Estimating Absolute Rates of Synonymous and Nonsynonymous Nucleotide Substitution in Order to Characterize Natural Selection and Date Species Divergences

Tae-Kun Seo,† *Hirohisa Kishino,*‡ *and Jeffrey L. Thorne*†

†Bioinformatics Research Center, Box 7566, North Carolina State University; and ‡Laboratory of Biometrics, Graduate School of Agriculture and Life Sciences, University of Tokyo, Bunkyo-ku, Tokyo, Japan

The rate of molecular evolution can vary among lineages. Sources of this variation have differential effects on synonymous and nonsynonymous substitution rates. Changes in effective population size or patterns of natural selection will mainly alter nonsynonymous substitution rates. Changes in generation length or mutation rates are likely to have an impact on both synonymous and nonsynonymous substitution rates. By comparing changes in synonymous and nonsynonymous rates, the relative contributions of the driving forces of evolution can be better characterized. Here, we introduce a procedure for estimating the chronological rates of synonymous and nonsynonymous substitutions on the branches of an evolutionary tree. Because the widely used ratio of nonsynonymous and synonymous rates is not designed to detect simultaneous increases or simultaneous decreases in synonymous and nonsynonymous rates, the estimation of these rates rather than their ratio can improve characterization of the evolutionary process. With our Bayesian approach, we analyze cytochrome oxidase subunit I evolution in primates and infer that nonsynonymous rates have a greater tendency to change over time than do synonymous rates. Our analysis of these data also suggests that rates have been positively correlated.

Introduction

Much attention has been paid to synonymous and nonsynonymous substitution amounts and their ratio (Nei and Gojobori 1986; Goldman and Yang 1994; Muse and Gaut 1994; Ina 1995; Ohta 1995; Zhang, Rosenberg and Nei 1998; Yang 1998; Yang and Nielsen 1998, 2000, 2002; Yang and Bielawski 2000; Yang and Swanson 2002), but less emphasis has been placed on the actual values of nonsynonymous and synonymous rates. This is unfortunate because absolute rates of nonsynonymous and synonymous substitutions can provide more insight into evolutionary mechanism than can the amounts or their ratio. A change in the ratio may be due to alteration of the nonsynonymous or synonymous substitution rate or both. With the ability to estimate the actual nonsynonymous and synonymous rates, the factors involved in a change to the evolutionary process can be more readily identified.

For estimation of substitution rates in chronological time units (e.g., years, millions of years, etc.), the most difficult obstacle is that only the amount of evolution (i.e., the product of rate and time) can be inferred solely from molecular sequence data. This confounding of rates and times necessitates the addition of information external to the molecular sequence data if actual nonsynonymous and synonymous substitution rates are to be accurately determined. Conventionally, this external information has taken the form of fossil data that calibrates a "molecular clock" that is assumed to specify a common rate of evolution on all branches of a phylogeny. Recently, the sampling times of quickly evolving organisms such as RNA viruses have been exploited to calibrate the molecular clock on some genealogies (Rambaut 2000;

Drummond and Rodrigo 2000). Regardless of how calibration is achieved, when the goal is to determine how substitution rates have changed over time, the molecular clock assumption of constant rates has little utility. There have been proposals for analyzing sequence data with "local clocks" that operate on only a portion of an evolutionary tree (Yoder and Yang 2000). This local clock approach is attractive, but there may be some difficulty in determining how many local clocks should be employed and in determining which branches should be governed by which local clocks.

Sanderson (1997; see also Sanderson 2002, 2003) introduced a pioneering approach that permits different rates in different lineages and that lacks the model specification difficulties of the local clock approaches. Later, Bayesian approaches for allowing the rate of evolution to change over time were developed (Thorne, Kishino, and Painter 1998; Huelsenbeck, Larget, and Swofford 2000; Kishino, Thorne, and Bruno 2001; Thorne and Kishino 2002; Aris-Brosou and Yang 2002). An advantage of the Bayesian implementations is their ability to incorporate both the full likelihood and prior information. Here, we adapt previously proposed Bayesian approaches (Thorne, Kishino, and Painter 1998; Kishino, Thorne, and Bruno 2001; Thorne and Kishino 2002) to a codon model of evolution. In this way, absolute synonymous and nonsynonymous substitution rates as well as divergence times can be estimated. In addition, the tendency for nonsynonymous rates to change over time can be compared and contrasted with the tendency for synonymous rates to change over time. We also introduce a test of whether nonsynonymous and synonymous substitution rates change in a correlated fashion over time, and we discuss the strengths and weaknesses of this test.

Our new methodology is applied to a group of cytochrome oxidase subunit I (COXI) primate sequences. Cytochrome oxidase is a protein complex that affects aerobic energy metabolism by playing an important role in the final step of the electron transport system. Cytochrome

oxidase is composed of 13 subunits. Subunits 1, 2, and 3 are encoded in the mitochondrion, and the others are encoded in the nucleus. The molecular evolution of some of the subunits has been investigated in detail and, despite the different genomic origins of the subunits that have been studied, a consistent pattern among subunits is that anthropoids have experienced a higher ratio of non-synonymous to synonymous change than have non-anthropoids (Adkins and Honeycutt 1994; Adkins Honeycutt, and Disotell 1996; Wu et al. 1997, 2000; Schmidt, Goodman, and Grossman 1999; Andrews and Easteal 2000; Grossman et al. 2001; Wildman, Goodman, and Grossman 2002). One suggested explanation of this pattern is that there has been an increase in size of the neocortex during anthropoid evolution and that the increased needs of the neocortex for energy may be responsible for a change in aerobic energy metabolism in anthropoids. Selection for an alteration to aerobic energy metabolism could have led to the higher nonsynonymous to synonymous substitution ratio of cytochrome oxidase in anthropoids (see Grossman et al. 2001).

By analyzing COXI, we obtain divergence time estimates similar to those obtained in a previous study that analyzed total mitochondrial genomes with a treatment that did not distinguish between synonymous and non-synonymous changes (Cao et al. 2000). Our inferred pattern of nonsynonymous rate increase in the anthropoid group is consistent with previous analyses based on different methods (Andrews and Easteal 2000; Wu et al. 2000). An advantage of our Bayesian method over traditional approaches that are based on pairwise distances is that it can localize rate changes to relatively specific portions of a phylogeny. For COXI, we tentatively conclude that nonsynonymous rates have been comparatively high throughout the anthropoid group and not just on the branch that was immediately ancestral to the anthropoid group.

Methods
Codon Model

Consider a slight modification of the codon substitution model introduced by Goldman and Yang (1994). The modification allows the expected proportion of nonsynonymous substitutions to vary among branches and has been previously suggested (Yang 1998). With this modification, the instantaneous rate of substitution $q_{ij}^{(m)}$ from codon i to codon j on branch m is

$$q_{ij}^{(m)} = \begin{cases} 0, & \text{for more than one nucleotide} \\ & \quad \text{substitution between i and j} \\ u^{(m)}\pi_j, & \text{for synonymous transversion} \\ u^{(m)}\kappa\pi_j, & \text{for synonymous transition} \\ u^{(m)}\omega^{(m)}\pi_j, & \text{for nonsynonymous transversion} \\ u^{(m)}\omega^{(m)}\kappa\pi_j, & \text{for nonsynonymous transition,} \end{cases}$$

(1)

where π_j is the frequency of codon j, κ differentiates between transitions and transversions, and $\omega^{(m)}$ distinguishes between nonsynonymous and synonymous substitutions for branch m. In contrast to the $\omega^{(m)}$ parameter,

the $u^{(m)}$ parameter affects the chronological rates of both synonymous and nonsynonymous substitutions on branch m. In equation (1), the π_i parameters and κ are common to all branches.

The average rate on a branch m of synonymous and nonsynonymous substitution per chronological time unit will be respectively denoted $r_s^{(m)}$ and $r_n^{(m)}$. These rates can be defined as

$$r_s^{(m)} = \sum_i \pi_i \sum_{j \neq i} q_{ij}^{(m)}$$

$$i \rightarrow j: \text{synonymous substitution} \qquad (2)$$

and

$$r_n^{(m)} = \sum_i \pi_i \sum_{j \neq i} q_{ij}^{(m)}$$

$$i \rightarrow j: \text{nonsynonymous substitution.} \qquad (3)$$

The synonymous and nonsynonymous branch lengths for a branch m with chronological time duration $t^{(m)}$ are

$$b_s^{(m)} = r_s^{(m)} t^{(m)} \qquad (4)$$

and

$$b_n^{(m)} = r_n^{(m)} t^{(m)}. \qquad (5)$$

The total length of branch m (i.e., the expected number of substitutions per codon) is then

$$b_s^{(m)} + b_n^{(m)} = r_s^{(m)} t^{(m)} + r_n^{(m)} t^{(m)}. \qquad (6)$$

Knowing $b_s^{(m)} + b_n^{(m)}$ and $\omega^{(m)}$ along with κ and the π_i parameters allows determination of $b_s^{(m)}$ and $b_n^{(m)}$. This is because κ, the π_i parameters, and $\omega^{(m)}$ define the $q_{ij}^{(m)} / u^{(m)}$ values. The $q_{ij}^{(m)} / u^{(m)}$ values then allow calculation of $r_s^{(m)} / u^{(m)}$ and $r_n^{(m)} / u^{(m)}$ via equations (2) and (3). Equation (6) can be used to solve for $u^{(m)}t^{(m)}$. Finally, $u^{(m)}t^{(m)}$ with $r_s^{(m)} / u^{(m)}$ determines $b_s^{(m)}$ (eq. 4) and $u^{(m)}t^{(m)}$ with $r_n^{(m)} / u^{(m)}$ determines $b_n^{(m)}$ (eq. 5).

Likewise, knowing $b_s^{(m)}$ and $b_n^{(m)}$ along with κ and the π_i parameters allows determination of $b_s^{(m)} + b_n^{(m)}$ and $\omega^{(m)}$. This is because κ and the π_i parameters are sufficient to define $r_s^{(m)} / u^{(m)}$. With equation (4), $u^{(m)}t^{(m)}$ can be expressed in terms of $r_s^{(m)} / u^{(m)}$ and $b_s^{(m)}$. Using $u^{(m)}t^{(m)}$ and equation (5), the value of $r_n^{(m)} / u^{(m)}$ follows. Finally, the value of $\omega^{(m)}$ can be found by applying equation (3).

We stress this ability to interconvert between the two sets of parameters because the $(\pi_i, \kappa, b_s^{(m)} + b_n^{(m)}, \omega^{(m)})$ parameterization is more natural for calculation of likelihoods whereas the $(\pi_i, \kappa, b_s^{(m)}, b_n^{(m)})$ parameterization is more natural when considering our model of synonymous and nonsynonymous rate change. Ability to interconvert between these two parameterizations also allows the invariance property of maximum likelihood estimators (e.g., Hogg and Craig 1995, p. 265) to be exploited so that estimates under one parameterization can therefore be converted to be estimates under the other. Maximum likelihood estimates are relevant because, as described below, our Bayesian approach involves approximating the likelihood surface needed for posterior density estimation by Taylor expansion around the maximum likelihood estimates.

Model of Nonsynonymous/Synonymous Rate Change

Our model of rate evolution is similar in spirit to that of Kishino, Thorne, and Bruno (2001). When evolutionary lineages are separated by only small amounts of time, their evolutionary rates are likely to be similar. When lineages are distantly related, their rates are more likely to substantially differ. In other words, evolutionary rates are autocorrelated.

Rates of evolution are likely to undergo a nearly continuous process of change on a lineage. Because molecular sequence data contain only limited information regarding this nearly continuous process of change, the average evolutionary rate on a branch is approximated in our treatment by the mean of the rate of evolution at the nodes that begin and end the branch. If rates of evolution changed in a deterministic linear fashion between beginning and ending nodes of a branch, our approximation would be exact. However, we simply view this connection between node rates and average rates on a branch as a computationally simplifying assumption. With this approximation, the average rates of evolution on branches and the expected amounts of evolution on branches both depend completely on rates of evolution at the nodes of a tree.

To model the rates of evolution at tree nodes, a nonsynonymous rate variation parameter v_n and a synonymous rate variation parameter v_s are introduced. The case where $v_n = v_s = 0$ represents the biologically unrealistic case of a molecular clock. These rate variation parameters are restricted to non-negative values and, as they increase, the expected departure from a molecular clock becomes more extreme. Given the nonsynonymous rate r_n^i and the synonymous rate r_s^i at a node i that begins a branch of time duration t, the logarithms of the rates at the node j that ends the branch are assumed to follow a bivariate normal distribution of the form:

$$\begin{pmatrix} \log r_n^j \\ \log r_s^j \end{pmatrix} \sim N\left(\begin{pmatrix} \log r_n^i \\ \log r_s^i \end{pmatrix}, \begin{pmatrix} v_n t & 0 \\ 0 & v_s t \end{pmatrix} \right). \quad (7)$$

With this parameterization, the expected value of $r_n^j(r_s^j)$ is larger than $r_n^i(r_s^i)$, even though the expected value of $\log r_n^j(\log r_s^j)$ equals $\log r_n^i(\log r_s^i)$. The difference gets bigger when $v_n(v_s)$ or t is large. To eliminate such differences, Kishino, Thorne, and Bruno (2001) forced the expected value of rates at the end of branches to be equal to their values at the beginning of branches. Because the parameterization in equation (7) simplifies our proposed test of concordance between synonymous and nonsynonymous rates (see below), we do not make any attempt to eliminate such differences here.

As a further simplification, we set the a priori covariance of $\log r_n^j$ and $\log r_s^j$ to be zero in equation (7). Future implementations could easily permit this covariance to be nonzero a priori or might even define a prior distribution for a parameter representing the covariance. This hierarchical approach may be worthwhile because some features of the evolutionary process such as generation time or mutation rate might simultaneously affect both nonsynonymous and synonymous rates when they change. Although we model synonymous and non-

synonymous rate change as occurring independently, we describe a procedure that uses sequence data to detect correlation in changes of these rates.

We assume the topology relating the sequences is known. Following Thorne, Kishino, and Painter (1998), our procedure requires an outgroup to root the ingroup, but no attempt is made to model rate evolution in the outgroup taxa. With the model of rate evolution described by equation (7), the prior density of rates at a node is defined in terms of the rates at its parental node, and this leaves the prior distribution for the rates at the ingroup root node to be specified. In our implementation, the ingroup root node has priors for nonsynonymous and synonymous rates that are specified by two independent gamma distributions. For v_n and v_s, we specify priors that are exponential distributions.

For every rooted tree, our implementation arbitrarily selects one branch that emanates from the ingroup root node. On this branch, the nonsynonymous and synonymous rates at the end of the branch are forced to be identical to the corresponding rates at the ingroup root node. In previous work (Kishino, Thorne, and Bruno 2001), this forcing of no rate change along one branch has made Markov chain Monte Carlo (MCMC) approximation of posterior rate densities computationally tractable. With a similar motivation, the forcing of no rate change was included here to achieve computational tractability with the MCMC procedure described below. With the exception of this single branch, evolution of rates on all branches is governed by equation (7).

Prior Distribution

For the Bayesian estimation of chronological rates of nonsynonymous and synonymous substitutions, a prior distribution of node times must be specified. The details of the prior distribution for node times are identical to those described in Kishino, Thorne, and Bruno (2001), and only an outline is provided here. For the case where there are no constraints on node times due to fossil or other information, our implementation uses a gamma distribution to describe the time of the ingroup root node. Given the ingroup root time, the prior distribution for the times of all other internal nodes are jointly distributed according to a modified Dirichlet distribution (Kishino, Thorne, and Bruno 2001). Representing the node times (including the ingroup root) by a vector $\mathbf{T}$, the modified Dirichlet distribution and the gamma distribution of the ingroup root time determine the prior density of the node times $p(\mathbf{T})$.

Often, fossil or other evidence suggests that a certain node time may exceed some value. Likewise, evidence external to the molecular sequence data may indicate that a node is more recent than some date. Sanderson (1997) demonstrated that constraints on node times can be valuable for estimating dates of divergence. When a set $\mathbf{C}$ of constraints on node times is available, we condition upon these constraints and use this conditional prior $p(\mathbf{T} \mid \mathbf{C})$ to estimate divergence times and evolutionary rates.

Posterior Distribution and MCMC

Following Yang (1998; see also Pederson, Wiuf, and Christiansen 1998), we have explored various parameterizations of the codon frequencies in terms of frequencies of the four nucleotide types. One parameterization forces these nucleotide frequencies to be identical at all three codon positions, and another allows different nucleotide frequencies at the first, second, and third codon positions. However, for all results presented here, we estimate probabilities of the different codon types by their empirical frequencies. We then treat these estimated codon frequencies as well as the maximum likelihood estimate of κ as if their values were known with certainty. Because we neglect uncertainty regarding the κ and π_i parameters, we omit them from subsequent notation and focus on approximating the posterior distribution $p(\mathbf{T}, r_n, r_s, v_n, v_s \mid \mathbf{X}, \mathbf{C})$. We note that the values of $\mathbf{T}$, $\mathbf{r}_n$, and $\mathbf{r}_s$ are sufficient to determine $\mathbf{b}_n$ and $\mathbf{b}_s$. In turn, $\mathbf{b}_n$ and $\mathbf{b}_s$ along with the known κ and π_i parameters determine (or can be determined by) ω and the sum $\mathbf{b}_n + \mathbf{b}_s$.

For given sequence data $\mathbf{X}$, the posterior distribution of unknown parameters is

$$
\begin{aligned}
&p(\mathbf{T}, r_n, r_s, v_n, v_s \mid \mathbf{X}, \mathbf{C}) \\
&= \frac{p(\mathbf{T}, r_n, r_s, v_n, v_s, \mathbf{X} \mid \mathbf{C})}{p(\mathbf{X} \mid \mathbf{C})} \\
&= \frac{p(\mathbf{X} \mid \mathbf{T}, r_n, r_s) p(r_n, r_s \mid \mathbf{T}, v_n, v_s) p(\mathbf{T} \mid \mathbf{C}) p(v_n) p(v_s)}{p(\mathbf{X} \mid \mathbf{C})},
\end{aligned}
$$

(8)

where $\mathbf{T}$, $\mathbf{r}_n$, and $\mathbf{r}_s$ respectively represent the vectors of the divergence times, the nonsynonymous rates r_n^i, and the synonymous rates r_s^i. Because $p(\mathbf{X} \mid \mathbf{T}, r_n, r_s) = p(\mathbf{X} \mid \mathbf{b}_n, \mathbf{b}_s)$, equation (8) can be rewritten as

$$
\begin{aligned}
&p(\mathbf{T}, r_n, r_s, v_n, v_s \mid \mathbf{X}, \mathbf{C}) \\
&= \frac{p(\mathbf{X} \mid \mathbf{b}_n, \mathbf{b}_s) p(r_n, r_s \mid \mathbf{T}, v_n, v_s) p(\mathbf{T} \mid \mathbf{C}) p(v_n) p(v_s)}{p(\mathbf{X} \mid \mathbf{C})}.
\end{aligned}
$$

(9)

Except for the treatment of $p(\mathbf{X} \mid \mathbf{b}_n, \mathbf{b}_s)$, the Metropolis-Hastings implementation that we use to sample from the posterior distribution $p(\mathbf{T}, r_n, r_s, v_n, v_s \mid \mathbf{X}, \mathbf{C})$ is a straightforward extension of that described earlier for estimating divergence times and evolutionary rates with less realistic models of sequence change (Thorne, Kishino, and Painter 1998; Kishino, Thorne, and Bruno 2001). The $p(\mathbf{X} \mid \mathbf{b}_n, \mathbf{b}_s)$ term in equation (9) is a likelihood and can be calculated with Felsenstein's pruning algorithm (Felsenstein 1981). Because the pruning algorithm is computationally demanding, we instead approximate $p(\mathbf{X} \mid \mathbf{b}_n, \mathbf{b}_s)$ up to a proportionality constant with a Taylor expansion of the log-likelihood function around the maximum likelihood estimates of $\mathbf{b}_n$ and $\mathbf{b}_s$ (see *Appendix*). Key to this approximation is the ability, as described earlier, to interconvert between the $(\mathbf{b}_n + \mathbf{b}_s, \omega, \pi_i, \kappa)$ parameterization that is convenient for calculating likelihoods and the $(\mathbf{b}_n, \mathbf{b}_s, \pi_i, \kappa)$ parameterization used in the $p(\mathbf{X} \mid \mathbf{b}_n, \mathbf{b}_s)$ term of equation (9).

We note that the approximation of $p(\mathbf{X} \mid \mathbf{b}_n, \mathbf{b}_s)$ that is implemented here is an improvement on the multivariate normal one of earlier studies (Thorne, Kishino, and Painter 1998; Kishino, Thorne, and Bruno 2001; Thorne and Kishino 2002). The one here is equivalent to the multivariate normal treatment when maximum likelihood branch length estimates are not close to 0. The approximation of $p(\mathbf{X} \mid \mathbf{b}_n, \mathbf{b}_s)$ in earlier studies (Thorne, Kishino, and Painter 1998; Kishino, Thorne, and Bruno 2001; Thorne and Kishino 2002) treated maximum likelihood branch length estimates that are close to 0 in an ad hoc fashion. Although the approximation here is similar to those applied elsewhere (e.g., see Gelman et al. 1995, p. 95; also see the supplementary information of Korber et al. 2000 at http://www.santafe.edu/~btk/science-paper/bette.html), we describe it in detail (see *Appendix*).

To approximate the likelihood function $p(\mathbf{X} \mid \mathbf{b}_n, \mathbf{b}_s)$ up to a constant of proportionality, we need to numerically evaluate the first and second derivatives with respect to branch lengths of the log-likelihood function around its maximum. Our experience has been that numerical stability of second derivative approximations is more difficult to achieve with codon models than with simpler models that have independent evolution among nucleotide positions. For this reason, we evaluated second derivatives with the strategy described in the *Appendix*. Our experience is that this strategy is numerically stable even when synonymous branch lengths are rather large.

Correlated Change of Synonymous and
Nonsynonymous Rates: Test of Concordance

In addition to inferring the pattern of changes in chronological rates of synonymous and nonsynonymous substitutions over time, it is of particular interest to examine whether changes in the synonymous and non-synonymous rates are correlated. Because mutation rate and generation length affect both synonymous and non-synonymous rates, changes in either mutation rate or generation length could yield changes in both these rates. Consider the null hypothesis that the synonymous rate and the nonsynonymous rate change independently. Let $r_n^{i,j}$ and $r_s^{i,j}$ be respectively the nonsynonymous and synonymous rates at node i in the jth sample from the Markov chain used in the MCMC approximation of the posterior distribution. We define the concordance measure

$$
\delta(i,j) := \begin{cases} 1 & \text{if} (r_n^{i,j} - r_n^{i_p,j})(r_s^{i,j} - r_s^{i_p,j}) > 0 \\ 0 & \text{otherwise,} \end{cases}
$$

where i_p is the parent node of i. The concordance measure takes the value 1 if both synonymous and nonsynonymous rates increase on a branch or if both decrease on a branch. Otherwise, the concordance measure has the value 0. Then, the posterior probability of the concordance for the branch between i and i_p is estimated as

$$
\widehat{q(i)} = \frac{1}{N} \sum_{j=1}^{N} \delta(i,j),
$$

(10)

where N is total number of MCMC samples. The average among branches of these estimated concordance probabilities is

$$S = \frac{1}{B} \sum_{i=1}^{B} \widehat{q(i)}, \qquad (11)$$

where B is the total number of branches on which synonymous and nonsynonymous rates could change. If S is close to 1, synonymous and nonsynonymous rates are inferred to change in the same direction on most branches of the tree. However, a value of S close to 1 does not necessarily indicate strong evidence of a positive correlation between synonymous and nonsynonymous rate changes, because a tree consisting of relatively few branches could yield a data set where S is close to 1. With a tree that has few branches, it is plausible that synonymous and nonsynonymous rates changed in the same direction on these few branches just by chance rather than because of a general tendency for these rate changes to be positively correlated.

To evaluate the null hypothesis that concordance and discordance are equally likely, we need to approximate the null distribution of our test statistic S. If the null hypothesis is true, the expected value of S is 0.5 and the null distribution of S can be simulated. The key to our simulation strategy is that $q(i)$ represents the posterior probability of concordance for branch i. Likewise, $1 - q(i)$ represents the posterior probability of discordance. When the null hypothesis is true, the observation that the posterior probability of concordance has some particular value would be exactly as surprising as the observation that the posterior probability of discordance takes that particular value. In other words, the distribution of $q(i)$ among branches i is expected to be symmetric about 0.5 when the null hypothesis is true. Consider a procedure that randomly determines a value $q(i)^*$ by setting $q(i)^* = \widehat{q(i)}$ with probability 0.5 and by setting $q(i)^* = 1 - \widehat{q(i)}$ otherwise. The quantity

$$S^* = \frac{1}{B} \sum_{i=1}^{B} q(i)^*$$

is then approximately a sample from the null distribution of S. A large number of values of S^* can then be simulated. If the value of S exceeds a proportion $(1 - \alpha)$ of the S^* realizations, then the null hypothesis can be rejected at a significance level α in favor of the alternative hypothesis that synonymous and nonsynonymous rates change in a concordant fashion.

Results

We illustrate our approach by analyzing cytochrome oxidase subunit I (COXI) from 19 primates along with two rodent outgroup species. Specifically, the represented primates and their GenBank accession numbers are as follows: *Homo sapiens* (J01415); *Pan troglodytes* (D38113); *Pan paniscus* (D38116); *Gorilla gorilla* (D38114); *Pongo pygmaeus pygmaeus* (D38115); *Pongo pygmaeus abelii* (X97707); *Hylobates lar* (NC002082); *Theropithecus gelada* (AF312702); *Cercopithecus ae-*

thiops (AF312703); *Saimiri ustus* (AF312704); *Saimiri sciureus* (AF312705); *Tarsius bancanus* (Af312706); *Perodicticus potto* (AF312707); *Propithecus verreauxi* (Af312708); *Macaca sylvanus* (NC002764); *Cebus albifrons* (NC002763); *Papio hamadryas* (NC001992); *Nycticebus coucang* (NC002765); *Lemur catta* (NC004025). The two rodent outgroup species are these: *Mus musculus* (V00711); *Rattus norvegicus* (X14848). Sequence data were manually edited after being aligned with the default option of the Clustal W software (Thompson, Higgins, and Gibson 1994). All gaps in the alignment were removed prior to subsequent analyses.

Maximum Likelihood Estimation of Branch Lengths

The COXI sequences were assumed to be related by topology of Purvis (1995). The π_i parameters were estimated with empirical codon frequencies, and other parameters $(\kappa, \omega^{(m)}, b_n^{(m)}, b_s^{(m)})$ were estimated via the maximum likelihood method. Figure 1*a* depicts the nonsynonymous branch length estimates, and figure 1*b* shows the synonymous branch length estimates. Although some of the branches have maximum likelihood estimates that exceed one expected synonymous change per codon, the likelihoods for these maximum likelihood estimates are still substantially higher than when the synonymous branch lengths are forced to be infinite. Therefore, we do not detect any saturation for synonymous substitution amounts in these sequences.

Prior Distribution

The prior distribution of node times was set to satisfy the constraint that African apes and humans diverged from Orangutans between 13 and 18 MYA (see Cao et al. 2000). Our approach for setting prior distributions on node times (Kishino, Thorne, and Bruno 2001) also needs the mean and standard deviation of a gamma distribution to be specified. This gamma distribution would represent the prior distribution of the ingroup root time if there were no constraints on node time. We selected a gamma distribution with mean 81.5 and standard deviation 5.0. This prior mean and standard deviation were intentionally designed so as to be very similar to the best estimate and 95% confidence interval of the primate root as reported in a recent study of fossil data (Tavaré et al. 2002). Because of the constraints on the Human + African ape/Orangutan divergence, our actual prior distribution for time since the common ancestor of the 19 primate species had a mean of approximately 80.1 and an approximate 95% credibility interval of (71.1, 90.0).

To set the prior distributions for the synonymous and nonsynonymous rates at the ingroup root, our practice is to calculate the estimated numbers of synonymous and nonsynonymous substitutions separating the root and each of the ingroup tips. The median among tips of the number of synonymous substitutions per codon was 4.87, and the median for nonsynonymous changes was 0.105. By dividing these median values by the prior mean of the ingroup root time, we get $4.87/81.5 \simeq 0.0598$ synonymous substitutions per codon per million years and
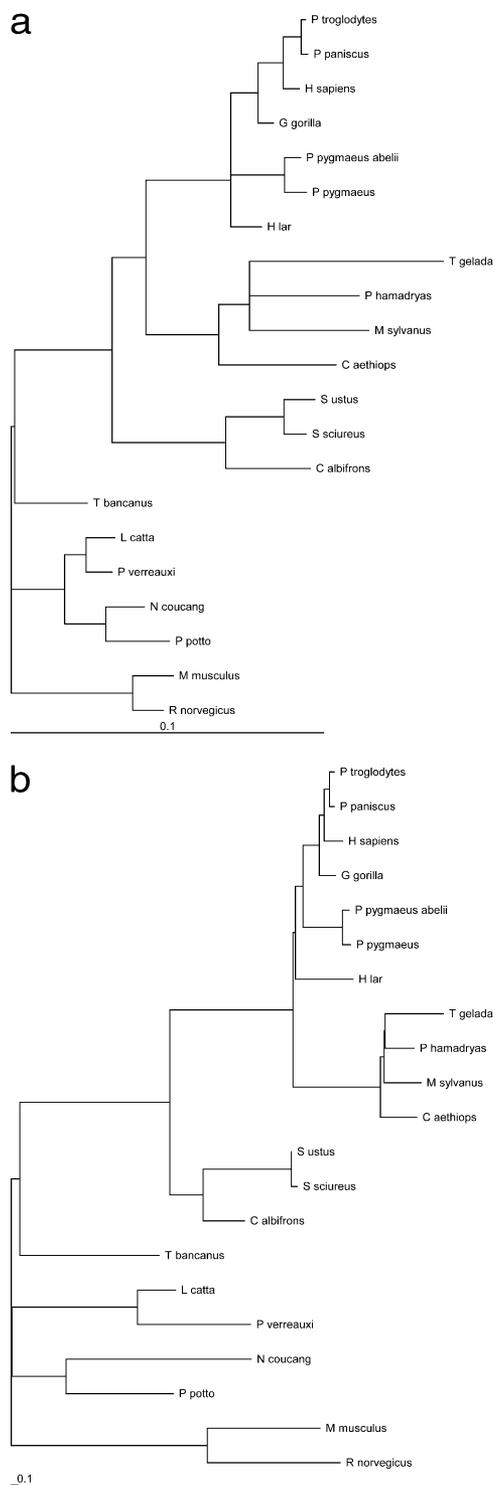
a

b

Fig. 1.—Maximum likelihood estimates of synonymous and nonsynonymous branch lengths from COXI primate data. *a*. Non-synonymous lengths. *b*. Synonymous branch lengths.

$0.105/81.5 \simeq 0.00129$ nonsynonymous substitutions per million years. We then set the median of an exponential prior distribution for the synonymous rate at the ingroup root node to be 0.0598 and the median of an exponential prior distribution for the nonsynonymous rate to be 0.00129. Technically, we are violating the Bayesian framework by using our data to specify the prior distributions for rates. However, the impact of this violation is expected to be insubstantial because the constraints on node times help to separate rates and times and because these exponentially distributed priors for rates have relatively large standard deviations. For the prior for both ν_n and ν_s, we employed an exponential distribution with a mean expected increase of 0.01 in the variance of the logarithm of evolutionary rates per million years.

Metropolis-Hastings Approximation

Posterior distributions were approximated by implementing a Metropolis-Hastings algorithm (Metropolis et al. 1953; Hastings 1970). As with the implementations described for simpler models of sequence change (Thorne, Kishino, and Painter 1998; Kishino, Thorne, and Bruno 2001), our Metropolis-Hastings algorithm consists of cycling through various ways of changing parameter values. To check the convergence of our Metropolis-Hastings algorithm to the posterior distribution, multiple separate Metropolis-Hastings runs from different initial parameter states were performed. All Metropolis-Hastings runs consisted of 500,000 burn-in cycles followed by one sample of the Markov chain for each 1,000 cycles until a total of 10,000 samples had been obtained.

Posterior Distribution

Figure 2 displays some of the results of our Bayesian analysis. It is interesting to compare figures 1 and 2 to inspect how relatively long estimated branch lengths can result in longer time duration of branches and/or higher evolutionary rates. When the distribution of the posterior mean of concordance statistic S is simulated according to the null hypothesis of independent synonymous and nonsynonymous rate change, the frequency of S^* values that exceed 0.5234 is 0.0309. Therefore, if the alternative hypothesis had been a tendency for concordant rate changes, the associated P value would be about 0.0309. Because our alternative hypothesis was, instead, that the null of independent rate change was false, we find only marginal significance and the P value for this two-tailed test to be 0.0618.

With the posterior distribution of the ν_n and n_s parameters, we can also quantify tendencies for nonsynonymous and synonymous rates to change over time. The estimated posterior mean and 95% credibility interval for ν_n is 0.01402 and (0.005233, 0.02984) whereas ν_s yields 0.004247 and (1.161×10^{-4}, 0.01588). Because of the large overlap between the 95% credibility intervals for ν_n and ν_s, it is tempting to conclude that the evidence is weak that ν_n exceeds ν_s. However, the posterior distributions of ν_n and ν_s are correlated. Out of 10,000 samples from the joint posterior distribution of ν_n and ν_s, the value of ν_n exceeded ν_s 9287 times. Based on the 10,000 samples, the median of ν_n / ν_s is 4.500 and the 95% credibility interval is (0.6433, 115.38). In contrast, the prior median of ν_n / ν_s must be 1 because the priors of ν_n and ν_s are identical. Explicit calculation shows that the 95% prior interval for ν_n / ν_s is (0.0256, 39.0). Based on
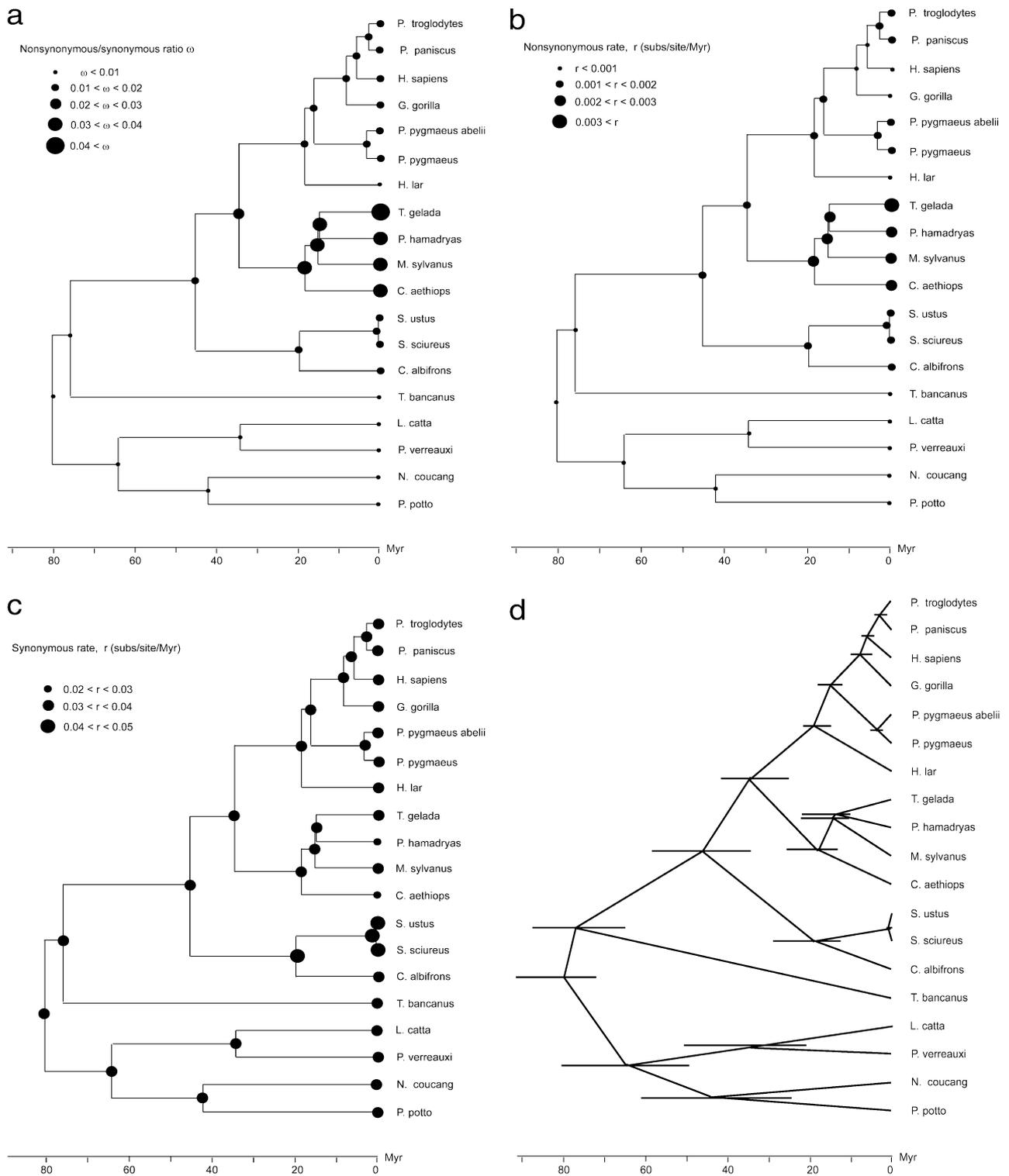
FIG. 2.—Bayesian parameter estimates from COXI primate data. Posterior means of divergence times are used to depict the internal node times. *a*. Posterior median of ω. *b*. Posterior median of expected number of nonsynonymous substitutions per codon per million years. *c*. Posterior medians of expected number of synonymous substitutions per codon per million years. *d*. Posterior means and 95% credible intervals of divergence times.

the difference between the prior and the estimated posterior distribution of ν_n / ν_s, the data indicate that nonsynonymous rates have changed more than synonymous rates, at least when rate change is assessed on a log-scale.

The large value of ν_n seems to be mostly due to the difference of nonsynonymous rates between anthropoids (from *Pan troglodytes* to *Tarsius bancanus* in figure 2) and non-anthropoid groups. Figure 2*b* shows that the
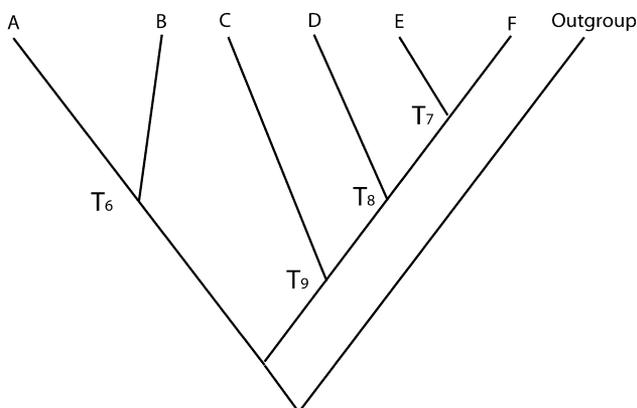
Fig. 3.—Topology and node labels used in the simulations.

nonsynonymous rates of non-anthropoids are generally lower than for anthropoids, whereas there seems to be no clear difference in synonymous rates between the two groups (fig. 2c). The elevated nonsynonymous rates in the anthropoid group seem to be responsible for the elevated ratio (ω) of nonsynonymous to synonymous rates in this group. The estimated values of ω are especially high in the Old World monkey group (*T. gelada*, *P. hamadryas*, *M. sylvanus*, and *C. aethiops*), as has been previously noted (Andrews and Easteal 2000).

Simulations

We have also assessed the performance of our codon-based procedure with simulations, and we report some of our findings here. Consider six ingroup taxa and one outgroup taxon that are topologically related as depicted in figure 3. In all simulation scenarios, 100 different sequence data sets with sequence lengths of 500 codons were generated.

To randomly assign sequences at the ingroup root, the 60 non-termination codons of the mammalian mitochondrial genetic code were assumed to be equally probable. Thereafter, the codon substitution model described in equation (1) was employed. The instantaneous rate matrix was parameterized by setting $\kappa = 5.0$ and $\pi_i = 1/60$ for all non-termination codons i.

Synonymous and nonsynonymous branch lengths in the ingroup were determined for each simulated data set by having evolutionary rates change according to our model. The value of $\nu_n = 0.01402$ and $\nu_s = 0.004247$ were selected for the simulations by adopting the estimates from the COXI data. In each case, the nonsynonymous rate at the ingroup root was 0.00112 nonsynonymous substitutions per codon per million years. The value 0.00112 was selected because it is the median of the estimated nonsynonymous rates among ingroup nodes in the COXI analysis. For every simulated data set, the synonymous rate at the ingroup root was 0.01 synonymous substitutions per codon per million years. The value 0.01 is less than the median (0.03675) of the estimated synonymous rates for the COXI analysis, but we chose it because we found that some simulated data sets with the rate 0.03675 at the root were completely saturated for synonymous changes. Saturation is more likely when the number of ingroup

taxa is six (as in these simulations) than when the number is larger (as in the COXI analysis).

Possible saturation of synonymous changes also motivated our artificial treatment of the branch connecting the ingroup root and the outgroup taxon. The outgroup branch lengths were not randomly simulated. Instead, the synonymous and nonsynonymous branch lengths were arbitrarily set to 0.2 and 0.0224.

The node times used in the simulations varied among scenarios and are listed in table 1. Scenario A has all branches exist for a relatively long time. The remaining three scenarios represent situations where one or two branches have a very short time duration. We selected the latter three scenarios to determine the impact of short branch lengths on our divergence time estimation procedure.

One difference between the techniques used to simulate and analyze the sequence data is that the analysis procedure constrained one branch emanating from the ingroup root to have identical beginning and ending rates, whereas the simulation procedure did not enforce this constraint. For the analyses, the prior distribution for the ingroup root time would have been a gamma distribution with mean 80 MYA and standard deviation 5 Myr, except that the ingroup root time was tightly constrained by being forced to exist between 79 and 81 MYA. Conditional upon the ingroup root time, the prior distribution of interior node times was the modified Dirichlet distribution of Kishino, Thorne, and Bruno (2001). As with the COXI analysis, prior distributions for synonymous (nonsynonymous) rates at the ingroup root were determined by finding the median among tips of the sums of synonymous (nonsynonymous) branch lengths from root to tip. These synonymous and nonsynonymous medians were then divided by 80 Myr to determine the medians of the exponentially distributed priors for synonymous and nonsynonymous rates. The ν_n and ν_s priors were identical to those employed in the analysis of the actual COXI data.

All simulated data sets were analyzed with our MCMC method. The burn-in period, the number of proposal cycles between MCMC samples, and the number of MCMC samples were set to 100,000, 1,000, and 1,000, respectively. Although convergence to the posterior distribution was not rigorously assessed in the simulation analyses, Markov chains of these lengths seem to yield adequate approximations.

Because all data sets were generated with identical values for synonymous and nonsynonymous ingroup root rates and identical values of ν_s and ν_n, overly general interpretations of these simulation results need to be avoided. In addition, the treatment of the outgroup branch is artificial. Therefore, these simulation results may be superior to those obtained in actual studies.

However, the simulation results do bode well for actual studies. The posterior means of node times are relatively unbiased and the general tendency is for the 95% credibility intervals to include the true node times (table 1). Also, the posterior intervals for the times tend to be substantially more narrow than are the 95% prior intervals.

Part of the motivation for examining Scenarios B–D (table 1) was to investigate whether our treatment of

Table 1
Performance of Divergence Time Estimation Procedure in Simulations

		T_6	T_7	T_8	T_9
	Prior	39.90 (2.31, 77.86)	19.93 (0.715, 56.58)	39.78 (7.73, 72.69)	59.96 (23.20, 79.41)
Scenario A	True	40	20	40	60
	Estimated	40.59 (31.14, 50.55)	20.64 (13.51, 29.74)	40.89 (30.83, 51.18)	60.72 (51.20, 69.84)
	Coverage	92	96	93	93
Scenario B	True	40	1	40	60
	Estimated	41.06 (31.57, 51.00)	1.149 (.3669, 2.479)	41.58 (31.30, 52.13)	60.06 (50.62, 69.09)
	Coverage	90	98	93	95
Scenario C	True	40	20	59	60
	Estimated	40.11 (30.77, 49.89)	20.32 (13.69, 28.38)	58.27 (49.01, 66.90)	61.00 (52.33, 69.31)
	Coverage	91	92	97	92
Scenario D	True	40	0	40	60
	Estimated	41.18 (31.37, 51.30)	.112 (.0027, .4477)	40.84 (30.30, 51.59)	59.85 (50.11, 69.18)
	Coverage	92	0	95	92

NOTE.—The columns labeled T_6, T_7, T_8, and T_9 correspond to the topology depicted in figure 3. In the row labeled 'Prior', the MCMC prior means of each node time are listed. In the subsequent rows, results for the different simulation scenarios are given. For each scenario, the first line lists the true node time. The second line contains the average among the one hundred posterior mean node times, followed in parentheses by the averages of the lower and upper values of the corresponding 95% credibility intervals. Among the one hundred 95% credibility intervals, the third line reveals how many intervals contained the true node time.

$p(\mathbf{X} \mid \mathbf{b}_n, \mathbf{b}_s)$ (see *Appendix*) was adequate when branch lengths are extremely short. The simulation results indicate that the performance of the divergence time estimation procedure is satisfactory even in the presence of one or two extremely short nonsynonymous branch lengths. Scenario D represents the extreme situation where the time T_7 is 0. With $T_7 = 0$, the branches terminating at tips E and F must have synonymous and nonsynonymous lengths that are both zero. As $T_7 = 0$ is at the boundary of the possible times for this node, we view the results for Scenario D as satisfactory even though the 95% credibility intervals for T_7 do not contain the true time.

Discussion

For determining whether two genes relating some group of taxa change evolutionary rates over time in a correlated fashion, Thorne and Kishino (2002) introduced a hypothesis test that uses the rank correlation among nodes of the evolutionary rates of the two genes. Our hypothesis test for comparing changes over time in synonymous and nonsynonymous rates relies upon a concordance statistic. The concordance statistic is based upon differences in synonymous and nonsynonymous rates between the beginning and ending of a branch. The concordance statistic is therefore not specially designed to capture trends in rates over long periods of time that encompass many successive branches on an evolutionary tree. Instead, the concordance statistic might be better suited for detecting relationships between rates that operate on a shorter time scale. In contrast, the rank correlation test of Thorne and Kishino (2002) might be well designed for detecting long-term tendencies for correlated rate change.

Both testing strategies may be subject to possible problems when the model for change in evolutionary rates over time differs greatly from reality. Our model for rate change and that of Thorne and Kishino (2002) describe a process whereby rates change in relatively gradual fashion rather than by infrequent but large jumps. If a large and relatively sudden change in evolutionary rates does occur, estimates of evolutionary rates at nodes near the position on the tree where the sudden change occurred may not completely reflect the sudden change. Instead, the estimated evolutionary rates at nearby nodes may show evidence of a more gradual and smoother change. For the hypothesis test presented here, this smoothing effect could cause the posterior estimates of concordance or discordance to change in a correlated way among neighboring branches. Unfortunately, this sort of correlation is not incorporated by the hypothesis test presented here. A similar sort of phenomenon could affect the procedure of Thorne and Kishino (2002) for examining whether different genes experience rate change in a correlated fashion. To lessen the possible problems that stem from smoothing of estimate rates, one possibility would be to employ models of evolutionary rate change where large jumps are more probable.

By application of relative rate tests (Wu and Li 1985; Muse and Weir 1992; Andrews, Jermin, and Easteal 1998), it has been suggested that COXI in the anthropoid group has an accelerated nonsynonymous substitution rate relative to non-anthropoid primates (Andrews and Easteal 2000; Wu et al. 2000). Relative rate tests are not designed or well-suited to provide rich descriptions of the patterns of rate evolution over time. For example, an observation that a relative rate test does not detect a statistically significant difference in rates among anthropoid lineages should not be construed to be an indication that these anthropoid lineages have evolved at identical rates. Our study (see fig. 2a and 2b) suggests that the nonsynonymous rate of COXI has been relatively accelerated in all anthropoid lineages rather than in solely the branch that ends with the most recent common ancestor of anthropoids. As noted by Andrews and Easteal (2000), this acceleration seems most extreme in the Old World monkeys and appears to be present but more mild in the Hominoids.

Our emphasis here is on COXI synonymous and nonsynonymous rate evolution and on the statistical methodology for studying it. Our COXI estimates of divergence times for primates should not be regarded as definitive. Other recent studies that date divergences of

primates and their close relatives (Cao et al. 2000; Yoder and Yang 2000; Stauffer et al. 2001; Glazko and Nei 2003) have been based on multiple genes and therefore presumably yielded more accurate dating. When divergence time estimation of primates is the goal, one attractive possibility would be to analyze the multigene data sets of earlier studies with an extension to multiple genes of our codon-based procedure. Thorne and Kishino (2002) introduced a divergence time estimation technique for multigene data sets that allows each individual gene to have its own trajectory, over time, of evolutionary rates, and that technique could be adapted to the codon-based framework used here.

Although our focus for COXI is not on divergence time estimation, the results for the node representing the most recent common ancestor of the squirrel monkeys *S. ustus* and *S. sciureus* warrant mention. The posterior mean estimate for the time since this node is recent (1.057 MYA) and its 95% credibility interval (between 0.4335 and 2.147 MYA) is short (see also figure 2*d*). We view the dating of this node as especially unreliable. The source of the problem seems to be the peculiar maximum likelihood estimates of branch lengths surrounding this node. The branch from the node to *S. sciureus* yielded maximum likelihood estimates of about 0.087 synonymous changes per codon and 0.008 nonsynonymous changes per codon. In contrast, the branch from the node to *S. ustus* yielded estimates of 0 synonymous changes per codon and about 0.010 nonsynonymous change per codon (see figure 1*a* and 1*b*). The estimated 0 synonymous changes from the node to *S. ustus* is unusual because it is less than the estimated amount of nonsynonymous change.

Generation time and population size are closely related to the evolutionary rate (Ohta 1987, 1995; Ohta and Tachida 1990). Whereas the substitution rate (i.e., the fixation rate) per generation of neutral mutants depends on only the mutation rate per generation, the substitution rate per generation of non-neutral mutants depends not only on the mutation rate but also on the effective population size and on natural selection. The substitution rate of non-neutral mutants per generation is negatively correlated with effective population size (Ohta 1972, 1987; Ohta and Tachida 1990). If we measure evolutionary rate in units of chronological time (month, year, million years, etc.), it is also important how many generations correspond to one chronological time unit. Generally, species with large effective population sizes have short generations (Chao and Carr 1993). This weakens the negative correlation between effective population size and substitution rate when substitution rate is measured in chronological units rather than in generations (Ohta 1995). This interplay between effective population size, generation time, and substitution rate can explain Kimura's (1983) findings that chronological rates of protein evolution have been relatively constant even among evolutionary lineages representing quite variable generation times (Chao and Carr 1993). In contrast, substitution rates of neutral mutants are subject to only generation-time effects if these rates are measured in units of chronological time.

For each of 49 genes, Ohta (1995) sampled a sequence from a primate, an artiodactyl, and a rodent. She then estimated amounts of synonymous and nonsynonymous substitution and observed generally higher variation in synonymous substitution amounts than in nonsynonymous substitution amounts. Our inference that the nonsynonymous rate of COXI varied more than the synonymous rate differs from Ohta's general finding (see Figure 1 and Table 3 of Ohta 1995). Although Ohta quantified rate variation differently than we do here, our examination of the estimated amounts of synonymous and nonsynonymous substitution for the 49 genes in Table 3 of Ohta (1995) suggests that the techniques for quantifying the rate variation are not responsible for the difference between our COXI inference about rate variation and Ohta's general finding.

One explanation for the different results may be that our analysis compared lineages within primates whereas Ohta's analysis contrasted more distantly related groups (primates, artiodactyls, and rodents). The tendency for species with shorter generation times to have larger effective population sizes may be stronger when distantly related groups are compared than when species within a relatively closely related group such as primates are compared. Another possible explanation for the difference between our COXI results and the overall picture reported by Ohta (1995) is that selection may have played a particularly important role in the molecular evolution of COXI primate genes. As mentioned in the *Introduction*, genes involved in aerobic energy metabolism may have experienced a difference in selective pressures between anthropoid and non-anthropoid lineages (Grossman et al. 2001).

Here, we introduced a method to obtain estimates of divergence times and nonsynonymous and synonymous evolutionary rates. With our method, it is possible to quantify the variation of nonsynonymous and synonymous rate over time. Because it can better separate evolutionary times and rates than can conventional procedures for studying patterns of nonsynonymous and synonymous substitutions, we believe our method provides useful tools for illuminating the history and process of molecular evolution.

An important future direction will be to extend this approach to multigene data sets. A multigene analysis would help reveal whether the discrepancy between our COXI results and those of Ohta (1995) are an anomaly or a general feature of the evolution of genes involved in aerobic energy metabolism within primates. By helping to disentangle lineage effects and gene effects, a multigene approach would also lead to improvements in the ability to characterize the impact on evolution of natural selection, effective population size, and generation time.

perform these analyses is freely available from T.-K. Seo (seo@statgen.ncsu.edu).

Appendix
Approximation of the Likelihood Function

Consider a vector of parameters $\boldsymbol{\theta}$ with maximum likelihood estimate $\hat{\boldsymbol{\theta}}$. The likelihood function will be $\prod_{i=1}^{N} f(x_i \mid \boldsymbol{\theta})$, where x_i represents the three aligned nucleotide columns of the ith codon and where N is the number of codons. With a Taylor expansion of the log-likelihood function around $\hat{\boldsymbol{\theta}}$, we can obtain the following relation:

$$
\sum_{i=1}^{N} \log f(x_i \mid \boldsymbol{\theta})
$$
$$
\simeq \sum_{i=1}^{N} \log f(x_i \mid \hat{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \sum_{i=1}^{N} \frac{\partial}{\partial \boldsymbol{\theta}} \log f(x_i \mid \hat{\boldsymbol{\theta}})
$$
$$
+ \frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \cdot \sum_{i=1}^{N} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \log f(x_i \mid \hat{\boldsymbol{\theta}}) \cdot (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}).
$$
$$(12)$$

By exponentiating both sides of equation (12), it becomes possible to approximate the likelihood function as

$$
\prod_{i=1}^{N} f(x_i \mid \boldsymbol{\theta})
$$
$$
\simeq \prod_{i=1}^{N} f(x_i \mid \hat{\boldsymbol{\theta}}) \times \exp\left\{ (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \sum_{i=1}^{N} \frac{\partial}{\partial \boldsymbol{\theta}} \log f(x_i \mid \hat{\boldsymbol{\theta}}) \right\}
$$
$$
\times \exp\left\{ \frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \cdot \sum_{i=1}^{N} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \log f(x_i \mid \hat{\boldsymbol{\theta}}) \cdot (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \right\}.
$$
$$(13)$$

When no maximum likelihood estimate of branch lengths is 0, $\hat{\boldsymbol{\theta}}$ is not on the boundary of the parameter space, and the second term of equation (12) is zero; then

$$
\prod_{i=1}^{N} f(x_i \mid \boldsymbol{\theta})
$$
$$
\simeq \prod_{i=1}^{N} f(x_i \mid \hat{\boldsymbol{\theta}})
$$
$$
\times \exp\left\{ \frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \cdot \sum_{i=1}^{N} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \log f(x_i \mid \hat{\boldsymbol{\theta}}) \cdot (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \right\},
$$
$$(14)$$

and this leads to the multivariate normal approximation employed in previous studies (Thorne, Kishino, and Painter 1998; Kishino, Thorne, and Bruno 2001; Thorne and Kishino 2002). When $\hat{\boldsymbol{\theta}}$ is on the boundary of the parameter space, the first derivative may not be zero and, for this reason, we do not rely on equation (14).

To evaluate the final term of equation (13), the determination of second derivatives is required. When the log-likelihood surface is relatively flat, numerical evalua-

tions of second derivatives can be unstable. We have found that a strategy of estimating second derivatives in terms of the first derivatives of individual codon log-likelihoods works well. The second derivative approximation is based on the relation (e.g., Hogg and Craig 1995, p. 373)

$$
E\left[\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \log f(X \mid \boldsymbol{\theta}) \right]
$$
$$
= -E\left[\left(\frac{\partial}{\partial \boldsymbol{\theta}} \log f(X \mid \boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}^T} \log f(X \mid \boldsymbol{\theta}) \right) \right]. \quad (15)
$$

This leads to

$$
\sum_{i=1}^{N} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \log f(x_i \mid \boldsymbol{\theta})
$$
$$
\simeq -\sum_{i=1}^{N} \frac{\partial}{\partial \boldsymbol{\theta}} \log f(x_i \mid \boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}^T} \log f(x_i \mid \boldsymbol{\theta}), \quad (16)
$$

because the left side of equation (16) estimates the left side of equation (15) multiplied by N and the right side of equation (16) estimates the right side of equation (15) multiplied by N. The right side of equation (16) is not computationally very burdensome to evaluate and it is numerically relatively stable because only first derivatives are involved (e.g., see Porter 2002). Substituting equation (16) into the third term of equation (13), produces the approximation that we used in our analysis.

$$
\prod_{i=1}^{N} f(x_i \mid \boldsymbol{\theta})
$$
$$
\simeq \prod_{i=1}^{N} f(x_i \mid \hat{\boldsymbol{\theta}}) \times \exp\left\{ (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \sum_{i=1}^{N} \frac{\partial}{\partial \boldsymbol{\theta}} \log f(x_i \mid \hat{\boldsymbol{\theta}}) \right\}
$$
$$
\times \exp\left\{ -\frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \cdot \sum_{i=1}^{N} \frac{\partial}{\partial \boldsymbol{\theta}} \log f(x_i \mid \hat{\boldsymbol{\theta}}) \right.
$$
$$
\left. \times \frac{\partial}{\partial \boldsymbol{\theta}^T} \log f(x_i \mid \hat{\boldsymbol{\theta}}) \cdot (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \right\}. \quad (17)
$$

Literature Cited

Adkins, R. M., and R. L. Honeycutt. 1994. Evolution of the primate cytochrome c oxidase subunit II gene. J. Mol. Evol. **38**:215–231.

Adkins, R. M., R. L. Honeycutt, and T. R. Disotell. 1996. Evolution of eutherian cytochrome c oxidase subunit II: heterogeneous rates of protein evolution and altered interaction with cytochrome c. Mol. Biol. Evol. **13**:1393–1404.

Andrews, T. D., and S. Easteal. 2000. Evolutionary rate acceleration of cytochrome c oxidase subunit I in simian primates. J. Mol. Evol. **50**:562–568.

Andrews, T. D., L. S. Jermiin, and S. Easteal. 1998. Accelerated evolution of cytochrome b in simian primates: adaptive evolution in concert with other mitochondrial proteins? J. Mol. Evol. **47**:249–257.

Aris-Brosou, S., and Z. Yang. 2002. Effects of models of rate evolution on estimation of divergence dates with special reference to the metazoan 18S ribosomal RNA phylogeny. Syst. Biol. **51**:703–714.

Cao, Y., M. Fujiwara, M. Nikaido, N. Okada, and M. Hasegawa. 2000. Interordinal relationships and time-scale of eutherian

evolution as inferred from mitochondrial genome data. Gene **259**:149–158.

Chao, L., and D. E. Carr. 1993. The molecular clock and the relationship between population size and generation time. Evolution **47**:688–690.

Drummond, A., and A. G. Rodrigo. 2000. Reconstructing genealogies of serial samples under the assumption of a molecular clock using serial-sample UPGMA. Mol. Biol. Evol. **17**:1807–1815.

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

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 1995. Bayesian data analysis. Chapman & Hall, 2-6 London.

Glazko, G. V., and M. Nei. 2003. Estimation of divergence times for major lineages of primate species. Mol. Biol. Evol. **20**:424–434.

Goldman, N., and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. **17**:32–43.

Grossman, L. I., T. R. Schmidt, D. E. Wildman, and M. Goodman. 2001. Molecular evolution of aerobic energy metabolism in primates. Mol. Phylogent. Evol. **18**:26–36.

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

Hogg, R. V., and A. T. Craig. 1995. Introduction to mathematical statistics. Fifth edition. Prentice-Hall, Eaglewood Cliffs, N.J.

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

Ina, Y. 1995. New methods for estimating the numbers of synonymous and nonsynonymous substitutions. J. Mol. Evol. **40**:190–226.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, U.K.

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.

Korber, B., M. Muldoon, J. Theiler, F. Gao, R. Gupta, A. Lapedes, B. Hahn, S. Wolinsky, and T. Bhattacharya. 2000. Timing the ancestor of the HIV-1 pandemic strains. Science. **288**:1757–1759.

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

Muse, S., and B. Gaut. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol. Biol. Evol. **11**:715–724.

Muse, S. V., and B. S. Weir. 1992. Testing for equality of evolutionary rates. Genetics **132**:269–276.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

Ohta, T. 1972. Population size and rate of evolution. J. Mol. Evol. **1**:305–314.

———. 1987. Very slightly deleterious mutations and the molecular clock. J. Mol. Evol. **26**:1–6.

———. 1995. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. J. Mol. Evol. **40**:56–63.

Ohta, T., and H. Tachida. 1990. Theoretical study of near neutrality. I. Heterozygosity and rate of mutant substitution. Genetics. **126**:219–229.

Pedersen, A.-M., K. C. Wiuf, and R. B. Christiansen. 1998. A codon-based model designed to describe lentiviral evolution. Mol. Biol. Evol. **15**:1069–1081.

Porter, J. 2002. Efficiency of covariance matrix estimators for maximum likelihood estimation. J. Bus. Econ. Stat. **20**:431–440.

Purvis, A. 1995. A composite estimate of primate phylogeny. Phil. Trans. R. Soc. Lond. Ser. B. **348**:405–421.

Rambaut, A. 2000. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. Bioinformatics **16**:395–399.

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

Sanderson, M. J. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. **19**:101–109.

Sanderson, M. J. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics **19**:301–302.

Stauffer, R. L., A. Walker, O. A. Ryder, M. Lyons-Weiler, and S. B. Hedges. 2001. Human and ape molecular clocks and constraints on paleontological hypotheses. J. Hered. **92**:469–474.

Schmidt, T. R., M. Goodman, and L. I. Grossman. 1999. Molecular evolution of the COX7A gene family in primates. Mol. Biol. Evol. **16**:619–626.

Tavaré, 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 **18**:726–729.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

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.

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

Wildman, D. E., W. Wu, M. Goodman, and L. I. Grossman. 2002. Episodic positive selection in ape cytochrome c oxidase subunit IV. Mol. Biol. Evol. **19**:1812–1815.

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.

Wu, W., M. Goodman, M. I. Lomax, and L. I. Grossman. 1997. Molecular evolution of cytochrome oxidase subunit IV: evidence for positive selection in simian primates. J. Mol. Evol. **44**:477–491.

Wu, W., T. R. Schmidt, M. Goodman, and L. I. Grossman. 2000. Molecular evolution of cytochrome c oxidase subunit I in primates: is there coevolution between mitochondrial and nuclear genomes? Mol. Phylogenet. Evol. **17**:294–304.

Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. **15**:568–573.

Yang, Z., and P. Bielawski. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. **15**:496–503.

Yang, Z., and R. Nielsen. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J. Mol. Evol. **46**:409–418.

———. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. **17**:32–43.

———. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol. Biol. Evol. **19**:908–917.

Yang, Z., and W. J. Swanson. 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Mol. Biol. Evol. **19**:49–57.

Yoder, A. D., and Z. Yang. 2000. Estimation of primate speciation dates using local molecular clocks. Mol. Biol. Evol. **17**:1081–1090.

Zhang, Z., H. F. Rosenberg, and M. Nei. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. USA **95**:3708–3713.