A Phylogenetic Model for Investigating Correlated Evolution of Substitution Rates and Continuous Phenotypic Characters

Nicolas Lartillot*[,1] and Raphaël Poujol[1]

[1]Département de Biochimie, Centre Robert Cedergren, Université de Montréal, Montréal, Québec, Canada

***Corresponding author:** E-mail: nicolas.lartillot@umontreal.ca.

Associate editor: Asger Hobolth

Abstract

The comparative approach is routinely used to test for possible correlations between phenotypic or life-history traits. To correct for phylogenetic inertia, the method of independent contrasts assumes that continuous characters evolve along the phylogeny according to a multivariate Brownian process. Brownian diffusion processes have also been used to describe time variations of the parameters of the substitution process, such as the rate of substitution or the ratio of synonymous to nonsynonymous substitutions. Here, we develop a probabilistic framework for testing the coupling between continuous characters and parameters of the molecular substitution process. Rates of substitution and continuous characters are jointly modeled as a multivariate Brownian diffusion process of unknown covariance matrix. The covariance matrix, the divergence times and the phylogenetic variations of substitution rates and continuous characters are all jointly estimated in a Bayesian Monte Carlo framework, imposing on the covariance matrix a prior conjugate to the Brownian process so as to achieve a greater computational efficiency. The coupling between rates and phenotypes is assessed by measuring the posterior probability of positive or negative covariances, whereas divergence dates and phenotypic variations are marginally reconstructed in the context of the joint analysis. As an illustration, we apply the model to a set of 410 mammalian cytochrome b sequences. We observe a negative correlation between the rate of substitution and mass and longevity, which was previously observed. We also find a positive correlation between $\omega = dN/dS$ and mass and longevity, which we interpret as an indirect effect of variations of effective population size, thus in partial agreement with the nearly neutral theory. The method can easily be extended to any parameter of the substitution process and to any continuous phenotypic or environmental character.

Key words: comparative method, independent contrasts, molecular dating, life-history evolution, Markov chain Monte Carlo, Bayesian statistics.

Introduction

Phylogenetic comparative methods are an essential tool in ecological and evolutionary analyses. One of their essential aims is to empirically investigate the evolutionary variations of phenotypic characters and life-history traits and to test hypotheses about the underlying evolutionary mechanisms. A large diversity of empirical questions can be addressed, for instance, concerning the existence of specific trends in the direction of evolution or the shape and the intensity of the correlations among phenotypic characters and life-history traits (Harvey and Pagel 1991). When measuring correlations between characters, the problem is often reduced to performing linear regressions, possibly after applying some transformation to the variables under investigation (e.g., a logarithmic transformation). An implicit assumption behind this approach is that the variables, once transformed, are linearly correlated. This assumption either may be justified on empirical grounds, as in the case of allometric relations between life-history traits and body mass (Calder 1984), equivalent to linear relations upon logarithmic transformation, or may be retrospectively assessed using goodness-of-fit tests.

Because species are phylogenetically related, the data obtained in a given set of taxa cannot be considered as independent, and therefore simple regression analysis does not apply. To address this problem, general methods accounting for phylogenetic dependencies have been proposed, the most popular being the method of independent contrasts (Felsenstein 1985). The independent contrast method is statistical in essence, relying on a probabilistic model assuming that the continuous characters under study evolve along the lineages of the phylogenetic tree according to a multivariate Brownian diffusion process. The framework is generally used to test the null hypothesis that the characters are not correlated or to estimate the covariance matrix specifying the correlation structure among the phenotypic characters, corrected for phylogenetic inertia. The idea has been reformulated in a generalized linear model framework (Martins and Hansen 1997), revised to account for intraspecific variation (Houseworth et al. 2004; Felsenstein 2008), extended to processes other than the Brownian diffusion process (Butler and King 2004), and extensively applied in ecology and evolution (Harvey and Pagel 1991; Martins and Hansen 1996; Pagel 1999; Garland et al. 2005).

Comparative analyses need not only be restricted to phenotypic and ecological traits but can also be applied to parameters of the substitution process, such as the substitution rate, the ratio of nonsynonymous to synonymous substitutions $\omega = dN/dS$, or the nucleotide or amino acid composition of sequences. The correlations between substitution parameters and phenotypic or life-history traits uncovered in this manner may provide fundamental

empirical clues about the mechanisms of molecular evolution. For example, the fact that the mitochondrial dN/dS negatively correlates with body size (and thus indirectly with population size) in mammals was interpreted as an indirect confirmation of the nearly neutral theory (Popadin et al. 2007). Many other studies have investigated the influence of metabolism, body size, generation time, or longevity on rate variation in nuclear or mitochondrial DNA in mammals, animals, or plants (Martins and Palumbi 1993; Ohta 1993; Bromham et al. 1996; Li et al. 1996; Gillooly et al. 2005; Thomas et al. 2006; Fontanillas et al. 2007; Lanfear et al. 2007; Welch, Bininda-Edmonds, et al. 2008).

Correlations between molecular, phenotypic, and ecological characters can also be extrapolated back in time, thus allowing inference of the history of phenotypic evolution based on ancestral sequences reconstructed using phylogenetic methods. In this direction, correlations between temperature and the guanine-cytosine (GC) content of ribosomal RNA stems, or the amino acid composition of the proteome, were used to infer the history of variations of the optimal growth temperature along the tree of eubacteria and archaea (Galtier et al. 1999; Boussau et al. 2008).

All these are but a few examples suggesting that further development of the comparative method, so as to jointly analyze morphological traits and substitution parameters in one single unified statistical framework, would provide essential empirical leverage for a more global understanding of evolution encompassing the molecular, phenotypic, and ecological dimensions. However, several methodological points need to be addressed so as to optimize power and avoid potential pitfalls.

A first issue is how to deal with uncertainty. Proceeding sequentially, first estimating the substitution parameters, then assessing their correlation with phenotypic characters, and finally extrapolating the correlation onto ancestral nodes, raises a potentially important problem of error propagation. The estimated correlation coefficients should ideally integrate the uncertainty about the substitution parameters and the divergence times. Conversely, the reconstructed phenotypic histories should account for the uncertainty about the estimated correlation coefficients and the substitution parameters. All these problems can in principle be naturally dealt with in a joint estimation framework. In this direction, a Bayesian method for estimating covariance between phenotypic characters while accounting for the uncertainty concerning the topology and the branch lengths of the tree has been developed previously (Huelsenbeck and Rannala 2003). This method could be extended to allow for more general covariance analyses including phenotypic characters as well as substitution parameters. Of note, joint estimation would also have the advantage of allowing for cross talk between components of the model, something which would otherwise not be possible in a sequential method, and which would allow the reconstruction of phenotypic and molecular histories, or the estimation of divergence times (Welch and Waxman 2008), to borrow strength from each other via the inferred covariations.

Second, what should normally be compared with the phenotypic characters of a given set of species are the instant values of the rate of substitution in those species at present time (Welch and Waxman 2008). In practice, what is often measured is the average substitution rate on the terminal branches of the phylogenetic tree. In principle, using sufficiently closely related taxa would allow as high a time resolution as desired. However, for fixed sequence length, this would be at the expense of the accuracy of the estimates of the substitution rates. A more satisfactory approach to the estimation of instant rates of substitution is to model rates as continuously evolving parameters. This has been first proposed in the context of molecular dating methods, using Brownian diffusion processes (Thorne et al. 1998; Kishino et al. 2001; Thorne and Kishino 2002; Yang and Rannala 2006; Lepage et al. 2007; Rannala and Yang 2007). Most often, the rate has implicitly been considered as a nuisance variable, modeled not so much for its own sake, but rather for integrating out the effects of its variations on divergence date estimates. Nevertheless, relaxed-clock models naturally provide, as a by-product, an estimate of the instant value of the rates, in particular at the leaf nodes of the tree, and those rates can in principle be used as the variables to be regressed against the continuous phenotypic traits of interest.

Third, unlike phenotypic characters, rates, and more generally molecular evolutionary parameters, can be inferred also at internal nodes of the phylogenetic tree, using nonstationary models of substitution. For an optimal statistical power, the correlation between substitution parameters should therefore be assessed in a more general framework, which would not exclusively rely on the values estimated at the leaves. In this direction, Seo et al. (2004) extended the idea of the Brownian relaxed clock to model the continuous variations of the rates of synonymous and of nonsynonymous substitution. In their analysis, they modeled these rates as two independent log-normal Brownian diffusion processes, although they suggested that modeling them as a general bivariate process would be a straightforward generalization of the approach.

In the present article, we develop a model combining the ideas of Felsenstein (1985), Huelsenbeck and Rannala (2003), and Seo et al. (2004). In this model, all the molecular evolutionary parameters of interest (of total number K) and all the phenotypic characters, environmental variables or life-history traits (of total number L), are modeled as one single multivariate diffusion process of dimension $K+L$. The covariance matrix of this multivariate process and the divergence times of the underlying phylogeny are considered as unknown. The history of the multivariate process, the covariance matrix, and the divergence times are jointly estimated in a Bayesian framework, using Monte Carlo estimation methods. We illustrate the method by applying it to the analysis of the correlations between the rate of substitution, $\omega = dN/dS$, together with several phenotypic characters and life-history traits (generation time, mass, and longevity) in a multiple alignment of cytochrome b sequences in mammals.

Methods

Definitions and Notations

The method relies on a combination of aligned coding sequences and phenotypic characters for a set of P taxa. The alignment $\mathbf{D}$ is made of P coding sequences of length $3N$ nucleotides (N codon positions), and the phenotypic data are summarized in an $L \times P$ matrix $\mathbf{C}$ such that C_i^j is the value taken by phenotypic character $i = 1, \ldots, L$ in taxon $j = 1, \ldots, P$. In the following, taxa, branches, and nodes will always be referred to using upper indices and phenotypic characters using lower indices.

The taxa are related through a rooted bifurcating phylogenetic tree. The topology is assumed known and will never explicitly appear in the equations. Variables associated to the nodes of the tree are upper indexed by $j = 0, \ldots, 2P-2$, with the convention that root has index 0, leaf nodes have index $j = 1, \ldots, P$, and internal nodes have index $j = P+1, \ldots, 2P-2$. If $j > 0$, we refer to the index of the node immediately ancestral to node j as j_{up}. Similarly, upper indices $j = 1, \ldots, 2P-2$ will be used to refer to branches, with the convention that branch j is the branch immediately ancestral to node j. The divergence times are noted $\mathbf{T} = (T^j)_{j=0,\ldots,2P-2}$. They are relative to the age of the root, that is, $T^0 = 1$ and $T^j = 0$ for the leaf nodes ($j = 1, \ldots, P$). For $j > 0$, $\Delta T^j = T^{j_{\text{up}}} - T^j$ is the time interval represented by the branch leading to node j.

Codon Substitution Model

We consider a codon substitution process, such as originally proposed by Muse and Gaut (1994). First, a general time-reversible Markov process is defined at the nucleotide level, by a 4×4 instant rate matrix R specifying a rate of transition between any pair (n_1, n_2) of nucleotides:

$$R_{n_1 n_2} = \frac{1}{Z} \rho_{n_1 n_2} \pi_{n_2},$$

where Z is the normalization constant, ρ is the set of relative exchangeability parameters constrained to sum to 1 (5 degrees of freedom), and π is the set of nucleotide equilibrium frequencies (3 degrees of freedom). The rate of substitution between any pair of codons (b_1, b_2) differing only at one position and with respective nucleotides n_1 and n_2 at that position is then defined to be equal to

$$Q_{b_1 b_2} = \lambda_S R_{n_1 n_2} \quad \text{if } b_1 \text{ and } b_2 \text{ are synonymous,}$$
$$Q_{b_1 b_2} = \lambda_N R_{n_1 n_2} \quad \text{if } b_1 \text{ and } b_2 \text{ are nonsynonymous.}$$

The rate of substitution between any two codons differing at more than one position is assumed to be equal to 0 (Muse and Gaut 1994). The parameters λ_S and λ_N are the rates of synonymous and nonsynonymous substitution.

Alternatively, one can define $\omega = \lambda_N / \lambda_S$ and express the rates of substitution in terms of λ_S and ω:

$$Q_{b_1 b_2} = \lambda_S R_{n_1 n_2} \quad \text{if } b_1 \text{ and } b_2 \text{ are synonymous,}$$
$$Q_{b_1 b_2} = \lambda_S \omega R_{n_1 n_2} \quad \text{if } b_1 \text{ and } b_2 \text{ are nonsynonymous.}$$

Both formulations will be explored in the following, as they are not strictly equivalent in the context of the present work. They will be called the (λ_S, λ_N) and the (λ_S, ω) parameterizations.

Multivariate Process

The L phenotypic characters are assumed to evolve continuously along the lineages of the phylogenetic tree. In addition, some of the parameters of the substitution process (hereafter called substitution parameters) also vary continuously along the lineages. In the specific case developed in this article, there are two such parameters: either λ_S and ω or λ_S and λ_N. In more general settings, one could consider the variations of any set of K independent substitution parameters. The substitution process (which is here homogeneous across sites) will then be described by an instant rate matrix $\mathbf{Q}$, itself depending on the K substitution parameters: $\mathbf{Q} = \mathbf{Q}(y_1, \ldots, y_K)$. We wish to model the variations in time of y_k, for $k = 1, \ldots, K$, and to correlate these variations with those of the L continuous phenotypic characters.

Accordingly, we define a multivariate diffusion process $X(t)$, of dimension $M = K + L$, running along the branches of the tree. The mth component of $X(t)$ is noted $X_m(t)$ for $m = 1, \ldots, M$. By convention, the first K components of the process describe the variations of the K substitution parameters, and the last L map to the phenotypic characters. Note that we might have to impose a transformation on the phenotypic characters and the substitution parameters. In the following, the continuous characters that we will consider are life-history traits such as body mass, longevity, and generation time for which a logarithmic transformation is justified based on known allometric relations (Calder 1984). In the case of the synonymous and nonsynonymous substitution rates, we follow Seo et al. (2004) and impose a logarithmic transformation also in their case. Thus,

$$X_1(t) = \ln \lambda_S(t),$$
$$X_2(t) = \ln \omega(t),$$
$$X_{l+2}(t) = \ln C_l(t), \quad l = 1, \ldots, L,$$

in the (λ_S, ω) parameterization or

$$X_1(t) = \ln \lambda_S(t),$$
$$X_2(t) = \ln \lambda_N(t),$$
$$X_{l+2}(t) = \ln C_l(t), \quad l = 1, \ldots, L,$$

in the (λ_S, λ_N) formulation.

The stochastic process $X(t)$ is assumed to be multivariate Brownian. The use of a Brownian motion entails several important assumptions. First, it is a Markovian process. Second, it does not display any trend in the direction of its variations. And third, the rate of change per unit of time is constant and is completely determined by an $M \times M$ symmetric definite covariance matrix Σ. Between time t and time $t + \mathrm{d}t$, the process undergoes a random increment drawn from a normal distribution of mean 0 and variance $\Sigma\, \mathrm{d}t$. The total variation of the process over a finite time t can be seen as an infinite sum of such normally distributed random increments and is thus also normally distributed of mean 0 and variance Σt (Felsenstein 1985):

$$X(t) - X(0) \sim N(0, \Sigma t). \tag{1}$$

The Brownian motion does not have a stationary distribution.

As a way of explicitly referring to the actual meaning of each entry of the covariance matrix Σ, in the following, we will use a bra–ket notation (Dirac 1982). For instance, assuming that we have only $K = 1$ phenotypic character, and that we are working under the (λ_S, ω) parameterization, the entries of the matrix will be noted

$$\Sigma = \begin{pmatrix} \langle \lambda_S, \lambda_S \rangle & \langle \lambda_S, \omega \rangle & \langle \lambda_S, C_1 \rangle \\ - & \langle \omega, \omega \rangle & \langle \omega, C_1 \rangle \\ - & - & \langle C_1, C_1 \rangle \end{pmatrix}.$$

Strictly speaking, these entries correspond to the covariances between the logarithm of the variations of each pair of variables. However, for short, we will more simply refer to them as the covariances between the variables.

The Phylogenetic Multivariate Process

Because we cannot instantiate the process at all times along the phylogeny, we only consider the values of $X(t)$ at the nodes of the tree. We note X^j the instant value of the process at node j, and $\mathbf{X} = (X^j)_{j=1,\ldots,2P-2}$ the set of values at all nodes except the root. The Brownian process is Markovian, and we can therefore express the joint probability of $\mathbf{X}$, given the initial value X^0 at the root, as follows:

$$p(\mathbf{X} \mid X^0, \mathbf{T}, \Sigma) = \prod_{j=1}^{2P-2} p(X^j \mid X^{j_{\text{up}}}, \Delta T^j, \Sigma), \qquad (2)$$

where the finite-time transition probabilities are given by equation (1).

Conditioning the last L dimensions of the multivariate process $X(t)$ on the values observed at leaf nodes (i.e., in extant species) is straightforward, as it is done by setting: $X^j_{l+K} = \ln C^j_l$ for $l = 1, \ldots, L$ and $j = 1, \ldots, P$. To introduce a coupling of the first K components of the multivariate process with the substitution process, one would like to set: $y_k(t) = e^{X_k(t)}$ at all times. Doing this would define a nonstationary substitution process, of instant rate matrix $\mathbf{Q}(t) = \mathbf{Q}(y_1(t), \ldots, y_K(t))$. However, we cannot instantiate the values of $X(t)$ at all times, but only at the nodes of the tree, and integrating the likelihood over all possible realizations of $X(t)$ conditional on the values at the nodes seems in most cases completely intractable. Instead, we make the approximation consisting in assuming a constant value for y^j_k along a given branch j, equal to some average of the values of the process at both ends. Doing this for all k defines a constant rate matrix on branch j: $\mathbf{Q}^j = \mathbf{Q}(y^j_1, \ldots, y^j_K)$, and thus, the substitution model reduces to a set of branch-specific substitution matrices $\mathbf{Q}^j$.

There are several possible ways the averages over branches can be computed. First, a very simple approximation, already used in most implementations of relaxed clock models (Thorne et al. 1998; Lepage et al. 2007), consists of taking the "arithmetic" average:

$$y^j_k = \frac{1}{2}\left(e^{X^{j_{\text{up}}}_k} + e^{X^j_k}\right).$$

An alternative method can be proposed. In the case of the Brownian process, the most likely path (or geodesic) going from $X^{j_{\text{up}}}$ at time $T^{j_{\text{up}}}$ to X^j at time T^j is the straight line, and therefore, it would make sense to take the mean value of $e^{X(t)}$ along this geodesic, which is equal to:

$$y^j_k = \frac{e^{X^{j_{\text{up}}}_k} - e^{X^j_k}}{X^{j_{\text{up}}}_k - X^j_k}.$$

We will refer to this latter averaging method as the "geodesic" average. Its main advantage, compared with the arithmetic average, is to more properly account for the convexity of the exponential function. Both approximations are admittedly crude, but because they are quite different in their formulation, using both of them in turn, and comparing the results, will allow some check of the robustness of the method to these finite-time approximations.

Priors

We set a uniform prior on relative divergence times. Analyses were performed with and without calibrations. Without calibrations, divergence dates are simply measured relative to the root as in Lepage et al. (2007). With calibrations, we proceed as in Kishino et al. (2001), that is, we use a gamma density for the age of the root, and conditional on this age, we impose a uniform density on relative ages, truncated so as to be compatible with the intervals specified by the calibrations.

A uniform Dirichlet distribution was imposed on the nucleotide frequencies π and the nucleotide exchangeabilities ρ, a truncated uniform prior defined on $[-100, 100]$ for the root state X^0, and an inverse Wishart prior distribution on the covariance matrix Σ, parameterized by $\Sigma_0 = \kappa I_M$, where I_M is the identity matrix of dimension M, and with $q = M + 1$ degrees of freedom. As for κ, we tried two different values, $\kappa = 1$ and $\kappa = 10$, and we checked that the results were not sensitive to this choice.

The inverse Wishart distribution can be defined as follows (Mardia et al. 1979). If one samples q independent and identically distributed multivariate normal random variables of dimension M, $Z_i \sim N(0, \Sigma_0^{-1})$ for $i = 1, \ldots, q$ and computes the scatter matrix

$$M = \sum_{i=1}^{q} Z_i Z_i',$$

then, $\Sigma = M^{-1}$ is, by definition, distributed according to an inverse Wishart of mean Σ_0 and with q degrees of freedom: $\Sigma \sim W^{-1}(\Sigma_0, q)$. The probability density is

$$p(\Sigma \mid \Sigma_0, q) \propto |\Sigma_0|^{\frac{q}{2}} |\Sigma|^{-\frac{q+M+1}{2}} e^{-\frac{1}{2}\mathrm{tr}(\Sigma_0 \Sigma^{-1})},$$

where we have dropped numerical constants. The choice of the inverse Wishart is motivated by the fact that it is conjugate to the multivariate normal distribution, a property that is key to the efficiency of the estimation strategy.

An alternative version of the model is obtained by enforcing all nondiagonal entries of the covariance matrix to be equal to 0. This can be seen as an alternative prior on Σ, with support restricted to the set of diagonal matrices. To make this "diagonal" model as close as possible to the fully

"covariant" model introduced thus far, the prior on the entries of the diagonal matrices can be chosen to be the same as the marginal priors of those same entries in the inverse Wishart. Technically, these are inverse gamma distributions of shape parameter $\alpha = 2$ and scale parameter $\beta = \kappa/2$ (Mardia et al. 1979).

Markov Chain Monte Carlo Sampling

Samples from the posterior distribution are obtained by Markov chain Monte Carlo (MCMC). For divergence times, internal nodes are taken one by one (in an order defined by a recursive traversal of the tree). For node j, a simple additive move is applied to T^j within the constraints defined by immediately upstream and downstream nodes. The two vectors ρ and π are updated using a simple move constrained so as to keep the sum of all components equal to 1 (Lartillot 2006). In the simplest version of this mechanism, we randomly choose a pair of two entries of the vector to be resampled (say, the two entries π_a and π_b of π), and set $x = \pi(a) + \pi(b)$, and $y = \pi(a)$. We then propose $y' = y + \epsilon(U - 0.5)$. If y' falls outside of the interval $[0, x]$, we reflect it back. Finally, we set $\pi'(a) = y'$ and $\pi'(b) = x - y'$, thus preserving the total sum of the two stationary probabilities. The Hastings ratio is 1. A generalized version of this sum-constrained mechanism randomly draws d nonoverlapping pairs of entries of the vector to be resampled, and simultaneously proposes a compensated move independently on each pair.

Concerning the covariance matrix Σ and the multivariate process X, we implemented two update schemes. The first is a simple alternation between Metropolis–Hastings updates of X conditional on Σ, and conversely, of Σ conditional on X (and all other parameters). Specifically, for the multivariate process, internal nodes of the tree are visited one by one. For node j, one among three types of moves are proposed, each with probability 1/3. According to move number 1, one entry of X^j is chosen uniformly at random, and, if this entry is not clamped, an additive move is performed on it (otherwise, nothing happens). According to move number 2, all entries of X^j (if not clamped) are moved by a same random amount $\epsilon(U - 0.5)$. According to move number 3, all entries that are not clamped are simultaneously moved, each by a different random amount $\epsilon(U_m - 0.5)$ for $m = 1, \ldots, M$. Concerning the covariance matrix Σ, an entry l, m, such that $l \leqslant m$ is chosen uniformly at random, an additive move is proposed: $\Sigma'_{lm} = \Sigma_{lm} + \epsilon(U - 0.5)$, and the symmetry of the matrix is restored by setting Σ'_{ml} equal to Σ'_{lm}. The move is immediately refused if the resulting matrix is not positive definite. Otherwise, the Metropolis decision rule is applied. Note that this is equivalent to setting a posterior density of 0 on all symmetric but not definite positive matrices.

However, this simple alternate update scheme is not efficient due to potentially strong correlations between X and Σ in their joint distribution. A much more efficient approach relies on the conjugate relation between the inverse Wishart and the multinomial distributions to analytically integrate away the covariance matrix Σ.

To see this, we first make a change of variables so as to define the branchwise independent contrasts $\mathbf{Y} = (Y^j)_{j=1,\ldots,2P-2}$:

$$Y^j = \frac{X^j - X^{j_{\text{up}}}}{\sqrt{\Delta T^j}}.$$

These contrasts are i.i.d. from a multivariate normal distribution:

$$Y^j \sim N(0, \Sigma).$$

With this change of variable, any density defined on $\mathbf{X}$ corresponds to a density on $\mathbf{Y}$ according to

$$p(\mathbf{X}) = p(\mathbf{Y})|J|,$$

where $|J|$ is the Jacobian of the transformation:

$$|J| = \prod_{j=1}^{2P-2} (\Delta T^j)^{-\frac{M}{2}}.$$

Next, we rely on the fact that the inverse Wishart distribution is conjugate to the multivariate normal distribution. We can write, up to a normalization constant, the prior on Σ:

$$p(\Sigma|\Sigma_0) \propto |\Sigma|^{-\frac{q+M+1}{2}} e^{-\frac{1}{2}\text{tr}(\Sigma_0 \Sigma^{-1})}, \quad (3)$$

and the sampling probability of $\mathbf{Y}$:

$$p(\mathbf{Y}|\Sigma) \propto \prod_{j=1}^{2P-2} \frac{1}{\sqrt{|\Sigma|}} e^{-\frac{1}{2} Y'^j \Sigma^{-1} Y^j}, \quad (4)$$

$$\propto |\Sigma|^{-(P-1)} e^{-\frac{1}{2}\text{tr}(A\Sigma^{-1})}, \quad (5)$$

where we define the sample covariance matrix

$$A = \sum_{j=1}^{2P-2} Y^j Y'^j. \quad (6)$$

By Bayes theorem, the posterior on Σ, conditional on a particular realization of X (and thus Y) is proportional to the product of equations (3) and (5):

$$p(\Sigma|\mathbf{Y}, \Sigma_0) \propto |\Sigma|^{-\frac{q+M+2P-1}{2}} e^{-\frac{1}{2}\text{tr}((\Sigma_0 + A)\Sigma^{-1})}, \quad (7)$$

from which we see that the posterior is also an inverse Wishart, of parameter $\Sigma_0 + A$ and with $q + 2P - 2$ degrees of freedom.

By identification, the marginal probability density on $\mathbf{Y}$ can be obtained

$$p(\mathbf{Y}|\Sigma_0, q) = \int p(\mathbf{Y}|\Sigma) p(\Sigma|\Sigma_0, q) \mathrm{d}\Sigma, \quad (8)$$

$$\propto \frac{|\Sigma_0|^{\frac{q}{2}}}{|\Sigma_0 + A|^{\frac{q+2P-2}{2}}}, \quad (9)$$

where we have dropped unimportant numerical constants. Finally, the marginal probability of X is obtained by multiplying equations (9) and (3):

$$p(\mathbf{X}|\Sigma_0, q) = p(\mathbf{Y}|\Sigma_0, q)|J|. \quad (10)$$

The two important results coming out from this mathematical derivation are contained in equations (7) and (10). Equation (10) tells us that we can compute the probability of a particular configuration of the stochastic process **X** integrated over all possible values of Σ. This means that we can devise a MCMC sampler working on the reduced parameter space ($\mathbf{X}, \mathbf{T}, \pi, \rho$), not including Σ. The equilibrium distribution of this sampler is the marginal posterior distribution:

$$p(\mathbf{X}, \mathbf{T}, \pi, \rho | D, C, \Sigma_0, q) \propto p(D, C | \mathbf{X}, \rho, \pi) \; p(\mathbf{X} \mid \Sigma_0, q) p(\mathbf{T}) p(\pi) p(\rho), \quad (11)$$

where $p(\mathbf{X}|\Sigma_0, q)$ is given by equation (10). This posterior distribution is conditional on the multiple alignment D, the matrix of continuous characters C, but also on the constants of the model Σ_0 and q.

In a second step, for each value of **X** obtained from this reduced MCMC (burn-in excluded), a scatter matrix A can be computed (eq. 6), and a value of Σ can then be sampled from the distribution given by equation (7). Because, at equilibrium, the values of **X** sampled from the reduced MCMC are from the posterior distribution:

$$\mathbf{X} \sim p(\mathbf{X}|D, C, \Sigma_0, q), \quad (12)$$

and because Σ is conditionally independent of the data (D and C) given X:

$$p(\Sigma, \mathbf{X}|D, C, \Sigma_0, q) = p(\Sigma|\mathbf{X}, \Sigma_0, q)(\mathbf{X}|D, C, \Sigma_0, q), \quad (13)$$

sampling Σ from equation (7) results in Σ and **X** being sampled from their joint posterior distribution. In practice, our sampler works on the reduced parameter vector, but resamples Σ on the fly, each time before saving a new parameter configuration.

Data Augmentation

In addition to the conjugate sampling method just described, the overall MCMC framework relies on data augmentation (Lartillot 2006; Mateiu and Rannala 2006; de Koning et al. 2010). At any time, a complete substitution history (or mapping) is specified, for all sites and over the whole tree, and all the update mechanisms described above are performed conditional on the current mapping. Periodically, the mapping is refreshed, that is, it is resampled conditional on the current parameter values, using a combination of two algorithms described elsewhere (Nielsen 2002; Rodrigue et al. 2008). This MCMC strategy allows for drastic simplifications of the computations (Lartillot 2006; Mateiu and Rannala 2006). First, it does not use the pruning algorithm (Felsenstein 1981) except for resampling the substitution mappings. This is a substantial advantage in the present case, where the state space of the substitution process has size 61. Second, the probability of the substitution mapping depends on fairly compact sufficient statistics (total waiting time in each possible codon, number of transitions between each pair of codons, and in the case of the root, total number of occurrences of each codon), that have here to be computed separately for each branch, but can be summed over all sites of the alignment (Lartillot 2006; Mateiu and Rannala 2006; de Koning et al. 2010). The cost of all update mechanisms except the data augmentation step itself are therefore virtually independent of the number of sites, which is a great advantage for long sequences. Using fast-access associative-array representations of the 61 vectors and 61×61 arrays of sufficient statistics, which are potentially sparse for smaller alignments, makes the implementation efficient across the whole range of sequence length.

The frequency at which the substitution mapping is refreshed, now the limiting step, is tuned so that the MCMC sampler spends between one-tenth and one-half of the total computing time refreshing mappings, whereas the rest of the time is distributed over all other update operators. In practice, all parameters are each resampled several hundred times between each update of the substitution mapping, and one point is saved before each such update. The burn-in is determined visually, and the chain is run for approximately 1,000 points. Each analysis was run at least twice independently.

The implementation was checked using three different methods (Blanquart and Lartillot 2006): 1) the program was run using alternative sampling methods (using conjugate or regular sampling for the covariance matrix, using sufficient statistics or directly recomputing the probability of the substitution histories), and we checked that the equilibrium distributions obtained under the different methods were indistinguishable; 2) the MCMC was run with no data to visually check whether the model was indeed sampling from the prior; and 3) 100 replicates were simulated from the prior, using the (λ_S, ω) parameterization and assuming one continuous character, and reanalyzed under the model so as to obtain for each replicate a sample of 1,000 points approximately from the posterior. For each replicate, and for a series of eight summary statistics, the true value of the statistic was ranked against the sample from the posterior. The 100 ranks thus obtained (expressed in percentiles) should follow a uniform distribution (Huelsenbeck and Rannala 2004), which we visually checked, and quantitatively assessed by the Kolmogorov uniformity test. The summary statistics were the the total length of the tree, the mean value of omega along the tree, and the six independent entries of the 3×3 covariance matrix.

Posttreatment

Once a sample approximately from the posterior is obtained, marginal estimates of any parameter of the model are readily computed. Concerning the reconstructed chronology and phenotypic histories, we estimate, for each node of the tree, a posterior mean and a 95% credibility interval for its date and for each phenotypic character. The same argument applies to substitution parameters.

Concerning the covariance matrix Σ, for each entry, we simply report the posterior mean. For nondiagonal entries $k \neq l$, the correlation coefficient is defined as follows:

$$r_{kl} = \frac{\Sigma_{kl}}{\sqrt{\Sigma_{kk}\Sigma_{ll}}}.$$

The reported correlation coefficients are obtained by applying this formula separately for each point sampled approximately from the posterior distribution and then taking the average. The posterior probability (pp) of a positive correlation ($r_{kl} > 0$) is also estimated based on the observed frequency at which the r_{kl} parameter was found to be positive.

Finally, the slope of the linear regression between two components k and l can be estimated. Here, we use the major axis method, which estimates the slope of the major axis of the bivariate ellipsoid formed by the joint distribution of the two variables of interest (Harvey and Pagel 1991):

$$\frac{\partial X_l}{\partial X_k} = \frac{\Sigma_{ll}^2 - \Sigma_{kk}^2 + \sqrt{(\Sigma_{ll}^2 - \Sigma_{kk}^2)^2 + 2\Sigma_{kl}^2}}{2\Sigma_{kl}}.$$

This slope is computed for each point sampled from the posterior, thus giving a distribution from which an average and a 95% credibility interval are then immediately obtained.

For multiple regressions, one is interested in knowing the covariance between k and l, for constant m, which is given by

$$\Sigma_{kl;m} = \Sigma_{kl} - \frac{\Sigma_{km}\Sigma_{lm}}{\Sigma_{mm}}.$$

Again, the formula is applied for each point from the posterior so as to obtain a distribution from which to compute a mean and a pp for assessing significance.

The software program runs under the Linux or MacOS operating system. It is freely available from our Web site http://www.phylobayes.org.

Data Set

We analyzed an alignment of cytochrome b sequences of 410 therian species, 29 marsupials, and 381 placentals (1,146 nucleotide positions, or equivalently, $N = 382$ codon positions) obtained from Nabholz et al. (2008). Three life-history traits were investigated: age of female at maturity, taken as a proxy for generation time; adult weight, as a proxy for body mass; and maximum recorded lifespan, as a proxy for longevity. The values of these three characters were obtained from the AnAge database (de Magalhaes and Costa 2009). The 410 sequences correspond to all marsupials and placentals of the initial data set of Nabholz et al. (2008) for which the three life-history traits are documented in the AnAge database. We also extracted from this large alignment a reduced data set restricted to carnivores (67 taxa). We used the fossil calibrations reported in Nabholz et al. (2008) except for the five involving taxa absent from our data set (i.e., *Bradypus*/*Dasypus*, *Apodemus*, *Gerbillus*, *Bolomys*/*Acodon*, and *Neofiber*/*Ondatra*). The prior on the age of the root was defined as an exponential of mean 150 My.

Simulations

A first series of simulations was conducted to assess whether the method is able to recover reasonable estimates of the covariance matrix in practical situations. To ensure realism, the simulations were based on a tree, a set of divergence times, and a mutation rate matrix estimated on the carnivore data set. In addition, and as an attempt to address the problem of the discretization error induced by the finite-time averages computed along each branch, the simulations were performed using a more sophisticated version of the model in which each branch is subdivided in 50 small segments of equal length. The simulation proceeds step-by-step successively along each segment, which results in smaller discretization errors, and is therefore closer to the ideal situation in which the codon substitution process is supposed to change continuously along the branches.

For 50 replicates, random covariance matrices were sampled from the prior (using $\kappa = 1$) and were used to simulate a complete history of the multivariate process along the tree, and from this realization of the process, a codon alignment of 342 coding positions (1,146 aligned nucleotides) and a data matrix for one continuous character, using the (λ_S, ω) model with arithmetic averages as the simulation model. To avoid numerical problems and unrealistic substitution rates, the replicates for which one of the branches had a length or a value of $\omega > 5$ were discarded. The simulated data sets were then analyzed under the (λ_S, ω) model. To test the effects of the approximations due to the finite-time averages taken over branches, we analyzed the data using either the arithmetic or the geodesic averaging schemes. For each replicate, and each covariance parameter of the 3×3 covariance matrix, the 95% credibility interval obtained under each model was compared with the true value.

In theory, when the replicates have been simulated and analyzed under the same model, and with the same prior, the Bayesian credibility intervals have a simple frequentist interpretation, namely, that 5% of the true values are expected to fall outside the 95% credibility intervals (Huelsenbeck and Rannala 2004). In the present case, because the simulation and estimation model are different, and because we condition the simulations, but not the analysis, on a fixed, predefined, chronogram and a fixed mutation matrix, we do not expect this property to strictly hold. On the other hand, as long as the parameters that are fixed across simulations are not too atypical a priori, the frequentist property is expected to hold approximately. Another point of interest is how far from the true value the mean estimated covariances will be, which can be checked visually.

In both cases, using either the arithmetic or the exponential averaging method, a strong correlation between the true and the estimated covariances is observed (fig. 1). The choice between linear or geodesic averaging seems to have a rather small influence on the estimation (compare fig. 1*B*, *D*, *F* with *A*, *C*, *E*), although the geodesic method appears to be slightly more accurate. Specifically, the 95% credibility intervals encompass the true value except in 10% of the cases (16 out of 150 estimated covariances) when using the arithmetic average and 6% of the cases (9 out of 150) when using the geodesic average.

A second series of simulations was aimed at assessing the rate of false positives of the method. When estimating a covariance matrix on a true data set, one naturally wants to assess how confident to be about the fact that the covariance

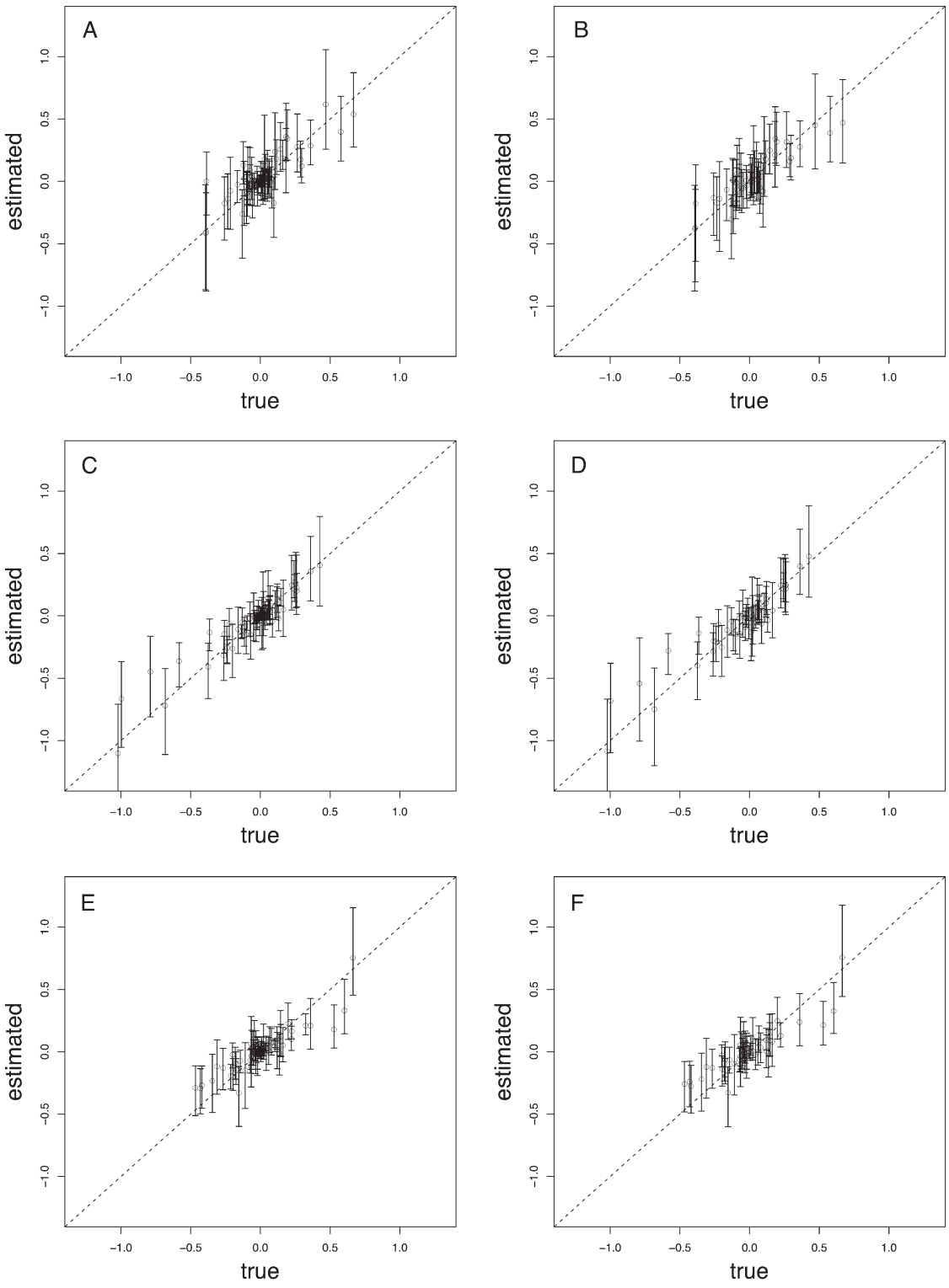

FIG. 1. Comparison between true value (x axis), posterior mean and 95% credibility interval (y axis) for the three covariance parameters of the model (A, B: $\langle \lambda_S, \lambda_N \rangle$, C, D: $\langle \lambda_S, C_1 \rangle$, E, F: $\langle \lambda_N, C_1 \rangle$). A, C, E: arithmetic averages, B, D, F: geodesic averages (see text for details).

between two parameters of interest is indeed positive (or negative). In a Bayesian framework, the pp that the covariance between the two parameters of interest is positive is supposed to measure exactly this confidence. Note that, by symmetry, the prior probability of a positive covariance is 0.5, and therefore, the model does not a priori favor any particular direction.

In principle, the pp is not to be interpreted in frequentist terms, that is, 1 — pp is not supposed to be an equivalent of the P value of a frequentist test in which the null hypothesis would be that the covariance is in fact equal to zero. Nevertheless, it is natural to expect that the method does not produce false positives too often, that is, does not often give a high pp for a positive or a negative covariance, when

Table 1. Rate of False Positives.[a]

	α				
Averaging Method	**0.100**	**0.050**	**0.010**	**0.001**	**0.0001**
Arithmetic	0.050	0.022	0.002	0.001	0.000
Geodesic	0.049	0.021	0.000	0.000	0.000

[a]Frequency, over 100 simulations under the diagonal model at which the posterior probability of a positive covariance is less than $\alpha/2$ or greater than $1-\alpha/2$ (see text for details).

applied to data that have in fact been simulated under a null covariance model.

To assess this on a more empirical ground, we first estimated the parameters of the diagonal model (i.e., with all covariances set to 0) on the carnivore data set and with the three continuous life-history traits (generation time, mass, and longevity). We then resimulated data under the posterior predictive distribution, that is, we simulated 100 replicates of the data set, each replicate consisting of a codon alignment of 342 coding positions (1,146 aligned nucleotides) and a set of continuous phenotypic characters always under the assumption of no correlation between the $M = 5$ components of the process. Next, we applied the fully covariant model on each replicate and measured the pp of a positive covariance between each $M(M-1)/2 = 10$ pairs of entries of the multivariate process. In this way, we can assess the frequency at which pps are more extreme than a given threshold. Because we do not have any prior expectation about the sign of the covariance, for a given threshold α, we measure the frequency at which either $\text{pp} > 1-\alpha/2$ or $\text{pp} < \alpha/2$.

The results are presented in table 1 for several values of α. Whether the data are simulated and tested under the same model or whether different approximation schemes are used for simulation and analysis, the test, as seen in a frequentist perspective, seems slightly conservative (i.e., the rate of false positives at the α level appears to be less than α). The specific approximation scheme does not seem to have a strong impact on the behavior of the test. A point of great practical importance is that, for a very low threshold ($\alpha = 0.0001$), no false positives were seen among the 100 replicates, thus for all 1,000 covariances tested. This means that, if anything, the method does not seem to result in apparently strongly significant, albeit in fact spurious, correlations. Altogether, although more extensive simulations and more definitive theoretical results would probably be needed to add further weight to this conclusion, the present empirical analysis suggests that we can be confident in the pps associated with the observed correlations.

Results

To illustrate the method, we applied it to two alignments of cytochrome b sequences of 67 carnivores and 410 therian mammals (Nabholz et al. 2008). The phenotypic or life-history characters were generation time, mass, and longevity, and the substitution parameters were the rates of synonymous substitution λ_S and the ratio of nonsynonymous over synonymous substitution ω.

Covariance Analysis

The estimated covariance matrix is reported in table 2 together with the correlation coefficients and the pp for each nondiagonal entry to be positive.

In therians, mass, generation time, and longevity are strongly and positively correlated with each other ($\text{pp} > 0.99$). The rate of synonymous substitution λ_S is negatively correlated with mass ($\text{pp} < 0.01$) and with longevity ($\text{pp} = 0.01$). No correlation is observed with generation time ($\text{pp} = 0.30$). Similarly, ω is positively correlated with mass ($\text{pp} > 0.99$), with longevity ($\text{pp} = 0.99$), but again not with generation time ($\text{pp} = 0.35$).

Table 2. Covariance Analysis for Carnivores (left) and for Therians (right) under the (λ_S, ω) Parameterization.[a]

	Carnivores					Therians				
Covariance	λ_S	ω	**Maturity**	**Mass**	**Longevity**	λ_S	ω	**Maturity**	**Mass**	**Longevity**
λ_S	0.93	−0.25	−0.01	0.08	−0.06	0.59	−0.15	−0.03	−0.30*	−0.07*
ω	—	1.09	0.28	0.90*	0.13	—	1.02	−0.03	0.58*	0.13*
Maturity	—	—	0.98	0.95*	0.18*	—	—	0.81	0.77*	0.19*
Mass	—	—	—	4.31	0.38*	—	—	—	4.54	0.61*
Longevity	—	—	—	—	0.31	—	—	—	—	0.34
Correlation	λ_S	ω	**Maturity**	**Mass**	**Longevity**	λ_S	ω	**Maturity**	**Mass**	**Longevity**
λ_S	—	−0.24	−0.01	0.04	−0.11	—	−0.19	−0.04	−0.18*	−0.16*
ω	—	—	0.24	0.41*	0.23	—	—	−0.03	0.27*	0.22*
Maturity	—	—	—	0.46*	0.33*	—	—	—	0.40*	0.37*
Mass	—	—	—	—	0.33*	—	—	—	—	0.49*
Posterior Prob.[b]	λ_S	ω	**Maturity**	**Mass**	**Longevity**	λ_S	ω	**Maturity**	**Mass**	**Longevity**
λ_S	—	0.11	0.47	0.60	0.21	—	0.02	0.30	<0.01*	0.01*
ω	—	—	0.93	0.99*	0.94	—	—	0.35	>0.99*	0.99*
Maturity	—	—	—	>0.99*	>0.99*	—	—	—	>0.99*	>0.99*
mass	—	—	—	—	>0.99*	—	—	—	—	>0.99*

[a]Covariances estimated using the geodesic averaging procedure, and $\kappa = 10$. Asterisks indicate a posterior probability of a positive covariance smaller than 0.025 or greater than 0.975.
[b]Posterior probability of a positive covariance.
*Posterior probability >0.975 or <0.025.

Table 3. Covariance Analysis for Therians, under the (λ_S, ω) Parameterization and using Fossil Calibrations.[a]

	Therians				
Covariance	λ_S	ω	**Maturity**	**Mass**	**Longevity**
λ_S	0.77	−0.21*	−0.04	−0.40*	−0.09*
ω	—	1.07	−0.04	0.66*	0.16*
Maturity	—	—	0.99	0.90*	0.22*
Mass	—	—	—	5.23	0.69*
Longevity	—	—	—	—	0.39
Correlation	λ_S	ω	**Maturity**	**Mass**	**Longevity**
λ_S	—	−0.24*	−0.05	−0.20*	−0.16*
ω	—	—	−0.04	0.28*	0.25*
Maturity	—	—	—	0.40*	0.36*
Mass	—	—	—	—	0.48*
Posterior Prob.[b]	λ_S	ω	**Maturity**	**Mass**	**Longevity**
λ_S	—	0.01*	0.27	<0.01*	0.01*
ω	—	—	0.33	>0.99*	0.99*
Maturity	—	—	—	>0.99*	>0.99*
Mass	—	—	—	—	>0.99*

[a]Covariances estimated using the geodesic averaging procedure, and $\kappa = 10$. Asterisks indicate a posterior probability of a positive covariance smaller than 0.025 or greater than 0.975.
[b]Posterior probability of a positive covariance.
*Posterior probability >0.975 or <0.025.

In carnivores ω is also correlated with mass (pp $>$ 0.99), marginally with longevity (pp $=$ 0.94) and, unlike in therians, marginally also with generation time (pp $=$ 0.93). On the other hand, in carnivores, λ_S does not seem to correlate with any of the three life-history traits (table 2). Using either the geodesic or the arithmetic averaging procedure or using $\kappa = 1$ or $\kappa = 10$ for the inverse Wishart prior did not seem to have any influence on the inference (not shown).

Using fossil calibrations, in the case of therians, led to a global enhancement of the estimated covariance matrix (table 3). In particular, the variance per unit of time of λ_S is larger by nearly 50%, which clearly indicates that the variations of the mutation rate in mitochondrial DNA are underestimated when divergence dates are not properly calibrated as previously suggested (Nabholz et al. 2008). Interestingly, the calibrated analysis also yields a significantly negative correlation between λ_S and ω, which was not observed in the analysis without calibrations. All other estimates are very similar, whether or not calibrations are used (table 3).

An analysis was also conducted under the (λ_S, λ_N) parameterization (table 4). The results are concordant with those obtained under the (λ_S, ω) parameterization, that is, λ_S does not correlate with life-history traits and λ_N correlates with mass and marginally with longevity and generation time in carnivores. In therians, a negative correlation between λ_S and mass and longevity is recovered. As for λ_N, it shows a marginal positive correlation with mass and longevity. Of interest, λ_S and λ_N are found to be positively correlated in therians (pp $=$ 0.99) and marginally in carnivores (pp $=$ 0.92).

Some of the methods of standard linear regression and analysis of variance have a direct equivalent in the present case. In particular, the slope of the pairwise relation between two variables can be estimated (see Methods). For instance, in the case of therians, the slope of the logarithmic variations of generation time versus mass is estimated at 0.20, with a 95% credibility interval (95% CI) at [0.16,0.25]. In the case of longevity as a function of mass, we obtain 0.14 (95% CI [0.11,0.17]). The estimated slopes were very similar, with or without calibrations, under $\kappa = 1$ or 10, and using the arithmetic or the geodesic averaging method. They are smaller than the coefficients of 0.25 and 0.20 often reported for these allometric scaling relations (Calder 1984). On the other hand, a direct linear regression on the life-history traits of the 410 therian taxa yields a slope of 0.22 for generation time versus mass and of 0.17 for longevity versus mass, which suggests that the discrepancy may come from

Table 4. Covariance Analysis for Carnivores and Therians under the (λ_S, λ_N) Parameterization.[a]

	Carnivores					Therians				
Covariance	λ_S	λ_N	**Maturity**	**Mass**	**Longevity**	λ_S	λ_N	**Maturity**	**Mass**	**Longevity**
λ_S	1.04	0.29	−0.03	0.07	−0.07	0.62	0.30*	−0.02	−0.32*	−0.08*
λ_N	—	1.13	0.26	0.91*	0.08	—	1.18	−0.05	0.28	0.06
Maturity	—	—	0.98	0.94*	0.18*	—	—	0.82	0.78*	0.20*
Mass	—	—	—	4.31	0.38*	—	—	—	4.56	0.61*
Longevity	—	—	—	—	0.31	—	—	—	—	0.34
Correlation	λ_S	λ_N	**Maturity**	**Mass**	**Longevity**	λ_S	λ_N	**Maturity**	**Mass**	**Longevity**
λ_S	—	0.27	−0.03	0.03	−0.13	—	0.35	−0.03	−0.19*	−0.17*
λ_N	—	—	0.25	0.41*	0.13	—	—	−0.05	0.12	0.09
Maturity	—	—	—	0.46*	0.33*	—	—	—	0.40*	0.37*
Mass	—	—	—	—	0.33*	—	—	—	—	0.49*
Posterior Prob.[b]	λ_S	λ_N	**Maturity**	**Mass**	**Longevity**	λ_S	λ_N	**Maturity**	**Mass**	**Longevity**
λ_S	—	0.92	0.44	0.58	0.17	—	0.99*	0.34	<0.01*	<0.01*
λ_N	—	—	0.93	0.99*	0.81	—	—	0.29	0.95	0.88
Maturity	—	—	—	>0.99*	0.99*	—	—	—	>0.99*	>0.99*
Mass	—	—	—	—	>0.99*	—	—	—	—	>0.99*

[a]Covariances estimated using the geodesic averaging procedure, and $\kappa = 10$. Asterisks indicate a posterior probability of a positive covariance smaller than 0.025 or greater than 0.975.
[b]Posterior probability of a positive covariance.
*Posterior probability >0.975 or <0.025.

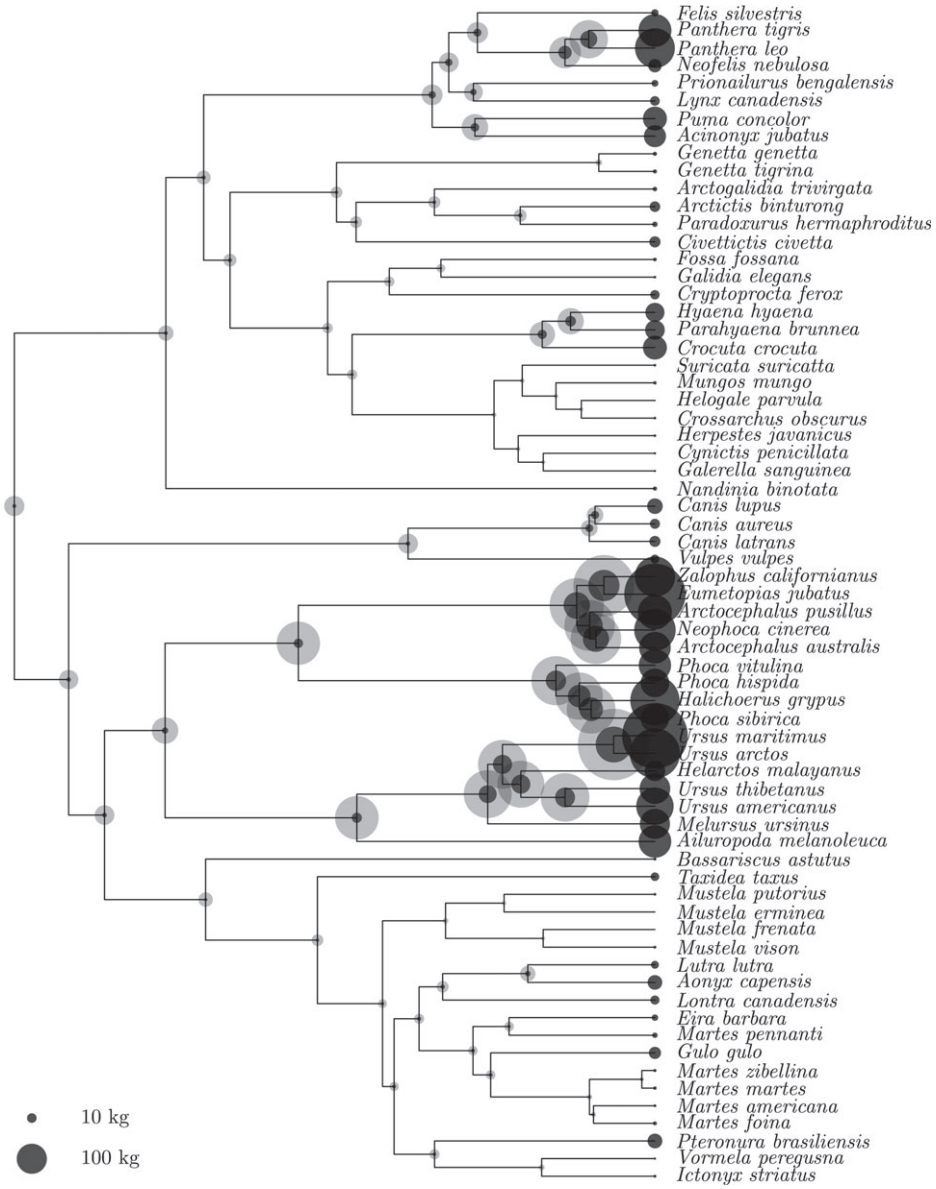

FIG. 2. Reconstruction of the evolution of body mass in carnivores. Disk area is proportional to body mass. Boundaries of the 95% credibility interval at each node are represented by the dark- and light-shaded disks.

the particular taxonomic level or sampling presently considered and not from the model.

Another quantity of interest is the square of the correlation coefficient between two components r_{kl}^2, which expresses how much of the total variation of l is explained by k and vice versa. Here, the squared correlation coefficient is $(-0.18)^2 = 0.032$ between λ_S and mass, and $(-0.27)^2 = 0.073$ between ω and mass (table 2). In other words, the variations of body weight explain 3% of the variations of the rate of substitution and 7% of the variations of ω. Mass and longevity are the most strongly correlated characters, explaining 25% of the variation. In therians, the correlation coefficient between λ_S and λ_N (table 4) is 0.35, that is, the rate of synonymous substitution explains $(0.35)^2 = 12\%$ of the variations in the rate of nonsynonymous substitutions.

Multiple regression analysis can also be used (see Methods), for instance, to attempt to discriminate between mass and longevity as the primary factor correlating with λ_S (table 4). Under constant mass, no residual correlation is observed between longevity and λ_S (pp $= 0.16$). In contrast, under constant longevity, the covariance between λ_S and mass is still significantly negative (pp $= 0.02$). Thus, according to the present analysis, mass, and not longevity, seems to be the main explanatory variable for substitution rate variations in therians.

Phylogenetic Reconstruction of the Phenotypes

The divergence times and the variations in body mass along the phylogeny were reconstructed under the fully covariant model for carnivores (fig. 2) and for therians (not shown). Overall, the 95% credibility intervals are large. Specifically, the common ancestor of carnivores is inferred to have an adult body mass between 1.8 and 24.7 kg. From there, the evolutionary trends are very different, depending on

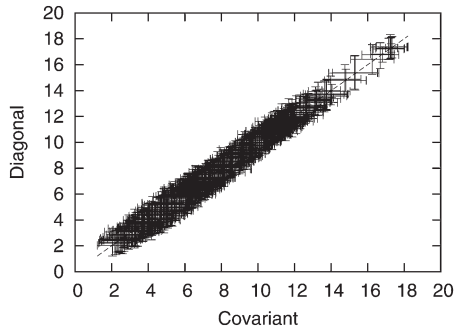

FIG. 3. Comparison between inferred ancestral masses under the covariant (x axis) and the diagonal (y axis) model for the therian data set. Error bars correspond to the marginal 95% credibility intervals at each node.

the suborders considered: mass progressively increases in the lineage leading to Ursidae and Pinnipedia, being estimated at 46 (95% credibility interval [11,8,210]) kg in their common ancestor, and decreases to <1 kg in Herpestidae (mongooses).

To measure the impact of potential interactions between molecular and phenotypic reconstructions through the covariance matrix, an inference was also conducted under the diagonal model, obtained by constraining the nondiagonal entries of the matrix to be equal to zero. To quantify the differences between the two reconstructions, we computed, for each node, a deviation index as follows:

$$z^j = \frac{2(E_{\mathrm{cov}}[\ln C^j] - E_{\mathrm{diag}}[\ln C^j])}{\sqrt{V_{\mathrm{cov}}[\ln C^j]} + \sqrt{V_{\mathrm{diag}}[\ln C^j]}}, \qquad (14)$$

where $E[\cdot]_{\mathrm{cov}}$ and $E[\cdot]_{\mathrm{diag}}$ are the sample means, $V[\cdot]_{\mathrm{cov}}$ and $V[\cdot]_{\mathrm{diag}}$ the sample variances, under the covariant and the diagonal models, and C^j is the value reconstructed at node j for the character of interest. This deviation is loosely analogous to a z-score, although it is not meant as a measure of significance, but only as a heuristic measure of the difference between the estimates obtained under the two models.

The differences between the reconstructions inferred under the covariant and the diagonal models are small (fig. 3). In therians, the first ten highest deviation indices are all positive, between 0.9 and 1, and all of them fall in the group of Cricetidae. Thus, there is a signal in the multiple sequence alignment indicating that early Cricetidae may have been larger than what is inferred just based on the phenotypic data and the phylogenetic tree. For instance, the ancestor of Cricetidae is inferred to have a mass of 151 (95% CI [58,376]) grams under the covariant model instead of 97 (95% CI [32,279]) grams under the diagonal model. An opposite trend is observed in carnivores, with a maximum deviation index of -0.5 for the ancestor of Ursidae, suggesting that, in this case, covariance between substitution rates and body mass results in a downward correction of body mass for the ancestor of Ursidae. Likewise, the most recent common ancestor of Ursidae and Pinnipedia, which comes third ($z = -0.4$), is inferred with a mass of 46 (95% CI [11,8210]) kg under the covariant model instead of the 73 (95% CI [18,317]) kg found under the diagonal model. In all cases, however, the differences between the covariant and the diagonal model are small compared with the credibility intervals, and may just as well be a stochastic fluctuation, or a consequence of inaccurate divergence time reconstruction.

Discussion

Comparative analyses of molecular and phenotypic characters are a key aspect of molecular evolutionary studies. In this direction, what we propose here is the first fully integrated method dealing with the nuisances caused by phylogenetic dependences and by the various sources of uncertainty about the phenotypic and molecular history.

Probably, the most immediate advantage of the method developed here is its practical simplicity. Essentially, the entire procedure reduces to a one-step analysis in which all the available evidence is given as an input and estimates of all potentially interesting aspects of the problem (covariances, divergence times, phenotypic histories) are obtained as the output. The pps offer a simple and natural method for evaluating the significance of the observed correlations.

Joint estimation of all parameters in a Bayesian framework has another important advantage. When estimating a parameter of interest, in particular the covariance matrix, the uncertainty about all other parameters (e.g., on reconstructed rates of substitution or on divergence times) is automatically accounted for (Huelsenbeck et al. 2000; Huelsenbeck and Rannala 2003; Gelman et al. 2004). This is particularly important in analyses of single genes such as cytochrome b for which the sampling error associated with the estimation of substitution rates is potentially large. Integrating over divergence times may also be important, as errors on branch lengths have been shown to result in inflated type I errors (Diaz-Uriarte and Garland 1998).

On the other hand, accounting for the uncertainty about the nuisance parameters by integrating over the prior raises the issue of prior sensitivity. In the present case, one can point out at least two components of the model for which prior sensitivity may be an issue. First, divergence times are potentially more sensitive to the prior (Inoue et al. 2010) than previously suggested (Lepage et al. 2007), indicating that the present framework should be extended to accommodate alternative priors on divergence dates, in particular the birth–death prior (Yang and Rannala 2006), and that the robustness of the analysis to the choice of the prior should be thoroughly assessed.

Another point concerning prior sensitivity is the choice of κ, the prior mean variance parameter for each component of the multivariate process. Defining a sensible value for κ is particularly difficult because we have absolutely no relevant prior information about the scale of the rate of change of the substitution parameters nor of the phenotypic characters. In the present case, we just checked that the analysis was robust with respect to the choice of $\kappa = 1$ or 10, which is probably good enough given that the posterior mean values obtained for the variance parameters are within this range (table 2, diagonal coefficients). However, this approach is not totally satisfactory, conceptually

Table 5. Multiple Regression Analysis in Therians.[a]

	Constant Mass					Constant Longevity				
Covariance	λ_S	ω	**Maturity**	**Mass**	**Longevity**	λ_S	ω	**Maturity**	**Mass**	**Longevity**
λ_S	0.74	−0.17	0.03	—	−0.04	0.75	−0.18*	0.01	−0.24*	—
ω	—	0.98	−0.15	—	0.07	—	0.99	−0.13	0.37*	—
Maturity	—	—	0.83	—	0.10*	—	—	0.86	0.51*	—
Mass	—	—	—	—	—	—	—	—	4.00	—
Longevity	—	—	—	—	0.30	—	—	—	—	—
Posterior Prob.[b]	λ_S	ω	**Maturity**	**Mass**	**Longevity**	λ_S	ω	**Maturity**	**Mass**	**Longevity**
λ_S	—	0.04	0.67	—	0.16	—	0.02*	0.56	0.02*	—
ω	—	—	0.04	—	0.94	—	—	0.07	0.99*	—
Maturity	—	—	—	—	>0.99*	—	—	—	>0.99*	—

[a] Covariances estimated using the geodesic averaging procedure, fossil calibrations, and $\kappa = 10$. Asterisks indicate a posterior probability of a positive covariance smaller than 0.025 or greater than 0.975.
[b] Posterior probability of a positive covariance.
*Posterior probability >0.975 or <0.025.

speaking. An alternative would be to work in a hierarchical Bayes framework, and use an uninformative prior, such as Jeffreys' prior (Jeffreys 1961), possibly accommodating different values κ_m for each component $m = 1, \ldots, M$ of the process. Another approach, more in the spirit of empirical Bayes, would consist in optimizing the marginal likelihood of the overall model according to the hyperparameter κ (or the hyperparameters κ_m).

More fundamentally, the choice of an inverse Wishart distribution as our prior on the covariance matrix was motivated exclusively by computational arguments. The inverse Wishart is conjugate to the multivariate normal distribution, thus allowing us to integrate away the covariance matrix from the MCMC sampler (see Methods). In practice, the improvement brought by the conjugate sampling method seems to be essentially dependent on the dimension M of the multivariate process, which is expected, given that the number of independent parameters represented by the covariance matrix increases as M^2. Thus, the improvement is minor for $M = 3$ (i.e., two substitution parameters combined with one continuous character), significant for $M = 5$, with a burn-in three times as long under the nonconjugate sampling method than under the conjugate one, and essential for larger values of M: for $M > 10$, we were unable to obtain convergence using the nonconjugate method. This is also true if we increase the dimension of the process by having more substitution parameters allowed to vary along the lineages (not shown).

On the other hand, alternative priors could be imagined. In particular, conditional independence between certain pairs of variables could be modeled more directly, and perhaps more adequately, by allowing the corresponding nondiagonal entries of Σ^{-1} to be equal to zero with positive prior probability. This can be seen as a reformulation, in a comparative context, of covariance selection models (Dempster 1972; Dobra et al. 2004). Reversible jump Monte Carlo methods might have to be developed in order to sample from such models. Alternatively, under certain conditions, it might be possible to develop covariance selection priors while preserving conjugacy (Dawid and Lauritzen 1993; Roverato 2002; Letac and Massam 2007). In both cases, the model would offer a direct estimate of the marginal pp of conditional independence between each pair of variables, and may also have a better fit, thanks to its smaller effective number of parameters.

Apart from the question of the prior, our method makes several assumptions and approximations, all of which may deserve further discussion. First, we approximate the continuously changing substitution process by a piecewise constant process, using average substitution matrices, one for each branch. We have proposed two alternative approximations for these average matrices and checked by simulations that these approximations did not result in large estimation errors. On the other hand, our checks do not offer any guarantee that the approximation will be acceptable in all circumstances. A possibility would be to use less extreme discretization schemes, for instance, by sampling the values of the multivariate process at intermediate points along the branches and not only at their extremities. An acceptable compromise between granularity and computational cost may then be found empirically on a case-by-case basis.

Second, we have assumed that logarithmic transformations would be adequate to reduce the problem to one of estimating linear correlations between variables. In the present case, a logarithmic transformation is probably the most obvious choice to make in the case of life-history traits for which known allometric relations are equivalent to log-linear correlations between variables (Calder 1984). Concerning substitution rates, the case is less obvious, although the fact that rates can display variations on several orders of magnitude (Nabholz et al. 2008) strongly argues in favor of a change of variable akin to a logarithmic transformation. In principle, alternative transformations of the variables could be proposed and compared by computing Bayes factors, or alternatively, could be averaged over using a hierarchical model. An even more advanced approach would consist in developing nonparametric methods able to estimate the transformation directly from the data.

Finally, the assumptions implied by the use of a Brownian diffusion process, namely a constant rate of change, and an absence of trend in the direction of the changes could be

relaxed by implementing alternative stochastic processes, such as the Ornstein–Uhlenbeck process (Butler and King 2004), or burst models, allowing for a varying rate of phenotypic evolution in different regions of the tree (Cooper and Purvis 2010).

Insights into Molecular Evolutionary Mechanisms

We have introduced two alternative parameterizations of the covariance model in terms of either λ_S and ω or λ_S and λ_N. Ideally, because the process is multivariate normal, and because $\ln\omega = \ln\lambda_N - \ln\lambda_S$ is a log-linear relation preserving the normality of the process, the two representations should be equivalent, the two covariance matrices being related by the following change of variables:

$$\begin{aligned}\langle C,\omega\rangle &= \langle C,\lambda_N\rangle - \langle C,\lambda_S\rangle,\\ \langle \lambda_S,\omega\rangle &= \langle \lambda_S,\lambda_N\rangle - \langle \lambda_S,\lambda_S\rangle,\\ \langle \omega,\omega\rangle &= \langle \lambda_S,\lambda_S\rangle + \langle \lambda_N,\lambda_N\rangle - 2\langle \lambda_S,\lambda_N\rangle,\end{aligned}$$

where C is any phenotypic character. On the other hand, because the prior on Σ is not invariant by this change of variable, for finite data, the result of the estimation will depend on the chosen representation. Which representation is more convenient depends on the question being addressed. The prior is centered on the diagonal model and thus is neutral with respect to positive or negative covariance among the substitution parameters and the phenotypic characters. Therefore, the choice should mainly depend on which variables we consider as a priori independent.

The justification of the (λ_S,ω) parameterization is mechanistic. Assuming that selection on synonymous substitutions is negligible, variations of λ_S will mostly be due to variations of the mutation rate λ and will be independent of population size. On the other hand, if the mutation rate is not too high, so that interferences between nonneutral polymorphisms are negligible, ω will be independent of the mutation rate λ and will be equal to the fraction of effectively neutral nonsynonymous mutations. This fraction is expected to be a decreasing function of effective population size as slightly deleterious mutations that would otherwise be nearly neutral in species with a small effective size may find themselves purified away in species with a larger effective size (Ohta 1973, 1974; Kimura 1979, 1983; Welch, Eyre-Walker, et al. 2008). Using the (λ_S,ω) parameterization therefore amounts to assuming that such a nearly neutral model applies to the data at hand.

On the other hand, if we are more suspicious about the validity of the nearly neutral model, in particular if we suspect that nonsynonymous substitutions may not be limited by the mutation rate, but instead by ecological adaptive opportunities (Gillespie 1991), then considering λ_S and λ_N as the two a priori independent variables may turn out to be more adequate. Using the (λ_S,λ_N) parameterization can also be seen as a way of testing whether λ_S and λ_N are indeed positively correlated, as would be expected under the nearly neutral model, or more generally if the mutation rate is limiting. For those reasons, we think it is important to propose a software program in which the two alternative parameterizations are available.

The positive correlation observed between λ_S and λ_N in our analyses (table 4) is consistent with the nearly neutral model. The correlation is significant, albeit perhaps a bit weak, with the variations of λ_S explaining only 12% of those of λ_N in therians. This may simply be due to a lack of power, owing to the small size of the alignment (small number of positions). Alternatively, it could be the consequence of adaptive phenomena partially decoupling λ_N from λ_S. More extensive analyses, in particular using longer sequences, would be needed here.

Also in favor of the nearly neutral model, we see a positive correlation between ω and mass and longevity, both at the level of one single order (carnivores) and at the more global scale of therians (table 2). This correlation can be interpreted as an indirect effect of variations of effective population size, itself negatively correlated with mass and longevity, in agreement with previous observations (Weinreich 2001; Popadin et al. 2007). What may appear more intriguing is that population size is known to also be correlated with generation time in mammals (Chao and Carr 1993). Yet, in the present case, we do not observe any correlation between ω and generation time in therians, and we see it only marginally in carnivores (table 2).

We also observe a negative correlation between λ_S and mass and longevity, which is consistent with a previous analysis (Nabholz et al. 2008). Two alternative explanations can be proposed for this correlation. First, it could be an indirect effect of metabolism, larger animals having a lower metabolism (Martins and Palumbi 1993; Gillooly et al. 2005). However, several analyses have already questioned the metabolic rate hypothesis, suggesting that the correlation with metabolic rate was at best indirect, being mediated by a body size effect (Bromham et al. 1996; Lanfear et al. 2007). Alternatively, the mutation rate could be under adaptive regulation, linked to the necessity of restricting mitochondrial somatic mutations in large and long-living mammals (Speakman 2005; Nabholz et al. 2008). The question could be further investigated under the present framework, and using multiple regression, to discriminate between mass and metabolic rate.

Finally, we do not observe any correlation between λ_S and generation time, neither in carnivores nor in therians, and whether or not fossil calibrations are used. A strong generation time effect has often been reported previously, but mostly for nuclear sequences (Li and Tanimura 1987; Li et al. 1996). In contrast, the generation time effect was found to be weaker in mitochondrial sequences (Nabholz et al. 2008). Nevertheless, the fact that we could not observe any correlation between either λ_S or ω and generation time, despite the fact that such correlations would be plausible, should probably be further investigated.

Reconstructing Phenotypic Evolution

Reconstructing phenotypic evolution based on a joint analysis of phenotypic and molecular data is one of the most exciting perspectives opened by the present method. Joint estimation implies that potentially relevant information is shared across the different components of the parameter

vector. Via the covariance matrix, a potentially interesting cross talk may therefore occur between substitution rates and divergence times (Welch and Waxman 2008) or between rates and the reconstructed phenotypic history.

In the present case, however, we have not seen much influence of the covariance structure of the model on divergence times nor on phenotypic reconstructions. This could be due to several factors, although a likely explanation in the present case is indicated by the squared correlation coefficients estimated for this data set (tables 2 and 3). Because λ_S explains only around 3%, and ω about 7%, of the variations of mass and longevity, we should not expect λ_S and ω to have a strong influence on the phenotypic reconstruction. Such a weak coupling between rates and life-history traits could be an intrinsic property of the evolution of mammalian mitochondrial sequences, that is, rates may have other hidden determinants that happen to dominate their overall fluctuations. Alternatively, it could be a consequence of the large uncertainty associated with the estimation of substitution rates, itself a consequence of the small number of aligned positions used in the present study. In the latter case, a comparative analysis conducted with the entire proteome of mammalian mitochondrial genomes should lead to higher correlations and may therefore help reveal significant interaction between rates and phenotypes.

Perspectives

Our observations concerning the pattern of molecular and phenotypic evolution in mammals are very preliminary. Their aim was merely to introduce the method, and it is clear that, in order to draw more definitive conclusions about the several points of biological interest raised above, much more ambitious analyses need to be conducted.

First, not only are the variations of λ_S and of ω for cytochrome b subject to large stochastic errors, due to the small number of aligned positions, but they may also express specific adaptations of cytochrome b in particular lineages. Because variations of life-history traits are expected to have genome-wide consequences on the pattern of molecular evolution, one possible approach would consist in integrating the signal over several genes, so as to average out gene-specific idiosyncrasies and recover only the global trends. The power of comparative analyses also crucially depends on a dense taxonomic sampling, and therefore the two criteria, many genes and many taxa, should ideally be met simultaneously.

Second, mitochondrial sequences are particularly saturated (Brown et al. 1979), owing to a very high mutation rate in mammals, and more generally in metazoans. Saturation has probably eroded a significant proportion of the molecular evolutionary signal in the deeper part of the mammalian tree. Using the less saturated nuclear sequences, and more generally, analyzing several genetic units subject to different evolutionary regimes should significantly increase the power of the comparative approach.

Finally, the method could be extended to many other potentially interesting substitution parameters. For instance, investigating the correlation between the transition–transversion ratio, the equilibrium GC content, and phenotypic characters may help discriminate between alternative hypotheses about the determinants of the mutation pressure or the population genetic mechanisms underlying the substitution process. More generally, the method developed here could in principle be used for investigating a wide diversity of potential correlations between phenotypes and sequences, thereby providing many stimulating empirical observations helping us to better understand the mechanisms of molecular evolution and to reconstruct the evolution of phenotypic and life-history traits.

Acknowledgments

We wish to thank Benoît Nabholz, Sylvain Glémin, and Nicolas Galtier for providing the data, Nicolas Rodrigue, Fredrik Ronquist, and two anonymous reviewers for their useful comments on the manuscript. We also thank the Réseau Québecois de Calcul de Haute Performance for computational resources. This work was funded by the Natural Science and Engineering Research Council of Canada.

References

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

Boussau B, Blanquart S, Lartillot N, Gouy M. 2008. Parallel adaptations to high temperatures in the Archaean eon. *Nature* 456:942–945.

Bromham L, Rambaut A, Harvey PH. 1996. Determinants of rate variation in mammalian DNA sequence evolution. *J Mol Evol.* 43:610–621.

Brown WM, George M Jr, Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci U S A.* 76:1967–1971.

Butler MA, King AA. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am Nat.* 164:683–695.

Calder WA. 1984. Size, function and life history. Cambridge: Harvard University Press.

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

Cooper N, Purvis A. 2010. Body size evolution in mammals: complexity in tempo and mode. *Am Nat.* 175:727–738.

Dawid AP, Lauritzen SL. 1993. Hyper Markov laws in the statistical analysis of decomposable graphical models. *Ann Stat.* 21:1272–1317.

de Koning AP, Gu W, Pollock D. 2010. Rapid likelihood analysis on large phylogenies using partial sampling of substitution histories. *Mol Biol Evol.* 27:249–265.

de Magalhaes JP, Costa J. 2009. A database of vertebrate longevity records and their relation to other life-history traits. *J Evol Biol.* 22:1770–1774.

Dempster AP. 1972. Covariance selection. *Biometrics* 28:157–175.

Diaz-Uriarte R, Garland T Jr. 1998. Effects of branch length errors on the performance of phylogenetically independent contrasts. *Syst Biol.* 47:654–672.

Dirac PAM. 1982. The principles of quantum mechanics. Oxford: Oxford University Press.

Dobra A, Hans C, Jones B, Nevins JR, Yao G, West M. 2004. Sparse graphical models for exploring gene expression data. *J Multivar Anal.* 90:196–212.

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

Felsenstein J. 1985. Phylogenies and the comparative method. *Am Nat.* 125:1–15.

Felsenstein J. 2008. Comparative methods with sampling error and within-species variation: contrasts revisited and revised. *Am Nat.* 171:713–725.
Fontanillas E, Welch JJ, Thomas JA, Bromham L. 2007. The influence of body size and net diversification rate on molecular evolution during the radiation of animal phyla. *BMC Evol Biol.* 7:95.
Galtier N, Tourasse N, Gouy M. 1999. A nonhyperthermophilic common ancestor to extant life forms. *Science* 283:220–221.
Garland TJ, Bennett AF, Rezende EL. 2005. Phylogenetic approaches in comparative physiology. *J Exp Biol.* 208:3015–3035.
Gelman A, Carlin JB, Stern HS, Rubin DB. 2004. Bayesian data analysis. Boca Raton (FL): Chapman and Hall/CRC.
Gillespie JH. 1991. The causes of molecular evolution. Oxford: Oxford University Press.
Gillooly JF, Allen AP, West GB, Brown JH. 2005. The rate of DNA evolution: effects of body size and temperature on the molecular clock. *Proc Natl Acad Sci U S A.* 102:140–145.
Harvey PH, Pagel MD. 1991. The comparative method in evolutionary biology. Oxford: Oxford University Press.
Houseworth EA, Martins EP, Lynch M. 2004. The phylogenetic mixed model. *Am Nat.* 163:84–96.
Huelsenbeck JP, Rannala B. 2003. Detecting correlation between characters in a comparative analysis with uncertain phylogeny. *Evolution* 57:1237–1247.
Huelsenbeck JP, Rannala B. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Syst Biol.* 53:904–913.
Huelsenbeck JP, Rannala B, Masly JP. 2000. Accommodating phylogenetic uncertainty in evolutionary studies. *Science* 288:2349–2350.
Inoue JG, Donogue PCJ, Yang Z. 2010. The impact of the representation of fossil calibration on Bayesin estimation of species divergence times. *Syst Biol.* 59:74–89.
Jeffreys H. 1961. Theory of probability. Oxford: Oxford University Press.
Kimura M. 1979. Model of effectively neutral mutations in which selective constraint is incorporated. *Proc Natl Acad Sci U S A.* 76: 3440–3444.
Kimura M. 1983. The neutral 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.
Lanfear R, Thomas JA, Welch JJ, Brey T, Bromham L. 2007. Metabolic rate does not calibrate the molecular clock. *Proc Natl Acad Sci U S A.* 104:15388–15393.
Lartillot N. 2006. Conjugate Gibbs sampling for Bayesian phylogenetic models. *J Comput Biol.* 13:1701–1722.
Lepage T, Bryant D, Philippe H, Lartillot N. 2007. A general comparison of relaxed molecular clock models. *Mol Biol Evol.* 24:2669–2680.
Letac G, Massam H. 2007. Wishart distributions for decomposable graphs. *Ann Stat.* 35:1278–1323.
Li W-H, Ellsworth DL, Krushkal J, Chang BH-J, Hewett-Emmett D. 1996. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Mol Phylogent Evol.* 5:182–187.
Li W-H, Tanimura M. 1987. The molecular clock runs more slowly in man than in apes and monkeys. *Nature* 326:93–96.
Mardia KV, Kent JT, Bibby JM. 1979. Multivariate analysis. San Diego (CA): Academic Press.
Martins AP, Palumbi SR. 1993. Body size, metabolic rate, generation time, and the molecular clock. *Proc Natl Acad Sci U S A.* 90:4087–4091.
Martins EP, Hansen TF. 1996. The statistical analysis of interspecific data: a review and evaluation of phylogenetic comparative methods. In: Martins E, editor. Phylogenies and the comparative method in animal behavior. Oxford: Oxford University Press. p. 22–75.
Martins EP, Hansen TF. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am Nat.* 149:646–667.
Mateiu L, Rannala B. 2006. Inferring complex dna substitution processes on phylogenies using uniformization and data augmentation. *Syst Biol.* 55:259–269.
Muse SV, Gaut BS. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitutions, with applications to chloroplast genome. *Mol Biol Evol.* 11:715–724.
Nabholz B, Glémin S, Galtier N. 2008. Strong variations of mitochondrial mutation rate across mammals—the longevity hypothesis. *Mol Biol Evol.* 25:120–130.
Nielsen R. 2002. Mapping mutations on phylogenies. *Syst Biol.* 51:729–739.
Ohta T. 1973. Slightly deleterious mutant substitutions in evolution. *Nature* 252:315–354.
Ohta T. 1974. Mutational pressure as the main cause of molecular evolution and polymorphisms. *Nature* 252:351–354.
Ohta T. 1993. An examination of the generation-time effect on molecular evolution. *Proc Natl Acad Sci U S A.* 90:10676–10680.
Pagel M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877–884.
Popadin K, Polishchuk LV, Mamirova L, Knorre D, Gunbin K. 2007. Accumulation of sightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals. *Proc Natl Acad Sci U S A.* 104:13390–13395.
Rannala B, Yang Z. 2007. Inferring speciation times under an episodic molecular clock. *Syst Biol.* 56:453–466.
Rodrigue N, Philippe H, Lartillot N. 2008. Uniformization for sampling realizations of Markov processes: applications to Bayesian implementations of codon substitution models. *Bioinformatics* 24:56–62.
Roverato A. 2002. Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Scand J Stat.* 29:391–411.
Seo TK, Kishino H, Thorne JL. 2004. Estimating absolute rates of synonymous and nonsynonymous nucleotide substitution in order to characterize natural selection and date species divergences. *Mol Biol Evol.* 21:1201–1213.
Speakman JR. 2005. Correlations between physiology and lifespan—two widely ignored problems with comparative studies. *Aging Cell* 4:167–175.
Thomas JA, Welch JJ, Woolfit M, Bromham L. 2006. There is no universal molecular clock for invertebrates, but rate variation does not scale with body size. *Proc Natl Acad Sci U S A.* 103:7366–7371.
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.
Weinreich DM. 2001. The rates of molecular evolution in rodent and primate mitochondrial DNA. *J Mol Evol.* 52:40–50.
Welch JJ, Bininda-Edmonds ORP, Bromham L. 2008. Correlates of substitution rate variation in mammalian protein-coding sequences. *BMC Evol Biol.* 8:53.
Welch JJ, Eyre-Walker A, Waxman D. 2008. Divergence and polymorphism under the nearly neutral theory of molecular evolution. *J Mol Evol.* 67:418–426.
Welch JJ, Waxman D. 2008. Calculating independent contrasts for the comparative study of substitution rates. *J Theor Biol.* 251: 667–678.
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.