Bayesian approaches are gaining popularity in phylogenetics (Rannala and Yang, 1996; Li et al., 2000) in part because of the ease with which one may compare a growing number of competing evolutionary hypotheses using Bayesian model selection (Huelsenbeck et al., 2001; Suchard et al., 2001, 2002). Bayesian approaches are used to estimate the posterior distribution of unknown model parameters given the observed phylogenetic data using Bayes theorem. Given a statistical model $M$ with parameter space $\theta$ and data $X$, Bayes theorem states that the posterior distribution of $\theta$ is proportional to the sampling density of the data given $\theta$, commonly referred to as the model likelihood, multiplied by the prior distribution of $\theta$:

$$p(\theta | X, M) = \frac{f(X | \theta, M)q(\theta | M)}{m(X | M)},$$

(1)

where the reciprocal proportionality constant $m(X | M) = \int_\theta f(X | \theta, M)q(\theta | M)d\theta$ is the marginal likelihood of the data $X$ given model $M$.

One may describe various evolutionary models by restricting one or more parameters to specific ranges of values and then may compare these different models using a Bayes factor, the ratio of the posterior probability of one model to that of another (the prior odds) divided by the ratio of the prior probabilities (the prior odds). The Bayes factor is a measure of the change in support for one model versus another given the data and is the Bayesian analogue of the likelihood ratio test (LRT). LRTs have been used very effectively in phylogenetics (Huelsenbeck and Rannala, 1997) but can be remiss in that the data are sparse and the space of possible evolutionary trees is discrete so that standard likelihood asymptotics may not apply (Goldman, 1993; Sinsheimer et al., 1996; Whelan and Goldman, 1999). Bayes factors have fewer difficulties in discrete spaces, because probability mass functions can be naturally substituted for continuous distributions, and do not rely on large sample asymptotics. Most importantly, when utilizing increasingly large data sets, a frequentist-based approach to model selection, such as the LRT, will almost always reject the restricted model in favor of a larger one, whereas the Bayes factor remains more conservative.

We recently published an article on Bayesian inference in phylogenetics that included tests of molecular clocks (Suchard et al., 2001). A molecular clock assumes that the rate of evolutionary change remains constant over time and across taxa, such that the evolutionary distances between any two taxa and their most recent common ancestor (MRCA) are equal. Imposing this restriction facilitates the incorporation of fossil evidence into analyses and allows divergence times to be estimated. Like many previous approaches to testing molecular clocks (Takezaki et al., 1995; Huelsenbeck et al., 2000), our tests give weight to only a subset of possible trees. Specifically, we limit ourselves to trees in which a particular taxon is assumed to have diverged first and then conduct the test conditional on the modal tree assuming this outgroup. Although in general using these tests may pose little bias, there are situations where there is uncertainty in the outgroup, particularly among organisms where historical information is sparse or absent, and incorrect inference may result. As an example, we consider inference among closely related organisms where the fossil record is incomplete, such as primates. To assume an outgroup in these cases may imprecisely estimate the variability of test statistics used for inference by limiting the parameter space and can lead to erroneous results (Taylor et al., 1996).
Here, we use an irreversible model of evolutionary change. This model allows for inference of the location of the root, the MRCA for all taxa under study. Our primary aim, however, is not to estimate this position because this task cannot be accomplished with great precision (Yang, 1994; Huelsenbeck et al., 2002). Rather, we use the irreversible model to explicitly integrate over all possible root positions, so that neither the tree nor an outgroup need be specified in the prior when testing molecular clocks. We then outline our prior assumptions, discuss inference methods based on Bayes factors, and lay out the necessary tools to calculate or estimate induced priors. An induced prior describes the marginal prior distribution of a function of model parameters that results from assuming a joint prior on the parameters. For example, assume \( \theta_1 \) and \( \theta_2 \) are two model parameters and we are interested in testing the sharp hypothesis that \( \theta_1 - \theta_2 = c \) for some constant \( c \). If we assume a joint prior \( q(\theta_1, \theta_2) \), then the relevant induced prior is \( q(\theta_1 - \theta_2) \). In complicated models, such as the graphical models in phylogenetics, determining the induced prior is not always a straightforward activity. The induced prior is a necessary component of the Bayes factor between two models related by a sharp hypothesis. We have employed these tools with a data set commonly used to report on new phylogenetic methodology and briefly discuss the general implications of our model.

**Statistical Model**

**Data and Notation**

We used aligned DNA or RNA sequences to determine the evolutionary histories relating \( N \) organisms. We assumed an irreversible continuous-time Markov chain model for evolutionary change at equilibrium, characterized in terms of an unknown evolutionary history and an infinitesimal rate matrix. Here, an evolutionary history consists of a rooted, bifurcating tree \( \tau_N \) where root node \( R \) is the MRCA of all \( N \) extant taxa and a vector of edge weights (branch lengths) \( b \in T \) for \( b = 1, 2, \ldots, 2N - 2 \). There exist

\[
E_N = \frac{(2N - 3)!}{2^{N-2}(N-2)!}
\]

different rooted trees connecting \( N \) taxa (Felsenstein, 1978).

**Rootable Model**

We used a restricted parameterization of the infinitesimal rate matrix \( Q \) introduced by Schadt et al. (1998) (SSL98) that consists of eight free, nonnegative parameters and is given by

\[
Q_{SSL98} = (q_{s_0, s_1}) = \begin{pmatrix}
-a & b/2 & c/2 \\
-d & b/2 & c/2 \\
e/2 & f/2 & g \\
e/2 & f/2 & h
\end{pmatrix},
\]

where the minus sign represents minus the sum of the remaining row elements and \( s_0, s_1 \in \{A,G,C,T/U\} \) represent the initial and final nucleotide states. The stationary distribution \( \pi = (\pi_A, \pi_G, \pi_C, \pi_{T/U}) \) of SSL98 was given by Schadt et al. (1998). Only the product \( b_0 \times Q \) enters the likelihood calculation, so without loss of generality, we constrain Trace(\( Q \)) = -1, enforcing \( a + b + c + d + e + f + g + h = 1 \).

Although we assume stationarity, SSL98 does not necessarily satisfy detailed balance, meaning there exists at least one pair \( s_0 \neq s_1 \) where

\[
\pi_{s_0} q_{s_0, s_1} \neq \pi_{s_1} q_{s_1, s_0}.
\]

The SSL98 model is therefore irreversible and can be used to infer the root.

Yang (1994) proposed a more general irreversible model (Y94) than SSL98. Y94 contains 12 parameters in its infinitesimal rate matrix \( Q_{Y94} \), and Huelsenbeck et al. (2002) employed this model in a Bayesian framework. Under Y94, calculations of finite time transition probabilities require numerical determination of the possibly complex eigensystem of \( Q_{Y94} \). This determination must be repeated for each new set of parameter values examined and can become computationally expensive. An advantage of SSL98 is that analytic expressions for the finite time transition probabilities have been determined (Schadt et al., 1998). Use of these expressions eases the computational burden.

The reversible model proposed by Tamura and Nei (1993) (TN93) is nested within SSL98 by setting

\[
ae = df, bg = ch.
\]

The model proposed by Hasegawa et al. (1985) (HKY85) is further nested within these models by additionally setting

\[
bd = ch.
\]

TN93 and HKY85 at equilibrium satisfy detailed balance and therefore may not be used to infer the location of root without further restrictions among the branch lengths (Felsenstein, 1981).

**Bayesian Methods**

**Model Prior**

Let \( \theta = (\tau_N, T, \mu, \Lambda) \) be the model parameter space under SSL98, where \( \mu \) is a hyperparameter used to define the prior for \( T \) and \( \Lambda = (a, b, c, d, e, f, g, h) \). We employ a vague yet completely proper prior on \( \theta \) by setting

\[
\tau_N \sim \text{Uniform over all trees with } N \text{ taxa},
\]

\[
tb \mid \mu \sim \text{Exponential}(\mu),
\]

\[
\mu \sim \text{Inverse-gamma}(\alpha, \beta), \text{ and}
\]

\[
\Lambda \sim \text{Dirichlet}(1, 1, 1, 1, 1, 1, 1, 1),
\]

where \( \alpha = 2.1 \) and \( \beta = 1.1 \). The notation \( x \sim Y \) denotes that the random variable \( x \) is drawn from distribution \( Y \). In Bayesian model specification, a random
variable $x$ can represent either an estimable model parameter, e.g. $\tau_N$, $b_i$, $\mu$, and $\Lambda$, or the observed data themselves. When $x$ represents the data, the resulting expression specifies the model likelihood. However, when $x$ is a model parameter, the expression describes a prior distribution.

The priors on $\tau_N$ and $\Lambda$ are flat. Our choice of $\alpha$ and $\beta$ results in the hyperparameter $\mu$ having expectation 1 and a rather large variance of 10, because we know little about its tendencies a priori. The resulting marginal prior on $b_i$ is diffuse on $(0, \infty)$ but remains integrable. The usual uninformative Jeffrey’s prior on $b_i$ is $1/0_b$. This latter prior is improper, precluding calculation of the Bayes factors necessary to test molecular clocks. The Inverse-gamma distribution allows computation of the reciprocal moments of $\mu$ that will be needed later.

**Inference**

We employed Bayes factors to infer the appropriateness of modeling restrictions within $\Lambda$ and among branch lengths. A Bayes factor $B_{10}$ in favor of model 1 ($M_1$) over model 0 ($M_0$) is also the ratio of the marginal likelihood of $M_1$ over the marginal likelihood of $M_0$ (Kass and Raftery, 1995).

When $M_0$ is nested within $M_1$, so that the parameter space of $M_1$ is $(\omega, \phi)$ and the parameter space of $M_0$ is $(\omega = \omega_0, \phi)$, where $\omega_0$ is a known constant and the assumed priors under $M_0$ and $M_1$ satisfy $q(\phi|M_0) \propto q(\omega = \omega_0, \phi|M_1)$, then we may estimate $B_{10}$ using the Savage–Dickey ratio (Verdinelli and Wasserman, 1995),

\[
\frac{1}{B_{10}} = B_{01} = \frac{p(\omega = \omega_0 | X, M_1)}{q(\omega = \omega_0 | M_1)},
\]

where $q(\omega = \omega_0 | M_1)$ is the prior and $p(\omega = \omega_0 | X, M_1)$ is the posterior of $\omega$ under $M_1$, both evaluated at $\omega_0$.

We sampled from model posteriors using reversible-jump Markov chain Monte Carlo (MCMC) computation (Green, 1995). MCMC methods provide an algorithm to generate a Markov chain that explores the parameter space of a given statistical model and whose stationary distribution equals the model posterior (Metropolis et al., 1953; Hastings, 1970). Metropolis et al. (1953) suggested generating this chain by breaking each step into two stages. In the first stage, an application-specific transition kernel proposes a new destination state by stochastically perturbing the current chain state. In the second stage, the sampler accepts or rejects the proposal based on the Metropolis–Hastings criterion. If accepted, the chain moves to the destination state in the parameter space; otherwise, the chain state remains in place during this chain step. Reversible-jump MCMC extends these methods to sample across a countable number of nonnested parameter spaces. As an example, branch lengths do not necessarily retain definition from tree to tree, so jumps between trees move the chain from one nonnested parameter space to another. From the posterior sample, we estimated the posterior ordinate of $\omega$ in Equation 8 using either parametric or nonparametric kernel density estimators. Most of the details of the sampling and density estimation procedures were given by Suchard et al. (2001). One key exception is how we propose new root locations. This proposal is accomplished by randomly selecting a new target branch and randomly locating the new root along this branch. We accepted or rejected the proposal using the Metropolis–Hastings criterion.

**Mathematical Developments**

Critical to Bayes factor estimation between competing models is the calculation of the induced prior evaluated at the nested restriction in Equation 8. Here, we develop the induced prior for restrictions among branch lengths given equally likely evolutionary trees. We first outline an algorithm to draw a random tree from this set. Such an algorithm is paramount to Monte Carlo estimation of the induced priors. We then determine the induced prior distribution of a branch length restriction useful in testing molecular clocks.

**Drawing a Random Tree Using Recursion**

To draw a random evolutionary tree $\tau_N$ with $N \geq 2$ labeled external nodes drawn from the uniform distribution $q(\tau_N) = 1/E_N$, we base our algorithm on a recursive enumeration procedure (Felsenstein, 1978). We start with a tree $\tau_2$ containing taxa 1 and 2 connected together by a single internal node, the root. For steps $s = 3, 4, \ldots, N$, we draw a random node $I_s$ uniformly from the set of $2s – 3$ current internal nodes in $\tau_{s-1}$ and attach taxon $s$ as the new sibling to $I_s$, where the labeling of taxa is unimportant. Sprouting taxon $s$ from $\tau_{s-1}$ to form $\tau_s$ involves the addition of both the external node representing taxon $s$ and a new internal node $P_s$ that becomes the immediate parent to both taxon $s$ and $I_s$. If $I_s$ is not the root in $\tau_{s-1}$, then the parent of $I_s$ in $\tau_{s-1}$ becomes the parent of $P_s$ and grandparent of $I_s$ in $\tau_s$. Otherwise, $P_s$ becomes the new root in $\tau_s$.

We prove this algorithm draws from the uniform distribution by induction. For $N = 2$, there exists exactly one tree $\tau_2$. If we assume $\tau_N \sim q(\tau_N) = 1/E_N$, then there exist $2N – 1$ possible new trees $\tau_{N+1}$ that we could create from a single $\tau_N$, each with probability $1/(2N – 1)$ given our algorithm. Therefore,

\[
q(\tau_{N+1}|\tau_N) = \frac{1}{2N – 1}; \quad \text{and}
\]

\[
q(\tau_{N+1}) = q(\tau_{N+1}|\tau_N)q(\tau_N)
\]

\[
= \frac{1}{2N – 1}q(\tau_N)
\]

\[
= \frac{2N – 1}{2N – 2}(N – 2)! \left(\frac{2N – 2}{2N – 2}\right)
\]
\[2^{N-1}(N-1)! = \frac{(2N-1)!}{(2N-1)!} = \frac{1}{E_{N+1}}. \quad (9)\]

Removing the final root location also generates an unrooted evolutionary tree drawn uniformly from all unrooted trees of size \(N\).

**Induced Prior Distribution of the Distance of Two Taxa from Their MRCA**

Assuming equally likely trees, we derive the prior marginal density of the height difference \(\Delta_{ij}\) evaluated at 0 for any two arbitrary taxa \(i\) and \(j\) drawn from a random tree of fixed size \(N\). The height difference \(\Delta_{ij} = h_i - h_j\), where \(h_i\) and \(h_j\) are the sum of branch lengths measured from the MRCA of taxa \(i\) and \(j\) to taxa \(i\) and \(j\) themselves. Under a molecular clock, \(\Delta_{ij} = 0\). Because we are considering only two taxa, the MRCA of \(i\) and \(j\) is not necessarily the root. Assuming \(b_0 \sim \text{Exponential}(\mu)\),

\[q(\Delta_{ij} = 0|n_i, n_j, \mu) = \frac{(n_i + n_j - 2)!\Gamma(n_j - \frac{1}{2})}{2^{(n_i-n_j+1)}\sqrt{\pi}(n_i - 1)!(2n_j - 2)!} \left(\frac{1}{\mu}\right)^{n_i + n_j - 1}, \quad (10)\]

where \(n_i\) is the level of taxon \(i\) and \(n_j\) is the level of taxon \(j\) from their MRCA in a given tree (Suchard et al., 2001), and \(\Gamma(\cdot)\) is the gamma function. The level \(n_k\) of an arbitrary node \(k\) from a given MRCA counts the number of edges traversed from the node to the MRCA, such that the level of the MRCA is 0 and the children of a node at level \(k\) are at level \(k+1\).

Integrating Equation 10 against the prior \(q(n_i, n_j, \mu) = q(n_i, n_j)q(\mu)\) returns the sought-after marginal density of \(q(\Delta_{ij})\) evaluated at 0 for use in a pairwise molecular clock Bayes factor test.

We determine \(q(n_i, n_j)\) given our prior \(q(\tau_N)\) using a recursion argument. Without loss of generality, we arbitrarily assume taxa \(i\) and \(j\) are \(A\) and \(B\), the first two taxa connected in \(\tau_2\) and, starting recursively, calculate the probability of randomly generating a tree from \(q(\tau_N)\) with \((n_A, n_B) = (k, \ell)\). There exist three independent ways of generating a tree with \((n_A, n_B) = (k, \ell)\) from \(q(\tau_N)\) given a tree with one less taxon drawn from \(q(\tau_{N-1})\). First, we start with a tree where \((n_A, n_B) = (k, \ell)\) and add the \(N\)th taxon as nearest neighbor to any node in \(\tau_{N-1}\) that is not along the path from \(A\) or \(B\) to their MRCA. Given that there are \(2N-3\) possible nearest neighbors, this occurs with probability \((1-(n_A + n_B)/(2N-3))\).

Second, we start with a tree where \((n_A, n_B) = (k - 1, \ell)\) and add the \(N\)th taxon as nearest neighbor to any node along the path from \(A\) to the MRCA with probability \((n_A - 1)/(2N-3)\). Third and likewise, the proceeding event occurs for taxon \(B\).

Let \(P_{n_A, n_B, N}\) be the probability that \(A\) is at level \(n_A\) and \(B\) is at level \(n_B\) away from their MRCA in a tree containing \(N\) taxa. Then, trivially, \(P_{1,1} = 1\), as there is only one tree connecting \(A\) and \(B\), who are both at level 1 away from their MRCA, the root. Further, \(P_{n_A, n_B, N} = 0\) if \(n_A = 0, n_B = 0\) or \(n_A + n_B > N\). Summing the conditional probabilities yields the recurrence relationship

\[P_{n_A, n_B, N} = P_{n_A, n_B, N-1} \left(1-\frac{n_A + n_B}{2N-3}\right) + P_{n_A-1, n_B, N-1} \left(\frac{n_A - 1}{2N-3}\right) + P_{n_A, n_B-1, N-1} \left(\frac{n_B - 1}{2N-3}\right). \quad (11)\]

Given an Inverse-gamma(\(\alpha, \beta\)) prior on \(\mu\),

\[q(\Delta_{ij} = 0) = \int_{n_i} \int_{n_j} \int_{\mu} q(\Delta_{ij}|n_i, n_j, \mu)q(\mu)d\mu dn_i dn_j = \alpha^{n_i + n_j} \beta^{n_i + n_j-1} (n_i + n_j - 2)!\Gamma(n_i - \frac{1}{2}) \left(\frac{1}{\mu}\right)^{n_i + n_j - 1} \int_{n_i, n_j=1}^{\mu \cdot \frac{n_i + n_j}{\mu}} P_{n_i, n_j, N}. \quad (12)\]

**Testing Restrictions among Branch Lengths**

The existence of a molecular clock between two taxa \(i\) and \(j\) restricts \(\Delta_{ij}\) to be exactly 0. To test the hypothesis that \(\Delta_{ij} = 0\) unconditional on tree and without assuming an outgroup, we estimated the posterior density \(p(\Delta_{ij} = 0|X)\) from a sample of \(\Delta_{ij}\) drawn using our rMCMC algorithm. The posterior \(p(\Delta_{ij}|X)\) may be nonnormal and multimodal with different modes corresponding to alternative root locations. We controlled for these factors during density estimation by first simulating \(p(\tau_R|X)\) and recalling that

\[p(\Delta_{ij} = 0|X) = \sum_{\tau_R} p(\Delta_{ij} = 0|\tau_R, X)p(\tau_R|X). \quad (13)\]

We then exploited Occam’s window (Madigan and Raftery, 1994). Occam’s window states that a reasonable estimate of \(p(\Delta_{ij} = 0|X)\) can be obtained by averaging over only the fraction of all possible \(\tau_R\) that possess nominal posterior probability. We independently estimated \(p(\Delta_{ij} = 0|\tau_R^*, X)\) for each \(\tau_R^*\) such that \(p(\tau_R^*|X) \geq \epsilon\), using a \(k\)th nearest neighbor density estimator. A \(k\)th nearest neighbor density estimator is a multivariate, nonparametric method that works by determining the minimum volume \(V\) of a hypersphere necessary to contain the \(k\) closest samples to the point of interest \(x\). The estimated density at \(x \in (nN), \) where \(n\) is the total sample size. If the true density at \(x\) is locally linear, then this method is unbiased. Usually, \(k \approx \sqrt{n}\) is chosen (Loftsgaarden and Quesenberry, 1965). The conditional simulations insure
that there are sufficient posterior samples of $\Delta_{ij}$ for each $\tau_{R}^{i}$ to estimate the density accurately. The marginal posterior density is

$$p(\Delta_{ij} = 0|X) \approx \sum_{\tau_{R}^{i}} p(\Delta_{ij} = 0|\tau_{R}^{i}, X)p(\tau_{R}^{i}|X).$$

Dividing Equation 14 by the prior given in Equation 12 yields the Bayes factor in favor of a molecular clock. pairwise or multitaxon molecular clock comparisons may also be used as a diagnostic to identify subsets of taxa that do or do not support a local molecular clock.

A global molecular clock among $N$ taxa makes $N - 1$ restrictions, $h_{i} = h_{j}$ $\forall$ $i, j = 1, 2, \ldots, N$, where $h_{i}$ is the distance from the inferred root to taxon $i$. We reparametrized these equalities into a vector of $N - 1$ differences $H$ and estimate $p(H = 0|X)$ by averaging the conditional densities over trees, similar to Equation 14.

We estimated the prior ordinate $q(H = 0)$ by drawing $\tau_{N}^{(p)}$ using the algorithm presented here. We drew $\mu^{(p)}, T^{(p)}|\mu^{(p)}$ from our priors and calculated $H^{(p)}$ for each $k$th nearest neighbor density estimator. The Bayes factor in favor of a molecular clock is

$$B_{MC} = \frac{p(H = 0|X)}{q(H = 0)}.$$**

Numerical Example: Primate Mitochondrial DNA

To illustrate these tests, we analyzed a set of mitochondrial DNA sequences taken from nine primates (human, chimpanzee, gorilla, orangutan, gibbon, macaque, squirrel monkey, tarsier, and lemur) (Brown et al., 1982). The data set comprises the protein coding regions for subunits 4 and 5 of the enzyme NADH-dehydrogenase and three tRNAs and contains 888 sites after the removal of alignment gaps. We previously used these data in testing molecular clocks with the HKY85 model in which we assumed the lemur was the outgroup (Suchard et al., 2001).

Figure 1 depicts the modal rooted tree for the primate data set under SSL98. Branch lengths in the figure are the posterior means, and approximately equal distances of the anthropoids from the root is suggestive of a local molecular clock. Care should be taken in making this inference because no measures of branch length uncertainty are depicted.

The inferred root position accounts for 93.2% of all primates is assumed to lie in one of two positions. The first position is along the branch that partitions the lemur from the remaining taxa, and the second position is along the branch that partitions both prosimians from the remaining taxa (Schwartz, 1986). Thus, the estimated posterior probability for the root position is consistent with prior evidence.

Matrix 16 presents the posterior means and SDs of the evolutionary parameters in $\Lambda$:

$$Q_{SSL98} = \begin{pmatrix}
- & 0.082(0.006) & 0.074(0.006) & 0.026(0.005) \\
0.131(0.016) & - & 0.074(0.006) & 0.026(0.005) \\
0.071(0.004) & 0.005(0.002) & - & 0.241(0.012) \\
0.071(0.004) & 0.005(0.002) & 0.193(0.012) & -
\end{pmatrix}$$

The posterior means and SDs of the reparameterization of $(a, b, c, d, e, f, g, h)$ into $\pi$ are $\pi_{A} = 0.308(0.012), \pi_{C} = 0.122(0.009), \pi_{C} = 0.278(0.011)$, and $\pi_{T} = 0.292(0.012)$.

Approximately equal distances (heights) from the MRCA for all anthropoids ($N = 7$) in Figure 1 offers an indication of support for a local molecular clock within this group. However, caution is necessary, because branch
length variability is not reported in the figure. In the following paragraphs, we formally test the appropriateness of the SSL98 model and the validity of the existence of a molecular clock while taking this variability into consideration.

Testing the Appropriateness of SSL98 over HKY85

We tested the appropriateness of the SSL98 model for these data by estimating the Bayes factor in favor of SSL98 over HKY85. Unlike the irreversible model fit test proposed by Yang (1994), the Bayes factor approach does not require that we condition on a fixed tree or root position. Previously, we showed that the HKY85 model provides a better description of the data than do more reduced models (Suchard et al., 2001). We estimated the numerator in the Savage–Dickey ratio using a multivariate normal approximation, because marginal posterior histograms of these parameters are approximately normal in appearance (Fig. 2).

Using the three restrictions given in Equations 5 and 6, we estimated the $\log_{10}$ posterior density at the restriction to be $-9.22$. Via Monte Carlo simulation of length $P = 100,000$ under the prior $g(\lambda) \sim \text{Dirichlet}(1, 1, 1, 1, 1, 1)$, we estimated the $\log_{10}$ induced prior density at the restriction to be $3.54$. The $\log_{10}$ Bayes factor in favor of SSL98 against HKY85 was $12.76$, offering decisive support (Kass and Raftery, 1995) in favor of SSL98.

Pairwise Molecular Clock Comparisons

Based on the distribution of trees, we set $\epsilon = 0.01$ and calculated the $\log_{10}$ Bayes factors in favor of a molecular clock two taxa at a time (Table 1). All $\log_{10}$ Bayes factors for the lemur and any other taxa and all but two for the tarsier and any other taxa are less than $-1$, providing strong evidence against a molecular clock within the prosimians and between the prosimians and the remaining taxa. Mostly positive $\log_{10}$ Bayes factors between the anthropoids suggests a local molecular clock; however, these estimates are highly correlated and will be considered jointly.

Joint Molecular Clock Tests

Via Monte Carlo simulation of length $P = 100,000$, we estimated $g(H = 0)$ to be $0.076$ for $N = 9$ and $0.227$ for $N = 7$. The SDs between 10 independent runs were 0.004 for $N = 9$ and 0.032 for $N = 7$, showing small simulation variability.

Among all primates, we estimated $\log_{10} p(H = 0 | X) = -5.73$, resulting in a $\log_{10}$ Bayes factor of $-4.61$ in favor of a molecular clock. This calculation decisively rejects (Kass and Raftery, 1995) the molecular clock model among all primates. Among the anthropoids, we estimated $\log_{10} p(H = 0 | X) = 1.80$. The $\log_{10}$ Bayes factor was $2.44$ and offers decisive support for a local molecular clock.

Remarks

The $\log_{10}$ Bayes factor of $-4.61$ is substantially stronger than the $-1.3$ estimate of Suchard et al. (2001). This difference results from the better fit using the larger model and from the fact that previously we were forced to ignore a large portion of the information from the lemur data by fixing the lemur as the outgroup and calculating a Bayes factor between the remaining eight taxa. Inclusion of the lemur in the analysis adds substantially greater weight against a molecular clock (Table 1). Additionally, the nonreversible substitution model provides a significantly better description of the data than did the HKY85 model.

Although Bayesian approaches are gaining popularity in phylogenetics, little research has focused on how prior choice affects the prior induced on parameter restrictions that nest competing phylogenetic models. The
induced prior illustrates how the prior and modeling assumptions affect inference about other modeling assumptions; the Bayes factor used to compare these models is highly dependent on the induced prior at the nested restriction (Verginelli and Wasserman, 1995). In general, researchers should pay more attention to prior choice sensitivity when using Bayesian phylogenetics. A general comparison between Bayes factor and LRT molecular clock tests is in progress.

ACKNOWLEDGMENTS

This work was partially supported by a HHMI predoctoral fellowship and NIH training grants GM08042 and GM08185 (M.A.S.) and USPHS grant CA16042 (J.S.S.). Java source code used to calculate the finite time transition probabilities under the SSL98 irreversible model is available at http://www.biomath.medsch.ucla.edu/msuchard/software.html. The primate data set is available at http://www.biomath.medsch.ucla.edu/msuchard/datasets.html.

REFERENCES


First submitted 19 April 2002; reviews returned 6 August 2002; final acceptance 27 August 2002

Associate Editor: Bruce Rannala