLEs, either with or without the clock, the Bayes estimates of the siamang divergence date (15–17 MYA) are younger than the MLEs (about 36–37 MYA). Those MLEs appear too old, but they involve very large sampling errors.

Application to the 18S rRNA Data Set

To test the different Bayes models of rate evolution and demonstrate the important effect of rate change on date estimation, we reanalyzed the nuclear-encoded 18S rRNA genes from 39 metazoan species (Bromham et al., 1998), rooted by a fern, Polypodium. As reviewed by Cooper and Fortey (1998), the time of origin of the animal phyla has been controversial. A common view, based on the fossil records, holds that the early Cambrian (ca. 545 MYA) was characterized by accelerated evolution, marking an "explosion" of the metazoan phyla (e.g., Valentine et al., 1996). In particular, the divergence between protostomes and deuterostomes is thought to have occurred between 530 and 600 MYA. However, molecular studies such as that of Bromham et al. (1998) produced estimates as far back as about 1,200 MYA, almost twice as old.

The sequences consist of 1,710 nucleotides. Gaps were removed from the alignment, and the data set was analyzed under the HKY85 + Γ model of nucleotide substitution, with the transition: transversion rate ratio and the shape parameter of the Γ distribution set to their MLEs obtained without the clock ($\hat{\kappa} = 3.46$ and $\hat{\alpha} = 0.373$). The tree topology was fixed (Fig. 4), according to Nielsen (1995).

The molecular clock assumption was rejected by the LRT; the test statistic is $2\Delta\ell = 2[-13948.10 - (-14{,}381.85)] = 867.50$, $P < 0.01$. The shape of the ML tree under no clock (not shown) indicated very variable rates among lineages, which may preclude traditional analyses either by ML local clocks (Yoder and Yang, 2000) or by linearizing the tree (Takezaki et al., 1995).

The Bayes analysis was conducted by drawing the hyperparameters of the BDP prior for times from uniform distributions, $\lambda \sim U(0, 15)$, $\mu \sim U(0, 5)$, and $\rho \sim U(0, 0.001)$. We used the ED, SLN, and OUP models of rate change. MCMC runs included a burn-in period of 10^5 steps, after which 10^5 samples were collected every 100 accepted states. We averaged the posterior estimates over eight calibration points given by Bromham et al. (1998): Collembola–Pterygota, 390 MYA (1); Aranaea–Scorpionida, 405 MYA (2); Arachnida–Merostomata, 520 MYA (7); Cephalochordata–Chordata, 530 MYA (8); Coelacanth–Dipnoi/Tetrapoda, 418 MYA (3); Osteichthyes–Dipnoi/Tetrapoda, 428 MYA (4); Agnata–Gnathostoma, 510 MYA (6); Asteroidea–Echinoidea, 500 MYA (5) (numbers 1–8 refer to Fig. 4).

Dates estimated with a Bayesian clocklike model, with a small variance for the prior on the rates, place the echinoderm–chordate and protostome–deuterostome divergences at 1,205 MYA (95% credible set: 1,062–1,341 MYA) and 1,450 MYA (95% credible set: 1,321–1,567 MYA), respectively. These estimates are slightly smaller than but very similar to the ones found by the original authors (cf. Bromham et al., 1998: Fig. 2).

To relax the molecular clock hypothesis, the ED, SLN, and OUP models of rate change have been evaluated with an empirical Bayes phase to estimate the hyperparameters. The results are summarized in Table 3. The posterior mean L_k^{post} under each model and parameterization is provided only for the values around the maximum L_k^{post}, although an extensive search was carried out to make sure there were no other optima. Again, OUP explains the data better than does any other model. The maximum L_k^{post} is around $\beta = 1$ and $\sigma^2 = 10$, but the probability surface in this region is almost flat (Table 3). The estimated σ^2 for SLN is around 10. This large value is consistent with the large statistic in the LRT of the clock and indicates that rates are more variable than in the small hominoid data set.

The estimates of the divergence times under the ED model, summarized in Figure 4, are very similar to those under SLN and OUP (not shown). The time estimates are consistent with the fossil records (e.g., Conway Morris, 1998), or with linearized analyses performed on many genes (Ayala et al., 1998). This latter analysis, of 18 protein-coding genes, produced estimated divergence dates at 628 ± 76 MYA for the echinoderm–chordate split and at 736 ± 65 MYA for the protostomes–deuterostome separation. Our estimates, for a single gene,

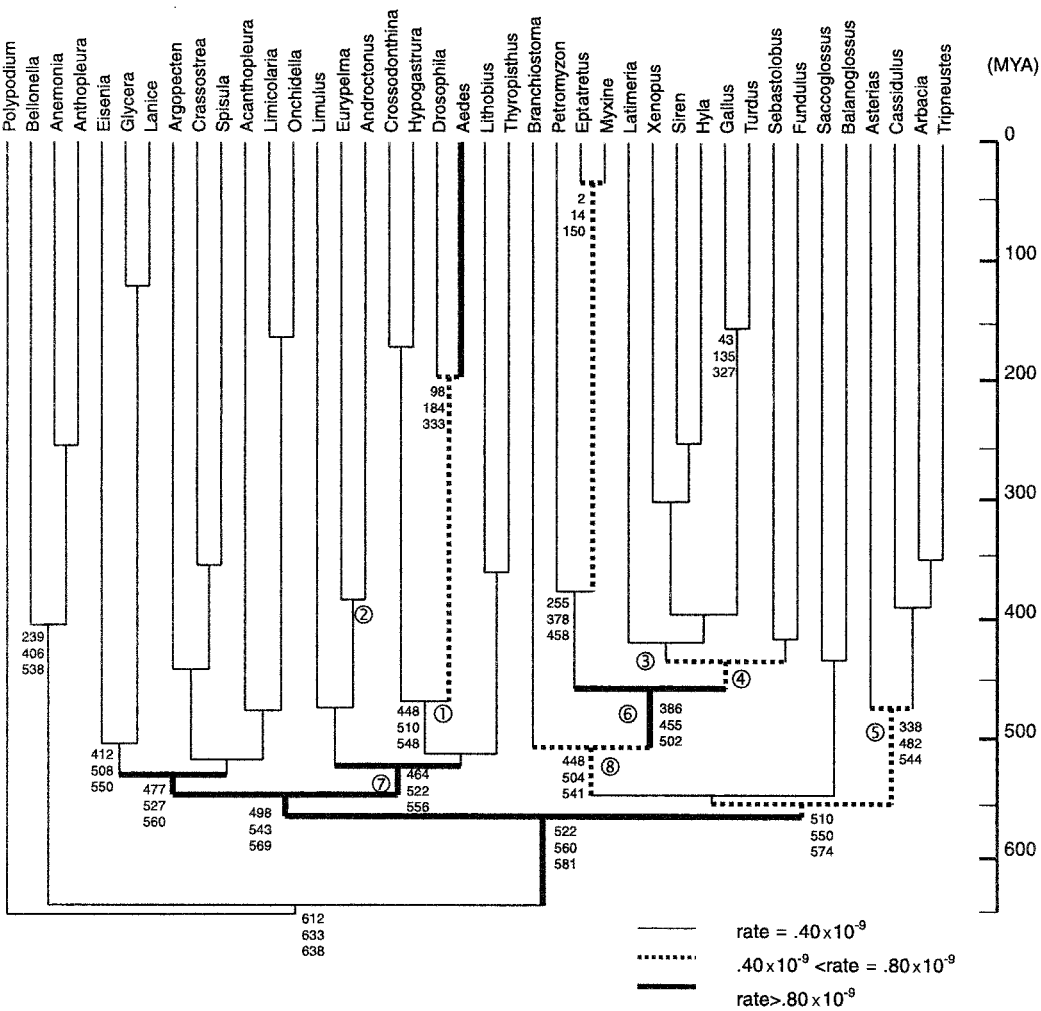

FIGURE 4. The posterior estimates of divergence times for 39 metazoan species under the ED model of rate change. Calibration points from fossil dates are indicated by circled numbers (1–8). Estimates were obtained under the HKY85 + Γ model of nucleotide substitution. Branch lengths are scaled to time, and the thickness of a branch indicates the evolutionary rate (expected number of substitutions per site per 10^9 years). Numbers at internal nodes represent, from top to bottom, the lower limit of the 95% credible set, the time estimate, and the upper limit of the 95% credible set.

are respectively 550 MYA (95% credible set: 510–574 MYA) and 560 MYA (95% credible set: 522–581 MYA) under ED and 579 MYA (498–608 MYA) and 595 MYA (519–616 MYA) under OUP ($\beta = 0.1$ and $\sigma^2 = 10$). Because there is no need to eliminate outlying taxa, all the available information in the gene is taken into account in the Bayesian approach.

Possible biases must be taken into account when we interpret the results of the Bayesian analysis. First, we used a fixed tree topology, although uncertainty exists regarding the evolutionary history of the metazoan phyla. The effect of the uncertain phylogeny on date estimation deserves consideration, although Yoder and Yang (2000) suggested that plausible topologies gave similar speciation date estimates. Second, in the hominoid data set Bayesian inference may be sensitive to the hyperparameter σ^2 of the prior model of rate change. Similar effects were found in the metazoan data set. For example, date estimates under the clock assumption were drastically different from those presented in Figure 4. In this regard, estimates of dates under different models were similar when optimum values of σ^2 were used in each model of rate change. Third, the results presented

TABLE 3. Fit of different models of rate change to the metazoan 18S rRNA sequences.

Model	β	σ^2	$\log L_k^{post}$
ED	n.a.[a]	n.a.	−13992.64
SLD	n.a.	1	−14020.02
		10	−14000.03
		20	−14001.52
		40	−14009.06
OUP	0.01	1	−13984.80
		10	−13976.05
		20	−13978.62
		40	−13976.80
	0.1	1	−13984.86
		10	−13975.52
		20	−13976.97
		40	−13977.77
	1	1	−13991.27
		10	−13976.32
		20	−13975.22
		40	−13975.87

[a] n.a. = not applicable.

here were obtained from a single gene and should be taken with caution.

The rate estimates (Fig. 4) suggest that the metazoan 18S rRNA gene has a complex history, with high evolutionary rates during the Cambrian (between 550 and 500 MYA) for triploblastic animals and much lower rates, for diploblastic animals, which seem to be conserved to date. The episode of high evolutionary rate in the Cambrian was followed by a steep decline to a more or less steady rate for protostomes, whereas the pro-chordates underwent another burst at approximately late Ordovician/Silurian. Subsequent rate accelerations were detected for the branches leading to the Myxiniformes and the Diptera, with a burst for the Nematocera. The history of the 18S rRNA gene might therefore not be characterized as a mere decline of rates as suggested recently (see Bromham and Hendy, 2000), although the reasons for this episodic evolution (Gillespie, 1991) are not yet understood.

DISCUSSION

Analyses of both the hominoid and metazoan data sets suggest that date estimates are sensitive to the molecular clock hypothesis. Most molecular date estimates have been based on the simplifying assumption of the molecular clock, although some methods have been proposed to constrain a data set to conform to this hypothesis, for example by the relative rate test (Wu and Li, 1985) or tree linearization (Takezaki et al., 1995). Bromham et al. (2000) pointed out that the power to detect rate variation might not be very high, and as a result use of such tests to filter data might still lead to systematically biased date estimates. As demonstrated by our analysis of the metazoan data set, the Bayes approach offers a promising alternative to the problem, estimating divergence dates while detecting and accommodating possible rate variation.

The likelihood-based local-clock models (Yoder and Yang, 2000) are useful if prior information is available about which lineages might have different rates. For example, such models can be used for testing whether certain groups of species (e.g., primates versus rodents) have different evolutionary rates. When such information is unavailable, it is more natural to use a Bayes model of random rate change, although there is a greater computational cost. The hyperparameter σ^2 of the rate distribution controls the amount of rate variation. Although the approach appears most appropriate when rates change slowly over time or across branches, it can accommodate rapid rate changes with the use of large values of σ^2, as in the metazoan data set. Our results suggest that beyond a certain value, the hyperparameter σ^2 has little influence on the date estimates.

Our approach of estimating the hyperparameter σ^2 (or β and σ^2 in OUP) does not properly account for the uncertainty concerning those hyperparameters because the optimum values were treated as known when divergence dates were estimated. A full Bayes approach should integrate over β and σ^2. We attempted to apply such an approach to both data sets analyzed in this paper, averaging over uniform priors for β and σ^2 in the MCMC. However, the chain did not converge well, in particular regarding the marginal distributions of β and σ^2, and the probability surface was relatively flat for large values of σ^2. Simultaneous use of multiple calibration points might provide information about rates and thus help with the convergence of the MCMC in a full Bayes analysis.

A C program implementing methods discussed here is available at http://abacus.gene.ucl. ac.uk/stephane/

ACKNOWLEDGMENTS

We thank Mike Holder, Chris Simon, and two anonymous referees for constructive comments. S.A.B. thanks Hirohisa Kishino for stimulating discussions and helpful comments. This study was supported by a Biotechnological and Biological Sciences Research Council grant to Z.Y.

REFERENCES

AITKIN, M. 1991. Posterior Bayes factor. J. R. Stat. Soc. B 53:111–142.

AYALA, F. J., A. RZHETSKY, AND F. J. AYALA. 1998. Origin of the metazoan phyla: Molecular clocks confirm paleontological estimates. Proc. Natl. Acad. Sci. USA 95:606–611.

BICKEL, D. R. 2000. Implications of fluctuations in substitution rates: Impact on the uncertainty of branch lengths and on relative-rate tests. J. Mol. Evol. 50:381–390.

BROMHAM, L. D., AND M. D. HENDY. 2000. Can fast early rates reconcile molecular dates with the Cambrian explosion? Proc. R. Soc. Lond. B Biol. Sci. 267:1041–1047.

BROMHAM, L., D. PENNY, A. RAMBAUT, AND M. D. HENDY. 2000. The power of relative rates tests depends on the data. J. Mol. Evol. 50:296–301.

BROMHAM, L., A. RAMBAUT, R. FORTEY, A. COOPER, AND D. PENNY. 1998. Testing the Cambrian explosion hypothesis by using a molecular dating technique. Proc. Natl. Acad. Sci. USA 95:12386–12389.

CONWAY MORRIS, S. 1998. Metazoan phylogenies: Falling into place or falling to pieces? A palaeontological perspective. Curr. Opin. Genet. Dev. 8:662–667.

COOPER, A., AND R. FORTEY. 1998. Evolutionary explosions and the phylogenetic fuse. Trends Ecol. Evol. 13:151–156.

CUTLER, D. J. 2000. Estimating divergence times in the presence of an overdispersed molecular clock. Mol. Biol. Evol. 17:1647–1660.

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

GILKS, W. R., S. RICHARDSON, AND D. J. SPIEGELHALTER. 1996. Markov chain Monte Carlo in practice. Chapman & Hall, London.

GILL, P. E., W. MURRAY, AND M. H. WRIGHT. 1981. Practical optimization. Academic Press, London.

GILLESPIE, J. H. 1991. The causes of molecular evolution. Oxford Univ. Press, Oxford, U.K.

HASEGAWA, M., H. KISHINO, AND T. YANO. 1985. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.

HASTINGS, W. K. 1970. Monte Carlo sampling using Markov chains and their applications. Biometrika 57:97–109.

HORAI, S., Y. SATTA, K. HAYASAKA, R. KONDO, T. INOUE, T. ISHIDA, S. HAYASHI, AND N. TAKAHATA. 1992. Man's place in Hominoidea revealed by mitochondrial DNA genealogy. J. Mol. Evol. 35:32–43.

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

KARLIN, S., AND H. M. TAYLOR. 1981. A secondary course in stochastic processes. Academic Press, New York.

KASS, R. E., AND A. E. RAFTERY. 1995. Bayes factors. J. Am. Stat. Assoc. 90:773–795.

KENDALL, D. G. 1948. On the generalized birth-and-death process. Ann. Math. Stat. 19:1–15.

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.

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

NEE, S., R. M. MAY, AND P. H. HARVEY. 1994. The reconstructed evolutionary process. Philos. Trans. R. Soc. Lond. B Biol. Sci. 344:305–311.

NEI, M., AND S. KUMAR. 2000. Molecular evolution and phylogenetics. Oxford Univ. Press, Oxford, U.K.

NIELSEN, C. 1995. Animal evolution: Interrelationships of the living phyla. Oxford Univ. Press, Oxford, U.K.

RAFTERY, A. E. 1996. Hypothesis testing and model selection Pages 163–187 *in* Marko chain Monte Carlo in practice (W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, eds.). Chapman & Hall, Boca Raton.

RAFTERY, A. E., AND S. M. LEWIS. 1996. Implementing MCMC. Pages 115–130 *in* Markov chain Monte Carlo in practice (W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, eds.). Chapman & Hall, London.

RAMBAUT, A., AND L. BROMHAM. 1998. Estimating divergence dates from molecular sequences. Mol. Biol. Evol. 15:442–448.

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

SARICH, V. M., AND A. C. WILSON. 1967. Immunological time scale for hominid evolution. Science 158:1200–1203.

SOLTIS, P. S., D. E. SOLTIS, V. SAVOLAINEN, P. R. CRANE, AND T. G. BARRACLOUGH. 2002. Rate heterogeneity among lineages of tracheophytes: Integration of molecular and fossil data and evidence for molecular living fossils. Proc. Natl. Acad. Sci. USA 99:4430–4435.

TAKEZAKI, N., A. RZHETSKY, AND M. NEI. 1995. Phylogenetic test of the molecular clock and linearized trees. Mol. Biol. Evol. 12:823–833.

THOMPSON, E. A. 1975. Human evolutionary trees. Cambridge Univ. Press, Cambridge, U.K.

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.

VALENTINE, J. W., D. H. ERWIN, AND D. JABLONSKI. 1996. Developmental evolution of metazoan bodyplans: The fossil evidence. Dev. Biol. 173:373–381.

WU, C. I., AND W. H. LI. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA 82:1741–1745.

YANG, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. J. Mol. Evol. 39:306–314.

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

YODER, A. D., AND Z. YANG. 2000. Estimation of primate speciation dates using local molecular clocks. Mol. Biol. Evol. 17:1081–1090.

ZUCKERKANDL, E., AND L. PAULING. 1965. Evolutionary divergence and convergence in proteins. Pages 97–166 *in* Evolving genes and proteins (V. Bryson and H. J. Vogel, eds.). Academic Press, New York.

First submitted 1 August 2001; reviews returned 26 December 2001; final acceptance 18 February 2002
Associate Editor: John Huelsenbeck