A Compound Poisson Process for Relaxing the Molecular Clock

John P. Huelsenbeck,* Bret Larget† and David Swofford‡

*Department of Biology, University of Rochester, Rochester, New York 14627, †Department of Mathematics and Computer Science,
Duquesne University, Pittsburgh, Pennsylvania 15282 and ‡Laboratory of Molecular Systematics,
Smithsonian Museum Support Center, Suitland, Maryland 20746

ABSTRACT

The molecular clock hypothesis remains an important conceptual and analytical tool in evolutionary biology despite the repeated observation that the clock hypothesis does not perfectly explain observed DNA sequence variation. We introduce a parametric model that relaxes the molecular clock by allowing rates to vary across lineages according to a compound Poisson process. Events of substitution rate change are placed onto a phylogenetic tree according to a Poisson process. When an event of substitution rate change occurs, the current rate of substitution is modified by a gamma-distributed random variable. Parameters of the model can be estimated using Bayesian inference. We use Markov chain Monte Carlo integration to evaluate the posterior probability distribution because the posterior probability involves high dimensional integrals and summations. Specifically, we use the Metropolis-Hastings-Green algorithm with 11 different move types to evaluate the posterior distribution. We demonstrate the method by analyzing a complete mtDNA sequence data set from 23 mammals. The model presented here has several potential advantages over other models that have been proposed to relax the clock because it is parametric and does not assume that rates change only at speciation events. This model should prove useful for estimating divergence times when substitution rates vary across lineages.

THE molecular clock hypothesis states that the evolutionary rate of a gene is roughly constant among different lineages (Zuckerkandl and Pauling 1962). Zuckerkandl and Pauling (1965) also suggested that the substitution process is approximately Poisson. If the rate of substitution is constant across lineages, then the distances between species should be ultrametric (*i.e.*, all tips are an equal distance from the root of the tree). Furthermore, if the substitutions follow a Poisson process, then the variance and the mean of the number of substitutions that occur on different lineages in the same amount of time should be equal. Since the early 1970s, however, neither prediction has been shown to hold true; the variance to mean ratio of the number of substitutions is generally greater than one, suggesting that the substitution process is overdispersed (Ohta and Kimura 1971; Langley and Fitch 1973, 1974). Moreover, rates of substitution have been shown to vary across lineages (see Gillespie 1991).

Despite the observation that the molecular clock hypothesis does not perfectly explain the substitution process, it remains a powerful conceptual and analytical tool in evolutionary biology. Conceptually, the molecular clock provides a timescale, albeit an imperfect one, for evolution as well as a mechanistic description of the substitution process. As an analytical tool, the molecular clock has also proven useful. Many of the more interesting applications of phylogenies assume that the molecular clock holds. For example, statistical analysis of host-parasite cospeciation often assumes a molecular clock (Huelsenbeck *et al.* 1997). Also, the estimation of divergence times using molecular data depends upon (1) an accurate calibration date for at least one speciation event on the tree and (2) a constant substitution rate among lineages. Several recent studies have attempted to date the divergence times for eubacteria/eukaryotes (Doolittle *et al.* 1996), metazoa (Wray *et al.* 1996), birds (Cooper and Penny 1997), mammalian orders, and major lineages of vertebrates (Kumar and Hedges 1998) using the molecular clock assumption. To a large extent, the validity of any such analysis depends on how well the data conform to the clock assumption.

Several different approaches have been taken to accommodate rate variation among lineages. One method, commonly used for estimating phylogenetic trees, is to assign each branch of a phylogenetic tree its own rate parameter. However, this procedure does not allow estimation of divergence times of clades because rate and time are confounded. Another approach was proposed by Sanderson (1997), who used a nonparametric method for smoothing the rate differences across speciation events on the tree. Sanderson's method allows rates to vary on branches of the tree and also allows divergence times to be estimated. At the other

Corresponding author: John P. Huelsenbeck, Department of Biology, University of Rochester, Rochester, NY 14627.
E-mail: johnh@brahms.biology.rochester.edu

extreme, Thorne *et al.* (1998) proposed a parametric model for relaxing the clock. Like Sanderson (1997), they assume that rates are autocorrelated across speciation events; their model assigns new rates to descendent lineages from a lognormal distribution with the mean of the distribution equal to the rate of the ancestral lineage. The model we present here differs from the models proposed by Sanderson (1997) and Thorne *et al.* (1998) by allowing rates to change anywhere on the tree. Yet, the model introduces only two additional parameters over the strict molecular clock.

In this article, we propose a compound Poisson process for introducing rate variation across lineages on a tree. We assume that nucleotide substitutions occur along branches of the tree according to a Poisson process, as in the Markov substitution models widely used for phylogenetic inference. However, an independent Poisson process also generates events of substitution-rate change. At each of these events, the rate of substitution is changed by multiplying the current rate by a gamma-distributed random variable. We use Markov chain Monte Carlo when making Bayesian inferences.

METHODS

A compound Poisson process of rate variation: We assume that the phylogeny of a group of species can be represented by a rooted binary tree, an arbitrary example of which is shown in Figure 1. The tips are labeled n_1 to n_s, and the internal nodes are labeled n_{s+1} to n_{2s-1}, where s is the number of sequences. The root of the tree is always labeled n_{2s-1}. The times of the nodes are denoted $t = (t_1, t_2, \ldots, t_{2s-1})$. The branches are scaled such that the tips are at time 0 ($t_1 = t_2 = \ldots = t_s = 0$) and the root is at time 1 ($t_{2s-1} = 1$). The sum of the

scaled branch lengths is referred to as the total length of the tree, T. The total length of the tree of Figure 1, for example, is $T = 4.55$. If the branch lengths are multiplied by a parameter m, representing the expected number of substitutions per site on a single branch reaching from the root of the tree to the tips, the branch lengths of the tree are in terms of expected number of substitutions per site. Moreover, the branch lengths on the tree conform to a molecular clock; there has been no variation in rates across lineages with the result that all of the tips can be drawn to lie on a single time line. Figure 2 shows a rescaling of the tree of Figure 1 with $m = 0.17$.

We use a compound Poisson process to introduce rate variation across lineages of the tree. Events (positions on the tree at which substitution rate changes) occur according to a Poisson process with parameter λ. When an event occurs, the rate of substitution just prior to the event (m) is multiplied by a gamma-distributed random variable (taking value r) to produce a new substitution rate above the event (m'; $m' = mr$). The gamma distribution has density

$$\Gamma(r|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}r^{\alpha-1}e^{-\lambda r}, \quad r \geq 0, \qquad (1)$$

where α is the shape parameter and β is the scale parameter. One possible parameterization of the gamma distribution that we initially considered is to set $\alpha = \beta = \alpha_P$, so that r is $\Gamma(r|\alpha_P, \alpha_P)$. The mean rate of change, then, is $E[r] = 1$. A potentially unappealing property of such a model of rate change is that as the branch lengthens, substitutions stop happening along the branch. In such

Figure 1.—We assume that a rooted binary tree describes the genealogy of the sequences n_1 to n_s. This arbitrarily chosen tree illustrates the labeling of the nodes used in this article. The unscaled times of the tree have the tips at time 0 and the root at time 1. All other nodes on the tree have times between 0 and 1.
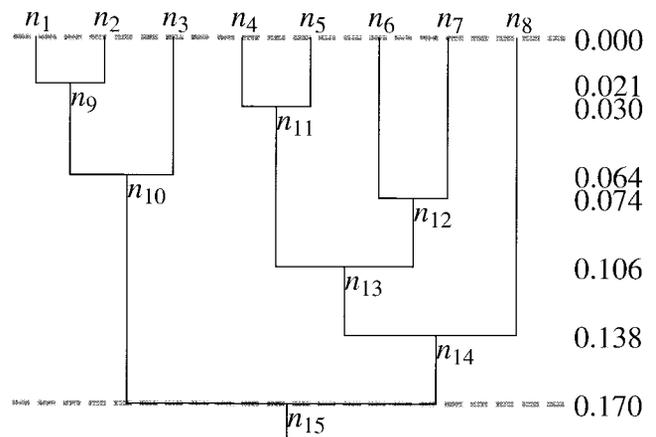
Figure 2.—The branch lengths that result when the tree of Figure 1 has all branches multiplied by $m = 0.17$. The rate of substitution on the tree, m, is in terms of expected number of substitutions per site that occur along a single lineage reaching from the root of the tree ($t = 1$) to the tips ($t = 0$). Multiplying the branch lengths in Figure 1 by m converts branch lengths into expected number of substitutions per site. Note that this tree has no rate variation among lineages. The branch lengths of the tree conform to a molecular clock.

a model, the rate along a branch is a base rate, m, multiplied by $\Pi_i^n r_i$, where n is the number of Poisson change points and r_i is $\Gamma(\alpha_P, \alpha_P)$. As the branch length grows, n becomes large and the product converges to 0 in probability.

A prior model that changes rates multiplicatively with independent multipliers must be carefully chosen so that the rate does not tend to either 0 or infinity in probability. To see this, let r_i, $i = 1, 2, \ldots$ be independent and identically distributed positive random variables and let $R_n = \Pi_{i=1}^n r_i$. Then for every $\varepsilon > 0$,

$$P[R_n < \varepsilon] = P\left[\prod_{i=1}^n r_i < \varepsilon\right]$$

$$= P\left[\sum_{i=1}^n \log R_i < \log \varepsilon\right].$$

Thus, the product of rates converges in probability to 0 if $E[\log r_i] < 0$ and increases without bound in probability if $E[\log r_i] > 0$.

In this article, we largely circumvent this problem by placing a prior on λ such that the number of change points on the tree does not become large. In addition, we define a one-dimensional family of gamma-distributed random variables with the property that $E[\log r_i] = 0$ by letting $\beta = e^{\psi(\alpha_P)}$, where ψ is the derivative of the logarithm of the gamma function:

$$\psi(\alpha) = \frac{d}{d\alpha}\log \Gamma(\alpha)$$

$$= \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$$

$$= \left(\int_0^\infty (\log x) x^{\alpha-1} e^{-x} dx\right) \Big/ \left(\int_0^\infty x^{\alpha-1} e^{-x} dx\right).$$

Here is a short derivation of this property. Suppose that $r \sim \Gamma(\alpha, \beta)$. Then r is equal in distribution to X/β where $X \sim \Gamma(\alpha, 1)$:

$$E[\log r] = E[\log(X/\beta)]$$

$$= \int_0^\infty (\log x) \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} dx - \log \beta$$

$$= \psi(\alpha) - \log \beta.$$

This is 0 precisely when $\beta = e^{\psi(\alpha)}$.

In this article, we multiply the rate at an event by a gamma-distributed random variable with density

$$g(r|\alpha_P) = \frac{e^{\psi(\alpha_P)}}{\Gamma(\alpha_P)} r^{\alpha_P-1} e^{-re^{\psi(\alpha_P)}}, \quad r \geq 0. \qquad (2)$$

The collection of events is denoted $e = (\xi, z, r)$, where ξ is the number of rate change events, z is the vector of event positions, and r is the vector of rate multipliers. When $\xi = 0$, z and r are empty. For $\xi > 0$, $z \in [0, T]^\xi$ and $r \in \Re^\xi$. There is a mapping that takes points from
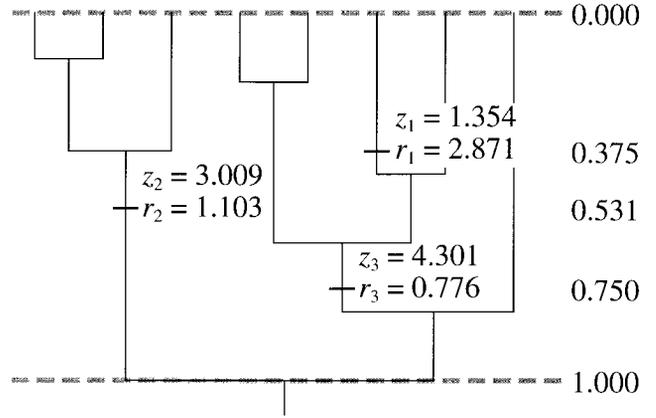
Figure 3.—The compound Poisson process discussed in this article places events of substitution-rate change on the tree according to a Poisson process. Associated with each event is a rate multiplier that multiplies the rate of substitution up to the event (m) by a gamma-distributed random variable (taking value r) to produce a new rate of substitution above the event (m'). In this example, which illustrates the process for the tree from Figure 1, three events of substitution-rate change occur. The rate multipliers are 2.871, 1.103, and 0.776 for events 1, 2, and 3, respectively. The z_i specify the event locations on the tree.

$[0, T]$ to points on the tree. Figure 3 illustrates the compound Poisson process for rate change events acting upon the tree of Figure 1. In this example, three events of rate change occur on the tree [$e = (3, z, r)$, $z = (1.354, 3.009, 4.301)$, $r = (2.871, 1.103, 0.776)$]. Associate each branch with the smaller number of its two nodes (that node that is furthest from the root of the tree) and partition the interval $[0, T]$ into segments associated with branches ordered from the lowest index to the highest index. For example, the branch descending from node n_1 is represented by the interval $[0, 0.125]$. Similarly, the branch descending from node n_6 is represented by the interval $(0.979, 1.416]$. The first event is located on this branch at $z_1 = 1.354$. Events 1 and 2 increase the rate of substitution whereas event 3 decreases the rate of substitution. Figure 4 shows the relative branch lengths in terms of expected number of substitutions per site. Note that the branch lengths of the tree no longer obey the molecular clock. The molecular clock is simply a special case of our model with $\lambda = 0$ and/or $\alpha_P = \infty$.

The prior distribution of $e = (\xi, z, r)$ given the tree and parameters α_P and λ is described by the probability measure

$$f(e|\tau, t, \lambda, \alpha_P) = e^{-\lambda T}\delta_0 + \sum_{\xi=1}^\infty \frac{e^{-\lambda T}(\lambda T)^\xi}{\xi!} \times \left(\frac{1}{T}\right)^\xi \times \prod_{i=1}^\xi g(r_i|\alpha_P),$$

$$(3)$$

where δ_0 is the point mass measure at $(0, \emptyset, \emptyset)$. This corresponds to picking a Poisson (λT) number of points, choosing their locations from mutually indepen-
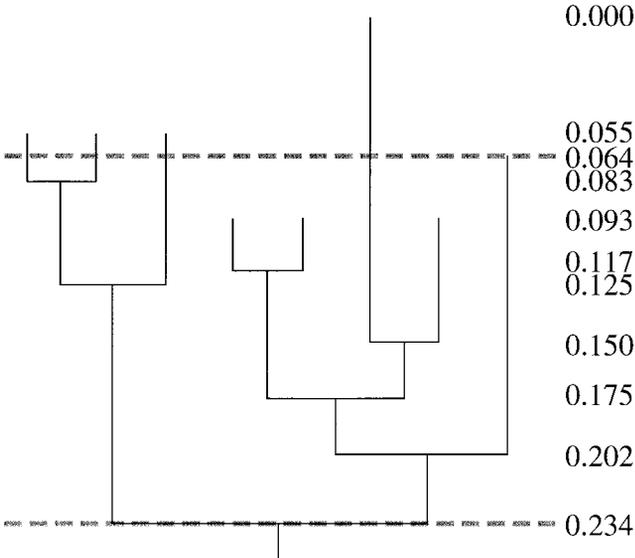
Figure 4.—The branch lengths in terms of expected number of substitutions per site when the substitution rates are modified using the compound Poisson process. The rates were modified using the three events from Figure 3 with a starting substitution rate of $m = 0.17$ (at the root of the tree). Note that the branch lengths of the tree no longer conform to the molecular clock.

dent uniform $[0, T]$ distributions, and attaching mutually independent gamma-distributed rate multipliers.

Model of DNA substitution: We assume that DNA sequences from homologous regions are available for species n_1 to n_s. Let $X = \{x_{ij}\}$ be the aligned nucleotide sequences, where $i = 1, 2, \ldots, s; j = 1, 2, \ldots, c;$ and c is the number of nucleotide sites per sequence. Each column of the data matrix $x_j = \{x_{1j}, \ldots, x_{sj}\}'$ specifies the nucleotides for the s sequences at the jth site.

As is usual for DNA sequences, we assume that substitutions occur according to a Poisson process with rate matrix Q. A general reversible rate matrix allows a different stationary frequency for the four nucleotides $\pi = (\pi_A, \pi_G, \pi_G, \pi_T)$ constrained to sum to one and six rates for the 12 substitution types

$$Q = \{q_{ij}\} = \begin{bmatrix} \cdot & \pi_C r_{AC} & \pi_G r_{AG} & \pi_T r_{AT} \\ \pi_A r_{AC} & \cdot & \pi_G r_{CG} & \pi_T r_{CT} \\ \pi_C r_{AG} & \pi_C r_{CG} & \cdot & \pi_T r_{GT} \\ \pi_A r_{AT} & \pi_C r_{CT} & \pi_G r_{GT} & \cdot \end{bmatrix} \quad (4)$$

(Yang 1994b). The diagonal of the rate matrix is specified such that the row sums are equal to zero. We add an additional constraint by rescaling so that $-\Sigma q_{ii}\pi_i = 1$; this means that branch lengths of the tree are in terms of expected number of substitutions per site, v. This model is reversible because it fulfills the reversibility criterion that $\pi_i q_{ij} = \pi_j q_{ji}$ for all i and j. Most commonly used models of DNA substitution are constrained to be reversible and are simply special cases of the model described here. In the analyses that follow, we use the

HKY85 model of DNA substitution (Hasegawa *et al.* 1984, 1985). This model allows different base frequencies and for a transition/transversion rate bias ($r_{AG} = r_{CT}$; $r_{AC} = r_{AT} = r_{CG} = r_{GT}$; $\kappa = r_{AG}/r_{AC}$). The transition probabilities are calculated as $P(v) = \{p_{ij}(v)\} = e^{Qv}$.

Among-site rate variation can be accommodated in several different ways. One potential method partitions the DNA sequence into different regions (*e.g.*, first, second, and third codon positions) and then estimates the rate separately for each partition, assuming that the rate within a partition is homogeneous [*i.e.*, instead of a single substitution rate (m), applying to all sites, there are multiple substitution rates (m_1, m_2, . . .)]. Another method assumes that the rate assigned to a site is a random variable. Typically, the gamma distribution, parameterized with the shape parameter equal to the scale parameter, is used to model rate variation across sites (Yang 1993, 1994a). The shape parameter of the gamma distribution for among-site rate variation is denoted α_R. In this article, we assume equal rates across sites or gamma-distributed rates across sites. Models that assume gamma-distribution rate variation are denoted "$+\Gamma$." Our implementation of gamma-distributed rate variation uses the discrete gamma approximation with four rate categories (Yang 1994a).

The probability of observing the states present at the ith site (x_i) is a sum over all possible assignments of nucleotides to the internal nodes of the tree. Suppose $y = \{y_k\}$ for $k = s + 1, \ldots, 2s - 1$ is a generic data vector at the internal nodes. Branch k of the tree has length v_k expected substitutions per site and ancestral node $\sigma(k)$. The transition probability from state i to state j along a branch with v expected substitutions is $p_{ij}(v)$. The initial substitution rate at the root is m. Then, the conditional probability of observing the data at the ith site given the tree and rate events is

$$f(x_i|\tau, t, e, m) = \sum_y \pi_{y_{2s-1}} \left(\prod_{k=1}^{s} p_{y_{\sigma(k)}x_k}(v_k) \right) \left(\prod_{k=s+1}^{2s-2} p_{y_{\sigma(k)}y_k}(v_k) \right).$$

$$(5)$$

The summation is over all possible combinations of nucleotide states that can be assigned to the internal nodes of the tree. The expected number of substitutions on branch k (v_k) is found by integrating the rate over the length of the branch, $v_k = \int_{\sigma(k)}^{t_k} r_k(u) \, du$, where $r_k(u)$ is the rate along branch k at time u. The rate is a step function for the compound Poisson process considered in this article. Figure 5 shows a single branch starting at $t_B = 0.40$ and ending at $t_E = 0.20$. There are two events of substitution-rate change along this branch, occurring at times 0.30 and 0.25. The expected number of substitutions per site along this branch, then, is

$$v = (0.40 - 0.30) \times 0.10 + (0.30 - 0.25)$$

$$\times 0.12 + (0.25 - 0.20) \times 0.06$$

$$= 0.019. \quad (6)$$

$$t_E = 0.20$$

$$m'' = 0.06$$

$$r_1 = 0.50 \qquad t_1 = 0.25$$

$$m' = 0.12$$

$$r_2 = 1.20 \qquad t_2 = 0.30$$

$$m = 0.10$$

$$t_B = 0.40$$
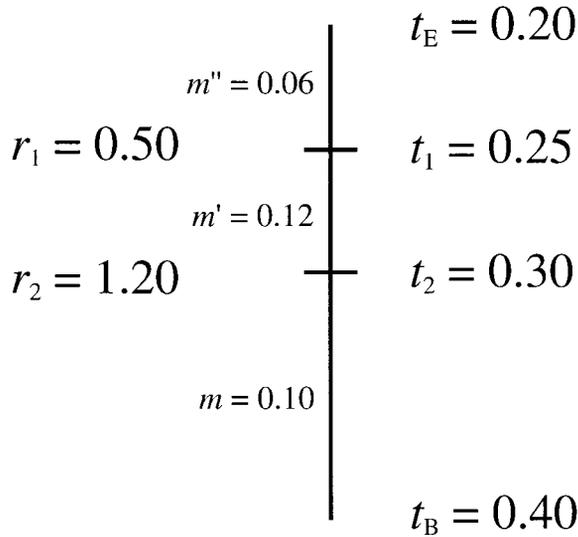
Figure 5.—An example of how the average number of substitutions per site is calculated for a branch when substitution rate changes through time. Here, the substitution rate changes according to the compound Poisson process used in this article. There are two events of substitution rate change along this branch. The first doubles the current substitution rate and the second halves the current substitution rate.

The substitution rate at the base of the branch is $m = 0.10$.

Assuming independence of the substitutions across sites, the conditional probability of observing the full sequence data set given the tree and rate events is the product of the probabilities of observing the sites:

$$f(X|\tau, t, e, m) = \prod_{i=1}^{c} f(x_i|\tau, t, e, m). \qquad (7)$$

Estimating m, λ, and α_P using Markov chain Monte Carlo: We wish to estimate the rate of molecular evolution at each branch through Bayesian estimation of the parameters λ, α_P, and m (where m is the rate of substitution at the base of the tree). The likelihood function for λ, α_P, and m is

$$\ell(\lambda, \alpha_P, m) = \int_{\varepsilon} f(X|\tau, t, e, m) \, dF(e|\tau, t, \lambda, \alpha_P) \, dF(t),$$

$$(8)$$

where the single integral denotes integration over all rate events and branch lengths consistent with the tree topology τ. Integration with respect to the probability measure is used to denote a summation for discrete random variables (such as the number of events on a tree or the topology of a tree) as well as integration for continuous random variables (such as the position of the events, the speciation times, and the gamma-distributed rates associated with events). The tree topology is considered to be fixed in this study, but the speciation times and the position and rates of events are treated as random variables. In a Bayesian analysis, we specify

a joint prior probability distribution for the parameters λ, α_P, and m along with the rate events and branch lengths.

All Bayesian inference arises from the joint posterior distribution of the parameters of interest (in this case, the posterior distribution of λ, α_P, and m). The posterior probability density of λ, α_P, and m is

$$f(\lambda, \alpha_P, m|X) = \frac{\ell(\lambda, \alpha_P, m) f(\lambda, \alpha_P, m)}{f(X)}, \qquad (9)$$

where

$$f(X) = \int_{\eta} \ell(\lambda, \alpha_P, m) \, dF(\lambda, \alpha_P, m)$$

and η is the space for λ, α_P, and m. We assume that λ, α_P, and m have independent priors. The parameters α_P and m have uniform priors on $[0, B_{\alpha P}]$ and $[0, B_m]$, respectively. We place an exponential prior on λ. The relative speciation times, t, which are scaled to be between 0 and 1 (see Figure 1), are distributed as the order statistics drawn from a uniform $(0, 1)$ distribution, conditional on agreement with the tree topology τ. There is no biological meaning to the uniformly distributed priors. However, in Bayesian analysis, such priors are often used in cases where there is little or no prior knowledge about the parameters. Using uninformative priors is a way to avoid biasing the results of the analysis; the posterior probability distribution will mainly be determined by the likelihood function. An exponential prior on λ decreases the probability of substitution rate histories with a large number of small rate-change events.

Calculation of the posterior probability density (Equation 8) involves evaluating high-dimensional integrals and summations. We use Markov chain Monte Carlo (MCMC) integration to estimate the posterior distributions of interest. Specifically, we used the Metropolis-Hastings-Green (MHG) algorithm (Green 1995; see also Geyer 1999), an extension of the Metropolis-Hastings algorithm for problems in which the dimension of the sample space changes. The MHG algorithm constructs a Markov chain by first proposing a new state and then moving to that state with probability R. The steps of the algorithm are as follows: (1) the current state of the chain is θ; (2) with probability density $f(\theta^*|\theta)$, a new state (θ^*) is nominated; (3) the acceptance probability of the nominated state is calculated,

$$R = \min\left\{1, \frac{f(\theta^*) f(\theta|\theta^*)}{f(\theta) f(\theta^*|\theta)}\right\}, \qquad (10)$$

where $f(\theta)$ is the target distribution (*i.e.*, Equations 7 or 8); and (4) a uniformly distributed random variable on the interval $[0, 1]$ is generated. If this random variable is $<R$, then the nominated state is accepted and becomes the current state of the chain ($\theta = \theta^*$). Otherwise, the chain remains in state θ. Steps 1–4 are repeated many times.

Several different mechanisms were used to update the state of the chain. These mechanisms included (1) adding an event to the tree, (2) deleting an event from the tree, (3) changing the position of an event on the tree, (4) changing the gamma-distributed random variable associated with an event on the tree, (5) changing the time of an internal node on the tree, (6) changing the substitution rate (m) at the base of the tree, (7) changing the Poisson parameter (λ), (8) changing the gamma parameter (α_P), (9) changing the transition/transversion rate ratio (κ), (10) changing the gamma shape parameter for among-site rate variation (α_R), and (11) changing the base frequencies (π).

The acceptance probabilities for the different moves take the form

$$R = \min\{1, \text{likelihood ratio}\} \times (\text{prior ratio})$$
$$\times (\text{proposal ratio}) \times (\text{Jacobian})\} \qquad (11)$$

(Green 1995). For move types 3–8, the standard Markov chain theory employed in the Metropolis-Hastings algorithm (Metropolis *et al.* 1953; Hastings 1970) applies. However, move types 1 and 2 add and delete an event, respectively. These move types involve a change in the dimensionality of the sample space. Hence, we construct reversible Markov chains for move types 1 and 2 that jump between parameter subspaces of different dimensionality using the methodology developed by Green (1995).

Move type 1: Adding an event to the tree: The prior ratio for the addition of a single point (z^*, r^*) to the current state $e = (\xi, z, r)$ is

$$[(e^{-\lambda T}(\lambda T)^{\xi+1}/(\xi + 1)!) \times (1/T)^{\xi+1} \times \prod_{i=1}^{\xi} g(r_i|\alpha_P)$$
$$\times g(r^*|\alpha_P)]/[(e^{-\lambda T}(\lambda T)^{\xi}/\xi!) \times (1/T)^{\xi} \times \prod_{i=1}^{\xi} g(r_i|\alpha_P)]$$
$$= \lambda g(r^*|\alpha_P)/(\xi + 1) \qquad (12)$$

and the proposal ratio is

$$\frac{d_{\xi+1} \times 1/(\xi + 1)}{b_\xi \times 1/(\xi + 1) \times 1/T \times g(r^*|\alpha_P)} = \frac{Td_{\xi+1}}{g(r^*|\alpha_P) b_\xi}, \qquad (13)$$

where $d_{\xi+1}$ and b_ξ are the probabilities of making a move that deletes one of $\xi + 1$ events or adds an event when there are currently ξ events, respectively. The Jacobian is 1. The acceptance probability, then, is

$$R = \min\left\{1, (\text{likelihood ratio}) \times \frac{\lambda Td_{\xi+1}}{(\xi + 1) b_\xi}\right\}. \qquad (14)$$

Move type 2: Deleting an event from the tree: The acceptance probability for the reverse step, deleting an event from the tree, has the same form as Equation 13, but with the ratio terms inverted. Detailed balance between the move types that add and delete an event on the tree is demonstrated in the appendix.

Move type 3: Changing the position of an event: Two different move types changed the position of an event on the

tree. The first move type (3a) was chosen with probability ϕ_{3a} and picked at random one of the ξ events on the tree. A new position was chosen uniformly on the interval $[0, T]$. The second move type (3b) was chosen with probability ϕ_{3b} and picked one of the ξ events on the tree at random and moved its current position by a small amount randomly drawn from the interval $[-\varepsilon_{3b}, +\varepsilon_{3b}]$. The acceptance probability for both of these moves is

$$R = \min\{1, (\text{likelihood ratio})\}. \qquad (15)$$

Move type 4: Changing the rate multiplier associated with an event: Two different move types changed the rate associated with an event on the tree. Both move types pick at random one of the ξ events. This randomly chosen event will have its rate multiplier changed from r to r^*. For the first move type (4a) a change to r^* is proposed with probability ϕ_{4a} such that $\log_e(r^*/r)$ is uniformly distributed on the interval $[-\frac{1}{2}, +\frac{1}{2}]$. The acceptance probability for this move is then

$$R = \min\{1, (\text{likelihood ratio}) \times (r^*/r)^{\alpha_P} e^{-(r^*-r)\alpha_P/\psi(\alpha_P)}\}$$
$$(16)$$

(Green 1995). The other move type (4b) is chosen with probability ϕ_{4b} and picks a new rate by drawing a random variable from the distribution $g(r^*|\alpha_P)$. The acceptance probability for this move type is simply the likelihood ratio (Equation 15).

Move type 5: Changing the time of an internal node on the tree: With probability ϕ_5, the time of an internal node was changed. The times of the internal nodes were updated as follows. First, an internal node of the tree was chosen at random (excluding the root node). The time of this node was increased or decreased by adding a uniformly distributed random variable drawn from the interval $[-\varepsilon_5, +\varepsilon_5]$. The acceptance probability for this move is

$$R = \min\left\{1, (\text{likelihood ratio}) \times \frac{e^{-\lambda T^*}(\lambda T^*)^\xi}{e^{-\lambda t}(\lambda T)^\xi} \times \frac{T}{T^*}\right\}, \qquad (17)$$

where T^* is the total tree length after the node time has been adjusted. A change to the speciation time on an internal node affects the length of a total of three branches (the ancestral branch and the two descendant branches). Often, events of rate change occur along these branches. The times of the rate-change events are maintained proportionally along the same branch they started on.

Move types 6, 7, 8, 9, and 10: Changing m, λ, α_P, κ, and α_R: The substitution rate (m), Poisson parameter (λ), gamma shape parameter (α_P), transition/transversion rate ratio (κ), and gamma shape parameter for among-site rate variation (α_R) were updated with probabilities ϕ_6, ϕ_7, ϕ_8, ϕ_9, and ϕ_{10}, respectively, by adding to the current value a uniformly distributed random variable on the interval $[-\varepsilon_6, +\varepsilon_6]$, $[-\varepsilon_7, +\varepsilon_7]$, $[-\varepsilon_8, +\varepsilon_8]$, $[-\varepsilon_9,$

$+\varepsilon_9]$, and $[-\varepsilon_{10}, +\varepsilon_{10}]$, respectively. The acceptance probabilities are

$$R = \min\{1, \text{ (likelihood ratio)}\} \qquad (18)$$

for the update of m, κ, and α_R,

$$R = \min\left\{1, \text{ (likelihood ratio)} \times \frac{e^{-\lambda^* T}(\lambda^* T)^{\xi}}{e^{-\lambda} T(\lambda T)^{\xi}}\right\} \qquad (19)$$

for the update of λ, where λ^* is the proposed state and λ is the current state, and

$$R = \min\left\{1, \text{ (likelihood ratio)} \times \prod_{i=1}^{\xi} g(r_i|\alpha_P^*)/g(r_i|\alpha_P)\right\} \qquad (20)$$

for the update of α_p, where α_P^* is the proposed state and α_P is the current state. We also considered gamma priors for m, λ, and α_P. The gamma prior has parameters a and b. The exponential prior for λ used in the analyses of this article has $a = 1$. A change to m^* is proposed such that $\log_e(m^*/m)$ is uniformly distributed on the interval $[-\frac{1}{2}, +\frac{1}{2}]$. The acceptance probability for this move is then

$$R = \min\{1, \text{ (likelihood ratio)} \times (m^*/m)^a e^{-b(m^*-m)}\}. \qquad (21)$$

The acceptance probabilities for changing λ and α_p are the same as Equation 21, with m and m^* replaced by λ and λ^* and α_p and α_P^*, respectively.

Move type 11: Changing the base frequencies: With probability ϕ_{11} a move was attempted that changed the equilibrium base frequencies, π. The sum of the base frequencies is constrained to equal 1 and new values are proposed from a Dirichlet distribution with expected values at the current values. The Dirichlet distribution is the natural conjugate prior for a multinomial distribution and has probability density

$$f(\pi|\alpha) = \frac{\Gamma(\alpha_0)}{\prod_{i \in S}\Gamma(\alpha_i)}\prod_{i \in S}\pi_i^{(\alpha_i-1)}, \qquad (22)$$

where S is the state space (A, C, G, or T), α_i is the Dirichlet parameter of the ith nucleotide, $\alpha_0 = \Sigma_{k \in S}\alpha_k$, and π_i is the frequency of the ith nucleotide. New base frequencies are drawn from the Dirichlet distribution with $\alpha_i = \pi_i\alpha_0$. We set $\alpha_0 = 100.0$ in all of the analyses of this study.

Changing parameters near the boundary of the parameter space: Note that for move types 3b, 5, 6, 7, 8, 9, and 10 that the parameter is changed by adding a uniformly distributed random variable from an interval $[-\varepsilon, +\varepsilon]$. When the parameter is restricted to an interval (l, h) and the proposed value is outside this interval, we reflect the excess back into the interval. Namely, if $\theta' = \theta + U$, where θ is the current parameter value (either z, t, m, λ, or α_P) and U is the uniformly distributed change, the proposed parameter value θ^* is

$$\theta^* = \begin{cases} l + (l - \theta'): \theta' < l \\ \theta': l \le \theta' \le h \\ h - (\theta' - h): \theta' > h. \end{cases} \qquad (23)$$

The proposal ratio [in (10)] is one for this type of proposal.

Estimating parameters using Bayesian inference: The chain was sampled every g_S generations. The posterior distributions of the parameters are obtained directly by noting the position of the chain and recording the values of m, λ, and α_P for each sampled state. The proportion of the time the chain stays in different intervals is an approximation of the posterior distribution.

The chain was burned in by discarding the first g_B generations of the Markov chain. The burn-in was performed to allow the chain to approach stationarity before states are sampled.

Validation of computer program: A computer program implementing the compound Poisson process for the HKY85 + Γ (Hasegawa *et al.* 1984, 1985; Yang 1993, 1994a) model of DNA substitution was written in C by one of us (J.P.H.; available via anonymous ftp to brahms.biology.rochester.edu or via the WWW at http://brahms.biology.rochester.edu). The advantage of using MCMC for Bayesian inference is that the sampled points are a valid (albeit dependent) sample from the posterior distribution: the Markov chain law of large numbers (theorem 3; Tierney 1994) states that posterior probabilities can be validly estimated from long-run sample frequencies. However, it is impossible to guarantee that an implementation of MCMC will converge for any particular problem (Geyer 1999). As Geyer (1999, p. 80) states, "MCMC is a complex mixture of computer programming, statistical theory, and practical experience. When it works, it does things that cannot be done any other way, but it is good to remember that it is not foolproof."

The computer program was validated in several ways: (1) likelihoods were checked against existing computer programs (*e.g.*, PAUP*; Swofford 1998); (2) the acceptance probabilities for the various move types were worked out independently by two of us; and (3) the program was run without any data. When the program is run without any data, then the likelihood ratio equals one and the chain should target the prior distribution. When the chain is run without data, the expectation and variance of the number of events on the tree should both be equal to λT. Moreover, the average rate associated with the events should equal $\alpha_P/e^{\psi(\alpha_P)}$ and the variance should equal $\alpha_P/e^{\psi(\alpha_P)}$. When the chain was run without data, the results satisfied these expectations. Also, multiple independent chains were started from different starting values of the parameters. The chains were examined to see if they converged to the same posterior probability distribution.

Analysis of DNA sequence data: We applied the method to a single DNA sequence data set. The data set included complete mitochondrial sequences from 23 mammals (Arnason *et al.* 1997). The tree topology was considered fixed in the analysis; the tree topology estimated using UPGMA (Sokal and Michener 1958) with the Jukes-Cantor distance (Jukes and Cantor 1969) was used in the analysis. All analyses were performed assuming the HKY85 model of DNA substitution (Hasegawa *et al.* 1984, 1985). This model allows different base frequencies as well as a transition/transversion rate bias. Parameters of the substitution model were estimated on the UPGMA tree using maximum likelihood (as implemented in PAUP*; Swofford 1998).

The Markov chain was run for at least 500,000 generations. Every 50th state of the chain was sampled (*i.e.*, $g_S = 50$). The sampled states were used to construct the posterior distribution for the variables m, λ, and α_P.

RESULTS

The molecular clock hypothesis of equal rates across lineages could be rejected for the mammalian mtDNA sequence data set. We used a likelihood-ratio test to examine the molecular clock hypothesis (Felsenstein 1981). The HKY85 + Γ model was assumed in the analysis. The null hypothesis is that the molecular clock holds; the likelihood is maximized under the constraint of equal rates across lineages (L_0). The alternative hypothesis relaxes the clock constraint by assigning a different rate to each of the lineages; the likelihood under the alternative hypothesis is L_1. The likelihood-ratio test statistic (twice the difference in the $\log_e$ likelihood, $-2 \log_e\Lambda$; $\Lambda = L_0/L_1$) between the null and alternative models is approximately χ^2 distributed with $s - 2$ degrees of freedom. The molecular clock hypothesis was rejected at the 5% level (mammalian mitochondrial sequences: $\log_e L_0 = -114,514.23$, $\log_e L_1 = -114,431.21$, $-2 \log_e\Lambda = 166.03$, $P < 0.00001$).

We assumed two different priors for λ for the mtDNA data. Figures 6 and 7 show the results of the analysis of the mammalian mitochondrial sequences assuming an exponential prior for λ with mean 1.0. Figure 6 shows the change in the $\log_e$ likelihood through time. The probability of observing the data increased for the first 10,000 generations of the chain. The $\log_e$ likelihood then stabilized, fluctuating around a value of $\sim -130,240$. We discarded the first $g_B = 50,000$ generations as the burn-in time of the chain. Figure 7 shows the posterior distributions for the parameters m and α_P. The estimates were $m = 0.33$ (0.32, 0.34) and $\alpha_P = 5.70$ (2.35, 10.82). The credibility interval for each variable was determined by taking the 2.5% tails of the posterior distributions.

Figures 8 and 9 show the results of analyses in which an exponential prior with mean 10.0 was assumed for λ. Estimates of α_P (the shape parameter of the compound Poisson process) and m changed when different parame-

ter values of the prior were used [$m = 0.33$ (0.32, 0.33); $\alpha_P = 63.89$ (30.00, 104.24)]. The posterior distribution of m became broader and the posterior distribution of α_P increased (there were more events, each with a smaller effect). However, the branch lengths were very similar regardless of the exact value of the exponential prior on λ. Figure 10 shows the relationship between the maximum-likelihood estimates of the branch lengths obtained assuming the HKY85 model of DNA substitution and the mean of the posterior distribution of each branch obtained under the compound Poisson process. In both cases, the compound Poisson process was able to accommodate rate variation across lineages of the tree (the slope was 1.00, with $r^2 = 0.999$ for both priors). We also examined the relationship between the compound Poisson process with exponential priors of means 1 and 10 (Figure 11). Again, there is a close relationship between the branch lengths regardless of the prior used for λ. The posterior distributions of branch lengths are quite robust to the two different priors used for λ.

The compound Poisson process for relaxing the molecular clock should prove practically useful for several applications. For example, estimation of divergence times for clades depends upon a calibration time for at least one speciation event on the tree and equal rates across lineages (though see Sanderson 1997). The compound Poisson process can be used to relax the molecular clock while at the same time allow estimation of divergence times of clades. An application of the method applied to the mammalian mitochondrial data is shown in Figure 12. The general approach is the same as that taken by Thorne *et al.* (1998), but we used the compound Poisson process discussed in this article instead of a lognormal distribution to relax the clock assumption. The estimation of divergence times required only one modification to the notation described
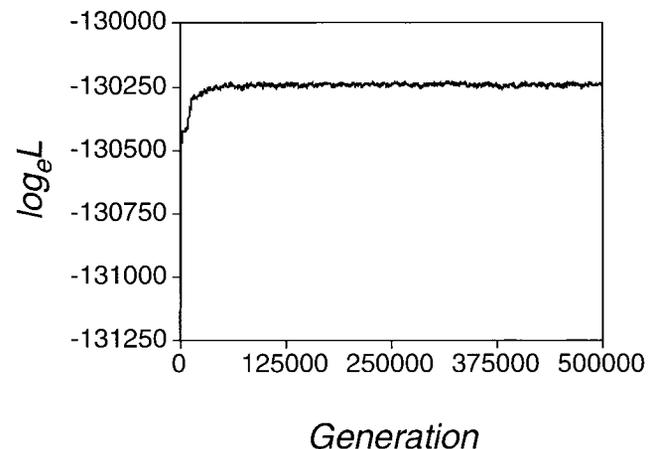
Figure 6.—The change in the $\log_e$ probability of observing the data (Equation 6) for the analysis of the mammalian mitochondrial sequences. The prior for λ was exponential with mean 1.

in this article; instead of rescaling branch times between 0 (at the tips) and 1 (at the root), the tips of the tree were considered to be time 0 and the times of other branches on the tree were in terms of millions of years before present. The HKY85 + Γ model of DNA substitution was assumed in the analysis; the gamma model fit the observed sequences better than a model that pooled sites according to codon position and allowed a different rate for each partition ($\log_e L = -114{,}431.21$ for HKY85 + Γ, $\log_e L = -117{,}091.83$ for HKY85 + SS). An exponential prior with mean 0.006 was placed on λ. The estimates of the substitution model parameters were $m = 0.010$ (0.008, 0.012), $\alpha_P = 7.58$ (2.60, 17.13), $\kappa = 13.67$ (13.00, 14.37), and $\alpha_R = 0.230$ (0.228, 0.240). The posterior distribution of the divergence time of ferungulates (carnivores, perissodactyls, artiodactyls, and cetaceans) and Xenartha (Edentata) is shown in Figure 12. Arnason *et al.* (1997) obtained an estimate of 86 mya for this speciation event using a calibration date of 60 million years for the cow-whale divergence. We used a calibration date of 56.5 million years for the cow-whale divergence that is based on the earliest occurrence of either artiodactyls or cetaceans. The estimate of the Ferungulate-Xenartha divergence is 92.5 mya (83.7, 99.6). The credibility interval overlaps the estimate obtained by Arnason *et al.* (1997). In Figure 12, we also indicate the divergence time estimates of the other clades on the tree, with the 95% credibility intervals.
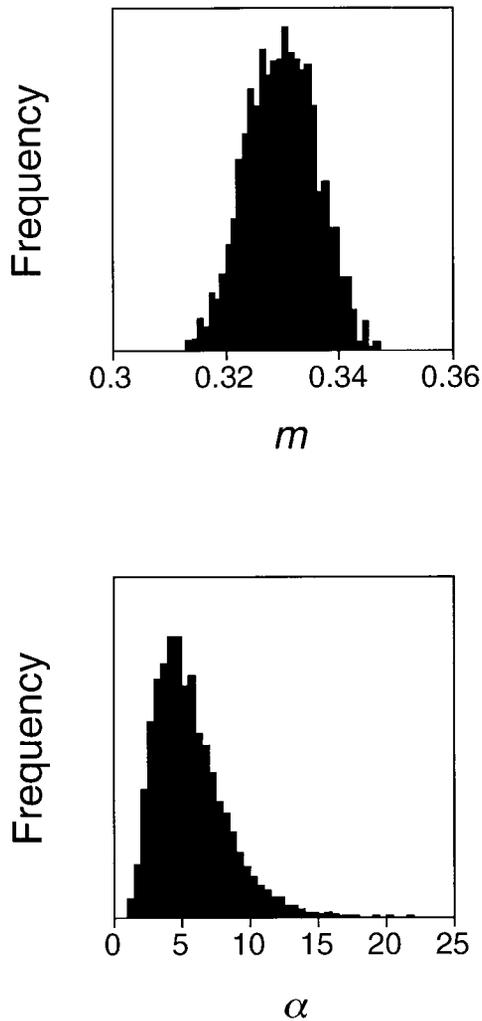
Two different priors were used in the analysis of the speciation times for the mammals. We used exponential priors with means 0.06 and 0.006 per million years. These numbers were chosen so that the means of the exponential priors were similar to the priors used earlier (0.06 × 150 mya = 10; 0.006 × 150 = 1). Figure 13 shows the relationship between estimates of speciation times for two different priors on λ for the mammal mtDNA data. We used exponential priors with means 0.06 (*x*) and 0.006 (*y*). The *x*-axis shows that speciation times differed by at most 0.2 for the two priors. The estimates of speciation times appear robust to the choice of prior used for λ.

DISCUSSION

The molecular clock hypothesis, although useful in evolutionary biology, is inaccurate in two important details; rates across lineages appear to vary and the substitution process deviates significantly from a Poisson process (Ohta and Kimura 1971; Langley and Fitch 1973, 1974; also see Gillespie 1991). This article considers only lineage effects on rate variation; that is, we only consider changes in the overall substitution rate that occur along different lineages through processes such as a change in the generation time. We do not relax

Figure 7.—The posterior probability distributions for the parameters *m* and α_P for the mammalian mitochondrial sequences. The prior for λ was exponential with mean 1.
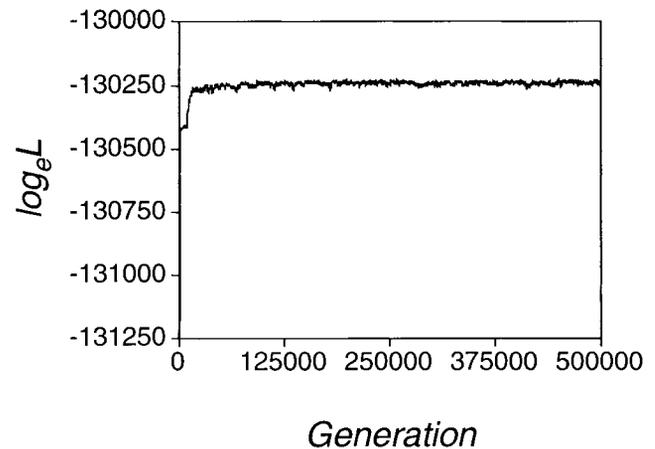
Figure 8.—The change in the $\log_e$ probability of observing the data (Equation 6) for the analysis of the mammalian mitochondrial sequences. The prior for λ was exponential with mean 10.
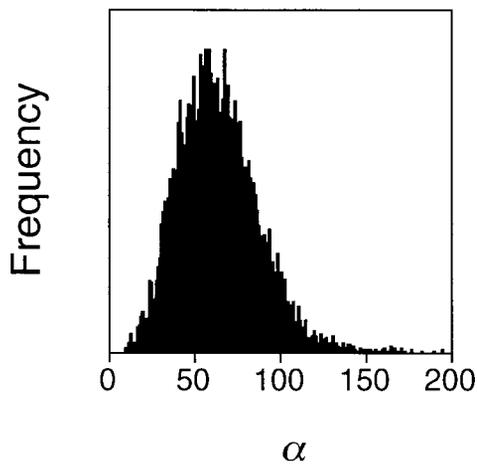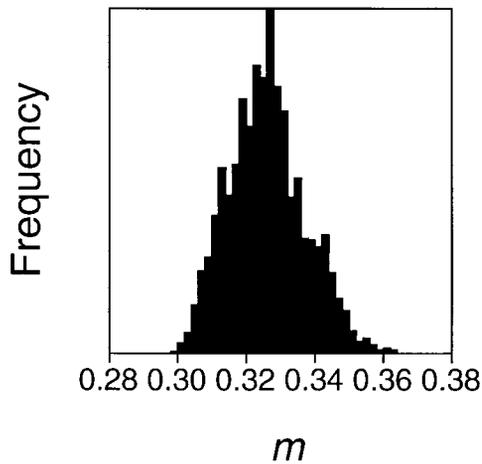
Figure 9.—The posterior probability distributions for the parameters m and α_P for the mammalian mitochondrial sequences. The prior for λ was exponential with mean 10.

Figure 10.—The Bayesian and maximum-likelihood estimates of the branch lengths for the mammalian mitochondrial sequences are highly correlated (slope = 1.0; r^2 = 0.999). Bayesian estimates of branch lengths were taken as the mean of the posterior distribution. Both sets of branch lengths were obtained under the HKY85 model of DNA substitution.

the assumption that the substitutions on a tree follow a Poisson process. It is important to relax the assumption that rates are constant across lineages because almost any study of the rate of substitution in DNA or amino acid sequences rejects the null hypothesis that rates across lineages are equal through time (Langley and Fitch 1974; Goodman *et al.* 1975; Wu and Li 1985). This is especially true now that biologists are using recently available software that allows easy testing of the molecular clock hypothesis (*e.g.*, PAML, Yang 1997; PAUP*, Swofford 1998).

Several approaches have been taken to relax the molecular clock. The most frequently used method for relaxing the rate-constancy assumption is to assign a fixed rate parameter to each branch of a phylogeny. This solution works well for the phylogeny problem where rates of substitution are often treated as nuisance parameters and the parameter of interest is the phylogeny of a group of organisms. However, allowing a rate

of substitution for each lineage on a phylogenetic tree does not allow estimation of divergence times or allow comparison of speciation times for different taxonomic groups. Sanderson (1997) made an important advance in the estimation of divergence times using molecular data by (1) showing how all of the sequences of a study can be used simultaneously to estimate the divergence time of a clade and (2) relaxing the molecular clock by smoothing the rate differences across speciation times on a tree. Similarly, Thorne *et al.* (1998) relaxed the molecular clock by assuming that rates change across speciation events. They assigned rates to descendant lineages by sampling rates from a lognormal distribu-
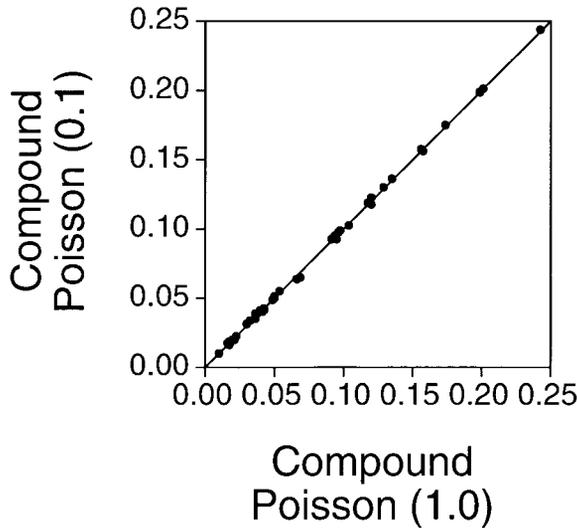
Figure 11.—The Bayesian estimates of branch lengths for the two exponential priors examined (1 and 10). The branch lengths are highly correlated (slope = 1.00; r^2 = 0.999). Bayesian estimates of branch lengths were taken as the mean of the posterior distribution. Both sets of branch lengths were obtained under the HKY85 model of DNA substitution.

tion. They evaluated the posterior distribution of speciation times using MCMC.

The compound Poisson process considered here models rate variation across lineages on a tree as a step function. There were several reasons we chose to model rate variation as a step function, with a major reason being analytical tractability. However, there are biologi-

cally plausible mechanisms that could cause rates to change in a nearly step-like manner. For example, any process that rapidly changes the substitution rate (such as a change in the proofreading mechanisms, a change in selective constraints, or a rapid change in the generation time) can be approximately described using a step function. One of the main advantages of treating rate variation as a compound Poisson process is that the model can introduce rate variation across lineages at any point on the phylogenetic tree; other methods that have been considered assume that substitution rate changes at speciation events (*e.g.*, Sanderson 1997; Thorne *et al.* 1998). Such models may be sensitive to taxon sampling. On the other hand, these solutions may prove to be more practical for data analysis because there are fewer parameters to estimate. The gamma distribution used to modify ancestral rates, however, is not so easily justified using as an argument biological plausibility. We chose to use a gamma distribution to modify rates because of its analytical tractability and flexibility. Other methods for modifying rates, however, may work just as well.

We uncovered significant rate variation across lineages for the mammalian mitochondrial DNA sequence data sets. The clock hypothesis could be rejected using a likelihood-ratio test (Felsenstein 1981). Importantly, the posterior distribution of the gamma parameter (α_P) was close to 0, which also suggests that the data were not consistent with the clock hypothesis.

Other models may also prove useful for relaxing the molecular clock assumption. For example, a doubly sto-
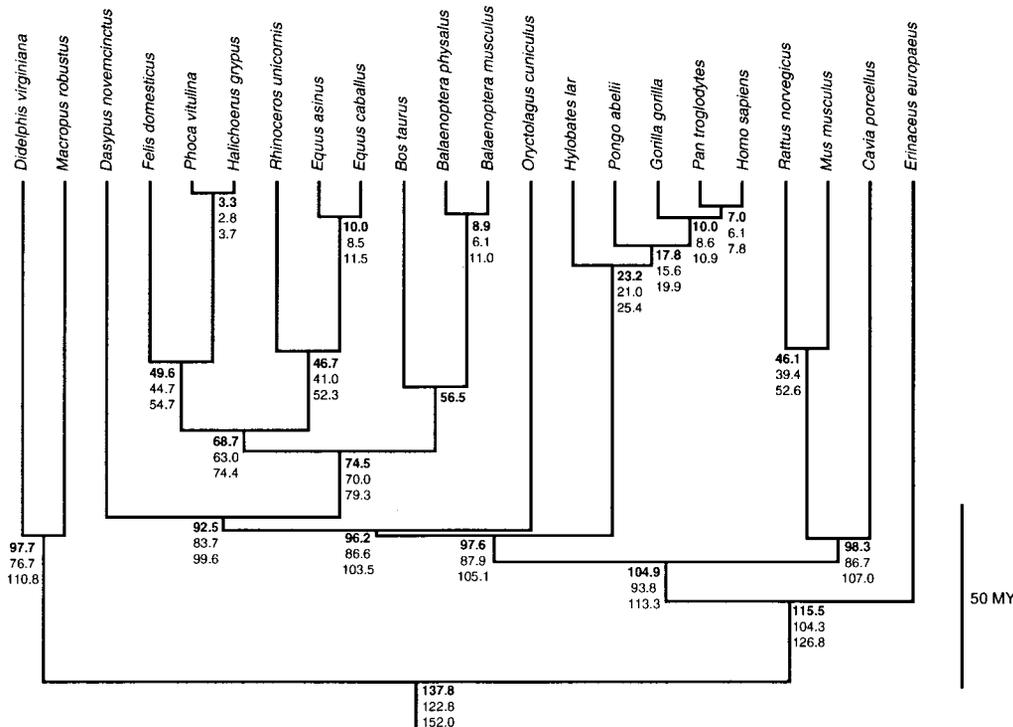
Figure 12.—The estimates of divergence times for 22 mammalian species. Estimates were made under the HKY85 + Γ model of DNA substitution. Numbers at internal nodes represent the point estimate of the divergence time (top) and the upper (middle) and lower (bottom) limit of the 95% credibility interval. An exponential prior with mean 0.006 was placed on λ.
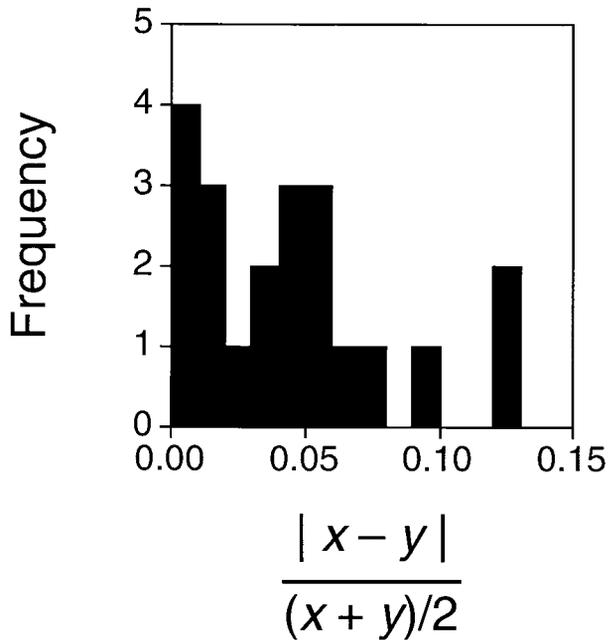
Figure 13.—The estimates of speciation times were similar regardless of the exponential prior placed on λ. Two exponential priors with means 0.06 and 0.006 were examined. The x-axis shows the proportion difference between the two estimates of speciation time. x and y represent the speciation times for the exponential priors with means 0.06 and 0.006, respectively.

to construct a more realistic probability distribution describing the divergence time of a clade.

One important result of this article is that we demonstrate that events can be placed onto a phylogenetic tree under a Poisson process and estimation performed using MCMC. The Poisson process is one of the most useful models in biology, and there are several possible extensions of the approach we describe in phylogenetics. For example, one could place events onto a tree according to a Poisson process with each event changing the synonymous/nonsynonymous rate ratio, the base frequencies, or whether the site has a positive rate or a rate of zero (i.e., the covarion model). The advantage of this approach is that it allows the rate of nucleotide substitution to change on a tree under a biologically reasonable model with the addition of a small number of parameters.

chastic (or Cox) process can be used to allow rates to vary through time (Gillespie 1991). In this process, the rate of substitution is drawn from some stationary distribution $r(u)$ (the rate at time u) with the transition probabilities calculated as they were in this article. Importantly, this process has a variance to mean ratio (index of dispersion) greater than one. One of the main difficulties of this process is the choice of the rate distribution. It is not clear what distribution would be most appropriate for modeling rate variation through time. Also, if the rate distribution has too many parameters, it may be difficult to estimate the parameters of the rate distribution while simultaneously allowing estimation of other interesting parameters (such as the divergence time of a clade).

We were able to estimate the divergence times of ferungulates (and other mammalian speciation events) using Bayesian inference under a relaxed molecular clock. In this study, we assumed one calibration date; the date of the speciation event leading to cows and whales was assumed to be exactly 56.5 mya. However, the Bayesian approach makes it easy to accommodate multiple calibration dates each with error. The error in a calibration date might be assumed to be uniformly distributed on some interval. However, this would only provide a starting point for characterizing the error in the calibration dates. Ideally, information on the density of fossil taxa around the calibration date would be used

LITERATURE CITED

Arnason, U., A. Gullberg and A. Janke, 1997 Phylogenetic analyses of mitochondrial DNA suggest a sister group relationship between Xenartha (Edentata) and Ferungulates. Mol. Biol. Evol. **14:** 762–768.

Cooper, A., and D. Penny, 1997 Mass survival of birds across the Cretaceous-Tertiary boundary: molecular evidence. Science **275:** 1109–1113.

Doolittle, F. F., D.-F. Feng, S. Tsar, G. Cho and E. Little, 1996 Determining divergence times of the major kingdoms of living organisms with a protein clock. Science **271:** 470–477.

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

Geyer, C. J., 1999 Likelihood inference for spatial point processes, pp. 79–140 in *Stochastic Geometry: Likelihood and Computation*, edited by O. E. Barndorff-Nielsen, W. S. Kendall and M. N. M. van Lieshout. Chapman and Hall/CRC, London.

Gillespie, J. H., 1991 *The Causes of Molecular Evolution.* Oxford University Press, Oxford.

Goodman, M., G. W. Moore and G. Matsuda, 1975 Darwinian evolution in the genealogy of hemoglobin. Nature **253:** 603–608.

Green, P. J., 1995 Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika **82:** 711–732.

Hasegawa, M., T. Yano and H. Kishino, 1984 A new molecular clock of mitochondrial DNA and the evolution of Hominoids. Proc. Japan Acad. Ser. B **60:** 95–98.

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

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

Huelsenbeck, J. P., B. Rannala and Z. Yang, 1997 Statistical test of host-parasite cospeciation. Evolution **51:** 410–419.

Jukes, T. H., and C. R. Cantor, 1969 Evolution of protein molecules, pp. 21–132 in *Mammalian Protein Metabolism*, edited by H. N. Munro. Academic Press, New York.

Kumar, S., and S. B. Hedges, 1998 A molecular timescale for vertebrate evolution. Nature **392:** 917–920.

Langley, C. H., and W. M. Fitch, 1973 The constancy of evolution: a statistical analysis of a and b haemoglobins, cytochrome c, and fibrinopeptide A, pp. 246–262 in *Genetic Structure of Populations*, edited by N. E. Morton. University of Hawaii Press, Honolulu.

Langley, C. H., and W. M. Fitch, 1974 An estimation of the constancy of the rate of molecular evolution. J. Mol. Evol. **3:** 161–177.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, 1953 Equations of state calculations by fast computing machines. J. Chem. Phys. **21:** 1087–1091.

Ohta, T., and M. Kimura, 1971 On the constancy of the evolutionary rate of cistrons. J. Mol. Evol. **1:** 18–25.

Sanderson, M., 1997 A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. **14:** 1218–1231.

Sokal, R. R., and C. D. Michener, 1958 A statistical method for evaluating systematic relationships. Univ. Kans. Sci. Bull. **28:** 1409–1438.

Swofford, D. L., 1998 *Phylogenetic Analysis Using Parsimony (* and other Methods).* Sinauer Associates, Sunderland, MA.

Thorne, J., H. Kishino and I. S. Painter, 1998 Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. **15:** 1647–1657.

Tierney, L., 1994 Markov chains for exploring posterior distributions. Ann. Stat. **22:** 1701–1762.

Wray, G., J. S. Levinton and L. Shapiro, 1996 Molecular evidence for deep Precambrian divergences among metazoan phyla. Science **274:** 568–573.

Wu, C.-I., and W.-H. Li, 1985 Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA **82:** 1741–1745.

Yang, Z., 1993 Maximum likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol. Biol. Evol. **10:** 1396–1401.

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

Yang, Z., 1994b Estimating the pattern of nucleotide substitution. J. Mol. Evol. **39:** 105–111.

Yang, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. CABIOS **15:** 555–556.

Zuckerkandl, E., and L. Pauling, 1962 Molecular disease, evolution, and genetic heterogeneity, pp. 189–225 in *Horizons in Biochemistry*, edited by M. Kasha and B. Pullman. Academic Press, New York.

Zuckerkandl, E., and L. Pauling, 1965 Evolutionary divergence and convergence in proteins, pp. 97–166 in *Evolving Genes and Proteins*, edited by V. Bryson and H. J. Vogel. Academic Press, New York.

Communicating editor: S. Tavaré

APPENDIX

Here we demonstrate detailed balance between the move types that add and delete an event from the tree. Following the notation of Green (1995), consider the move m "insert a point into position $\xi + 1$ in the sequence" and its dual return move "delete the point from position $\xi + 1$ in the sequence." We show that (from Green 1995; p. 714)

$$\int_A \pi(dx) \int_B q_m(x, dx') \alpha_{Pm}(x, x') = \int_B \pi(dx') \int_A q_m(x', dx) \alpha_{Pm}(x', x) \quad \text{(A1)}$$

holds for general Borel sets A and B. It is sufficient to show it holds when A and B are the products of open intervals in the spaces ξ and $\xi + 1$ events on the tree, respectively. When $\xi = 0$, an open interval is just the point $(0, \emptyset, \emptyset)$.

Let a generic point in A be $x = (\xi, z, r)$, where $z_i \in [0, T]$ and $r_i \in (0, \infty)$ for $i = 1, 2, \ldots, \xi$. The point $x' = (\xi + 1, (z, z^*), (r, r^*))$ is a generic point in B. Note that $q_m(x, dx') = 0$ if any of the arguments for the first ξ u's or r's in x' differ from the corresponding arguments in x (because m adds a point at the end of the sequence). This means that the double integrals will measure 0 over any points in A that are not in B restricted to the smaller dimension or points in B that are not in A expanded to the larger dimension. Therefore, it suffices when specifying A and B to assume that the open intervals for the first ξ dimensions of the z's and r's agree.

The purported acceptance probabilities for traversing the prior are

$$\alpha_{P\xi} = \alpha_{Pm}(x, x') = \min\left\{1, \frac{\lambda T d_{\xi+1}}{(\xi + 1) b_\xi}\right\} \quad \text{(A2)}$$

and

$$\beta_\xi = \alpha_{Pm}(x', x) = \min\left\{1, \frac{(\xi + 1) b_\xi}{\lambda T d_{\xi+1}}\right\} \quad \text{(A3)}$$

for adding and deleting an event, respectively. Note that the expressions depend on ξ only and may be factored out of the integral expressions.

Let $A = \{\xi\} \times \prod_{i=1}^{\xi}(\alpha_i, c_i) \times \prod_{i=1}^{\xi}(s_i, t_i)$ and $B = \{\xi + 1\} \times \prod_{i=1}^{\xi+1}(\alpha_i, c_i) \times \prod_{i=1}^{\xi+1}(s_i, t_i)$. This means that A is the set where $a_i < z_i < c_i$ and $s_i < r_i < t_i$ for generic $x \in A$ and there is a similar description for B. We have

$$\pi(dx) = \frac{e^{-\lambda T}(\lambda T)^\xi}{\xi!} \times \left(\frac{1}{T}\right)^\xi \times \sum_{i=1}^{\xi} g(r_i | \alpha_P) \times dx \quad \text{(A4)}$$

and

$$\pi(dx') = \frac{e^{-\lambda T}(\lambda T)^{\xi+1}}{(\xi + 1)!} \times \left(\frac{1}{T}\right)^{\xi+1}$$
$$\times \prod_{i=1}^{\xi} g(r_i | \alpha_P) \times g(r^* | \alpha_P) \times dx'. \quad \text{(A5)}$$

We also have

$$q_m(x, dx') = b_\xi\left(\frac{1}{\xi + 1}\right)\left(\frac{1}{T}\right) g(r^* | \alpha_P) \, dz^* \, dr^*$$
$$\times I\{x_i = x_i' : 1 \le i \le \xi\} \quad \text{(A6)}$$

and

$$q_m(x', dx) = d_{\xi+1}\left(\frac{1}{\xi + 1}\right)\left(\frac{1}{T}\right) \times I\{x_i = x_i' : 1 \le i \le \xi\}. \quad \text{(A7)}$$

The left-hand side of Equation 22 simplifies to

$$\int_A \pi(dx) \int_B q_m(x,\, dx') \alpha_{Pm}(x,\, x')$$

$$= \int_A \pi(dx) \int_{a_{\xi+1}}^{c_{\xi+1}} \int_{s_{\xi+1}}^{t_{\xi+1}} b_\xi \left(\frac{1}{\xi+1}\right)\left(\frac{1}{T}\right)$$

$$\times\, g(r^*|\alpha_P) \alpha_{P_\xi} dr^*\, dz^*$$

$$= \alpha_{P_\xi} \beta_\xi \left(\frac{1}{\xi+1}\right)\left(\frac{1}{T}\right) \times (c_{\xi+1} - a_{\xi+1})\,(G(t_{\xi+1}|\alpha_P)$$

$$-\, G(s_{\xi+1}|\alpha_P)) \int_A \pi(dx)$$

$$= \alpha_{P_\xi} \beta_\xi \left(\frac{1}{\xi+1}\right)\left(\frac{1}{T}\right)$$

$$\times\, (c_{\xi+1} - a_{\xi+1})\,(G(t_{\xi+1}|\alpha_P) - G(s_{\xi+1}|\alpha_P))$$

$$\times \int_A \frac{e^{-\lambda T}(\xi T)^\xi}{\xi!}\left(\frac{1}{T}\right)^\xi \prod_{i=1}^\xi g(r_i|\alpha_P)\, dx$$

$$= \alpha_{P_\xi} \times \left(\frac{b_\xi}{T}\right) \times \frac{e^{-\lambda T}\lambda^\xi}{(\xi+1)!} \times \prod_{i=1}^{\xi+1}(c_i - a_i)$$

$$\times \prod_{i=1}^{\xi+1} (G(t_i|\alpha_P) - G(s_i|\alpha_P)),$$

where $G(\cdot|\alpha_P)$ is the cumulative distribution function of $g(\cdot|\alpha_P)$. The right-hand side of equation A1 simplifies to

$$\beta_{\xi+1} \times \left(\frac{d_{\xi+1}}{\xi+1}\right) \times \frac{e^{-\lambda T}(\lambda T)^{\xi+1}}{(\xi+1)!} \times \prod_{i=1}^{\xi+1}(c_i - a_i)$$

$$\times \prod_{i=1}^{\xi+1} (G(t_i|\alpha_P) - G(s_i|\alpha_P)). \tag{A8}$$

The left-hand side over the right-hand side is

$$\left(\frac{\alpha_{P_\xi}}{\beta_{\xi+1}}\right) \times \left(\frac{(\xi+1)\,b_\xi}{\lambda T d_{\xi+1}}\right) = 1, \tag{A9}$$

so detailed balance is satisfied for the jump moves.

Note that

$$\left(\frac{\alpha_{P_\xi}}{\beta_{\xi+1}}\right) = \left(\frac{\lambda T d_{\xi+1}}{(\xi+1)\,b_\xi}\right) = 1, \tag{A10}$$

whether $\lambda T d_{\xi+1} > (\xi+1)\,b_\xi$ or $\lambda T d_{\xi+1} < (\xi+1)\,b_\xi$. A similar calculation for inserting/deleting a point in position i in the list of events would be similarly derived, differing only in labeling of subscripts.