Syst. Biol. 56(3):453–466, 2007

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150701420643

Inferring Speciation Times under an Episodic Molecular Clock

BRUCE RANNALA[1] AND ZIHENG YANG[2]

[1]*Genome Center and Section of Evolution and Ecology, University of California, Davis, California, USA*
[2]*Department of Biology, University College London, London, UK; E-mail: z.yang@ucl.ac.uk*

Abstract.— We extend our recently developed Markov chain Monte Carlo algorithm for Bayesian estimation of species divergence times to allow variable evolutionary rates among lineages. The method can use heterogeneous data from multiple gene loci and accommodate multiple fossil calibrations. Uncertainties in fossil calibrations are described using flexible statistical distributions. The prior for divergence times for nodes lacking fossil calibrations is specified by use of a birth-death process with species sampling. The prior for lineage-specific substitution rates is specified using either a model with autocorrelated rates among adjacent lineages (based on a geometric Brownian motion model of rate drift) or a model with independent rates among lineages specified by a log-normal probability distribution. We develop an *infinite-sites theory*, which predicts that when the amount of sequence data approaches infinity, the width of the posterior credibility interval and the posterior mean of divergence times form a perfect linear relationship, with the slope indicating uncertainties in time estimates that cannot be reduced by sequence data alone. Simulations are used to study the influence of among-lineage rate variation and the number of loci sampled on the uncertainty of divergence time estimates. The analysis suggests that posterior time estimates typically involve considerable uncertainties even with an infinite amount of sequence data, and that the reliability and precision of fossil calibrations are critically important to divergence time estimation. We apply our new algorithms to two empirical data sets and compare the results with those obtained in previous Bayesian and likelihood analyses. The results demonstrate the utility of our new algorithms. [Bayesian method; divergence times; MCMC; molecular clock.]

The molecular clock hypothesis postulates that the molecular evolutionary rate is constant over time (Zuckerkandl and Pauling, 1965) and provides a simple indirect means for dating evolutionary events. The expected genetic distance between sequences increases linearly as a function of the time elapsed since their divergence and fossil-based divergence dates can therefore be used to translate genetic distances into geological times, allowing divergence times to be inferred for species with no recent ancestor in the fossil record. The molecular clock hypothesis is often violated, however, particularly when distantly related species are compared, and such violations can lead to grossly incorrect species divergence time estimates (Bromham et al., 1998; Yoder and Yang, 2000; Adkins et al., 2003).

One approach to dealing with a violation of the clock is to remove sequences so that the clock approximately holds for the remaining sequence data. This may be useful if only one or two lineages have grossly different rates and can be identified and removed (Takezaki et al., 1995) but is difficult to use if the rate variation is more widespread. A more promising approach is to take explicit account of among-lineage rate variation when estimating divergence times. Variable-rates models have been the focus of much recent research, with both likelihood and Bayesian methodologies employed. In a likelihood analysis, prespecified lineages in the phylogeny are assigned independent rate parameters, estimated from the data (Kishino and Hasegawa, 1990; Rambaut and Bromham, 1998; Yoder and Yang, 2000). Recent extensions to the likelihood method (Yang and Yoder, 2003) allow the use of multiple calibration points and simultaneous analysis of data for multiple genes while accounting for their differences in substitution rates and in other aspects of the evolutionary process.

The Bayesian approach, pioneered by Thorne et al. (1998) and Kishino et al. (2001; see also Huelsenbeck et al., 2000; Drummond et al., 2006), uses a stochastic model of evolutionary rate change to specify the prior distribution of rates and, with a prior for divergence times, calculates the posterior distributions of times and rates. Markov chain Monte Carlo (MCMC) is used to make the computation feasible. Such methods build on the suggestion by Gillespie (1984) that the rate of evolution may itself evolve over time and may be considered as more rigorous implementations of Sanderson's rate-smoothing procedure (Sanderson, 1997; Yang, 2004). The algorithm was extended to analyze multiple genes (Thorne and Kishino, 2002). The method has been applied successfully to estimate divergence times in a number of important species groups, such as the mammals (Hasegawa et al., 2003; Springer et al, 2003), the birds (Pereira and Baker, 2006), and plants (Bell and Donoghue, 2005).

Thorne et al. (1998) used lower and upper bounds for node ages to incorporate fossil calibration information. With this prior, divergence times outside the bounds are impossible in the posterior, whatever the data. Biologists may often lack sufficiently strong convictions to apply such "hard bounds" and, in particular, fossils often provide good lower bounds (minimal node ages) but not good upper bounds (maximal node ages). However, the posterior can be sensitive to changes to the upper bounds. This observation prompted Yang and Rannala (2006) to implement arbitrary prior distributions for the age at a fossil calibration node. Such "soft bounds" may sometimes provide a more accurate description of uncertainties in fossil ages. Our implementation, however, assumed the molecular clock.

In this paper, we extend our previous model to relax the clock assumption. We implement two prior models that allow the evolutionary rate to vary over time or across lineages. The first assumes a geometric Brownian motion process of rate drift over time, the model implemented by Thorne et al. (1998) and Kishino et al. (2001). Rates are autocorrelated between ancestral and descendant lineages on the tree. The second is an independent-rates model, with no autocorrelation. Instead, branch-specific rates are independent variables drawn from a common distribution. If higher rates tend to change more than lower rates, there may be little autocorrelation of rates between ancestral and descendant lineages (Gillespie, 1984). In such cases, independent-rates models may be more flexible in accommodating large rate shifts that may occur during species radiations due to rapid range expansions, increasing effective population sizes, enhanced selection, etc.

We also develop an "infinite-sites theory" to understand the limit of divergence time estimation; even when the number of sites in the sequence approaches infinity, the errors in posterior time estimates will not approach zero because of inherent uncertainties in fossil calibrations and the confounding effect of rates and times in the sequence data. Yang and Rannala (2006) studied the analytical properties of this problem when the data consist of one locus and the molecular clock is assumed. Here, the theory is extended to the general case of variable rates and multiple loci. We use computer simulation to assess the information content of sequence data from multiple loci for estimation of divergence times (in the limit of infinite sequence length). We also analyze two empirical data sets for comparison with previous methods.

THEORY

The Bayesian Framework

Let s be the number of species. The topology of the rooted phylogenetic tree is assumed known and fixed. Sequence alignments are available at multiple loci, with the possibility that some loci are missing for some species. Let D be the sequence data and let $\mathbf{t}$ be the $s-1$ divergence times in the species tree. Let $\mathbf{r}$ represent the substitution rates for branches in the tree at all loci. Let g be the number of loci. At each locus, there are $2s-2$ branch rates, so $\mathbf{r}$ includes $(2s-2)g$ rates. Let θ denote parameters in the nucleotide substitution model as well as parameters in the prior for times $\mathbf{t}$ and rates $\mathbf{r}$.

Bayesian inference makes use of the joint conditional distribution

$$f(\theta, \mathbf{t}, \mathbf{r}|D) = \frac{f(D|\mathbf{t}, \mathbf{r})f(\mathbf{r}|\theta, \mathbf{t})f(\mathbf{t}|\theta)f(\theta)}{f(D)}$$

where $f(\theta)$ is the prior for substitution parameters, $f(\mathbf{t}|\theta)$ is the prior for divergence times, which incorporates fossil calibration information, $f(\mathbf{r}|\theta, \mathbf{t})$ is the prior for substitution rates for branches, and $f(D|\mathbf{t}, \mathbf{r})$ is the likelihood. The proportionality constant $f(D)$ involves integration over $\mathbf{t}$, $\mathbf{r}$, and θ. We construct a Markov chain whose states are $(\theta, \mathbf{t}, \mathbf{r})$ and whose steady-state distribution is $f(\theta, \mathbf{t}, \mathbf{r}|D)$. We implement a Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970). The general framework has been described before (see, e.g., Thorne et al., 1998). Given the current state of the chain $(\theta, \mathbf{t}, \mathbf{r})$, a new state $(\theta^*, \mathbf{t}^*, \mathbf{r}^*)$ is proposed through a proposal density $q(\theta^*, \mathbf{t}^*, \mathbf{r}^*|\theta, \mathbf{t}, \mathbf{r})$ and is accepted with probability

$$R = \min\left\{1, \frac{f(D|\mathbf{t}^*, \mathbf{r}^*)f(\mathbf{r}^*|\theta^*, \mathbf{t}^*)f(\mathbf{t}^*|\theta^*)f(\theta^*)}{f(D|\mathbf{t}, \mathbf{r})f(\mathbf{r}|\theta, \mathbf{t})f(\mathbf{t}|\theta)f(\theta)} \times \frac{q(\theta, \mathbf{t}, \mathbf{r}|\theta^*, \mathbf{t}^*, \mathbf{r}^*)}{q(\theta^*, \mathbf{t}^*, \mathbf{r}^*|\theta, \mathbf{t}, \mathbf{r})}\right\} \quad (1)$$

Note that $f(D)$ cancels in calculation of R. The proposal density $q(\cdot|\cdot)$ is flexible as long as it specifies an aperiodic and irreducible Markov chain. Calculation of the likelihood follows Felsenstein (1981) for models of one rate for all sites or Yang (1994) for models of variable rates among sites. The likelihood calculation is straightforward but expensive. The prior for divergence times $f(\mathbf{t}|\theta)$ is described in Yang and Rannala (2006). Our focus in this paper is the prior on rates $f(\mathbf{r}|\theta, \mathbf{t})$.

Prior Densities of Evolutionary Rates

Autocorrelated substitution rates on branches.—Thorne et al. (1998) and Kishino et al. (2001) used a geometric Brownian motion model to specify the prior for rates $f(\mathbf{r}|\theta, \mathbf{t})$. The logarithm of the rate drifts according to a Brownian motion process and the rate evolves according to a geometric Brownian motion process. Thus, given the rate r_A in the ancestor, the rate r time t later has a log-normal distribution. Kishino et al. (2001) force the expectation of the rate r to equal the ancestral rate r_A, that is $E(r|r_A) = r_A$, by applying a bias-correction term. The density of rate r given the ancestral rate r_A is

$$f(r|r_A) = \frac{1}{r\sqrt{2\pi t\sigma^2}} \exp\left\{-\frac{1}{2t\sigma^2}\left(\log(r/r_A) + \frac{t\sigma^2}{2}\right)^2\right\}, \quad 0 < r < \infty \quad (2)$$

Similarly, let $y = \log(r)$, $y_A = \log(r_A)$, and then $y|y_A \sim N(y_A - t\sigma^2/2,\ t\sigma^2)$. Here $N(m, s^2)$ stands for the normal distribution with mean m and variance s^2. Parameter σ^2 determines how fast the rate drifts across lineages (Thorne et al., 1998). The relevant parameter for evolutionary inference is the integral over the sample path of the rate on each branch. Kishino et al. (2001) formulated the model using rates for nodes, with the arithmetic average of the rates at ancestral and descendant nodes as a proxy for the average rate over the branch. We use instead the rate at the midpoint of a branch. This strategy simplifies the MCMC algorithm and makes it easier to compare the independent and autocorrelated rates models. The rate at the root, μ, is given a gamma prior with

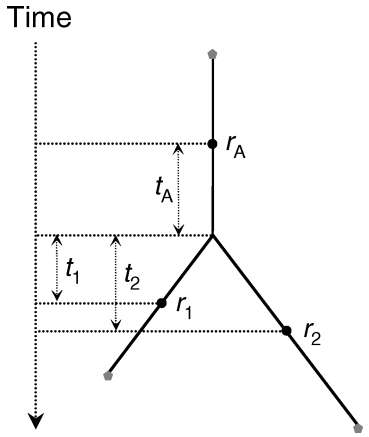

FIGURE 1. Definition of rates and times surrounding an internal node of a tree under our parameterization of the variable rates model. Parameter r_A is the rate at the midpoint of the ancestral branch, whereas r_1 and r_2 are the rates at the midpoints of the two descendent branches. The total time durations of the three branches are $2t_A$, $2t_1$, and $2t_2$.

parameters α_μ and β_μ. Because the autocorrelation model assumes that the geometric Brownian motion process is homogeneous, μ is also the mean rate across the tree. The parameter σ^2 is given a gamma prior with parameters α_{σ^2} and β_{σ^2}.

Here we derive the probability density of the rates at the midpoints of two descendant branches given the rate at the midpoint of an ancestral branch under a geometric Brownian motion model. We consider one locus first, and results are easily generalized to multiple loci. We refer to the rate at the midpoint of a branch as the rate for the branch; this is used to approximate the average rate for the whole branch. Each internal node that is not the root has two daughter branches and an ancestral branch (Fig. 1). For each such node, let r_1 and r_2 be the rates of the two daughter branches and r_A be the rate of the ancestral branch. Let the rate at the central node itself be r_0. To simplify the presentation, we let y_A, y_0, y_1, y_2 be the logarithms of rates r_A, r_0, r_1, r_2. Let half of the time length for the ancestral branch be t_A and half of the time length for the two daughter branches be t_1 and t_2, respectively. The density of r_1 and r_2 given r_A can be derived as follows.

$$y_0|y_A \sim N\left(y_A - t_A\sigma^2/2,\, t_A\sigma^2\right) \tag{3}$$

$$y_1,\, y_2|y_0,\, y_A \sim N_2\left[\begin{pmatrix} y_0 - \frac{1}{2}t_1\sigma^2 \\ y_0 - \frac{1}{2}t_2\sigma^2 \end{pmatrix}, \begin{pmatrix} t_1\sigma^2 & 0 \\ 0 & t_2\sigma^2 \end{pmatrix}\right] \tag{4}$$

Here $N_2(\mathbf{m}, \mathbf{S})$ is the bivariate normal distribution with mean vector $\mathbf{m}$ and variance-covariance matrix $\mathbf{S}$. Thus,

$$y_1,\, y_2|y_A \sim N_2(\eta,\, \Sigma), \tag{5}$$

where

$$\eta = \begin{bmatrix} y_A - \frac{1}{2}(t_A + t_1)\sigma^2 \\ y_A - \frac{1}{2}(t_A + t_2)\sigma^2 \end{bmatrix},\ \Sigma = \begin{bmatrix} t_A + t_1 & t_A \\ t_A & t_A + t_2 \end{bmatrix}\sigma^2 \tag{6}$$

Thus, by applying a variable transform from (y_1, y_2) to (r_1, r_2), we obtain

$$f(r_1, r_2|r_A) = f(r_1, r_2|y_A) = \frac{f(y_1, y_2|y_A)}{r_1 r_2} = \frac{\exp\left[-\frac{1}{2}(\mathbf{y} - \eta)'\Sigma^{-1}(\mathbf{y} - \eta)\right]}{2\pi|\Sigma|^{1/2}} \times \frac{1}{r_1 r_2} \tag{7}$$

where $\mathbf{y} = (y_1, y_2)' = (\log r_1, \log r_2)'$.

For the root, the rates of its two daughter branches, given the rate at the root (μ), have independent log-normal distributions. The density can be calculated using Equations (6) and (7) by setting $t_A = 0$ in Equation (6). From equation (7), the log rates y_1 and y_2, given the ancestral log rate y_A, have a correlation coefficient of

$$\rho_{12} = \frac{t_A}{\sqrt{(t_A + t_1)(t_A + t_2)}} \tag{8}$$

Note that our formulation of the model using rates for branches is similar to that of Thorne et al. (1998). However, as pointed out by Kishino et al. (2001), Thorne et al. (1998) failed to accommodate the correlation between the rates for two descendant branches given the rate at the ancestral branch. This correlation is due to the fact that both log rates y_1 and y_2 evolved from the common log rate y_0 at the central node (Fig. 1). It is explicitly accounted for in our implementation.

Independent and identically distributed substitution rates on branches.—We consider an alternative to the model of Thorne et al. (1998) for among-lineage rate variation that does not impose autocorrelation of rates among branches. The average rate over the ith branch is given by the log-normal density

$$f(r_i|\mu, \sigma^2) = \frac{1}{r_i\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\log(r_i/\mu) + 1/2\sigma^2\right]^2\right\},\ 0 < r_i < \infty \tag{9}$$

Parameter μ is the mean of the rate for all lineages and σ^2 is the variance of the log rate. Parameters μ and σ^2 are assigned gamma priors with parameters α_μ, β_μ and α_{σ^2}, β_{σ^2}, respectively.

We note that the overall rate parameter μ has the same interpretation under the autocorrelated and independent rates models. However, parameter σ^2 does not have the same interpretation; under the autocorrelated rates

model, the variance of the rate for a descendant branch given the rate for an ancestral branch depends on the elapsed time t, but under the independent rates model, the variance depends on σ^2 only.

Multiple loci.—To analyze multiple loci we assume that the jth locus has a set of branch-specific rates. Let $\mathbf{r} = \{r_{ij}\}$, where r_{ij} is the substitution rate on branch i at locus j. We ignore possible correlations in substitution rates among loci in the prior (although rates may be correlated in the posterior) and multiply the priors for rates across the loci. Each locus has the overall rate μ_j and variance parameter σ_j^2. Here the gamma prior on μ_j may be interpreted as a model of variable rates among loci, as may the gamma prior on σ_j^2.

Bayesian Inference with Infinite Number of Sites

Here we consider properties of the posterior distribution of divergence times when the number of sites tends to infinity, with the substitution rate either constant or variable across branches. This extends the analysis of Yang and Rannala (2006) of the asymptotic distribution in the case of a strict molecular clock and a single locus. The theory provides a way to quantify the residual uncertainty of divergence time estimates—uncertainty that cannot be further reduced by sequencing more sites for a gene or by sequencing more genes—and allows us to explore the limits of this kind of inference. We derive our theory for a specific phylogeny first and then describe how it is applied to any phylogeny.

Consider the phylogeny of $s = 4$ species shown in Fig. 2a. The parameters are the divergence times $\mathbf{t} = \{t_5, t_6, t_7\}$ and the branch-specific substitution rates $\mathbf{r} = \{r_1, r_2, r_3, r_4, r_5, r_6\}$. The likelihood of the sequence data depends on the branch lengths in the unrooted tree at each locus, measured by the expected number of substitutions: $b_1 = r_1 \times t_5$, $b_2 = r_2 \times t_5$, $b_3 = r_3 \times t_6$, $b_4 = r_4 \times t_7 + r_6 \times (t_7 - t_6)$, and $b_5 = r_5 \times (t_6 - t_5)$. Because we do not assume the clock, the root cannot be identified and the two branches joining nodes 7-6 and 7-4 in the rooted tree (Fig. 2a) are combined to become one branch of length b_4 in the unrooted tree (Fig. 2b). For each locus the prior on rates under both the autocorrelated rates and independent rates models is determined by two parameters: μ and σ^2.

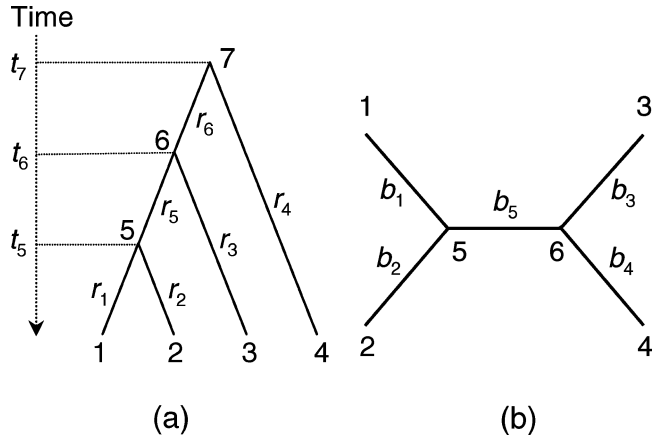

FIGURE 2. (a) Rooted tree for four species, showing three divergence times t_5, t_6, and t_7, and six branch rates $r_1, \ldots, r_6$. (b) The corresponding unrooted tree with five branch lengths measured in the expected number of substitutions: $b_1, \ldots, b_5$. Note that the position of the root (ancestor 7 in a) is not determined in the absence of a clock.

Single Locus

In the limit as the number of sites tends to infinity, the likelihood becomes a point mass with density concentrated at the true values of the branch lengths. The branch lengths can then be treated as observed data and it is possible to study the posterior density of the divergence times. Note that there are $3s - 1 = (s - 1) + 2 + (2s - 2)$ parameters in the model: $s - 1$ divergence times, μ and σ^2, and $2s - 2$ branch rates. Fixing the branch lengths in the unrooted tree by assuming an infinite number of sites reduces the dimension by $2s - 3$, leaving $s + 2$ parameters in the posterior density: the $s - 1$ divergence times, μ, σ^2, and one additional rate, which we arbitrarily choose to be the rate r_6 for the left daughter branch of the root. The other $2s - 3$ branch rates are analytically determined by the branch lengths on the unrooted tree.

The key to solving this problem is to transform the prior density of the original variables

$$f(t_5, t_6, t_7, r_6, \mu, \sigma^2, r_1, r_2, r_3, r_4, r_5)$$

to the prior density of the new set of variables

$$g(t_5, t_6, t_7, r_6, \mu, \sigma^2, b_1, b_2, b_3, b_4, b_5)$$

through a variable transform.

$$\begin{aligned}
t_5 &= t_5,\\
t_6 &= t_6,\\
t_7 &= t_7,\\
r_6 &= r_6,\\
\mu &= \mu,\\
\sigma^2 &= \sigma^2,\\
r_1 &= b_1/t_5,\\
r_2 &= b_2/t_5,\\
r_3 &= b_3/t_6,\\
r_4 &= [b_4 - r_6(t_7 - t_6)]/t_7 = (b_4 + r_6 t_6)/t_7 - r_6,\\
r_5 &= b_5/(t_6 - t_5)
\end{aligned} \tag{10}$$

The Jacobian determinant is

$$|J| = \left|\frac{\partial(t_5, t_6, t_7, r_6, \mu, \sigma^2, r_1, r_2, r_3, r_4, r_5)}{\partial(t_5, t_6, t_7, r_6, \mu, \sigma^2, b_1, b_2, b_3, b_4, b_5)}\right|$$

$$= \begin{vmatrix} 1 & 0 & 0 & 0 & 0\,0\,0\,0\,0\,0 \\ 0 & 1 & 0 & 0 & 0\,0\,0\,0\,0\,0 \\ 0 & 0 & 1 & 0 & 0\,0\,0\,0\,0\,0 \\ 0 & 0 & 0 & 1 & 0\,0\,0\,0\,0\,0 \\ 0 & 0 & 0 & 0 & 1\,0\,0\,0\,0\,0 \\ 0 & 0 & 0 & 0 & 0\,1\,0\,0\,0\,0 \\ -\frac{b_1}{t_5^2} & 0 & 0 & 0 & 0\,0\,\frac{1}{t_5}\,0\,0\,0 \\ -\frac{b_2}{t_5^2} & 0 & 0 & 0 & 0\,0\,0\,\frac{1}{t_5}\,0\,0 \\ 0 & -\frac{b_3}{t_6^2} & 0 & 0 & 0\,0\,0\,0\,\frac{1}{t_6}\,0 \\ 0 & \frac{r_6}{t_7} & -\frac{b_4+r_6t_6}{t_7^2} & \frac{t_6}{t_7} & -1\,0\,0\,0\,0\,0\,\frac{1}{t_7} \\ \frac{b_5}{(t_6-t_5)^2} & -\frac{b_5}{(t_6-t_5)^2} & 0 & 0 & 0\,0\,0\,0\,0\,0 \end{vmatrix}$$

$$= \frac{1}{t_5^2 t_6 t_7 (t_6 - t_5)} \tag{11}$$

The transformed density is

$$g(t_5, t_6, t_7, r_6, \mu, \sigma^2, b_1, b_2, b_3, b_4, b_5)$$
$$= f\left(t_5, t_6, t_7, r_6, \mu, \sigma^2, \frac{b_1}{t_5}, \frac{b_2}{t_5}, \frac{b_3}{t_6}, \frac{b_4 - r_6(t_7 - t_6)}{t_7}, \frac{b_5}{t_6 - t_5}\right)$$
$$\times |J|$$

Note that the upper-diagonal elements of the J matrix are all zero, whereas on the diagonal the reciprocals of the nonunity terms, $t_5, t_5, t_6, t_7, (t_6 - t_5)$ are the time durations of branches 1, 2, 3, 4, 5 in Fig. 2a; that is, of all the branches except the left daughter branch of the root node. These features are easily seen to hold in general and the Jacobian determinant for any tree of s species is

$$|J| = \left(\prod_{j \neq i}^{2s-2} T_j\right)^{-1}$$

where T_j is the time duration of the jth branch on the tree, with branch i, the left daughter branch of the root (branch 6 with rate r_6 in Fig. 2a), excluded. For a tree of s species, the posterior density of $\mathbf{t} = \{t_1, \ldots, t_{s-1}\}$, μ, σ^2 and r (the rate for the left daughter branch of the root), conditional on the branch lengths is

$$g(\mathbf{t}, r, \mu, \sigma^2 | \mathbf{b}) = \frac{g(\mathbf{t}, r, \mu, \sigma^2, \mathbf{b})}{\int_{\mathbf{t}} \int_r \int_\mu \int_{\sigma^2} g(\mathbf{t}, r, \mu, \sigma^2, \mathbf{b}) d\sigma^2 d\mu dr d\mathbf{t}} \tag{12}$$

where $g(.)$ is the transformed density as described above while the integral in the denominator is a normalizing constant. It is difficult to evaluate the integral analytically but relatively simple to develop an MCMC algorithm to generate the posterior density numerically and we use this strategy. Note that the evaluation of this density under the autocorrelation and independent rates models is essentially similar.

Multiple Loci

The above result can be readily extended to multiple loci if lineage-specific rates vary independently across loci, as assumed in our model. If there are K loci, the posterior density will have $s - 1 + 3K$ parameters: the $s - 1$ divergence times, and three parameters at each locus: μ, σ^2, and the rate r for the left daughter branch of the root. The transformed density is

$$g_K\left(\mathbf{t}, r^{(K)}, \mu^{(K)}, \sigma^{2(K)}, \mathbf{b}^{(K)}\right)$$
$$= f_K\left(\mathbf{t}, r^{(K)}, \mu^{(K)}, \sigma^{2(K)}, r^{-1}\left[\mathbf{b}^{(K)}\right]\right) \times \left(\prod_{j=1}^{2s-3} T_j\right)^K$$

where $r^{(K)}$ is a vector of the K rates across the K loci for the left daughter branch of the root, and $r^{-1}[\mathbf{b}^{(K)}]$ is the inverse function that maps the vector of branch lengths for each locus to the locus-specific rates. The joint posterior density given the branch lengths at the K loci is

$$g_K\left(\mathbf{t}, r^{(K)}, \mu^{(K)}, \sigma^{2(K)} | \mathbf{b}^{(K)}\right)$$
$$= \frac{g_K\left(\mathbf{t}, r^{(K)}, \mu^{(K)}, \sigma^{2(K)}, \mathbf{b}^{(K)}\right)}{\int_{\mathbf{t}} \int_{r^{(K)}} \int_{\mu^{(K)}} \int_{\sigma^{2(K)}} g_K\left(\mathbf{t}, r^{(K)}, \mu^{(K)}, \sigma^{2(K)}, \mathbf{b}^{(K)}\right) d\sigma^{2(K)} d\mu^{(K)} dr^{(K)} d\mathbf{t}}$$

An MCMC algorithm was implemented to evaluate this posterior density numerically.

Our prior model of rate change assumes that rates drift independently across loci. Nevertheless, the infinite-sites theory developed above applies if changes in rates are correlated across gene loci, as assumed by Kishino et al. (2001). The only change is that calculation of the prior density has to take into account the correlation structure in the model.

Similarly, the limiting posterior distribution can be derived for the case of multiple loci evolving under a global clock. This is an extension to the analysis of Yang and Rannala (2006), who considered one locus only. When the clock holds, the branch lengths are proportional across loci. The posterior density has one dimension, which may be the age of the root, as all other divergence times and branch rates are determined by the fixed branch lengths. It is noted that under the clock, increasing the number of loci does not lead to any reduction in the posterior credibility intervals (CIs).

Computer Simulation to Examine the Effect of the Number of Loci

Yang and Rannala (2006) conducted simulations to examine the impact of sequence length, fossil uncertainties, etc., on posterior time estimates, when hard and soft bounds are used to describe fossil uncertainties. A major result from that simulation is that typical sequence data sets appear to be nearly as informative as infinitely long sequences and most of the uncertainties in the posterior time estimates are therefore due to uncertainties in the fossils.

Here we conduct simulations to examine the effect of the number of loci on the precision of posterior time estimates when there are infinitely many sites at each locus, so that the branch lengths are estimated without error. We use the phylogeny of nine species shown in Fig. 3, which was used by Yang and Rannala (2006). Rate variation among lineages was modeled using a log-normal distribution with mean 1 and variance σ^2, corresponding to the independent rates model. At each locus, rates were sampled from the log-normal distribution and multiplied by the divergence times to generate branch lengths in units of expected number of substitutions. The data were then analyzed using the infinite-sites theory, as described above. We examined the effects of among-lineage rate variation, the number of loci, and the level of uncertainty in fossil calibrations on the cumulative width of the posterior CIs for divergence times (that is, the sum of the posterior CI widths for all node ages on the tree). This simulation design allows us to potentially examine the relative importance of three sources of uncertainty affecting divergence time estimation: fossil calibrations (improved by sampling more fossils with narrower strata ranges); uncertain locus-specific branch lengths due to finite sequence length (improved by sequencing more sites per locus); and among-lineage rate variation (improved by either sampling more loci or reducing the among-lineage rate variation at one or more loci).

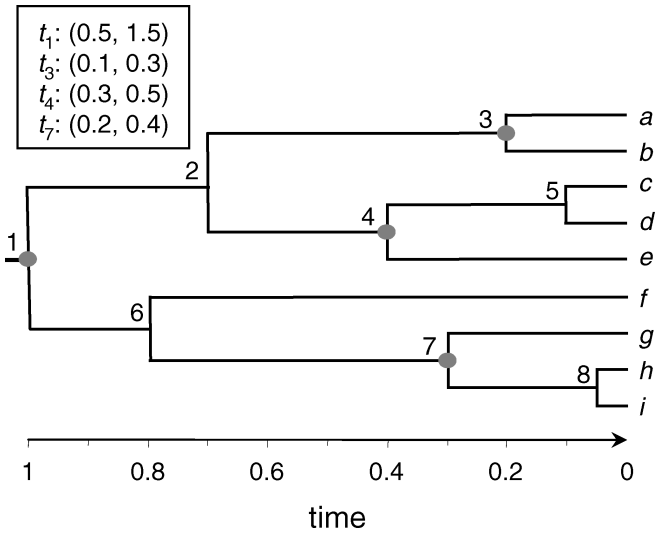

FIGURE 3. A tree of nine species used in computer simulation to examine the performance of divergence time estimation under variable-rates models. The true ages of nodes 1, 2, . . . , 8 are 1, 0.7, 0.2, 0.4, 0.1, 0.8, 0.3, and 0.05, as indicated by the time line. If one time unit is 100 million years, the age of the root will be 100 My. Fossil calibration information, in the form of minimal and maximal bounds, is available for nodes 1, 3, 4, and 7, as shown in the inset.

TABLE 1. The cumulative width of the 95% posterior CIs of divergence times when data of infinite sites are analyzed under the clock and variable-rates models. The sum of the widths of the 95% posterior CIs for the eight divergence times in the tree of Figure 3 is shown as a function of the number of loci, the model of rate evolution (C1 for global clock, C2 for the independent-rates model, and C3 for the correlated-rates model), and among-lineage rate variation reflected by σ, the standard deviation in the log-normal distribution used to simulate branch-specific substitution rates. In the third set of simulations, $\sigma_1 = 0.05$ for the first locus, whereas $\sigma_{\neq 1} = 0.25$ for all other loci.

No. loci	$\sigma = 0$	$\sigma = 0.05$,		$\sigma_1 = 0.05$, $\sigma_{\neq 1} = 0.25$		$\sigma = 0.25$		$\sigma = 0.50$	
	C1	C2	C3	C2	C3	C2	C3	C2	C3
1	0.40	0.68	1.37	1.44	2.18	2.46	2.64	2.93	2.89
5	0.40	0.47	0.68	1.22	1.76	1.64	1.77	1.78	2.00
10	0.40	0.41	0.56	1.06	1.43	1.19	1.57	1.46	1.42
30	0.40	0.38	0.42	0.54	0.89	1.14	1.01	1.04	0.95
50	0.40	0.38	0.35	0.50	0.76	0.51	0.75	0.87	0.77

The results are presented in Table 1. It is clear that when among-lineage rate variation exists at all loci, the width of the CIs converges (with increasing numbers of loci) to the width for a single locus under a perfect molecular clock. If even one locus exists that is highly clock-like, this locus tends to dominate inferences and including additional loci will have little effect.

In our previous study (Yang and Rannala, 2006) we showed that for perfectly informative (effectively infinite) sequences an exact linear relationship exists between the width of the CI and the mean of the posterior distribution of divergence times. Simulation results of Figs. 4 and 5 suggest that such a relationship exists also under the variable-rates models when the number of sites at each locus is infinite and the number of loci approaches infinity. It is noteworthy that increasing the number of loci has two effects. First, the relationship between the posterior mean and the posterior CI width becomes increasingly linear. Second, the slope of the regression line is reduced, until eventually it converges to the slope obtainable under a molecular clock with a single locus and an infinite number of sites. In the first set of simulations (Fig. 4), there is not much rate variation among lineages, and 30 loci appear to be close to the limit as the two regression lines in Figure 4 are very close. In the second set of simulations (Fig. 5), there is far more rate variation, and even 50 loci are far away from the limit.

ANALYSIS OF EMPIRICAL DATA SETS

We apply our new methods to two empirical data sets, for comparison with previous analyses. The first data set consists of mitochondrial protein-coding genes from seven species of apes (Fig. 6), compiled and analyzed by Cao et al. (1998). Because of the closeness of the species concerned, the molecular clock assumption

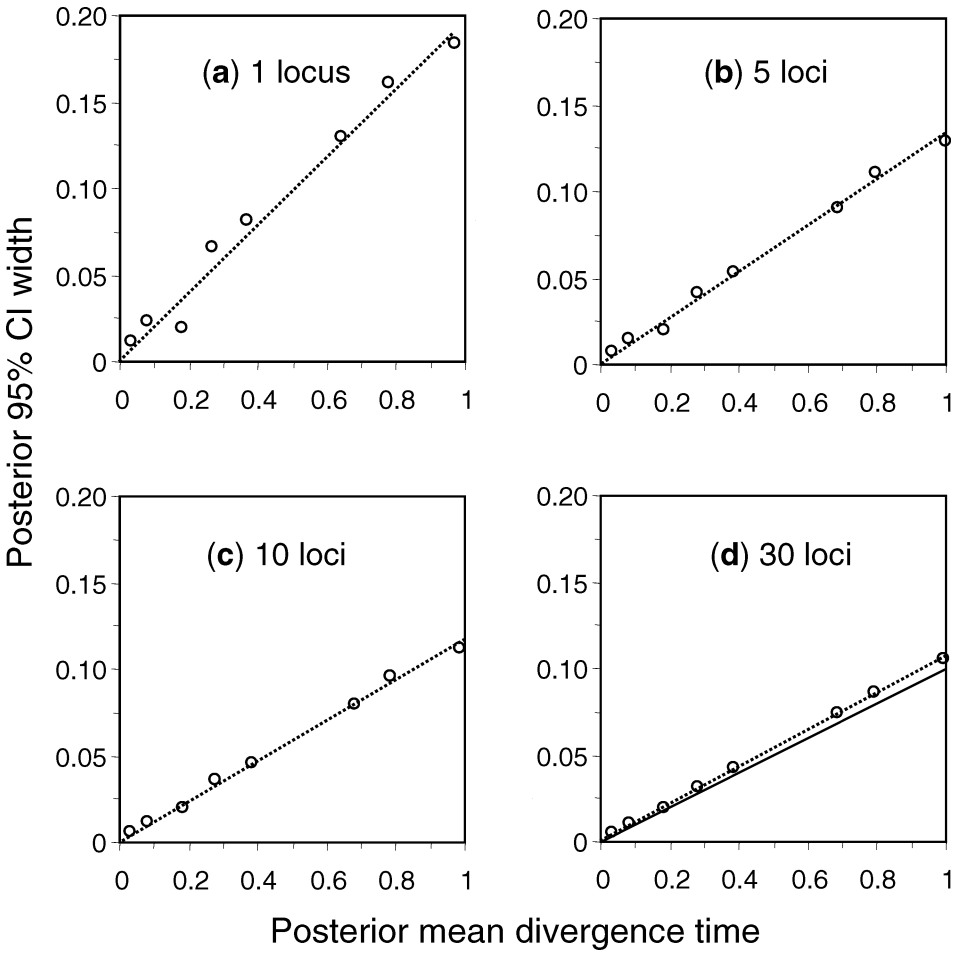

FIGURE 4. Widths of the 95% posterior CIs plotted against the posterior means of the divergence times for infinite-sites simulations using different numbers of loci. The parameter for rate-variation among loci was $\sigma = 0.05$. The independent rates model (clock 2) was used both to simulate and to analyze the data. Data (branch lengths) for 30 loci were simulated, and 1, 5, 10, or all 30 of them were analyzed to calculate results for the four plots. The solid line in plot (d) represents the results for one locus under the global clock, which is the limit for the variable-rates model when the number of loci reaches infinity.

is not seriously violated. The data set was analyzed by Yang and Rannala (2006) under the clock assumption. Here we apply the two variable-rates models for comparison. The second data set consists of nuclear genes from 38 species of the cat family (Fig. 7), analyzed by Johnson et al. (2006) using the method of Thorne et al. (1998) and Kishino et al. (2001), which relaxes the molecular clock assumption. Here we apply our methods for comparison.

Divergence Times of Apes

This data set consists of all twelve protein-coding genes encoded by the same strand of the mitochondrial genome from seven species of apes (Cao et al., 1998). The species phylogeny is shown in Fig. 6. See Cao et al. (1998) for the GenBank accession numbers. The alignment contains 3331 nucleotides at each codon position. We ignore the differences among the 12 genes but accommodate the huge differences among the codon positions. We use the HKY85+Γ_5 substitution model (Hasegawa et al., 1985), with a discrete gamma model of variable rates among sites, with five rate categories used (Yang, 1994). Each codon position has its own substitution rate, transition/transversion rate ratio κ, and gamma shape parameter α (Yang, 1996). Parameter κ is assigned the gamma prior $G(6, 2)$, whereas α has the gamma prior $G(1, 1)$. The data set is very informative about κ and α and the prior has little effect on the posterior for these parameters. The nucleotide frequencies are fixed at the observed frequencies. The overall substitution rate μ is assigned the gamma prior $G(2, 2)$ with mean 1 and variance 1/2. Here one time unit is 100 My, so the mean rate is one substitution per site per 10^8 years, which appears typical for an average mitochondrial rate in primates. The relatively large variance means that this prior is quite diffuse. Parameter σ^2 reflects the amount of rate variation across lineages or how seriously the molecular clock is violated. It is assigned a gamma prior $G(1, 10)$, with mean 0.1, variance 0.01, with the small mean to reflect

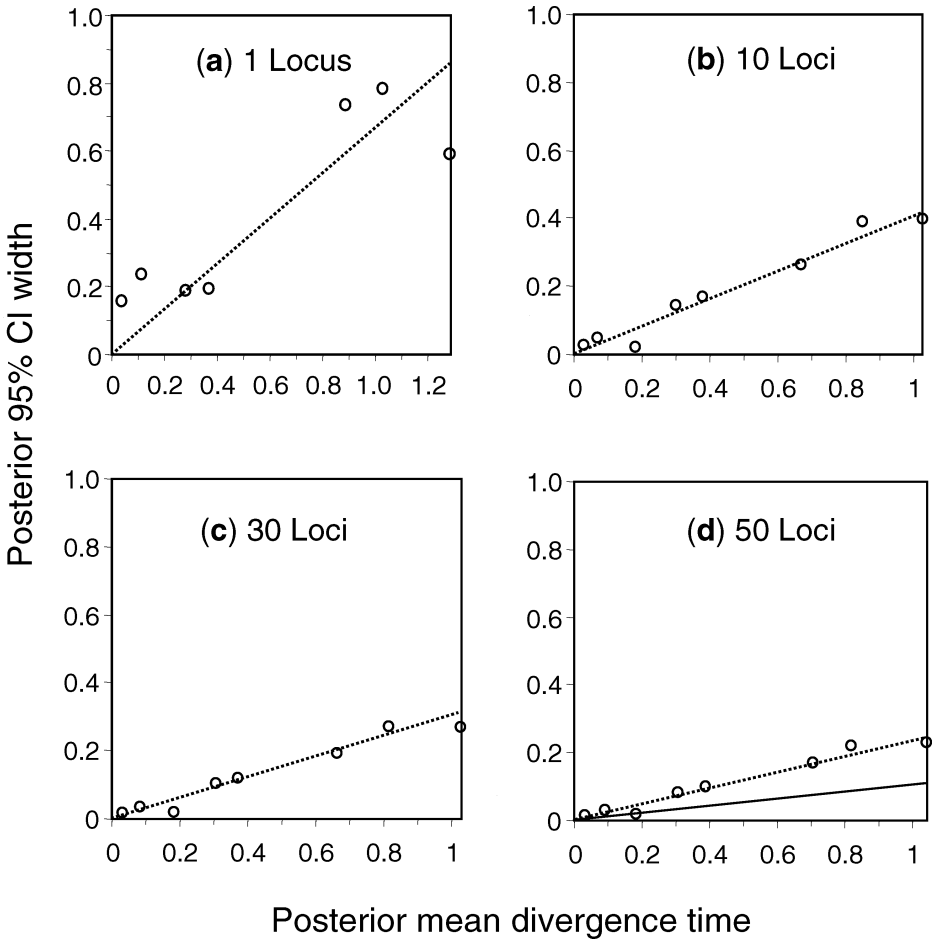

FIGURE 5. Widths of the 95% posterior CIs plotted against the posterior means of divergence times for infinite-sites simulations using 1, 5, 10, or 50 loci. The parameter for rate variation among loci is $\sigma = 0.25$. See legend to Figure 4 for more details.

the fact that the molecular clock roughly holds for these data. In the birth-death process with species sampling, we fix the birth and death rates at $\lambda = \mu = 2$, with the sampling fraction $\rho = 0.01$. For this data set, those priors are found to have only a minor effect on the posterior time estimates.

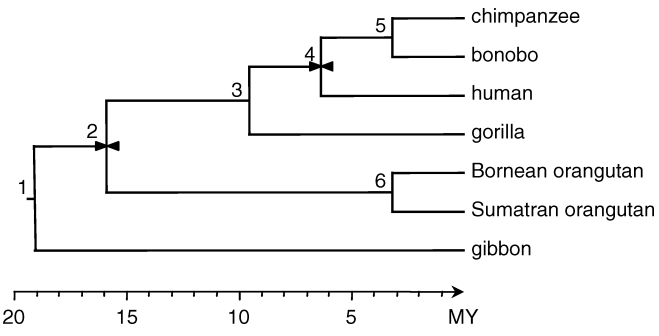

FIGURE 6. The tree of seven ape species for the mitochondrial data set of Cao et al. (1998). The branches are drawn to show posterior means of divergence times estimated under the autocorrelated-rates model (clock 3 in Table 2). Nodes 2 and 4 are used as fossil calibrations. The HKY85+Γ_5 substitution model was used to analyze the three codon positions simultaneously, with different parameters used to account for their differences.

The phylogenetic tree is shown in Figure 6. As in Yang and Rannala (2006), two fossil calibrations are used in our Bayesian analysis. The first is for the human-chimpanzee divergence, assumed to be between 6 and 8 My, with a most likely date of 7 My (Cao et al., 1998). A gamma prior $G(186.2, 2672.6)$, with one time unit representing 100 My, is used for the node age, with a cumulative tail probability of 5% determined according to the method of Yang and Rannala (2006). The second calibration is for the divergence of the orangutan from the African apes, assumed to be between 12 and 16 My, with a most likely date of 14 My (Raaum et al., 2005). The prior is specified as "$> 0.12 = 0.139 < 0.16$," and the gamma $G(186.9, 1337.7)$ is fitted, with tail probabilities 2.2 and 2.7%.

The MCMC was run for 40,000 iterations, after a burn-in of 4000 iterations. For each analysis, the MCMC algorithm was run at least twice using different starting values to confirm convergence to the same posterior.

Table 2 shows the posterior means and 95% CIs of the six divergence times. The posterior distributions for μ are shown as well. The posterior mean of μ suggests that

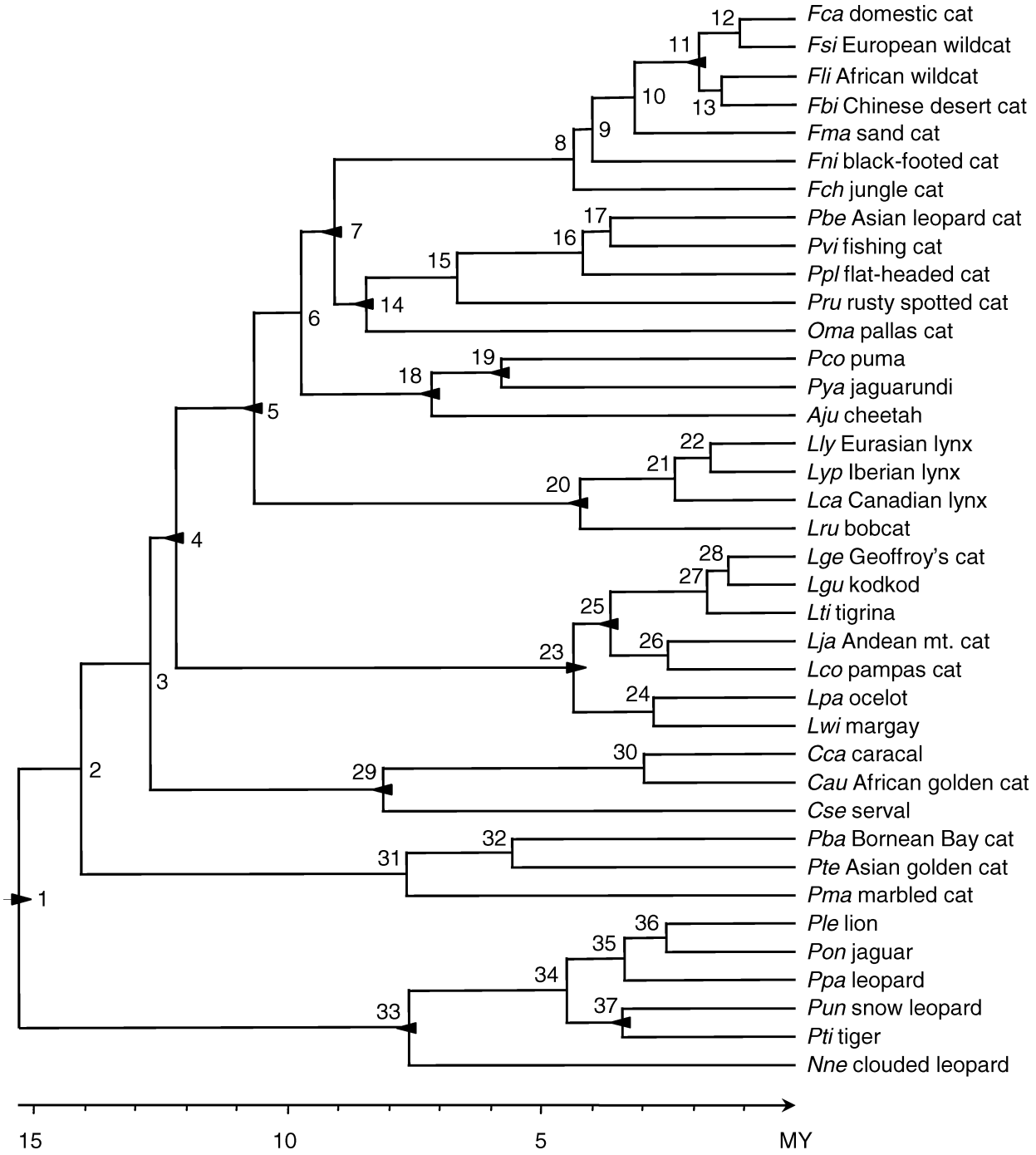

FIGURE 7. A phylogeny for 38 modern cat species, from Johnson et al. (2006). The branches are drawn in proportion to the posterior means of divergence times estimated under the HKY85+Γ_5 model and with autocorrelated rates (38 species, clock 3 in Table 3). Fourteen nodes have fossil calibration information, as indicated on the tree.

the substitution rate is about 0.50, 0.17, and 3.1×10^{-8} at first, second, and third codon positions, respectively. Posterior results for other parameters such as κ, α, σ^2 and the rates are not shown. The estimates under the independent rates (clock 2) and correlated rates (clock 3) models are rather similar to those under the clock model (clock 1). This similarity is expected as the molecular clock is not seriously violated for this data set. The posterior CIs are slightly wider under the variable-rates models than under the clock model, reflecting the relaxed assumptions of the model.

No fossil calibration is available at the root and no constraint is placed on the root age in the above analysis, following Yang and Rannala (2006). The means and 95%

TABLE 2. Posterior mean and 95% CIs of divergence times (My) and substitution rates estimated under different clock models for the mtDNA data. The three codon positions are analyzed as a combined data set under the HKY85 $+\Gamma_5$ model, with different parameters for each position. Divergence times are defined in Figure 1. The rates μ_1, μ_2, and μ_3 are for the three codon positions, respectively, measured by the expected number of substitutions per site per 100 My.

	Prior	Clock 1	Clock 2	Clock 3
t_1 (root)	1376 (268, 4857)	19.8 (17.5, 22.2)	19.9 (16.4, 24.1)	19.1 (16.5, 22.4)
t_2	14.0 (12.0, 16.1)	16.3 (14.6, 18.1)	15.8 (14.1, 17.7)	16.3 (14.6, 18.1)
t_3	10.5 (6.9, 14.5)	8.6 (7.6, 9.6)	9.0 (7.8, 10.3)	9.1 (7.9, 10.4)
t_4	7.0 (6.0, 8.0)	6.1 (5.5, 6.8)	6.3 (5.6, 7.1)	6.2 (5.5, 6.9)
t_5	3.5 (0.2, 7.0)	2.0 (1.8, 2.4)	2.3 (1.7, 3.0)	2.3 (1.9, 2.9)
t_6	7.0 (0.3, 14.0)	4.1 (3.5, 4.7)	4.4 (3.5, 5.6)	4.5 (3.6, 5.4)
μ_1	1.00 (0.12, 2.80)	0.49 (0.43, 0.57)	0.50 (0.41, 0.61)	0.49 (0.40, 0.58)
μ_2	1.00 (0.12, 2.80)	0.17 (0.14, 0.20)	0.17 (0.13, 0.22)	0.17 (0.14, 0.21)
μ_3	1.00 (0.12, 2.80)	3.11 (2.75, 3.51)	3.13 (2.58, 3.77)	3.22 (2.66, 3.86)

CIs for divergence times in the prior listed in Table 2 are obtained by running the MCMC without data; that is, by fixing $f(D|\mathbf{t}, \mathbf{r}) = 1$ in Equation (1). The prior mean for the root age from the birth-death process, at 1376 My, is huge. We found it in general beneficial to place a weak upper bound on the root age (maximal age) even if no fossil calibration is available at the root, especially when the molecular clock is seriously violated. For this data set, however, doing so had only a minor effect on posterior time estimates, perhaps because the clock roughly holds. For example, with the root age constrained to be <100 My, the prior mean and 95% CI for the root age became 58 My (17, 100). The posterior mean and 95% CI for the root age under clock 3 (correlated rates model) became 18.8 My (16.4, 21.9), very close to estimates in Table 2 obtained without placing any constraint on the root age.

The widths of the 95% posterior CIs for the eight node ages are plotted against their posterior means in Figure 8a. The slope of the regression line indicates the amount of uncertainty in posterior time estimates that cannot be removed by increasing sequence data; this is about 0.27 My of the width of the 95% posterior CI per My of divergence time. The considerable scatter around the regression line indicates that additional sequence data will very likely improve the precision of the estimates.

Divergence Times of Cats

The data set of Johnson et al. (2006) consists of 38 species of modern cats (family Felidae) plus seven outgroup species from feliform carnivoran families. The phylogeny of the 38 ingroup species is shown in Fig. 7, extracted from the phylogeny of Johnson et al. (2006). Following those authors, we used 30 nuclear genes (19 autosomal, 5 X-linked, and 6 Y-linked) for divergence time estimation. The results were compared with those obtained by Johnson et al. (2006) using the method of Thorne et al. (1998) and Kishino et al. (2001). Johnson et al. (2006) removed some small regions of the alignment. These appear reliable and are retained in our analysis. A total of 19,984 base pairs are included in the alignment. Some genes (loci) are missing in some species and the missing data are coded as question marks. See Johnson et al. (2006) for GenBank accession numbers.

Whereas the method of Thorne et al. (1998) requires outgroups, our method does not. Maximum likelihood

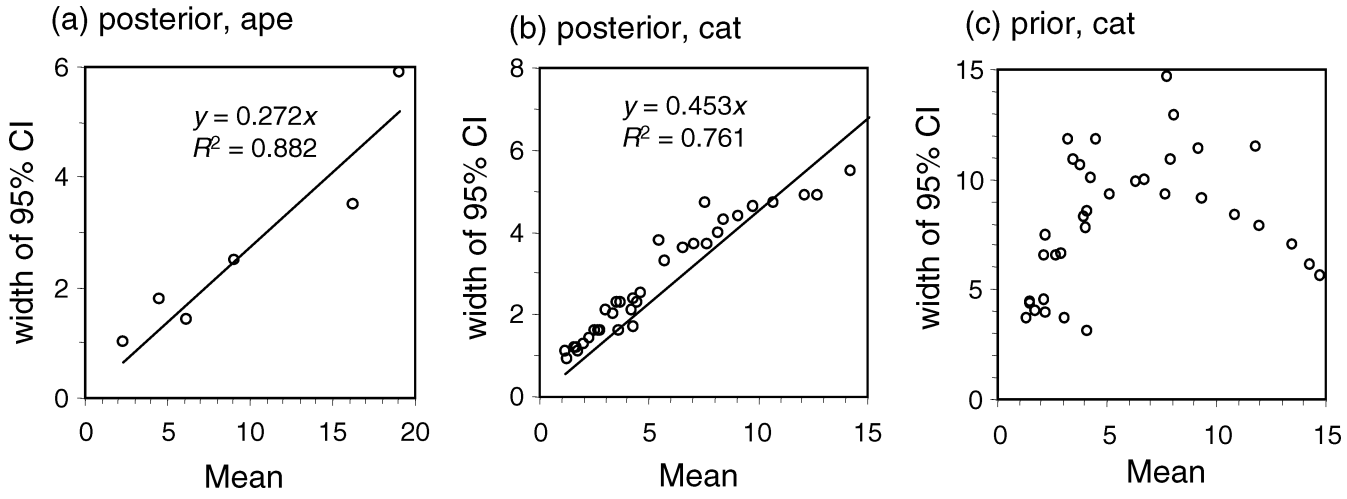

FIGURE 8. The 95% posterior CI widths plotted against the posterior means of divergence times in the analysis of the (a) ape and (b) cat data sets. The correlated-rates model (clock 3, 38 species in Table 3) was used in both analyses. The prior CI widths and prior means are plotted for the cat data in (c).

TABLE 3. Posterior mean and 95% CIs of divergence times (My) for the Felidae data of Johnson et al. (2006). Node numbers refer to those of Figure 7, following Johnson et al. (2006). "Fossil" represents fossil calibration information, specified as minimal or maximal ages for the node.

				39 species		38 species		
Node	Fossil	Johnson et al.	Prior	Clock 2	Clock 3	Clock 1	Clock 2	Clock 3
1	<16	10.8 (8.4, 14.5)	15.4 (12.4, 17.4)	12.6 (9.6, 16.0)	14.0 (10.1, 16.7)	15.2 (12.2, 17.1)	15.0 (12.0, 16.9)	15.3 (12.4, 17.1)
2		9.4 (7.4, 12.8)	14.8 (11.4, 17.1)	11.0 (8.3, 14.0)	12.2 (8.7, 15.0)	14.1 (11.2, 16.2)	14.0 (11.1, 16.1)	14.2 (10.9, 16.4)
3		8.5 (6.7, 11.6)	14.3 (10.6, 16.8)	9.9 (7.5, 12.7)	11.1 (7.9, 13.6)	12.8 (10.1, 14.6)	12.6 (9.9, 14.6)	12.7 (9.7, 14.6)
4	>5	8.1 (6.3, 11.0)	13.5 (9.4, 16.4)	9.4 (7.1, 12.0)	10.5 (7.5, 13.0)	12.2 (9.7, 14.1)	12.0 (9.4, 14.0)	12.1 (9.1, 14.0)
5	>5.3	7.2 (5.6, 9.8)	12.0 (7.7, 15.5)	8.3 (6.3, 10.7)	9.4 (6.8, 11.8)	10.8 (8.5, 12.4)	10.5 (8.2, 12.4)	10.7 (7.8, 12.5)
6		6.7 (5.3, 9.2)	10.9 (6.4, 14.9)	7.6 (5.8, 9.9)	8.7 (6.3, 10.9)	10.0 (7.9, 11.5)	9.7 (7.4, 11.5)	9.8 (6.9, 11.5)
7	>4.2	6.2 (4.8, 8.6)	9.3 (4.9, 14.0)	7.0 (5.3, 9.2)	8.1 (5.8, 10.2)	9.3 (7.3, 10.8)	9.0 (6.9, 10.7)	9.1 (6.2, 10.6)
8		3.4 (2.4, 4.9)	6.7 (2.4, 12.4)	4.1 (3.1, 5.5)	4.5 (3.1, 6.1)	5.4 (4.2, 6.4)	5.4 (4.0, 6.7)	4.5 (3.2, 5.5)
9		3.0 (2.2, 4.4)	5.1 (1.7, 11.1)	3.7 (2.7, 4.8)	4.1 (2.8, 5.4)	4.9 (3.8, 5.8)	4.8 (3.5, 6.0)	4.2 (2.9, 5.0)
10		2.5 (1.7, 3.7)	4.0 (1.3, 9.6)	3.0 (2.2, 4.0)	3.2 (2.1, 4.4)	4.1 (3.2, 5.0)	4.0 (2.9, 5.0)	3.4 (2.3, 4.3)
11	>1	1.4 (0.9, 2.2)	2.9 (1.0, 7.7)	1.8 (1.2, 2.5)	1.7 (1.2, 2.3)	2.3 (1.7, 2.9)	2.3 (1.6, 2.9)	2.0 (1.3, 2.6)
12		1.0 (0.6, 1.6)	1.5 (0.1, 4.4)	1.1 (0.7, 1.8)	1.1 (0.6, 1.5)	1.5 (1.0, 2.0)	1.5 (0.9, 2.1)	1.2 (0.7, 1.8)
13		1.2 (0.7, 1.9)	1.5 (0.1, 4.4)	1.4 (0.9, 2.0)	1.5 (1.1, 2.0)	1.9 (1.4, 2.4)	1.8 (1.3, 2.5)	1.6 (1.0, 2.2)
14	>1	5.9 (4.5, 8.2)	6.4 (1.9, 11.8)	6.5 (4.9, 8.6)	7.6 (5.5, 9.6)	8.7 (6.8, 10.1)	8.3 (6.3, 10.0)	8.4 (5.7, 10.0)
15		4.6 (3.4, 6.5)	4.1 (1.0, 9.6)	5.0 (3.6, 6.7)	6.0 (4.3, 7.9)	6.6 (5.1, 7.9)	6.4 (4.7, 7.9)	6.6 (4.5, 8.1)
16		2.9 (2.0, 4.3)	2.7 (0.4, 6.9)	3.2 (2.3, 4.4)	3.8 (2.7, 5.0)	4.3 (3.3, 5.2)	4.1 (2.9, 5.2)	4.3 (2.9, 5.3)
17		2.6 (1.7, 3.8)	1.5 (0.0, 4.4)	2.8 (1.9, 3.8)	3.4 (2.4, 4.5)	3.8 (2.9, 4.7)	3.6 (2.4, 4.6)	3.7 (2.4, 4.7)
18	>3.8	4.9 (3.9, 6.9)	7.7 (3.8, 13.1)	5.3 (4.0, 7.1)	6.2 (4.4, 8.1)	7.0 (5.5, 8.3)	6.7 (4.9, 8.3)	7.1 (5.0, 8.7)
19	>1.8	4.2 (3.2, 6.0)	4.1 (1.8, 9.6)	4.3 (3.1, 5.9)	5.0 (3.4, 6.7)	5.6 (4.3, 6.8)	5.5 (3.9, 7.0)	5.7 (3.9, 7.2)
20	>2.5	3.2 (2.5, 4.7)	8.0 (2.8, 13.7)	3.5 (2.5, 4.9)	3.8 (2.6, 5.3)	4.6 (3.5, 5.7)	4.4 (3.1, 5.7)	4.3 (3.1, 5.5)
21		1.6 (1.1, 2.6)	4.3 (0.7, 10.8)	1.8 (1.2, 2.6)	2.0 (1.3, 3.0)	2.3 (1.7, 3.0)	2.3 (1.5, 3.2)	2.3 (1.7, 3.1)
22		1.2 (0.7, 2.0)	2.1 (0.1, 6.6)	1.3 (0.8, 2.1)	1.5 (0.9, 2.2)	1.7 (1.2, 2.3)	1.7 (1.0, 2.6)	1.7 (1.1, 2.3)
23	<5	2.9 (2.0, 4.2)	4.1 (2.1, 5.2)	3.6 (2.6, 4.7)	3.7 (2.7, 4.9)	4.4 (3.4, 5.0)	4.5 (3.5, 5.2)	4.3 (3.3, 5.0)
24		1.6 (1.0, 2.4)	2.2 (0.1, 4.7)	2.4 (1.6, 3.4)	2.4 (1.6, 3.5)	2.9 (2.1, 3.6)	3.0 (2.0, 3.9)	2.8 (2.0, 3.6)
25	>1	2.4 (1.7, 3.6)	3.1 (1.2, 4.8)	2.9 (2.1, 3.9)	3.1 (2.2, 4.1)	3.7 (2.8, 4.4)	3.7 (2.8, 4.4)	3.6 (2.7, 4.3)
26		1.8 (1.2, 2.7)	1.8 (0.1, 4.1)	2.0 (1.3, 2.8)	2.2 (1.4, 3.0)	2.6 (1.8, 3.3)	2.5 (1.7, 3.4)	2.5 (1.8, 3.4)
27		0.9 (0.6, 1.5)	2.2 (0.4, 4.3)	1.4 (0.9, 2.1)	1.4 (0.9, 2.2)	1.7 (1.2, 2.3)	1.8 (1.2, 2.5)	1.8 (1.3, 2.4)
28		0.7 (0.4, 1.2)	1.3 (0.0, 3.7)	1.0 (0.6, 1.7)	1.1 (0.6, 1.7)	1.3 (0.9, 1.8)	1.3 (0.8, 1.9)	1.3 (0.9, 1.8)
29	>3.8	5.6 (4.1, 7.9)	9.2 (3.9, 15.3)	6.5 (4.6, 8.9)	7.2 (4.9, 9.4)	8.3 (6.4, 9.8)	8.2 (6.0, 10.3)	8.2 (6.1, 10.1)
30		1.9 (1.2, 2.9)	3.5 (0.2, 11.1)	2.3 (1.5, 3.5)	2.7 (1.7, 3.9)	2.9 (2.1, 3.8)	3.0 (2.0, 4.2)	3.0 (2.0, 4.1)
31		5.9 (4.3, 8.4)	7.8 (1.1, 15.8)	5.7 (4.0, 7.8)	6.7 (4.7, 8.7)	7.6 (5.9, 9.1)	7.3 (5.2, 9.3)	7.7 (5.7, 9.4)
32		4.3 (3.0, 6.4)	3.2 (0.1, 11.9)	3.9 (2.6, 5.7)	4.6 (3.0, 6.6)	5.2 (3.9, 6.6)	5.1 (3.5, 7.2)	5.5 (3.7, 7.5)
33	>3.8	6.4 (4.5, 9.3)	11.9 (4.8, 16.3)	6.4 (4.6, 8.9)	7.2 (4.8, 10.3)	7.7 (6.1, 9.0)	7.6 (5.7, 9.6)	7.6 (5.8, 10.5)
34		3.7 (2.4, 5.8)	8.1 (2.2, 15.1)	3.5 (2.6, 4.7)	4.3 (2.8, 6.1)	4.5 (3.5, 5.4)	4.4 (3.3, 5.6)	4.6 (3.5, 6.0)
35		2.9 (1.8, 4.6)	4.5 (0.7, 12.5)	2.7 (1.9, 3.7)	3.3 (2.2, 4.7)	3.4 (2.6, 4.2)	3.4 (2.4, 4.4)	3.5 (2.5, 4.8)
36		2.1 (1.2, 3.5)	2.2 (0.1, 7.5)	2.0 (1.3, 2.8)	2.4 (1.6, 3.5)	2.5 (1.8, 3.2)	2.5 (1.6, 3.4)	2.7 (1.9, 3.5)
37	>1	2.9 (1.8, 4.6)	3.8 (1.0, 11.7)	2.7 (1.9, 3.7)	3.2 (2.0, 5.0)	3.5 (2.7, 4.4)	3.4 (2.4, 4.4)	3.5 (2.5, 4.8)

estimates (MLEs) of branch lengths on the tree of all species without assuming the clock (results not shown) suggest that the seven outgroup species are very distantly related to the ingroup species, and the ingroup species evolve in a clock-like fashion. Thus we analyzed two data sets. The first one includes the 38 Felidae species as well as banded linsang (*Prionodon linsang*), with the six other, more distant, outgroups excluded. The results obtained from this data set under the independent rates (clock 2) and correlated rates (clock 3) models are summarized under the heading "39 species" in Table 3. All 16 calibrations of Johnson et al. (2006) are used. Fourteen of the calibrations place minimal or maximal ages for nodes on the ingroup tree and are shown in Fig. 7 and Table 3. The calibration at node 4, a minimal age of 5 My, is redundant as its daughter node (node 5) has a minimal age of 5.3 My. Two additional calibrations specify the minimal (28 My) and maximal (50 My) ages of the ancestor of *Prionodon linsang* and the Felidae clade.

The second data set includes the 38 ingroup species only, with all seven outgroup species excluded. Removal of linsang means that we can use only the 14 calibrations shown on Fig. 7 and Table 3. The results of this analysis are summarized under the heading "38 species" in Table 3.

The MCMC was run for 20,000 iterations, after a burnin of 500 iterations. For each analysis, the MCMC algorithm was run at least twice using different starting values to confirm convergence to the same posterior. We used both the JC69 (Jukes and Cantor, 1969) and HKY85+Γ_5 (Hasegawa et al., 1985; Yang, 1994) substitution models. They produced very similar estimates of divergence times, even though the log likelihood under HKY85+Γ_5 is higher than under JC69 by more than 1200 units. Simultaneous use of multiple calibrations appears to have made divergence time estimation rather robust to the assumed substitution model. We present the results under HKY85+Γ_5 only. Under the global clock (clock 1), the mean of the overall rate is 0.066 substitutions per site per 10^8 years, with the 95% posterior CI to be (0.057, 0.082). The posterior means and 95% CIs for the substitution parameters under HKY85+Γ_5 are 3.70 (2.88, 4.07) for κ and 0.22 (0.15, 0.51) for α.

We discuss the results from the 39-species data set first. The time estimates obtained under the two relaxed-clock models (clock 2 and clock 3) are similar, with the clock

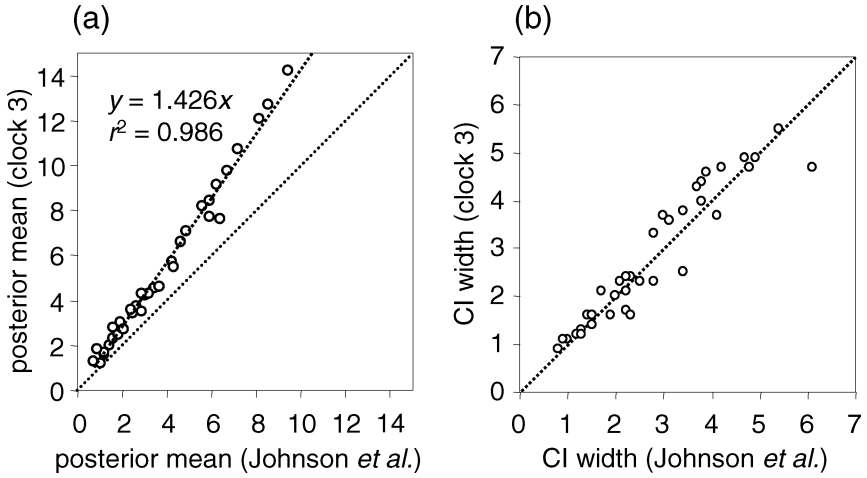

FIGURE 9. (a) The posterior means of divergence times estimated by Johnson et al. (2006) and in this study (38 species clock 3 in Table 3). (b) The 95% posterior CI widths of divergence times calculated in the two studies.

3 estimates being slightly older. However, estimates of ages for most nodes under both models are older than those obtained by Johnson et al. (2006). For example, the age of the ingroup root (node 1) is dated to 12.6 My with the 95% CI to be (9.6, 16.0) under clock 2 and 14.0 (10.1, 16.7) under clock 3. The corresponding estimates obtained by Johnson et al. are 10.8 (8.4, 14.5). The CIs obtained under different models have similar widths.

For the analysis of the data set of 38 ingroup species only, we used the global clock model, in addition to the independent rates (clock 2) and correlated rates (clock 3) models (Table 3). With only the ingroup species, the molecular clock is not seriously violated. Indeed, these three models produced very similar results. However, all age estimates are older than those obtained from the 39-species data set and much older than estimates obtained by Johnson et al. (2006), even though the 95% posterior CIs overlap among the analyses. The posterior means of divergence times obtained under the correlated rates model (clock 3) from the 38-species data set in our analysis are about 1.426 times as old as those of Johnson et al. (2006), with a strong correlation ($r = 0.993$) between the two sets of estimates (Fig. 9a), indicating systematic differences between the two analyses. No apparent trend can be seen concerning the 95% CI widths, apart from the fact that the CIs from our analysis assuming soft bounds tend to be wider (Fig. 9b). The reasons for the systematic differences between methods are not clear. We speculate that the following factors may be important. First, species sampling appears to have a large effect on time estimation in this data set, with inclusion of outgroup species producing younger estimates for node ages in the ingroup tree; time estimates obtained from the 38-species data set are younger than those from the 39-species data set, which are in turn younger than those from the complete data set of Johnson et al. (2006). We note that the linsang sequence is very divergent from all 38 ingroup sequences, while the other six outgroups are even more divergent. Second, we used soft bounds to specify fossil calibrations, while Johnson et al. (2006) used hard bounds. Soft bounds may be advantageous when not all fossil calibrations are reliable (Yang and Rannala, 2006). Nevertheless, our incorporation of fossil information at multiple calibration nodes does not appear to deal properly with the inherent constraints on node ages for ancestral and descendant nodes. See below for an extensive discussion of the differences between our implementation and that of Thorne et al. (1998) and Kishino et al. (2001).

The widths of the 95% posterior CIs for the 37 node ages are plotted against their posterior means in Fig. 8b. Interestingly the relationship between the CI width and posterior mean is approximately linear for small and moderate values of node ages (<10 My), whereas for older ages, the posterior CIs tend to be narrower than expected by the linear relationship. For comparison, a similar plot is shown in Fig. 8c using the prior 95% CI widths and prior means.

DISCUSSION

There are a number of similarities between the algorithm we have implemented here and the algorithm of Thorne et al. (1998) and Kishino et al. (2001). Both use geometric Brownian motion to model the change of evolutionary rate to relax the molecular clock (although we have implemented an independent rates model as well). Both can analyze multiple genes or site partitions simultaneously while accounting for their differences in evolutionary dynamics, and both accommodate missing loci in some species. Both can also use multiple calibration points simultaneously, accommodating fossil uncertainties through the use of prior distributions for node ages.

There are also a number of differences between the two implementations. Whether the differences are important to posterior time estimation is not very clear and may

depend on the number and nature of fossil calibrations, the seriousness of violation of the molecular clock, etc. Further tests using both real and simulated data sets are needed to fully understand the effects of those factors and the differences between the methods.

First, as discussed by Yang and Rannala (2006), Thorne et al. (1998) used hard bounds for fossil calibrations, while we implemented soft bounds and flexible distributions to describe fossil uncertainties. Such priors may prove useful for adequately incorporating uncertainties in fossil ages (Tavare et al., 2002; Benton and Donoghue, 2007). Fossil information appears to be critically important to divergence time estimation, and we expect that this difference in implementation may be important in some data sets.

Second, the two programs differ in the likelihood calculation. Thorne et al. (1998) used a two-step procedure to calculate the likelihood approximately. The first step is to calculate MLEs of the branch lengths on the unrooted phylogeny including both ingroup and outgroup species without assuming the clock. The variance-covariance matrix for branch lengths in the rooted ingroup tree is calculated, using the local curvature of the likelihood surface. The second step uses an MCMC algorithm for estimating divergence times on the rooted tree of the ingroup species, employing a multivariate normal density of branch lengths to approximate the likelihood. When the MLEs of some branch lengths are zero, as sometimes occurs on a large phylogeny, the normal approximation may not work well but, in general, the accuracy of the approximation is not well understood. We instead use a rooted tree for the ingroup only, without the need for outgroups, and use the pruning algorithm of Felsenstein (1981) to calculate the likelihood exactly. However, the computational cost of our approach is much greater, and our current algorithm does not appear computationally feasible for data sets with >100 species. It will be interesting to compare the two algorithms to assess the effects of the approximate likelihood calculation.

Third, Thorne et al. (1998) used rates for branches to implement the geometric Brownian motion model of evolutionary rate drift, whereas Kishino et al. (2001) formulated the model using rates for nodes. We used rates for branches to remove a parameter (the rate at the root) but properly accommodate the correlation of rates between two daughter branches given the rate for the mother branch. We suspect that this difference is technical and should have little effect on posterior time estimation. We note that calculation of the average rate for each branch is approximate in both algorithms; ideally, the length of a branch should be calculated as an integral over the sample path of the geometric Brownian motion process, or otherwise calculation of the transition probability from one nucleotide to another along the branch has to take explicit account of the fluctuating rate. In addition to the autocorrelated rates model, we have implemented a model of independent rates, which appears to produce sensible time estimates and may be useful for assessing the effect of the prior on rates on estimates of divergence times.

Fourth, the two implementations used different priors for divergence times for nodes with no fossil calibrations. Kishino et al. (2001) used a recursive procedure, proceeding from ancestral to descendant nodes. The age of the root is assigned a gamma prior. Then a Dirichlet density is used to break the path from an ancestral node to a tip into time segments, corresponding to branches on that path. The prior thus favors equally spaced branch time lengths. We used the birth-death process with species sampling (Yang, 1997) to specify the prior on times, calculating the prior density analytically. Use of such an analytically tractable prior enabled us to use flexible prior information on fossil calibrations. The parameters in the birth and death process, such as the speciation rate, extinction rate, and sampling fraction, affect the shape of the tree. By adjusting these parameter values, the prior can generate trees of different shapes, including bush-like trees with short internal branches as well as trees with long internal branches, and may be useful for assessing the influence of the divergence-time prior on posterior time estimates.

A similar MCMC algorithm has recently been described by Drummond et al. (2006), which incorporates arbitrary prior distributions for fossils and allows the evolutionary rate to drift over time. The variable-rates models are implemented using discretization. The continuous distribution of branch rates is approximated using as many discrete rates as there are branches on the tree. In the MCMC, two branches are chosen at random with their rates switched. Additional proposals alter the rates themselves. Technically, this does not appear to be a correct implementation of the variable-rates models. Consider the rooted tree of two species with two branches. Let the two possible rates from the discretized distribution be r_1 and r_2. Under the independent rates model, the following four rate assignments to the two branches should have equal probabilities (1/4): r_1r_1, r_1r_2, r_2r_1, r_2r_2, where r_ir_j means that branch 1 has rate r_i and branch 2 has rate r_j and so on. In the implementation of Drummond et al., the combinations r_1r_1 and r_2r_2 are not possible and only r_1r_2 and r_2r_1 are allowed, each assigned equal probability (1/2). One effect of this implementation is that a negative correlation between branch rates is introduced into the prior so that the model tends to underestimate possible positive autocorrelations in rates across branches. The effect should become minor in large trees with many branches. At any rate, we suggest that independent developments of multiple MCMC algorithms are very important to furthering our understanding of this complex estimation problem and will benefit empirical biologists who are interested in applying such methods to estimation of species divergence times.

Software Availability

The MCMC algorithm described in this paper has been implemented in the MCMCtree program in the PAML package (Yang, 1997). The variable clock = 1, 2, 3 represents the global clock, independent rates, and correlated rates models, respectively.

ACKNOWLEDGMENTS

We thank Jeff Thorne, Hirohisa Kishino, and Frank Anderson for many constructive comments. This study was supported by Canadian Institutes of Health Research (CIHR) grant MOP 44064, National Institutes of Health grant HG01988 (both to B.R.), Natural Environment Research Councils grant NE/C509974/1, and a travel grant from the National Science Foundation of China (both to Z.Y.).

REFERENCES

Adkins, R. M., A. H. Walton, and R. L. Honeycutt. 2003. Higher-level systematics of rodents and divergence time estimates based on two congruent nuclear genes. Mol. Phylogenet. Evol. 26:409–420.

Bell, C. D., and M. J. Donoghue. 2005. Dating the Dipsacales: Comparing models, genes, and evolutionary implications. Am. J. Bot. 92:284–296.

Benton, M., and P. J. Donoghue. 2007. Paleontological evidence to date the tree of life. Mol. Biol. Evol. 24:26–53.

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.

Cao, Y., A. Janke, P. J. Waddell, M. Westerman, O. Takenaka, S. Murata, N. Okada, S. Paabo, and M. Hasegawa. 1998. Conflict among individual mitochondrial proteins in resolving the phylogeny of eutherian orders. J. Mol. Evol. 47:307–322.

Drummond, A. J., M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:e88.

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

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

Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.

Hasegawa, M., J. L. Thorne, and H. Kishino. 2003. Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution. Genes Genet. Syst. 78:267–283.

Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their application. Biometrika 57:97–109.

Huelsenbeck, J. P., B. Larget, and D. Swofford. 2000. A compound Poisson process for relaxing the molecular clock. Genetics 154:1879–1892.

Johnson, W. E., E. Eizirik, J. Pecon-Slattery, W. J. Murphy, A. Antunes, E. Teeling, and S. J. O'Brien. 2006. The late Miocene radiation of modern Felidae: A genetic assessment. Science 311:73–77.

Kishino, H., and M. Hasegawa. 1990. Converting distance to time: Application to human evolution. Methods Enzymol. 183:550–570.

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. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. Equations of state calculations for fast computing machines. J. Chem. Phys. 21:1087–1092.

Pereira, S. L., and A. J. Baker. 2006. A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. Mol. Biol. Evol. 23:1731–1740.

Raaum, R., K. Sterner, C. Noviello, C.-B. Stewart, et al. 2005. Catarrhine primate divergence dates estimated from complete mitochondrial genomes: Concordance with fossil and nuclear DNA evidence. J. Human Evol. 48:237–257.

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.

Springer, M. S., W. J. Murphy, E. Eizirik, and S. J. O'Brien. 2003. Placental mammal diversification and the Cretaceous-Tertiary boundary. Proc. Natl. Acad. Sci. USA 100:1056–1061.

Takezaki, N., A. Rzhetsky, and M. Nei. 1995. Phylogenetic test of the molecular clock and linearized trees. Mol. Biol. Evol. 12:823–833.

Tavare, S., C. R. Marshall, O. Will, C. Soligo, and R. D. Martin. 2002. Using the fossil record to estimate the age of the last common ancestor of extant primates. Nature 416:726–729.

Thorne, J. L., and H. Kishino. 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst. Biol. 51:689–702.

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. 1994. Estimating the pattern of nucleotide substitution. J. Mol. Evol. 39:105–111.

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

Yang, Z. 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13:555–556.

Yang, Z. 2004. A heuristic rate smoothing procedure for maximum likelihood estimation of species divergence times. Acta Zool. Sin. 50:645–656.

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

Yang, Z., and B. Rannala. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol. Biol. Evol. 23:212–226.

Yang, Z., and A. D. Yoder. 2003. Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute-looking mouse lemur species. Syst. Biol. 52:705–716.

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 4 December 2006; reviews returned 8 January 2007; final acceptance 22 February 2007
Associate Editor: Frank Anderson