A General Comparison of Relaxed Molecular Clock Models

Thomas Lepage, David Bryant,† Hervé Philippe,‡ and Nicolas Lartillot§*

*Department of Mathematics and Statistics, McGill University, Montréal, Québec, Canada; †Department of Mathematics, University of Auckland, New Zealand; ‡Département de biochimie, Université de Montréal, Québec, Canada; and §Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier, UMR 5506, CNRS-Université de Montpellier 2, 161, rue Ada, 34392 Montpellier Cedex 5, France

Several models have been proposed to relax the molecular clock in order to estimate divergence times. However, it is unclear which model has the best fit to real data and should therefore be used to perform molecular dating. In particular, we do not know whether rate autocorrelation should be considered or which prior on divergence times should be used. In this work, we propose a general bench mark of alternative relaxed clock models. We have reimplemented most of the already existing models, including the popular lognormal model, as well as various prior choices for divergence times (birth–death, Dirichlet, uniform), in a common Bayesian statistical framework. We also propose a new autocorrelated model, called the "CIR" process, with well-defined stationary properties. We assess the relative fitness of these models and priors, when applied to 3 different protein data sets from eukaryotes, vertebrates, and mammals, by computing Bayes factors using a numerical method called thermodynamic integration. We find that the 2 autocorrelated models, CIR and lognormal, have a similar fit and clearly outperform uncorrelated models on all 3 data sets. In contrast, the optimal choice for the divergence time prior is more dependent on the data investigated. Altogether, our results provide useful guidelines for model choice in the field of molecular dating while opening the way to more extensive model comparisons.

Introduction

The dating of speciation events is one of the major objectives of evolutionary studies. The molecular clock hypothesis, that is, the constancy of the rate of evolution with time, when combined with the use of fossil records, allows the dating of branching events on phylogenetic trees. However, it has become clear that the strict molecular clock hypothesis is not biologically realistic and that the rate at which substitutions accumulate with time is subject to changes that need to be explicitly accounted for in molecular dating techniques.

Various methods have been proposed to relax the molecular clock, including nonparametric approaches (Sanderson 1997, 2002), local clocks (Yoder and Yang 2000), and Bayesian parametric models (Thorne et al. 1998; Huelsenbeck et al. 2000; Kishino et al. 2001; Aris-Brosou and Yang 2002, 2003; Drummond et al. 2006). Among these methods, the Bayesian parametric approach offers the opportunity of exploring a wide diversity of alternative models, each of which corresponds to specific assumptions concerning the shape of the tree and the way the rate of substitution changes with time.

In practice, many alternative probabilistic models of clock relaxation have indeed been proposed in this framework. They differ in various respects: some describe the rate as a continuous process (Kishino et al. 2001), others as a piecewise constant function of time (Huelsenbeck et al. 2000), and still others do not explicitly describe the rate process itself, but rather directly assign a mean rate parameter to each branch (Aris-Brosou and Yang 2003; Drummond et al. 2006). Models also differ in their dynamics: rates may be autocorrelated (Kishino et al. 2001; Aris-Brosou and Yang 2003), or not (Drummond et al. 2006). Finally, different priors have been proposed on the set of divergence times, including the uniform (Kishino et al. 2001; Drummond et al. 2006), Dirichlet (Thorne and Kishino 2002), and the birth–death (Aris-Brosou and Yang 2003; Yang and Rannala 2006) priors.

Which among this variety of available models yield reliable dates is still an open question. In particular, it is not totally clear whether autocorrelation is a feature that relaxation models should always account for (Drummond et al. 2006). A few controversies have also been raised about possible biases in certain cases (Aris-Brosou and Yang 2003) due to the relaxation model itself or to the prior on divergence times (Blair and Hedges 2005; Welch et al. 2005). More generally, the model seems to have a nonnegligible influence on the resulting divergence dates (Perez-Lozada et al. 2004; Smith et al. 2006), a fact that does not help sorting out the still open controversies raised by the clashes observed between fossils and molecular datings (Bromham et al. 1999; Smith and Peterson 2002). As these controversies sometimes bear on the very choice between recent and explosive radiation versus ancient and progressive diversification scenarios, for example, in the case of metazoans (Smith and Peterson 2002; Aris-Brosou and Yang 2003; Douzery et al. 2004; Peterson et al. 2004; Blair and Hedges 2005; Welch et al. 2005) or of mammals (Bromham et al. 1999; Springer et al. 2003), deciding between alternative models of clock relaxation is obviously an urgent question to be answered.

Remarkably, there have been few attempts thus far to compare the performances of available clock models. In some cases, models were validated by their ability to properly retrieve already known rates and dates from simulated data (Thorne and Kishino 2002; Ho et al. 2005; Drummond et al. 2006). However, the success of a model on simulated data may not always be representative of the model's relevance to the analysis of real data, which is more directly assessed by measuring the statistical fit of the model to the data at hand.

A common method for measuring the fit of a model M in a Bayesian framework consists in computing the marginal likelihood $P(D|M)$, that is, the marginal probability of generating the data D under the model of interest M (Jeffreys 1935). The relative fit of 2 models M_1 and M_2

Key words: relaxed clock, Bayes factor, molecular dating, CIR process, phylogeny, Markov chain Monte Carlo.

E-mail: nicolas.lartillot@lirmm.fr.

Mol. Biol. Evol. 24(12):2669–2680. 2007
doi:10.1093/molbev/msm193
Advance Access publication September 21, 2007

© The Author 2007. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org

Downloaded from https://academic.oup.com/mbe/article-abstract/24/12/2669/977266 by Imperial College London Library user on 09 October 2019

can then be assessed by the "Bayes factor" B_{12}, defined as the ratio of the marginal likelihoods of M_1 and M_2, $p(D|M_1)/p(D|M_2)$. Bayes factors present several advantages: they implicitly penalize models with higher dimensionality; furthermore, by using one single model as a global reference, they allow for a general comparison among an arbitrary number of models that are not necessarily nested.

Bayes factor computation involves the integration of the likelihood with respect to the parameters of the model (collectively referred to as θ) over the prior distribution $P(\theta|M)$:

$$p(D|M) = \int p(D|\theta, M)p(\theta|M)\mathrm{d}\theta.$$

Most of the time, the numerical estimation of this integral is a time-consuming task. One way to circumvent this problem is to integrate the likelihood over the posterior $p(\theta|D,M)$ instead of the prior distribution, a method called "posterior Bayes factors" (Aitkin 1991). Posterior Bayes factors have been used for relaxed clock model comparisons (Aris-Brosou and Yang 2003). However, they are regarded as very controversial (see answers to Aitkin 1991). In practice, they lead to an underestimation of the dimensionality penalty, hence favoring unnecessarily complex models. Bona fide Bayes factor estimation techniques require more intensive computations but may offer more reliable answers. One of these techniques, called thermodynamic integration (Ogata 1989), has been recently used in several phylogenetic studies (Lartillot and Philippe 2004, 2006; Blanquart and Lartillot 2006; Rodrigue et al. 2006).

In this work, we compare the fit of some of the most commonly used clock models, including the strict molecular clock, the lognormal (Thorne et al. 1998; Kishino et al. 2001), and the uncorrelated models (Drummond et al. 2006). The unconstrained model (whose prior is directly defined on the space of unrooted trees, without any reference to a global time scale) is used as the reference for our comparisons. We also introduce 2 new models: the "white-noise" and the "CIR" processes (Lepage et al. 2006). The former is a relaxed clock model with no autocorrelation. The latter, CIR, is similar to the Ornstein–Uhlenbeck (OU) model (Aris-Brosou and Yang 2003) but without many of the shortcomings of the OU identified by Welch et al. (2005). The CIR process mainly differs from the log normal in the fact that it possesses a stationary distribution. All these models were reimplemented in a single software and used with 3 alternative priors on divergence times (uniform, Dirichlet, and birth–death), so that the fit of all combinations of clock models and priors can be compared by thermodynamic integration. Our results provide strong evidence in favor of autocorrelated models such as the lognormal and the CIR models, in particular under rich taxonomic sampling. In contrast, the optimal choice concerning the prior seems more data set dependent.

Materials and Methods

General Framework for Relaxed Clock Models

Let D be the $N \times M$ matrix consisting of an alignment of N amino acid or nucleotide sequences, of length M. These sequences were obtained from a set of N taxa, which are related through a phylogenetic tree. The tree topology will be kept fixed throughout this work, and we will never explicitly mention it. We denote by $\mathbf{T} = (T_i)_{i=1,\ldots,N-2}$, the set of all divergence times on this topology. These divergence times are endowed with a prior $p(\mathbf{T}|\lambda)$ (where λ is a hyperparameter), to be specified later.

The evolutionary rate is described as a stochastic process $\mathbf{R}(t)$ running along the branches of the tree. The behavior of $\mathbf{R}$ will be described by its probability density function (pdf) $p(\mathbf{R}|\mathbf{T},\nu)$, where ν stands for the set of hyperparameters controlling the process (ν can possibly contain more than one parameter, e.g., decorrelation time and stationary variance, see below).

Suppose first the ideal case where one considers a completely specified realization of the rate process $\mathbf{R}$, at all times and along all the branches of the tree. In particular, $\mathbf{R}$ is specified at any time-point on the lineage between node i and its parent i_{up}. The "effective branch length" B_i is then the integral of $\mathbf{R}$ between T_i and $T_{i_{\mathrm{up}}}$:

$$B_i = B_i(\mathbf{R}, \mathbf{T}) = \int_{T_{i_{\mathrm{up}}}}^{T_i} R(s)\mathrm{d}s. \quad (1)$$

In our implementation, the divergence times are parameterized as relative times: the time at the root is 1 and that of the leaves is 0. Likewise, the rate process is defined on a relative scale (i.e., is constrained to have mean 1). To convert effective branch lengths $\mathbf{B}$ into expected numbers of substitutions per site, we multiply them by a scale factor μ. This yields the "absolute" effective lengths:

$$B_i' = \mu B_i(\mathbf{R}, \mathbf{T}) = \mu \int_{T_{i_{\mathrm{up}}}}^{T_i} R(s)\mathrm{d}s. \quad (2)$$

For short, we will denote the absolute effective lengths by $\mathbf{B}' = \mu\mathbf{B}$.

The absolute effective lengths, computed for all the branches using equation (2), can in turn be used to compute the likelihood of the alignment $p(D|\mathbf{B}') = p(D|\mu\mathbf{B}(\mathbf{R}))$ by using either the pruning algorithm (Felsenstein 1981) or the Gaussian multivariate approximation (Kishino et al. 2001). Combining this likelihood with the probability density of the realization of the rate process and with the priors on times and other parameters, and applying Bayes' theorem, yields the joint posterior density over the parameters of the model:

$$p(\mathbf{R}, \mathbf{T}, \lambda, \mu, \nu|D) = \frac{1}{p(D)} p(D|\mu\mathbf{B}(\mathbf{R}, \mathbf{T}))\, p(\mathbf{R}|\mathbf{T}, \nu)p(\mathbf{T}|\lambda)p(\lambda)p(\mu)p(\nu), \quad (3)$$

where

$$p(D) = \int p(D|\mu\mathbf{B}(\mathbf{R}, \mathbf{T}))p(\mathbf{R}|\mathbf{T}, \nu)\, p(\mathbf{T}|\lambda)p(\lambda)p(\mu)p(\nu)\mathrm{d}\mathbf{T}\mathrm{d}\mathbf{R}\mathrm{d}\lambda\mathrm{d}\mu\mathrm{d}\nu \quad (4)$$

is the normalization constant. Note that times ($\mathbf{T}$) and rates ($\mathbf{R}$) are confounded: they have an influence on the likelihood only through $\mathbf{B} = \mathbf{B}(\mathbf{R},\mathbf{T})$.

Downloaded from https://academic.oup.com/mbe/article-abstract/24/12/2669/977266 by Imperial College London Library user on 09 October 2019

However, apart from trivial cases, a fully specified realization of **R** along the phylogenetic tree would represent an uncountably infinite collection of values and is thus impossible to manage computationally. In practice, we restrict the description of **R** only at the nodes, so that $\mathbf{R} = \{R_i\}_{i=0,\ldots,2N-2}$ is now a discrete collection of rates, with the convention that R_0 is the rate at the root. Assuming that the rate process is Markovian, the joint probability density of the values of the process at the nodes is given by:

$$p\left(\mathbf{R}\middle|\mathbf{T},\nu\right)=p\left(R_0\right)\prod_{i=1}^{2N-2}p\left(R_i\middle|R_{i_{\mathrm{up}}},\Delta T_i,\nu\right), \qquad (5)$$

where $p(R|R',\Delta T,\nu)$ is the finite-time transition probability density and $\Delta T_i = T_i - T_{i_{\mathrm{up}}}$ is the time interval represented by the branch. The transition probability function is determined by the process and is available in closed form in many cases. Note that the choice of $p(R_0)$ is also part of the model's definition.

Another consequence of restricting the sampling of evolutionary rates at nodes is that we now have an incomplete account of $R(t)$ throughout the tree, more precisely between nodes: $B_i(\mathbf{R},\mathbf{T})$ is then no longer unambiguously defined by **R**, but is a random variable, whose density is conditional on the values R_i and $R_{i_{\mathrm{up}}}$ (as well as on the time duration ΔT_i and the hyperparameters ν). Finding a closed form for this density is impossible for most common types of stochastic processes, but the problem can be circumvented in 2 different ways.

First, one can ignore the stochasticity of B_i and approximate it as a deterministic function of R_i, $R_{i_{\mathrm{up}}}$, and ΔT_i (and possibly also of ν). The joint posterior distribution over rates, times, and hyperparameters then becomes:

$$p_{\mathrm{rigid}}(\mathbf{T},\mathbf{R},\nu,\mu,\lambda|D)=\frac{p(D|\mu B(\mathbf{R},\mathbf{T},\nu))p(\mathbf{R}|\mathbf{T},\nu)p(\mathbf{T}|\lambda)p(\mu)p(\nu)p(\lambda)}{p(D)}, \qquad (6)$$

which is similar to equation (3). We will denote this as the "rigid" approach. Kishino et al. (2001) adopted this strategy and estimated each branch length B_i as

$$B_i \approx \frac{\left(R_{i_{\mathrm{up}}}+R_i\right)\left(T_i-T_{i_{\mathrm{up}}}\right)}{2}.$$

A perhaps more robust way of estimating B_i is to make it equal to its expectation, that is,

$$B_i \approx E\left[\int_{T_{i_{\mathrm{up}}}}^{T_i} R(s)\,\mathrm{d}s\middle|R_i,R_{i_{\mathrm{up}}},\nu\right],$$

provided that this expectation is known or can easily and efficiently be approximated. This is what we do in the case of the CIR model (see below).

Alternatively, one can acknowledge the stochasticity of B_i and approximate its true distribution. The posterior distribution is then defined over rates, times, effective branch lengths, and hyperparameters. It reads as:

$$p_{\mathrm{flexible}}\left(\mathbf{B},\mathbf{T},\mathbf{R},\nu,\mu,\lambda|D\right)=\frac{p(D|\mu\mathbf{B})p(\mathbf{B}|\mathbf{R},\mathbf{T},\nu)p(\mathbf{R}|\mathbf{T},\nu)p(\mathbf{T}|\lambda)p(\lambda)p(\mu)p(\nu)}{p(D)}. \qquad (7)$$

Models that describe **B** as a random variable will be called "flexible" because the randomness in **B** is adding some flexibility to the model.

In certain cases, the distribution of B does not depend on the values of **R** at the nodes of the tree. This may be a true mathematical independence between the instantaneous values taken by the process and its integral over finite time intervals (e.g., the white-noise model described below) or, alternatively, just a matter of practical convenience (Aris-Brosou and Yang 2002; Drummond et al. 2006). Whatever the reason, in those "branchwise" models, the rates at nodes **R** disappear from the model's definition and the posterior distribution reduces to:

$$p_{\mathrm{branchwise}}\left(\mathbf{B},\mathbf{T},\nu,\mu,\lambda|D\right)=\frac{p(D|\mu\mathbf{B})p(\mathbf{B}|\mathbf{T},\nu)p(\mathbf{T}|\lambda)p(\lambda)p(\mu)p(\nu)}{p(D)}. \qquad (8)$$

Relaxed Clock Models

In this article, we examine 7 different relaxed molecular clock models.

STRICT: The strict molecular clock, defined as $R(t) = 1$ for all t, and $B_i = T_i - T_{i_{\mathrm{up}}}$.

UEXP: The "uncorrelated exponential" model (Drummond et al. 2006), in which branch lengths are distributed according to an exponential distribution of mean $\left(T_i - T_{i_{\mathrm{up}}}\right)$. This is a branchwise model, which amounts to considering branch-specific independent and exponentially distributed multipliers of mean 1. Note that Drummond et al. (2006) consider this mean as a parameter of the model, but this is not useful here as we also have a global scale parameter (μ).

UGAM: The "uncorrelated gamma" model (Drummond et al. 2006) is very similar to the UEXP model, except that the branches are now distributed according to a gamma distribution with mean $\left(T_i - T_{i_{\mathrm{up}}}\right)$ and variance $\sigma\left(T_i - T_{i_{\mathrm{up}}}\right)^2$. The UGAM model is also independent of the rate process and reduces to UEXP when $\sigma = 1$.

The 2 models UEXP and UGAM are directly defined in terms of the average rate over each branch. As such, they do not explicitly invoke a time process $R(t)$. In fact, the question of how to define a process $R(t)$ such that the resulting average rates over the branches of the tree would be distributed exactly as in UEXP or UGAM is not a trivial one. In particular, such a time process would display autocorrelations within the time window defined by each branch, so that, strictly speaking, UGAM and UEXP are uncorrelated only between branches. To make this point more evident, we define another uncorrelated model, which is totally uncorrelated at all times:

WN: In the white-noise model, the rate process is described as a pure white-noise process whose stationary

Downloaded from https://academic.oup.com/mbe/article-abstract/24/12/2669/977266 by Imperial College London Library user on 09 October 2019

distribution is a gamma of mean 1 and variance σ. The probability density $p(\mathbf{B}|\mathbf{R},\mathbf{T},\nu)$ is available analytically: branch lengths are independent gamma random variables of mean $(T_i - T_{i_{\text{up}}})$ and variance $\sigma(T_i - T_{i_{\text{up}}})$. Thus, even if WN is explicitly defined in terms of real time, it takes the form of a branchwise model, like UExp and UGam. In practice, the main difference between the WN model and the UGam model lies in the variance of $\mathbf{B}_i$: whereas it is quadratic with time in the UGam case, the WN has a linear variance, accounting for the fact that, by an averaging effect, the mean rate of an uncorrelated process is expected to have a smaller variance over longer branches.

LogN: In the popular lognormal model (Kishino et al. 2001), $R(t)$ varies according to a lognormal distribution with mean $R(0)$ and variance σt. The "autocorrelation parameter" σ indicates how much $R(t)$ is likely to depart from its initial value $R(0)$ (thus, here, $\nu = \sigma$). The LogN model is rigid as $B(t)$ is approximated by

$$B_i = \frac{(R_i + R_{i_{\text{up}}}) \times (T_i - T_{i_{\text{up}}})}{2}. \quad (9)$$

CIR: This model is based on the CIR process (Cox et al. 1985), which is defined as a mixture of "squared" OU processes. Note that here, we invoke squared OU processes and not directly OU processes as was proposed previously (Aris-Brosou and Yang 2003). This makes more sense as it preserves the positivity of the rate process and avoids the systematic bias (rates tending to be higher near the root) that had been pointed out in the model of Aris-Brosou and Yang (2002) by Welch et al. (2005).

The dynamics of the CIR process can be loosely described as a Brownian motion to which a spring-like constraint has been added: whenever the process, $R(t)$, drifts away from the mean value (which is here fixed to 1 to avoid identifiability problems), the spring pulls the process back towards 1, with a force proportional to the deviation $(1 - R(t))$. Hence, when the process is near 1, its behavior is close to a pure Brownian motion; on the other hand, when it is far away from 1 (i.e., very large values or close to zero), the spring effect dominates.

Apart from the stationary mean (of 1), the process has 2 additional parameters, θ and σ, so that in this case, $\nu = \{\theta, \sigma\}$. Given these 2 parameters, and given the initial value $R(0)$ of the process, the pdf of $R(t)$ is

$$p(R(t)|R(0),\theta,\sigma) = c\exp[-c(R(t) + R(0))\mathrm{e}^{-\theta t}] \times \left(\frac{R(t)}{R(0)\mathrm{e}^{-\theta t}}\right)^{\frac{\theta}{\sigma^2}-1/2} I_{\frac{2\theta}{\sigma^2}-1}(2c\sqrt{R(0)R(t)\mathrm{e}^{-\theta t}}),$$

where $c = \frac{2\theta}{\sigma^2}\left(1 - \mathrm{e}^{-\theta t}\right)^{-1}$ and $I_\nu(x)$ is the modified Bessel function of the first kind with parameter ν and argument x (Abramowitz and Stegun 1964). More informative about the process behavior are the mean and variance of $R(t)$ conditioned on $R(0)$:

$$\mathrm{E}[R(t)] = R(0)\mathrm{e}^{-\theta t} + (1 - \mathrm{e}^{-\theta t}), \quad (10)$$

$$\mathrm{Var}[R(t)] = R(0)\frac{\sigma^2}{\theta}\left(\mathrm{e}^{-\theta t} - \mathrm{e}^{-2\theta t}\right) + \frac{\sigma^2}{2\theta}(1 - \mathrm{e}^{-\theta t})^2. \quad (11)$$

We see here that the expected value of $R(t)$ tends to 1 for large values of t. More precisely, it follows an exponential decay of parameter θ; θ is, thus, the autocorrelation parameter of the CIR process (or equivalently, $1/\theta$ is the decorrelation time of the process). In addition, the variance of the process also converges to a finite value as t becomes large, namely $\sigma^2/2\theta$. In fact, as t goes to infinity, the process has a well-defined stationary distribution, which is a gamma distribution with mean one and variance $\sigma^2/2\theta$. This contrasts with the linearly increasing variance of the lognormal model, which does not have a stationary behavior. Finally, for small values of t, the variance of $R(t)$ is proportional to $R(0)\sigma^2$, so that σ will be designated as the "short-term" variance of the process. Note that, in order to define a valid CIR process, σ and θ have to be positive and such that $\sigma \leq 2\theta$.

To implement the CIR process, we cannot directly rely on the pdf of $B(t) = \int_0^t R(s)\mathrm{d}s$, which is not available in closed form. One possibility would be to use instead the approximation proposed for the lognormal model, that is, $(R(0) + R(t))t/2$. However, because of the convexity implied by the exponential decay of the mean value of the CIR (eq. 10), this estimate does not seem appropriate. Instead, we used the expectation of the process conditional on $R(0)$ and $R(t)$. In the case of the CIR, this expectation is analytically available and can be obtained as the first derivative of the moment-generating function (mgf), of $B(t)$:

$$\frac{\mathrm{d}M(z)}{\mathrm{d}z}|_{z=0} = E\left[B(t)|R(0),R(t)\right],$$

where

$$M(z) = E\left[\mathrm{e}^{zB(t)}|R(0),R(t)\right]$$

is the mgf (see Lepage et al. (2006) for a complete description and derivation). Note that the mgf of $R(t)$ in the lognormal case does not have a closed form.

Finally, as a reference model, we considered the unconstrained model: this model does not distinguish times from rates, but more simply considers effective branch lengths as independent exponential variables, of mean b. This is the most commonly used model in Bayesian phylogenetic inference, for example, when the problem is to estimate the topology (Huelsenbeck and Ronquist 2001).

Prior Choices

We implemented 3 different priors for divergence times:

The uniform (Unif) prior: Let $T_{i_{\text{up}}}$ and $T_{i_{\text{down}}}$ be the divergence times of the parent node and the oldest child node of T_i, respectively. The uniform prior posits that T_i is uniformly distributed in the interval $[T_{i_{\text{up}}}, T_{i_{\text{down}}}]$. In other words, $p(\mathbf{T})$ vanishes from equations (6), (7), and (10), provided that the order of divergence times conforms with the topology, and is otherwise equal to zero.

Downloaded from https://academic.oup.com/mbe/article-abstract/24/12/2669/977266 by Imperial College London Library user on 09 October 2019

The resulting prior can be seen as a joint distribution of uniform order statistics over a tree of variable (but always finite) length and is the simplest prior compatible with a fixed topology.

The Dirichlet (DIR) prior (Kishino et al. 2001; Thorne and Kishino 2002): Node ages are distributed according to a single-parameter Dirichlet distribution spanning the whole tree. The uniform prior is a particular case of the Dirichlet, with concentration parameter (α) equal to 1.

The birth–death (BD) process (Yang and Rannala 1997, 2006): The birth–death process has 3 hyperparameters: the birth rate λ, death rate μ, and sampling fraction ρ. There is, however, a lack of identifiability between the 3 parameters, as evaluating the prior at (λ, μ, ρ) gives the same value as if evaluated at $\left(\lambda + k, \mu + k, \frac{\lambda\rho}{\lambda+k}\right)$. We avoid such problems by restraining the parameter number to 2, with $p_1 = \lambda - \mu$ and $p_2 = \lambda\rho$ as our new parameterization. An exponential prior distribution of mean one is assigned to p_1 and p_2.

The prior assigned to hyperparameters θ and σ is a bivariate exponential distribution, with a support truncated by the condition $\sigma < 2\theta$ (see the definition of the CIR process). All hyperparameters are endowed with an exponential prior of mean 1, except for the scale parameter μ, which has a uniform prior over [0, 600], and α, the concentration parameter of the Dirichlet distribution, which is endowed with a uniform distribution over [0, 300].

Markov Chain Monte Carlo Sampling

Samples from the posterior distribution under each model are obtained using the method of Markov chain Monte Carlo (MCMC). Our implementation is an extension of a previously described program, PhyloBayes (Lartillot and Philippe 2004), which has been adapted to work both using classical pruning-based likelihood (Felsenstein 1981) or under the normal approximation (Kishino et al. 2001). As Bayes factors computations are CPU intensive, all results presented here rely exclusively on the normal approximation. We used the estbranches program (Thorne et al. 1998) to estimate the variance covariance matrix, under the JTT model of amino acid substitution (Jones and Taylor 1992). We did neither account for heterogeneities across sites nor across partitions of the data matrix.

The update mechanisms are all based on the Metropolis–Hastings algorithm (Metropolis et al. 1953; Hastings 1970), using additive moves to update each parameter in turn (the Hastings ratio is 1 in all such cases). In the case of flexible models, effective branch lengths have to be updated as well, which we do by using multiplicative moves: each branch is updated separately. The update consists in multiplying the branch by a random factor $m = e^{\delta[U-0.5]}$, where δ is a tuning parameter and U is a random number uniformly distributed in [0, 1]. The Hastings ratio of this move is m.

We also devised a so-called "identifiability move" between μ and **R**. This move was motivated by the observation that there is an intrinsic lack of identifiability, in the likelihood, between rates and the scale factor μ. In such a situation, the posterior distribution over μ and over the rates is still well defined and will be determined by the joint prior density on those parameters. However, the resulting MCMC sampling will be slow, unless specific compensatory moves between μ and the rates are implemented, leaving the effective branch lengths invariant. According to this update mechanism, μ is divided by a random factor $m = e^{\delta[U-0.5]}$ and, concomitantly, all the rates of the **R** vector are multiplied by that same factor m. The Hastings ratio is equal to m^{2N-3}.

A cycle of the MCMC sampler consists of a series of calls to each of the update mechanisms taken in turn. We usually let chains complete a burn-in period consisting of 100,000 cycles. When chains reach stationarity, we sample the posterior distribution every 1,000 cycles for models with no explicit rate variation (e.g., WN and Strict) and every 10,000 cycles for models with higher dimensionality (CIR and LogN), until a total of 1,000 points have been obtained.

Model Comparison

Bayes factors were evaluated using thermodynamic integration (Ogata 1989). In the present work, we follow the model–switch approach described in Lartillot and Philippe (2006). In brief, the method consists in defining a path between the 2 models' respective posterior distributions:

$$p_\beta(\theta) = p(\theta|D, M_1)^{1-\beta} p(\theta|D, M_2)^{\beta}. \qquad (12)$$

When β runs over [0, 1], the path starts at model M_1 and progressively lands over model M_2. Then, integrating $\ln p(D|\theta,M_2) - \ln p(D|\theta,M_1)$ over a quasi-static MCMC realization of such a path yields an estimate of the logarithm of the Bayes factor between the 2 models.

When performing a model–switch between each relaxed clock model and the deconstrained model, the variation of $E_\beta[\ln p(D|\theta,M_2) - \ln p(D|\theta,M_1)]$ can be of several orders of magnitude, mostly close the boundaries $\beta = 0$ and 1, which results in too large discretization errors under the model–switch scheme explained in Lartillot and Philippe (2006). To circumvent this difficulty, we here introduce a sigmoidal schedule for β. That is, if K is the number of steps of the integration, we set $\beta_0 = 0$, $\beta_K = 1$, and for for $0 < k < K$:

$$\beta_k = \frac{e^{\alpha x_k}}{e^{\alpha x_k} + e^{-\alpha x_k}},$$

where $x_k = k/K - 1/2$. In this way, the quasi-static MCMC spends most of the time in the vicinity of the boundaries, while making larger steps in the middle of the interval, where variations are small. We set $K = 10{,}000$, saving every 100 point.

In addition, to obtain a higher precision in the comparisons between the 3 alternative priors on divergence times, we also devised an additional model–switch scheme between either the Dirichlet or the birth–death prior and the uniform prior. These integrations were performed using a linear schedule, using $K = 10{,}000$, and saving every 100 points.

The error on the estimate includes the discretization error and the thermic lag but not the sampling error. By repeating the experiment 10 times in the case of one of the

Downloaded from https://academic.oup.com/mbe/article-abstract/24/12/2669/977266 by Imperial College London Library user on 09 October 2019

Table 1
Natural Logarithm of Bayes Factors for Relaxed Clock Models against the Unconstrained Model

Model	MOC	MAM	MITO10	MITO50	MITO100	MITO150
STRICT	[−3,072; −3,068]	[−160; −157]	[−178; −172]	[−517; −500]	[−956; −897]	[−1,919; −1,825]
UEXP	[−1.3; −0.6]	[18.8; 20.1]	[1.0; 1.3]	[−8.0; −6.9]	[−9.9; −7.0]	[−36.1; −31.8]
UGAM	[26.6; 27.2]	[53.0; 55.7]	[3.9; 4.0]	[14.6;16.8]	[34.6; 36.2]	[4.5; 12.0]
WN	[23.9; 24.6]	[48.8; 49.3]	[1.3; 1.6]	[9.0; 9.5]	[20.4; 21.3]	[−4.3; −3.0]
LOGN	[40.8; 43.1]	[61.1; 64.1]	[3.2; 3.7]	[23.5; 26.0]	[48.3; 56.8]	[36.7; 44.0]
CIR	[36.7; 40.5]	[58.9; 60.6]	[2.5; 3.0]	[20.0; 22.8]	[51.6; 57.4]	[51.4; 60.2]

data sets (MOC, see below), and based on a sample size of $K = 10{,}000$, we could estimate this sampling error to be of the order of a few units of logarithm (less than 3).

Data

The following data sets were considered:

MOC: A concatenation of 129 genes (30,399 positions), sampled in 36 eukaryotic species (Douzery et al. 2004). The slime mold *Dictyostelium* was taken as the outgroup.

MAM: A concatenation of 3 nuclear genes (3,700 positions) sampled in 60 mammalian species representing all mammalian orders, with 2 marsupials taken as outgroup species (Poux et al. 2006).

MITO: A concatenation of 11 protein-coding genes (ATP6, CO2, CYTB, ND2, ND4L, ND5, CO1, CO3, ND1, ND3, ND4) of the vertebrate mitochondrial genome (3,220 positions). This data set is made of a set of 155 osteichthyans and 5 chondrichthyans. We used the chondrichthyans as the outgroup and considered 4 different taxon samplings for the ingroup, obtained by randomly subsampling 10, 50, 100, and 150 osteichthyans. In each case, the topology was determined using the mtREV + gamma model in TreeFinder (Jobb et al. 2004).

Results and Discussion

Comparison of Alternative Rate Relaxation Models

We first compared alternative clock relaxations on 6 different data sets, by computing the Bayes factor between each clock model, that is, the strict clock (STRICT), the uncorrelated exponential (UEXP), gamma (UGAM), and white-noise (WN), as well as the autocorrelated lognormal (LOGN) and CIR models. The unconstrained model was used as a reference, and all comparisons were performed under a uniform prior on divergence dates. The results are shown in table 1 on a logarithmic scale. The most striking conclusions obtained from these comparisons are

1. Rejection of the molecular clock: We see that for all data sets, the molecular clock is overwhelmingly rejected in favor of the unconstrained model, with Bayes factors ranging from −157 to −3,072. This but just confirms many previous observations (Takezaki et al. 1995; Suchard et al. 2001; Douzery et al. 2004), indicating that the strict molecular clock is not adapted to describing the evolution of real sequences.
2. Presence of autocorrelation: For almost all data sets, the Bayes factors of the relaxed but completely uncorrelated models (WN, UGAM, UEXP) are significantly inferior to those of the autocorrelated relaxed clock models (LOGN and CIR), indicating that autocorrelation is an important feature of the variation of evolutionary rates across time. Only when the number of taxa is small (MITO10) are the uncorrelated models comparable to the autocorrelated models. Conversely, under dense taxonomic sampling (MITO150), the UExp and the WN models were so bad that they performed even worse than the unconstrained model. The fact that dense taxonomic sampling is necessary to discriminate between autocorrelated and nonautocorrelated models may explain previous results, in which no significant rate autocorrelation was detected in real data sets (Drummond et al. 2006).
3. Equivalence of autocorrelated models: For most data sets, the 2 autocorrelated models, LOGN and CIR, have comparable Bayes factors. Some discrimination between the 2 models seems to be possible at the margin, however: LogN is slightly favored in all cases, except on MITO100 (where the 2 models have equivalent scores) and on MITO150 (where CIR is significantly preferred over LOGN, table 1, last row). A possible explanation is that LOGN may perform better on smaller data sets because it is simpler (it has only one parameter, controlling the autocorrelation). Conversely, large taxonomic sampling would make it worth to use a more complex but also more flexible model such as the CIR that can capture both a decorrelation time and a stationary variance from the data. Note that, in the long term, the CIR has a well-defined behavior, that is, a stationary distribution, whereas in the LOGN model the rate has a variance that increases linearly with time. In general, this absence of stationarity for a stochastic process is seen as a potential problem (Lepage et al. 2006). Yet, the Bayes factors observed here tend to indicate that such theoretical shortcomings do not really affect the LogN model's performances in practice.

Dispersion Index and Rate Autocorrelation

Interestingly, UGAM, the uncorrelated gamma model (Drummond et al. 2006), appears to be the best among the 3 branchwise models, although it is less fit than LOGN and CIR. To understand why, a connection may be made between the present model comparisons and more classical tests of the molecular clock hypothesis, based on the experimental measure of the dispersion index.

Downloaded from https://academic.oup.com/mbe/article-abstract/24/12/2669/977266 by Imperial College London Library user on 09 October 2019

Table 2
Natural Logarithm of Bayes Factors for Models CIR and LogN, under the Birth–Death (BD) and the Dirichlet (Dir) Prior, against the Uniform Prior over Divergence Times

Model	MOC	MAM	MITO10	MITO50	MITO100	MITO150
CIR + Dir	[2.0; 2.6]	[−7.9; −6.3]	[−3.6; −2.3]	[−1.4; −1.0]	[2.1; 3.1]	[7.0; 8.5]
LogN + Dir	[2.5; 2.6]	[−7.6; −6.5]	[−5.4; −4.3]	[−2.8; −1.7]	[5.3; 5.8]	[13.4; 16.4]
CIR + BD	[−10.2; −8.1]	[−9.6; −6.3]	[−13.9; −12.9]	[−8.8; −6.9]	[13.3; 17.4]	[46.4; 53.8]
LogN + BD	[−20.5; −19.0]	[−10.3; −6.2]	[−13.9; −13.1]	[−8.4; −7.2]	[13.3; 17.1]	[46.6; 55.0]

The dispersion index $I(t)$ is defined as the ratio between the variance and the mean of the number of substitutions over a time interval of length t. It is equal to 1 in the case of a Poisson process of constant rate, which is why its measure has been used as a test of the molecular clock hypothesis, it considered as a corollary of the neutral theory (Kimura 1983; Gillespie 1991). Under rate variations, $I(t)$ is larger than one, being equal to:

$$I(t) = 1 + \frac{\mathrm{Var}[B(t)]}{E[B(t)]},$$

where $B(t)$ is the integral of the rate over the time interval (as defined in the Materials and Methods).

When $I(t) > 1$, the substitution process is said to be overdispersed (Cutler 2000; Gillespie 1991). Overdispersion is indeed observed in real sequences (Gillespie 1991): the dispersion index is significantly greater than 1 in many cases. Furthermore, is seems to be an increasing function of t (Ohta 1995), a pattern that can be reproduced only under processes of rate variations having sufficiently long-range autocorrelations (Bickel and West 1998).

In the present case, the theoretical dispersion index is equal to $1 + \sigma$ for WN and $1 + \sigma t$ for UGam. Thus, unlike WN, UGam has a dispersion index increasing with t, which is a typical feature of autocorrelated models. This illustrates the point stressed in the methods, namely, that UGam does indeed imply an autocorrelated rate process $R(t)$ within each time window defined by the branches of the tree. This also makes UGam more in accordance with experimentally observed values for the dispersion index (Ohta 1995; Bickel and West 1998), which may explain the better fit we obtain for UGam, compared with WN. Conversely, the fact that UGam has a lower fit than LogN and CIR indicates the presence of autocorrelation over times spanning several successive branches, something UGam is unable to capture.

Altogether, autocorrelation can be detected in 2 different ways: between branches, it can be directly observed. Within branches, it can only be indirectly inferred, through its net effect on the dispersion index. And the relative ordering of WN, UGam, and LogN and CIR in our Bayes factor evaluations provides evidence for the presence of autocorrelation at both the small and the large scales.

Comparison of Alternative Priors on Divergence Times

To compare the priors on divergence times, we computed the Bayes factors between either the Dirichlet (Dir) or the birth–death (BD) priors and the uniform prior, used as a reference. We performed these computations under the best 2 rate processes found in the previous section, that is, LogN and CIR.

For all data sets except MITO150, the Bayes factor between the Dir and Unif ranges at most a few natural units above or below zero, indicating that the Dirichlet is more or less equivalent to the uniform prior (table 2). In contrast, the performances of the birth–death prior strongly depend on the data set: it is rejected in favor of the uniform prior in all cases (ln BF < -6), except on MITO100 and MITO150, on which it is strongly favored over the 2 alternatives (ln BF > 45).

The reasons for such a large variation in the statistical fit are not totally clear: the lower fit of the birth–death prior for small data sets may be the result of a dimensional penalty (the birth–death has one more parameter than the Dirichlet). Alternatively, this could be due to the fact that the trees of MITO100 and 150 have long basal branches, separating densely sampled vertebrate classes, a pattern that may be particularly well fitted by a birth–death process. In contrast, the other 2 data sets (MOC, MAM) display a tree shape characterized by an early interphyletic or interordinal diversification, combined with a low phylogenetic redundancy (few closely related taxa). Additional observations would probably be needed to settle this issue.

Relative Divergence Times

We also looked at the effect of models on relative divergence dates. An example of chronogram is displayed on figure 1, using the CIR model on the MITO150 data set. No fossil records were used in this study, so dates are displayed on a relative scale where all leaves have an age of zero and the root an age of 100. As a consequence, this study will not provide any information on the effect of models on absolute dates and, in particular, on the root age. On the other hand, it allows us to investigate the discrepancies between models (given that large discrepancies in the relative node ages between 2 models will probably imply large discrepancies in the absolute dates as well), irrespective of all the issues raised about the way fossil calibration should be properly used (Yang and Rannala 2006).

These comparisons are congruent with the observations made in the previous section, based on Bayes factor evaluations. First, both the Strict (fig. 2*a*) and the WN models (fig. 2*b*) seem to yield relative age estimates significantly at odds with those obtained under the CIR. In contrast, relative ages under LogN and CIR are much more similar (fig. 2*c*).

Interestingly, there is a higher discrepancy between CIR and Strict than between WN and Strict (compare fig. 2*a* and *b*). This is particularly true for the nodes closer to the root (ages > 50), for which there is virtually no agreement between CIR and Strict (fig. 2*a*), whereas some

Downloaded from https://academic.oup.com/mbe/article-abstract/24/12/2669/977266 by Imperial College London Library user on 09 October 2019

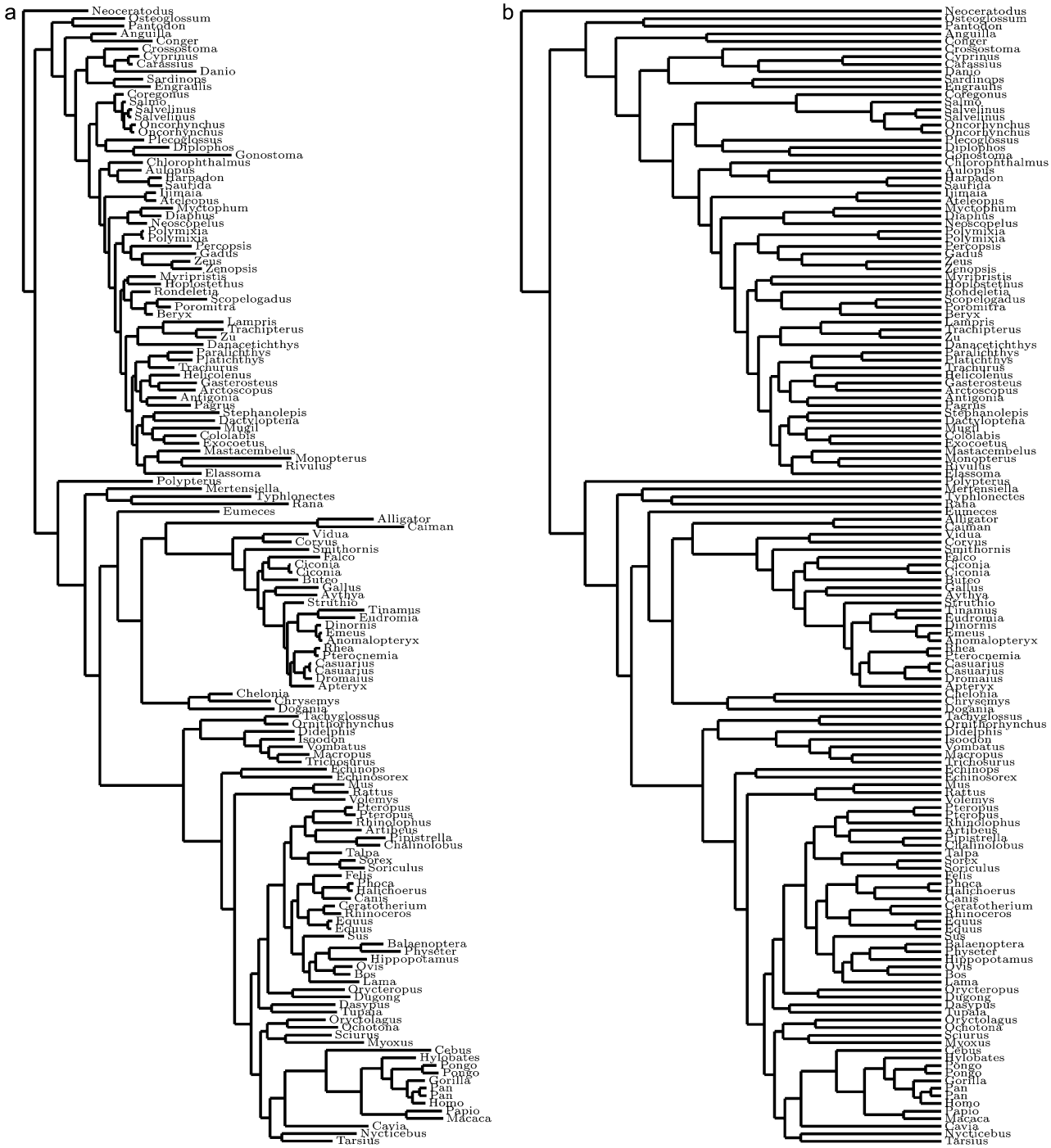

FIG. 1.—*a*) Phylogenetic tree representing the MITO150 data set (150 osteichthyans), with branch lengths determined by maximum likelihood (hence, measured in expected number of substitutions per site). (*b*) Relative divergence times under the MITO150 data set, under the CIR process and a uniform prior on divergence times.

overlap can be seen in the credibility intervals obtained under Strict and WN (fig. 2*d*). To get a more quantitative understanding, we computed the number of node age estimates under the STRICT model that were significantly departing from their counterpart in another given model *M* (i.e., fell outside the 95% credibility interval of the corresponding estimate under model *M*). We found only 45, 33, and 53 significantly departing nodes between the STRICT and the WN, UEXP, and UGAM models, respectively. On the other hand, the number of departing nodes went up to 100 between STRICT and CIR and to 93 between STRICT and LOGN.

This larger discrepancy between Strict and CIR, compared with WN and CIR, seems to have 2 different causes: first, WN is closer to the strict molecular clock than CIR in terms of posterior mean estimates (fig. 2*a* and *b*). But also, the estimates under CIR have a smaller uncertainty, with

Downloaded from https://academic.oup.com/mbe/article-abstract/24/12/2669/977266 by Imperial College London Library user on 09 October 2019

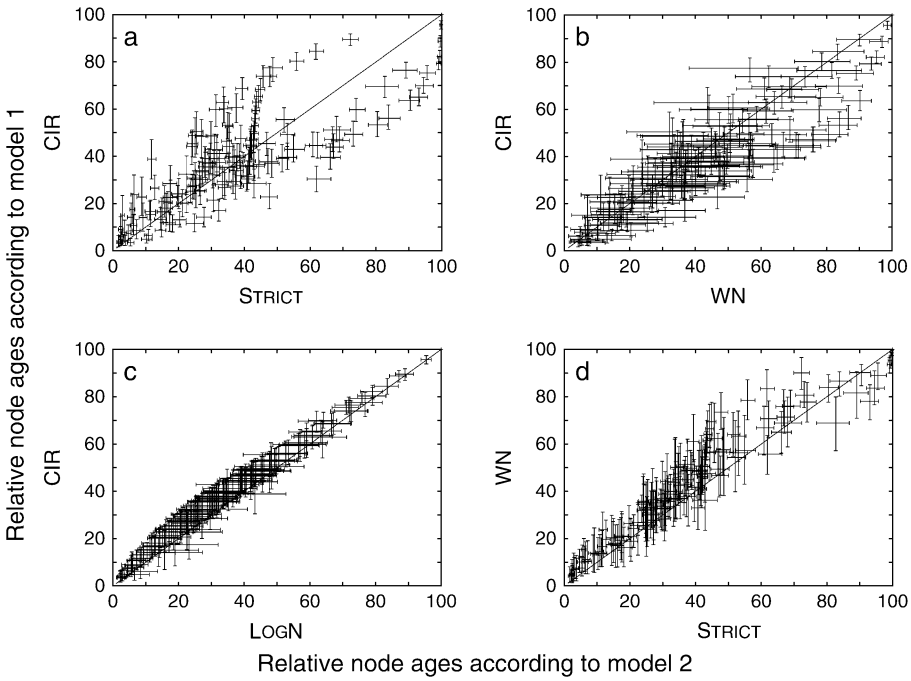

FIG. 2.—Relative node ages according to alternative relaxed clock models inferred in the case of the MITO150 data sets. Error bars represent the 95 credibility intervals. (*a*) STRICT against CIR, (*b*) WN against CIR, (*c*) LOGN against CIR, and (*d*) STRICT against WN.

credibility intervals about twice as small as those obtained under WN (compare fig. 2*a* and *b*) and more comparable to those obtained under the strict molecular clock (fig. 2*a*, horizontal bars). This makes some sense: thanks to autocorrelation, the rate, and time estimates at each node can rely on both local and distant information, which may help them to reduce their variance. Yet, whether such a high confidence corresponds to a reasonable measure of the uncertainty of the dating remains to be investigated.

We also investigated the effect of the prior on divergence times. It has been suggested that the birth–death prior might be too informative, to the point that in some cases (Aris-Brosou and Yang 2003) it was contributing to a larger extent to the final age estimates than the signal from the data (Welch et al. 2005). We therefore compared the prior and the posterior estimates of relative ages, using either the uniform or the birth–death prior (fig. 3*a* and *b*). We observe in both cases that, for most nodes, the posterior standard errors (SEs) are much smaller than the prior SEs, indicating that the 2 priors are not excessively informative in general.

Nodes near the root are an exception, however, and more strikingly so under the uniform prior: for old nodes, and compared with the birth–death prior, the uniform prior appears to be both more informative and more in contradiction with posterior intervals. In particular, it tends to push old nodes further toward the root (fig. 3*a*). This trend is also observed under the birth–death prior (fig. 3*b*) but is less pronounced. This could explain the better fit of the birth–death prior over the uniform one found in the case of MITO150 (table 2).

In any case, and irrespective of the exact prior used, the data seem to at least partially override the prior information as it provides age estimates that are quite robust (although not totally insensitive) to the exact choice of the prior (fig. 3*c* and *d*). Hence, the prior has an influence, but this influence is moderate, compared with that of the rate process (fig. 2).

Note that the birth–death prior appears here less informative than in Aris-Brosou and Yang (2003), as noted by Welch et al. (2005). This may be simply due to the fact that our priors on hyperparameters are more vague. Another point raised by Welch et al. (2005) is that the OU process, as implemented in Aris-Brosou and Yang (2003), imposes a general trend in the rate history. We therefore checked that the optimal model (CIR + BD) found under MITO150 (see fig. 1*b*) did not produce such patterns (fig. 4). Altogether, these checks indicate that the use of squared, instead of direct, OU processes and flexible enough hyperpriors on the parameters of the birth–death process does not seem to yield an excessively biased model for molecular datings.

Conclusion

In summary, our observations illustrate the high sensitivity of the estimated dates to the relaxation model used (fig. 2) and to a lesser extent also to the prior on divergence times (fig. 3). This confirms the need for a global comparison of relaxed clock models, so as to guide model choice in future molecular dating experiments.

A systematic evaluation of Bayes factors between alternative models and priors by practitioners in the field of molecular dating would, in the long run, yield a much more comprehensive perspective on the relative merits of these models and priors. To this end, the software presented in this article is made freely available (http://www.lirmm.fr/mab).

In this direction, our Bayes factor evaluations offer a series of indications. Concerning the relaxation process, our comparisons give a nearly unanimous answer in favor of autocorrelated models, which should therefore probably

Downloaded from https://academic.oup.com/mbe/article-abstract/24/12/2669/977266 by Imperial College London Library user on 09 October 2019

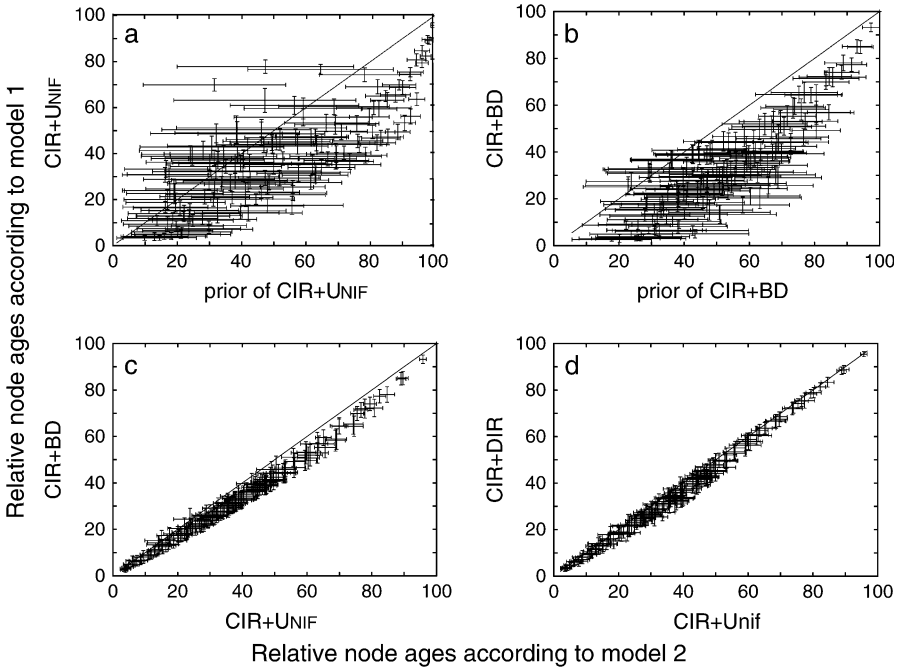

FIG. 3.—Comparisons of prior and posterior estimates of relative node ages using alternative priors on divergence times, in the case of the MITO150 data set. Error bars represent the 95 credibility intervals. (*a*) Prior versus posterior estimates under the CIR + UNIF model. (*b*) Prior versus posterior estimates under the CIR + BD model. (*c*) Posterior estimates under CIR + UNIF versus CIR + BD. (*d*) Posterior estimates of CIR + UNIF versus CIR + DIR.

be used on a systematic basis. This conclusion is assuring, given the current practice prevailing in the field of molecular dating. On the other hand, concerning the choice of the prior on divergence times, no clear guidelines can be deduced from our observations as the Bayes factors turn out to depend strongly on the data sets. More comparisons may be needed in this direction. Alternatively, a model averaging device using reversible jump (Green 1995) could be the most convenient way around the problem.

Our conclusions are exclusively based on relative age estimates, but they likely transpose to situations where one tries to estimate absolute ages using fossil calibrations. In this respect, the lesser variance observed under the autocorrelated models, compared with the uncorrelated ones (fig. 2*a* and *b*), will probably also be observed for absolute dates. The reason is again quite simple: relying on autocorrelation, at least through a model correctly capturing its properties, has the advantage of letting calibrations effectively inform the age estimates across larger distances over the tree. Note that Bayes factors could also be computed between alternative calibration settings, allowing one to measure the mutual agreement between multiple calibration points.

Otherwise, there are many other directions that were not considered in the present work: other models of rate relaxation can be imagined, among which piecewise constant clocks, which can be modeled using compound processes (Huelsenbeck et al. 2000) or fractal processes (Bickel and West 1998). Finally, the impact of the substitution model has not been investigated here. All these are important issues, which can all be addressed in the general framework of Bayesian model comparison that we have proposed here.

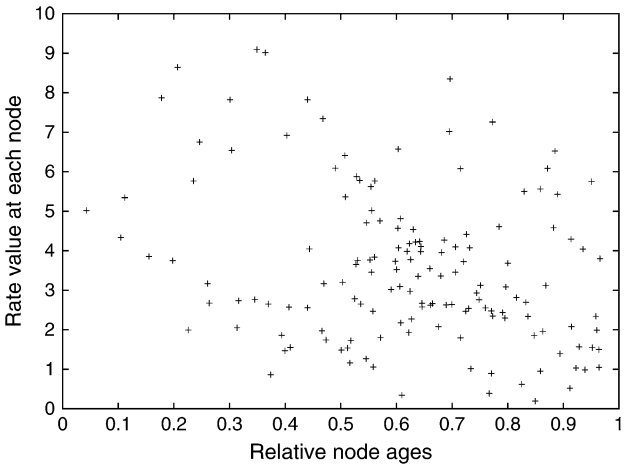

FIG. 4.—Mean posterior rate at each node as a function of its mean posterior relative age, for the MITO150 data set under the CIR + BD model.

Acknowledgments

We wish to thank Paul Tupper, Emmanuel Douzery, and 2 anonymous referees for comments on the manuscripts and the Réseau Québécois de Calcul de Haute Performance for computational resources. This work was supported by Génome Québec, the McGill Centre for Bioinformatics, the Canadian Institute for Advanced Research, the Canadian Chair Program, the Centre National de la Recherche Scientifique (through the ACI-IMPBIO Model-Phylo funding program), the french Agence Nationale pour la Recherche

Downloaded from https://academic.oup.com/mbe/article-abstract/24/12/2669/977266 by Imperial College London Library user on 09 October 2019

(MITOSYS project), the 60ème Comission permanente franco-québécoise, and the Robert-Cedergren Centre for Bioinformatics and Genomics.

Literature Cited

Abramowitz M, Stegun IA. 1964. Handbook of mathematical functions with formulas, graphs, and mathematical tables, 9th dover printing ed. New York: Dover Publications.

Aitkin M. 1991. Posterior Bayes factors. J R Stat Soc Ser B. 53:111–142.

Aris-Brosou S, Yang Z. 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.

Aris-Brosou S, Yang Z. 2003. Bayesian models of episodic evolution support a late precambrian explosive diversification of the Metazoa. Mol Biol Evol. 20:1947–1954.

Bickel DR, West BJ. 1998. Molecular evolution modeled as a fractal Poisson process in agreement with mammalian sequence comparison. Mol Biol Evol. 15:967–977.

Blair JE, Hedges SB. 2005. Molecular clocks do not support the Cambrian explosion. Mol Biol Evol. 22:387–390.

Blanquart S, Lartillot N. 2006. A Bayesian compound stochastic process for modeling nonstationary and nonhomogeneous sequence evolution. Mol Biol Evol. 23:2058–2071.

Bromham L, Philipps MJ, Penny D. 1999. Growing up with dinosaurs: molecular dates and the mammalian radiation. Trends Ecol Evol. 14:113–118.

Cox JC, Ingersoll JE, Ross SA. 1985. A theory of the term structure of interest rates. Econometrica. 53:385–408.

Cutler DJ. 2000. Estimating divergence times in the presence of an overdispersed molecular clock. Mol Biol Evol. 17: 1647–1660.

Douzery EJP, Snell EA, Bapteste E, Delsuc F, Philippe H. 2004. The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? Proc Natl Acad Sci USA. 101:15386–15391.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:699–710.

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

Gillespie JH. 1991. The causes of molecular evolution. Oxford: Oxford University Press.

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

Hastings WK. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika. 57: 97–109.

Ho SY, Phillips MJ, Drummond AJ, Cooper A. 2005. Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation phylogenetic tree. Mol Biol Evol. 22:1355–1363.

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

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 17:754–755.

Jeffreys H. 1935. Some tests of significance, treated by the theory of probability. Proc Camb Philos Soc. 31:203–222.

Jobb G, von Haeseler A, Strimmer K. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol. 23.

Jones D, Taylor WTJ. 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 8:275–282.

Kimura M. 1983. The neutral allele theory of molecular evolution. Cambridge: Cambridge University Press.

Kishino H, Thorne JL, Bruno W. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Mol Biol Evol. 18:352–361.

Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol. 21:1095–1109.

Lartillot N, Philippe H. 2006. Computing Bayes factors using thermodynamic integration. Syst Biol. 55:195–207.

Lepage T, Lawi S, Tupper PF, Bryant D. 2006. Continuous and tractable models for the evolutionary rate. Math Biosci. 199: 216–233.

Metropolis S, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. 1953. Equation of state calculation by fast computing machines. J Chem Phys. 21:1087–1092.

Ogata Y. 1989. A Monte Carlo method for high dimensional integration. Num Math. 55:137–157.

Ohta T. 1995. Synonymous and nonsynonymous substitution in mammalian genes and the nearly neutral theory. J Mol Evol. 40:56–63.

Perez-Lozada M, Hoeg J, Crandall K. 2004. Unraveling the evolutionary radiation of the thoracican barnacles using molecular and morphological evidence: a comparison of several divergence time estimation approaches. Syst Biol. 53: 244–264.

Peterson KJ, Lyons JB, Nowak KS, Takacs CM, Wargo MJ, McPeek MA. 2004. Estimating metazoan divergence times with a molecular clock. Proc Natl Acad Sci USA. 101: 6536–6541.

Poux C, Chevret P, Huchon D, De Jong WW, Douzery EJP. 2006. Arrival and diversification of caviomorph rodents and platyrrhine primates in South America. Syst Biol. 55: 228–244.

Rodrigue N, Philippe H, Lartillot N. 2006. Assessing site-interdependent phylogenetic models of sequence evolution. Mol Biol Evol. 23:1752–1775.

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

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

Smith AB, Peterson KJ. 2002. Dating the time origin of major clades: molecular clocks and the fossil record. Annu Rev Earth Planet Sci. 30:65–88.

Smith A, Pisani D, Mackenzie-Dodds J, Stockley B, Webster B, Littlewood T. 2006. Testing the molecular clock: molecular and paleontological estimates of divergence times in the echinoidea (echinodermata). Mol Biol Evol. 23:1832–1851.

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

Suchard MA, Weiss RE, Sinsheimer JS. 2001. Bayesian selection of continuous-time Markov chain evolutionary models. Mol Biol Evol. 18:1001–1013.

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

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

Thorne JL, Kishino H, Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol Biol Evol. 15:1647–1657.

Downloaded from https://academic.oup.com/mbe/article-abstract/24/12/2669/977266 by Imperial College London Library user on 09 October 2019

Welch JJ, Fontanillas E, Bromham L. 2005. Molecular dates for the "cambrian explosion": the influence of prior assumptions. Syst Biol. 54:672–678.
Yang Z, Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. Mol Biol Evol. 17:717–724.
Yang Z, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol Biol Evol. 23:212–226.
Yoder A, Yang Z. 2000. Estimation of primate speciation dates using local molecular clocks. Mol Biol Evol. 17:1081–1090.

Spencer Muse, Associate Editor

Accepted September 7, 2007

Downloaded from https://academic.oup.com/mbe/article-abstract/24/12/2669/977266 by Imperial College London Library user on 09 October 2019