Reconstructing the Phylogenetic History of Long-Term Effective Population Size and Life-History Traits Using Patterns of Amino Acid Replacement in Mitochondrial Genomes of Mammals and Birds

Benoit Nabholz1,*,†, Nicole Uwimana2,†, and Nicolas Lartillot2,3,*

1Institut des Sciences de l’Evolution, UMR 5554 CNRS, Université Montpellier II, France
2Département de Biochimie, Centre Robert Cederberg, Université de Montréal, Québec, Canada
3Laboratoire d’Informatique, de Robotique et de Microélectronique de Montpellier, UMR 5506, CNRS-Université de Montpellier 2, France
†These authors contributed equally to this work.
*Corresponding author: E-mail: bnabholz@univ-montp2.fr; nicolas.lartillot@umontreal.ca.
Accepted: May 20, 2013

Abstract

The nearly neutral theory, which proposes that most mutations are deleterious or close to neutral, predicts that the ratio of nonsynonymous over synonymous substitution rates (dN/dS), and potentially also the ratio of radical over conservative amino acid replacement rates (Kr/Kc), are negatively correlated with effective population size. Previous empirical tests, using life-history traits (LHT) such as body-size or generation-time as proxies for population size, have been consistent with these predictions. This suggests that large-scale phylogenetic reconstructions of dN/dS or Kr/Kc might reveal interesting macroevolutionary patterns in the variation in effective population size among lineages. In this work, we further develop an integrative probabilistic framework for phylogenetic covariance analysis introduced previously, so as to estimate the correlation patterns between dN/dS, Kr/Kc, and three LHT, in mitochondrial genomes of birds and mammals. Kr/Kc displays stronger and more stable correlations with LHT than does dN/dS, which we interpret as a greater robustness of Kr/Kc, compared with dN/dS, the latter being confounded by the high saturation of the synonymous substitution rate in mitochondrial genomes. The correlation of Kr/Kc with LHT was robust when controlling for the potentially confounding effects of nucleotide compositional variation between taxa. The positive correlation of the mitochondrial Kr/Kc with LHT is compatible with previous reports, and with a nearly neutral interpretation, although alternative explanations are also possible. The Kr/Kc model was finally used for reconstructing life-history evolution in birds and mammals. This analysis suggests a fairly large-bodied ancestor in both groups. In birds, life-history evolution seems to have occurred mainly through size reduction in Neoavian birds, whereas in placental mammals, body mass evolution shows disparate trends across subclades. Altogether, our work represents a further step toward a more comprehensive phylogenetic reconstruction of the evolution of life-history and of the population-genetics environment.

Key words: effective population size, life-history evolution, Markov chain Monte Carlo, Bayesian statistics, mitochondrial genome, nearly neutral model.

Introduction

Effective population size is a fundamental parameter of population genetics, shaping the structure of polymorphism within populations, and influencing patterns of divergence. Within populations, effective population size determines the force of genetic drift, and therefore influences the level of genetic diversity (Kimura 1983). In the long time scale, population size influences the probability of fixation of nonneutral mutations with small fitness effects (Ohta 1992). In particular, slightly deleterious mutations may behave as effectively neutral in small populations and, in contrast, be efficiently eliminated in large populations. This differential fixation probability could lead to variation in the substitution rate as a function of long-term effective population size (Ne; Ohta 1992).

Historically, the class of slightly deleterious mutations was introduced in the context of the nearly neutral theory (Ohta 1992).
1973), itself proposed as an extension of the strict neutral theory (Kimura 1968) including mutations with small fitness effects (Ohta 1992). The nearly neutral theory was originally invoked to explain why the correlation between substitution rate and generation time was more pronounced for synonymous substitutions than for nonsynonymous substitutions. The rationale was that an inverse relationship between population size and generation time could lead to a mutual compensation between the decrease in the mutation rate, and the increase in the fixation probability of slightly deleterious nonsynonymous mutations, in species with long generation times but small effective population sizes (Chao and Carr 1993; Ohta 1993). The presence of slightly deleterious mutations was subsequently identified as the likely reason for the excess of nonsynonymous polymorphism, compared with divergence, in mitochondrial protein-coding genes (Nachman 1998; Rand and Kann 1998; Gerber et al. 2001). Similarly, by comparing levels of polymorphism and divergence between species, Eyre-Walker et al. (2002) and Keightley et al. (2005) suggested that Hominidae may have experienced reduced selective constraints on their protein-coding genes compared with Muroidea. These results were interpreted in the light of the nearly neutral theory, based on the fact that hominids are believed to have a smaller long-term effective population size ($N_e$) than murids (Eyre-Walker et al. 2002; Keightley et al. 2005).

At larger evolutionary scale, several studies have attempted to explain between-species differences in the ratio of nonsynonymous ($d_N$) over synonymous ($d_S$) substitution rates by variation in population size. In particular, Johnson and Seger (2001) and Woolfit and Bromham (2005) have found that island birds species have, on average, a higher $d_N/d_S$ ratio than their mainland relatives. Woolfit and Bromham (2003) reported an increase in the substitution rate in the 16S rRNA gene of endosymbiotic organisms, compared with free living species. Domesticated populations, which probably experienced a severe bottleneck during the domestication process, frequently show an increased $d_N/d_S$ compared with wild progenitor species (Bjørnerfeldt et al. 2006; Wang et al. 2011). Finally, Popadin et al. (2007), Nikolaev et al. (2007), Lartillot and Delsuc (2012), and Romiguier et al. (2013) have found a positive correlation between body mass (or age at sexual maturity) and the $d_N/d_S$ ratio in nuclear and mitochondrial protein-coding genes of placental mammals. Relying on the well-known negative relationship between body mass and population density in mammals (Damuth 1987; White et al. 2007), they interpreted this correlation in terms of a difference in long-term effective population size between small and large-bodied mammals and, therefore, as empirical evidence in support of the nearly neutral theory. These results are particularly interesting because they suggest that long-term trends in life-history evolution across mammalian orders result in correlated changes in selection efficacy between groups and leave a distinctive trace at large evolutionary scale.

Currently, however, this result is yet to be replicated in another group of animals. Although less obvious, a relationship between population density and body mass also exists in birds (e.g., Nee et al. 1991) and, therefore, a relationship between life-history traits (LHT) and the efficacy of selection is also expected in this group. To date, evidence that the efficacy of selection could vary between birds is equivocal. Analyzing approximately 8,400 protein-coding genes, Nam et al. (2010) have failed to identify any difference in the average $d_N/d_S$ estimated in the chicken ($Gallus gallus$) lineage (which leads to the relatively large birds of the Galloanserae clade), compared with the zebra finch ($Taeniopygia guttata$) lineage (leading to the relatively small birds of the Passeriformes clade). Similarly, there is some debate as to whether island bird populations are more prone to accumulating slightly deleterious substitutions, compared with their mainland relatives (Johnson and Seger 2001; Woolfit and Bromham 2005; Wright et al. 2009).

The correlation between LHT and $d_N/d_S$ also suggests that large-scale phylogenetic reconstruction of $d_N/d_S$ might reveal long-term trends in effective population size, which in turn could be interpreted in terms of macroevolutionary patterns in life-history evolution across large phylogenetic groups. Currently, methods for reconstructing life-history or quantitative trait evolution along phylogenies (e.g., Venditti et al. 2011) do not directly use the substitution process. Yet, the correlation between substitution patterns and LHT suggests that potentially useful information about ancestral LHT is contained in molecular sequences (Lanfear et al. 2010) and can be integrated into methods for quantitative trait reconstruction (Lartillot and Delsuc 2012).

In practice, however, such integrated reconstruction methods require accurate estimation of ancestral substitution rates, both synonymous and nonsynonymous. Yet, this becomes increasingly difficult with increasing phylogenetic divergence. The accumulation of multiple substitutions at the same site over time, a process called mutational saturation, can easily erase the phylogenetic signal (Springer et al. 2001; Galewski et al. 2006) and lead to an underestimation of substitution rates. As synonymous substitutions accumulate more rapidly than nonsynonymous substitutions, this will lead to a differential saturation of $d_S$ and $d_V$, thus potentially resulting in biased estimates of their ratio, particularly along branches corresponding to long periods of time, ancient evolutionary lineages or fast evolving species. This problem is especially pronounced in the case of the fast evolving mitochondrial genomes (Ballard and Whitlock 2004; Galtier et al. 2006; Nabholz et al. 2008). On the other hand, using mitochondrial genomes might be interesting, because, as opposed to nuclear genomic sequences, the animals mitochondrial genome is largely free from recombination (Elson and Lightowlers 2006; Galtier, Nabholz, et al. 2009). In nuclear genomes, recombination induces biased-gene conversion phenomena (Duret and Galtier 2009), which can potentially impact $d_N$.
Materials and Methods

The Amino Acid Model

The framework introduced in Lartillot and Poujol (2011) is based on a codon model specified according to the formalism of Muse and Gaut (1994). Specifically, the rate of substitution between codons \( c_1 \) and \( c_2 \), differing at only one position, and with respective nucleotides \( n_1 \) and \( n_2 \) at that position, is given by the following equation:

\[
Q_{c_1, c_2} = \lambda R_{n_1, n_2} \text{ if synonymous,}
\]
\[
Q_{c_1, c_2} = \lambda R_{n_1, n_2} \omega \text{ if non-synonymous,}
\]

where \( R \) is a 4 \( \times \) 4, normalized, general time-reversible nucleotide substitution process. Substitution rates between non-nearest neighbor codons are equal to 0. This model is called MGdNdS in the following. The parameter \( \lambda \) represents the synonymous substitution rate, whereas \( \omega = dN/dS \) is the ratio of nonsynonymous over synonymous substitution rates. The \( R \) matrix is assumed to be constant in time, while \( \lambda \) and \( \omega \) vary among lineages in correlation with LHT (here, body size, female maturity, and maximum longevity) and, in some cases, other quantitative traits describing nucleotide compositional variation between taxa (discussed later). Note that, although the codon model is formally time-reversible for any fixed value of \( \omega \), the overall substitution process is not time-reversible, due to the variation in \( \omega \) through time. In the general case, a total of \( L \) quantitative traits are considered, jointly with \( K = 2 \) substitution parameters (\( \lambda \) and \( \omega \)), summing up to a total of \( M = K + L \) parameters, assumed to follow a \( M \)-dimensional log-normal Brownian process running along the lineages of the phylogeny and parameterized by a \( M \times M \) covariance matrix \( \Sigma \). Using a log-normal Brownian model amounts to assuming linear correlations between variations in the logarithm of the quantitative traits and the substitution parameters. This is in turn equivalent to making the hypothesis that all LHT and substitution parameters scale allometrically with body size. The parameters of the model (the matrix \( R \), the reconstruction of \( dS \), \( dN/dS \), and quantitative traits along the phylogeny, the covariance matrix \( \Sigma \) and divergence times) are jointly estimated using Bayesian MCMC methods (see Lartillot and Poujol 2011 for more details).

Based on this formalism, we now develop a model, which we call RadConsAA, directly specified at the amino acid level and capturing the variation among lineages in the ratio \( Kr/Kc \) of radical over conservative amino acid replacement rates. Following an analogy with equation (1), instead of a codon model, we propose a 20 \( \times \) 20 amino acid replacement process, such that the rate of substitution between amino acids \( a \) and \( b \) is given as follows:

\[
Q_{a_1, a_2} = \lambda R_{a_1, a_2} \text{ if conservative,}
\]
\[
Q_{a_1, a_2} = \lambda R_{a_1, a_2} \eta \text{ if radical,}
\]
where \( R \) is now a \( 20 \times 20 \), normalized, general-time reversible amino acid replacement matrix parameterized as follows: \( R_{ab} = \rho_{ab} p_{b} \), where \((\rho_{ab})_{1 \leq a < b \leq 20}\) is a set of 190 relative exchangeabilities, and \((p_{a})_{a=1...20}\) are the 20 equilibrium frequencies of the amino acid replacement process. Both \( \rho \) and \( \pi \) are free parameters of the model. The parameter \( \lambda \) is now the conservative substitution rate, and \( \eta \) is proportional to the ratio of conservative over radical amino acid replacement rates (\( Kr/Kc \)). As in the MGdNdS model, we let \( \lambda \) and \( \eta \) vary among lineages, jointly with \( L \) other quantitative traits, whereas \( \rho \) and \( \pi \) are assumed to be constant across the entire phylogeny. Note that, because the relative exchangeabilities are left unconstrained and are empirically estimated, unless additional constraints are specified, the parameters of the model are not identifiable. Specifically, defining, for any positive constant \( A \), \( \eta' = \eta A \), \( p'_{ab} = p_{ab}/A \), for all radical pairs of amino acids \((a,b)\) and \( p'_{ab} = p_{ab} \), for all conservative pairs will result in the same substitution process \( R \). To restore identifiability, we impose the arbitrary convention that \( \eta = 1 \) at the root of the tree. Therefore, as a measure of \( Kr/Kc \), \( \eta \) is defined only up to a proportionality constant. In a second step, reconstructions of \( Kr/Kc \) along the tree are scaled by displaying \( Kr/Kc \) only up to a proportionality constant. In a second step, reconstructability, we impose the arbitrary convention that \( \rho \) and \( \pi \) are free parameters of the model. The parameter \( \lambda \) is now a \( 20 \times 20 \) general time-reversible amino acid replacement matrix.

To classify amino acid substitutions as radical or conservative, we follow Sainudiin et al. (2005). First, amino acids are partitioned according to three criteria: volume (large amino acids: L, I, F, M, Y, W, H, K, R, E, and Q), polarity (polar amino acids: Y, W, H, K, R, E, Q, T, D, N, S, and C), or charge (positively charged amino acid: H, K, R; negatively charged amino acids: D, E and neutrally charged amino acids: A, N, C, Q, G, I, L, M, F, P, S, T, W, Y, and V). Next, any combination of criteria defines a model in which substitutions that do not conserve all of these criteria are considered as radical. In what follows, we consider four distinct models, corresponding to charge alone (RadConsAA-Chg), volume alone (RadConsAA-Vol) or polarity alone (RadConsAA-Pol), as well as the combination of volume and polarity (RadConsAA-PolVol).

To control for nucleotide compositional variation in the context of a RadConsAA analysis, we computed the empirical frequencies of A, C, G, and T at the third coding positions of the alignment (denoted as \( \pi_{C}, \pi_{G}, \pi_{A}, \text{and } \pi_{T} \), respectively. As most compositional variation seen in mammalian mitochondrial genomes is due to variation in \( \pi_{C} \) and \( \pi_{T} \) (Gibson et al. 2005), we considered these two log-transformed variables as additional quantitative traits, to be analyzed along with LHT (i.e., as two additional quantitative traits in the model). Then, multiple regression (as described later) allows estimation of the correlation between \( \eta \) and the three life-history variables, while controlling for compositional variation in C and T. Similarly, we also controlled for GC bias, GC skew, and AT skew, where GC bias is defined as \((\pi_{G} + \pi_{C})/ (1 - [\pi_{G} + \pi_{C}])\); GC skew as \(\exp([\pi_{G} - \pi_{C}]/[\pi_{G} + \pi_{C}])\) and AT skew as \(\exp([\pi_{A} - \pi_{T}]/[\pi_{A} + \pi_{T}])\). Finally, we also performed the analyses excluding NADH dehydrogenase 6 (ND6). In birds and mammals, ND6 is the only protein-coding gene located on the cytosine-rich light strand of the mitochondrial genome and therefore has a different nucleotide composition, compared with other mitochondrial protein-coding genes.

We used the following priors: a uniform prior on divergence times, with fossil calibrations (hard constraints), a fixed topology (see "Phylogenetic trees" section) and an inverse-Wishart prior on the covariance matrix, of parameter \( \Sigma_{0} = \text{Diag} (\kappa_{1} \ldots \kappa_{M}) \), where \( M \) is the dimension of the Brownian process. For each \( m = 1 \ldots M \), the prior on \( \kappa_{m} \) is a truncated log-uniform prior between \( 10^{-4} \) and \( 10^{3} \) (equivalently, each \( \kappa_{m} \) has a density proportional to \( 1/ \kappa_{m} \) over \([10^{-4}, 10^{3}]\)). Truncation ensures normalization of the posterior density. A uniform Dirichlet distribution is used for the priors on the relative exchange rates and the equilibrium frequencies of the general time-reversible matrix \( R \). Concerning the prior on the value of the Brownian process at the root of the tree, we use a uniform prior between \(-100 \) and \( 100 \) for the entries corresponding to the logarithm of the substitution parameters \( \lambda \) and \( \omega \), and a normal distribution of mean and variance determined empirically for the LHT in mammals (mean and standard deviation of 8.4 for the logarithm of the body mass in grams, of 5.8 for the logarithm of the age at sexual maturity in days, and 5.2 for the logarithm of the longevity in months) and in birds (mean and standard deviation of 4.6 for the logarithm of the body mass, 2.3 for the logarithm of the maturity and 5.9 for the logarithm of the longevity). These values were obtained using the mean of the logarithm of the current life-histories traits (discussed later). Finally, we used the linear approximation for branch-specific averages of \( \lambda \), \( \omega \), and \( \eta \) (as defined in Lartillot and Poujol 2011).

**Monte Carlo**

Under all settings, and for all data sets, MCMC runs were duplicated. Convergence was first checked visually. The results obtained from the two independent runs were very similar in all cases, except for the large placental data set (with 273 taxa), for which one run was stuck in a local optimum. Exceptionally, we ran two additional chains for this particular case. Convergence was then more quantitatively assessed using the tracecomp program in the PhyloBayes package (Lartillot et al. 2009), for estimating the effective sample size.
and the overlap between two independent runs, based on the following summary statistics: the logarithm of the likelihood and of the prior density, the mean substitution rate over the tree, the mean dN/dS or mean Kr/Kc over the tree, the entries of the covariance matrix and the age of the root.

### Covariance Analysis

Covariance analyses were conducted as in Lartillot and Poujol (2011), reporting mean posterior correlation coefficients, and assessing the strength of the statistical support by estimating the posterior probability (PP) of a positive or a negative correlation between the two traits of interest. Specifically, the marginal correlation coefficient between two entries \( k \) and \( l \) of the multivariate process (which can correspond to either substitution parameters or quantitative traits) is defined as follows:

\[
\hat{r}_{kl} = \frac{S_{kl}}{\sqrt{S_{kk}S_{ll}}}
\]

where \( \Sigma \) is the covariance matrix. Equation (3) is averaged over the sample obtained by MCMC, yielding the point estimate of the correlation coefficient. Statistical support is indicated by always reporting the estimated PP of a positive correlation. Therefore, a PP close to 1 indicates a high PP of a negative correlation, whereas a PP greater than 0.95 or smaller than 0.05 is called marginally supported. Note that this is not formally equivalent to a classical Bayesian point-hypothesis testing procedure. Testing the point hypothesis of no correlation would be done by associating a positive prior probability mass with the parameter configuration corresponding to a correlation coefficient exactly equal to 0, and then calculating the PP of this configuration. On the other hand, the present test is able to detect situations in which several such variables need to be controlled for, equation (5) is applied recursively. The Kr/Kc models introduced here have been implemented in the Coevol software program, available from www.phylobayes.org.

### Sequence Data, Multiple Sequence Alignments

In this study, we analyzed two mitochondrial data sets composed of the 13 protein-coding genes representing 11,442 bp for 92 bird species and 11,529 bp for 273 placental mammals. The bird data set includes the 80 species of Neoaves and Galloanserae used in Pacheco et al. (2011), combined with eight species of Rattites and Tinamiformes (Härdill et al. 1998; Haddrath and Baker 2001) and four near-complete passerines mitochondrial genomes (Nabholz et al. 2010). The placental data set included all the placental species for which the complete mitochondrial genomes was available as of September 2011 in NCBI/GenBank database (http://www.ncbi.nlm.nih.gov/genbank/). To perform the time consuming MGdNdS analysis, this data set was reduced to 194 species by randomly selecting one species per genus.

Sequences were downloaded from the NCBI nucleotide database (http://www.ncbi.nlm.nih.gov/, last accessed September 2011). For each protein-coding gene, amino acid sequences were aligned using MUSCLE (Edgar 2004). The alignments were checked by eye and manually corrected if needed, and then used as a template for aligning nucleotide sequences while respecting the underlying coding structure. All the species names, taxonomic information, gene accession numbers, and the alignments are available on Dryad (http://datadryad.org, doi:10.5061/dryad.72594).

### Phylogenetic Trees

For placental mammals, we followed Murphy et al. (2007) and grouped Afrotheria and Xenarthra in the so-called Atlantogenata, as the monophyletic sister group of Boreoeutheria (Laursiathera and Euarchontoglires; Madsen et al. 2001; Murphy et al. 2001). Within these clades, we respected the topology proposed by Deluc et al. (2002) for Xenarthra; Dubey et al. (2007), Flynn et al. (2005), Teeling et al. (2005), and Springer et al. (2004) for Laurasiatheria; and Blanga-Kanfi et al. (2009), Fabre et al. (2009), Huchon et al. (2002), and Perelman et al. (2011) for Euarchontoglires.

In the case of birds, some uncertainties remain regarding the relationship at the base of Neoavian clade (i.e., all the extant birds excluding Galloanserae and Palaeognathae, see e.g., Hackett et al. 2008; Pacheco et al. 2011; Pratt et al.
Here, we chose to apply a pragmatic approach by testing two alternative topologies: first, a phylogeny was estimated using a maximum likelihood method (RAxML with GTR + G substitution model, version 7.0.4, Stamatakis 2006). Second, we used the topology of Hackett et al. (2008), which represents the most up-to-date phylogeny obtained with nuclear markers (this topology is presented in supplementary fig. S1, Supplementary Material online). The results obtained with the two topologies being very similar, we only present the results obtained with the maximum likelihood topology, except if stated otherwise.

Fossil Calibrations
In the case of birds, following Benton et al. (2009), we constrained the Paleognathae/Neognathae divergence to lie between 66 and 86.5 Ma and the crown Passeriformes (Acanthisitta/other Passeriformes) to be younger than 65 Myr. The oldest known penguins (Sphenisciformes) were estimated at 60–61 Myr (Slack et al. 2006), which allows us to constrain the split between Sphenisciformes (penguins) and Procellariiformes (the “tube-nosed” seabirds: albatrosses and petrels) to be older than 62 Myr.

For placents, we used 13 fossil calibrations, all obtained from Benton et al. (2009), including calibrations of the crown Hominioidea, Hylobatidae (used only for the mitochondrial data set), Catarrhini, Primates, Lagomorpha, Glires, Rodentia, Rodentia (minus sciurids), Muridae, Carnivora, Laurasiatheria, Laurasiatheria (minus Lipotyphla), and Boreoeutheria (see corresponding paragraphs in Benton et al. 2009 for details). All the calibrations are presented in supplementary table S1, Supplementary Material online.

Life-History Traits
Body mass for all the bird species analyzed here were obtained from Dunning (2007). Maximum recorded lifespan (a proxy for longevity) and female age at sexual maturity were extracted from the AnAge database (de Magalhães and Costa 2009). In the case of birds, where longevity and age at sexual maturity were recorded, we used the mean traits at the genus level. In the case of placents, data for the three LHT were obtained from the PANTheria database (Jones et al. 2009). For birds, body mass was available for all the 92 species whereas maximum longevity and age at sexual maturity were available for 60 and 44 species, respectively. For placental mammals, body mass was available for 230 species, maximum longevity for 189 and age at sexual maturity for 183 species. The distribution of these traits is shown in supplementary figure S2, Supplementary Material online.

Results
Using a previously developed phylogenetic covariance model (Lartillot and Poujol 2011), and extending so as to estimate variation in the ratio of radical over conservative amino acid replacement rates (Kr/Kc), we analyzed the relationship between dN/dS, Kr/Kc, and three LHT (body mass, female age at sexual maturity, and maximum recorded lifespan, as a proxy for longevity) in the mitochondrial proteome (13 protein-coding genes) of 92 birds and 273 placental mammals.

The Relationship between dN/dS and LHT
In placents, we found a positive correlation between mitochondrial dN/dS and the three LHT (PP of a positive correlation > 0.96, table 1). This is well in line with previous studies using partially overlapping data sets (Popadin et al. 2007; Lartillot and Poujol 2011). Statistical support is lost for all LHT in partial correlations (i.e., controlling for dS and the other two LHT) although the effect of longevity remains moderately strong (PP = 0.94; table 1).

Unexpectedly, an opposite result was obtained in birds. In this group, the mitochondrial dN/dS displays a negative correlation with LHT. The correlation is supported in the case of longevity (PP = 0.01; table 1) and remains so in partial correlation (i.e., controlling for dS, body mass and maturity; PP = 0.01; table 1).

The Relationship between Mitochondrial Kr/Kc and LHT
The high mutation rate experienced by the mitochondrial genome (Ballard and Whitlock 2004; Galtier et al. 2006; Nabholz, Gémin, et al. 2008) may prevent accurate estimation of dN/dS. In our data set, there is a marked saturation visible when uncorrected pairwise divergence is plotted against phylogenetic distance (fig. 1). As is clear from figure 1, saturation is more pronounced on synonymous than on non-synonymous substitutions.

Table 1: Covariance between dN/dS and dS and LHT Using the Nucleotide Data Sets

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Body Mass</th>
<th>Longevity</th>
<th>Age at Sexual Maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marginal correlation*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placentalia 194</td>
<td>0.47 (0.99)**</td>
<td>0.51 (0.99)**</td>
<td>0.38 (0.98)**</td>
</tr>
<tr>
<td>Birds</td>
<td>−0.30 (0.07)</td>
<td>−0.63 (0.01)**</td>
<td>−0.18 (0.28)</td>
</tr>
<tr>
<td>Partial correlation*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placentalia 194</td>
<td>0.12 (0.83)</td>
<td>0.22 (0.94)</td>
<td>−0.08 (0.30)</td>
</tr>
<tr>
<td>Birds</td>
<td>0.04 (0.56)</td>
<td>−0.60 (0.01)**</td>
<td>0.17 (0.63)</td>
</tr>
</tbody>
</table>

aCorrelation coefficients corresponding to the marginal correlation between each pair of variables.

bCorrelation coefficients corresponding to the partial correlations.

**PP < 0.025 or > 0.975.
Fig. 1.—Pairwise uncorrected divergence against phylogenetic distance estimated on the bird (A) and on the placental (B) data sets. On both panels, gray dots correspond to measures based on amino acids and black triangles to measures based on nucleotides. Phylogenetic distances are estimated by maximum likelihood using CODEML (Yang 2007; mttarn.dat substitution matrix) for the amino acid data, and BASEML (GTR+GAMMA substitution model) for nucleotides.

saturation may also explain some of the unexpected correlations observed between dN/dS and LHT, in particular in birds. Specifically, although smaller birds are expected to have smaller effective population sizes and, therefore, lower dN/dS ratios, in the presence of saturation, their higher overall rate of substitution (Nabholz et al. 2009) will only weakly impact dS but more strongly impact dN. As a result, the apparent dN/dS will increase, whereas in fact the real dN/dS is undergoing a decrease. This effect is also predicted in the case of mammals, although in that case, the better taxonomic sampling, combined with a wider overall range of variation in body size, may be sufficient for ensuring a robust positive correlation between dN/dS and body mass.

Measures of the strength of selection exclusively relying on the less saturated nonsynonymous substitutions may be more reliable than dN/dS in the context of mitochondrial genome evolution. With this in mind, we constructed several models based on the ratio of radical over conservative amino acid substitution rates. Biochemical conservativeness can be defined in several different ways, depending on whether amino acid replacements resulting in a change in charge, volume, or polarity were considered as radical. Accordingly, we defined four different versions of the model taking into account the three different changes mentioned above plus the combination of polarity and volume (see Materials and Methods).

In placentals, whichever definition of biochemical conservativeness is used, in all cases, Kr/Kc correlates positively with LHT, thus mirroring the correlation patterns observed for dN/dS (Popadin et al. 2007). Among the four models, the one considering a change in polarity or in volume as a radical change (RadConsAA-PolVol) gives a consistently stronger support for a positive correlation between LHT and Kr/Kc (table 2). Using this model, the correlation with LHT is stronger than using dN/dS (table 3). In contrast, the model using charge as the criterion for discriminating radical and conservative amino acid substitutions (RadConsAA-Chg) results in a weaker covariance, with a PP above 0.95 obtained only for longevity (table 2). Using a different method, Popadin et al. (2007) also found that a definition of radical substitutions based on charge results in a weaker correlation between body-mass and Kr/Kc, compared with volume or polarity. In contrast to our study, however, Popadin et al. (2007) reported that the model based on polarity or on volume alone were slightly better than the model based on polarity and volume in combination. In partial correlation analyses, Kr/Kc is found to correlate primarily with age at sexual maturity for the complete data set, and with longevity for the reduced data set (with 194 species, table 3).

Interestingly, the results obtained on birds are at odds with the dN/dS analyses. Apart from the model based on charge, for which Kr/Kc displays negative but not supported correlations with LHT, all other models indicate a positive correlation between LHT and Kr/Kc (table 2). Under the model based on polarity and volume, support for a positive marginal correlation is found between body mass and Kr/Kc (PP = 0.98,
The marginal correlation with the other LHT is not significant, which may seem surprising given that all LHT are strongly correlated with each other. However, this could be explained by the fact that a substantial fraction of values is missing for these two LHT (35% and 48% of the entries for longevity and maturity, respectively). The effect of body mass is lost in partial correlations (table 3). Similar results were obtained using the alternative topology for the bird phylogeny (PP = 0.99, supplementary table S2, Supplementary Material online).

The Influence of Base Composition and Mutation Bias

Mitochondrial nucleotide composition, especially in Cytosine (C) and in Thymine (T), varies substantially between mammalian species (Gibson et al. 2005). In birds, we observed a similar pattern, with the frequency of C (πC) being highly variable between species (ranging from 0.32 to 0.51), whereas A and G show moderate variation (from 0.29 to 0.43 and from 0.03 to 0.10, respectively). As in the case of mammals, in birds, the frequency of C (πC) appears strongly negatively correlated with the frequency of T (πT), suggesting that the joint variation in the two frequencies, in C and in T, is primarily the consequence of an underlying variation in the relative rate of cytosine deamination (Reyes et al. 1998).

Nucleotide compositional variation between species could potentially affect our estimation of dN/dS and Kr/Kc. Variation in nucleotide composition may differentially influence synonymous and non synonymous rates, owing to the fact that the nucleotide content is more variable on synonymous than on non synonymous sites. For Kr/Kc, the potential impact of compositional variation is less obvious, but could occur in several subtle ways. For instance, codons for polar amino acids have a tendency to be GC-poor compared with codons encoding nonpolar amino acids, and thus, a shift in nucleotide composition could potentially result in a higher rate of radical amino acid changes. In this context, the existence of a correlation between nucleotide composition and LHT, such as longevity (Jobson et al. 2010), raises the possibility that nucleotide compositional variation may be responsible for at least part of the correlations observed between LHT and dN/dS or Kr/Kc (tables 1 and 2), thus calling for a more detailed investigation of the impact of nucleotide composition on our estimation.

Table 2
Marginal Covariance between Kr/Kc and LHT According to Alternative Definitions of Biochemical Conservativeness

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Classification</th>
<th>Body Mass</th>
<th>Longevity</th>
<th>Age at Sexual Maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammals (All)</td>
<td>Charge</td>
<td>0.19 (0.91)</td>
<td>0.43 (0.99)**</td>
<td>0.17 (0.83)</td>
</tr>
<tr>
<td></td>
<td>Polarity</td>
<td>0.29 (0.99)**</td>
<td>0.23 (0.94)</td>
<td>0.49 (0.99)**</td>
</tr>
<tr>
<td></td>
<td>Volume</td>
<td>0.30 (0.99)**</td>
<td>0.25 (0.95)*</td>
<td>0.47 (0.99)**</td>
</tr>
<tr>
<td></td>
<td>Polarity and volume</td>
<td>0.37 (0.99)**</td>
<td>0.43 (0.99)**</td>
<td>0.64 (0.99)**</td>
</tr>
<tr>
<td>Birds</td>
<td>Charge</td>
<td>−0.05 (0.42)</td>
<td>−0.28 (0.24)</td>
<td>−0.07 (0.41)</td>
</tr>
<tr>
<td></td>
<td>Polarity</td>
<td>0.24 (0.90)</td>
<td>0.06 (0.62)</td>
<td>−0.03 (0.44)</td>
</tr>
<tr>
<td></td>
<td>Volume</td>
<td>0.26 (0.80)</td>
<td>0.33 (0.82)</td>
<td>0.18 (0.69)</td>
</tr>
<tr>
<td></td>
<td>Polarity and Volume</td>
<td>0.37 (0.98)**</td>
<td>0.23 (0.84)</td>
<td>0.09 (0.59)</td>
</tr>
</tbody>
</table>

*PP > 0.95 or < 0.05.
**PP < 0.025 or > 0.975.

Table 3
Covariance between Kr/Kc, Kc, and LHT

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Classification</th>
<th>Body Mass</th>
<th>Longevity</th>
<th>Age at Sexual Maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marginal correlation^a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placentalia (All)</td>
<td></td>
<td>0.37 (-0.99)**</td>
<td>0.43 (-0.99)**</td>
<td>0.64 (-0.99)**</td>
</tr>
<tr>
<td>Placentalia (N = 194)</td>
<td></td>
<td>0.46 (-0.99)**</td>
<td>0.60 (-0.99)**</td>
<td>0.54 (-0.99)**</td>
</tr>
<tr>
<td>Birds</td>
<td></td>
<td>0.37 (0.98)**</td>
<td>0.23 (0.84)</td>
<td>0.09 (0.59)</td>
</tr>
<tr>
<td>Partial correlation^b</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placentalia (All)</td>
<td></td>
<td>−0.03 (0.39)</td>
<td>0.21 (0.87)</td>
<td>0.42 (-0.99)**</td>
</tr>
<tr>
<td>Placentalia (N = 194)</td>
<td></td>
<td>0.04 (0.57)</td>
<td>0.37 (0.97)*</td>
<td>0.22 (0.88)</td>
</tr>
<tr>
<td>Birds</td>
<td></td>
<td>0.30 (0.92)</td>
<td>0.03 (0.53)</td>
<td>−0.12 (0.36)</td>
</tr>
</tbody>
</table>

^aCorrelation coefficients corresponding to the marginal correlation between each pair of variables.
^bCorrelation coefficients corresponding to the partial correlations.
*PP > 0.95 or < 0.05.
**PP < 0.025 or > 0.975.

NOTE.—Change in polarity and volume are considered as radical substitutions.

The Influence of Base Composition and Mutation Bias

Mitochondrial nucleotide composition, especially in Cytosine (C) and in Thymine (T), varies substantially between mammalian species (Gibson et al. 2005). In birds, we observed a similar pattern, with the frequency of C (πC) being highly variable between species (ranging from 0.32 to 0.51), whereas A and G show moderate variation (from 0.29 to 0.43 and from 0.03 to 0.10, respectively). As in the case of mammals, in birds, the frequency of C (πC) appears strongly negatively correlated with the frequency of T (πT), suggesting that the joint variation in the two frequencies, in C and in T, is primarily the consequence of an underlying variation in the relative rate of cytosine deamination (Reyes et al. 1998).

Nucleotide compositional variation between species could potentially affect our estimation of dN/dS and Kr/Kc. Variation in nucleotide composition may differentially influence synonymous and non synonymous rates, owing to the fact that the nucleotide content is more variable on synonymous than on non synonymous sites. For Kr/Kc, the potential impact of compositional variation is less obvious, but could occur in several subtle ways. For instance, codons for polar amino acids have a tendency to be GC-poor compared with codons encoding nonpolar amino acids, and thus, a shift in nucleotide composition could potentially result in a higher rate of radical amino acid changes. In this context, the existence of a correlation between nucleotide composition and LHT, such as longevity (Jobson et al. 2010), raises the possibility that nucleotide compositional variation may be responsible for at least part of the correlations observed between LHT and dN/dS or Kr/Kc (tables 1 and 2), thus calling for a more detailed investigation of the impact of nucleotide composition on our estimation.
To do so, for the analyses using the MGdN/dS model, we allow the equilibrium GC (GC*) of the codon substitution model to vary continuously along the tree (see Materials and Methods and Lartillot 2013). Variation in GC* among lineages is then treated like variation in dS and dN/dS, such that the overall covariance analysis includes a total of three substitution parameters and three LHT. Multiple regression then allows controlling for GC* when estimating the correlation between dN/dS and LHT. More specifically, we controlled for both the equilibrium GC composition at all codon position (GC*) and the equilibrium GC at the third-codon position (GC3*), the latter being markedly more variable than the former (Gibson et al. 2005). Both models leads to very similar results; therefore, we present only the GC3* results hereafter.

The GC3* appears positively correlated with dN/dS in birds ($R = 0.57$, PP $> 0.99$) and in placental mammals ($R = 0.22$, PP $> 0.99$). In placental mammals, GC3* is positively correlated with all LHT (PP $> 0.99$). In birds, GC3* correlates negatively with maximal longevity ($R = -0.36$, PP $= 0.03$), but not with body mass and age at sexual maturity. When controlling for GC*, the correlation between dN/dS and LHT remains largely unchanged for birds (table 4). In placentals, the strength of the correlation is overall reduced, with body mass remaining the only correlate of dN/dS receiving a strong statistical support (PP $> 0.99$; table 4).

To control for nucleotide composition in the case of the RadConsAA models, we opted for a pragmatic method circumventing the fact that nucleotide frequencies are not explicit parameters of the substitution model. Specifically, for each taxon, we computed the frequencies of C and T ($\pi_C$ and $\pi_T$) at the third positions, and considered these (log-it transformed) frequencies as quantitative traits, to be included in our covariance analyses along with the three LHT. We then performed multiple regression analyses, so as to estimate the correlation of LHT with Kr/Kc while controlling for composition in C and T. Alternatively, we devised a control using GC bias, AT skew and GC skew (all log-transformed) instead of $\pi_C$ and $\pi_T$ (see Materials and Methods). These controls were conducted both on the original sequence alignments and on alignments from which the ND6 gene was excluded, as this gene, being located on the light strand, experiences a distinct mutation bias compared to other mitochondrial protein coding genes.

We find that $\pi_C$ and $\pi_T$ are, respectively, positively and negatively correlated with longevity in mammals (PP $> 0.99$ for $\pi_C$ and PP $< 0.01$ for $\pi_T$ for all the LHT). These results are in agreement with previous analyses (Jobson et al. 2010). We also observe that Kr/Kc is positively correlated with $\pi_C$ ($R = 0.35$, PP $> 0.99$) and negatively with $\pi_T$ ($R = -0.29$, PP $< 0.01$). This correlation could be due in part to a joint covariation of Kr/Kc and composition in C and T with LHT. However, controlling for LHT does not eliminate the statistical support for the correlation of Kr/Kc with nucleotide composition ($\pi_C$: $R = 0.24$, PP $= 0.99$; $\pi_T$: $R = -0.19$, PP $= 0.03$), suggesting instead that compositional variation has a direct impact on Kr/Kc estimation. On the other hand, biases induced by compositional variation do not entirely explain the correlation of Kr/Kc with LHT, as controlling for $\pi_C$ and $\pi_T$ does not reduce the strength the correlation between LHT and Kr/Kc (table 5). Similarly, controlling for GC bias, AT skew and GC skew does not have any impact on the strength of the correlation (PP $> 0.98$ for all LHT; table 5) nor does the exclusion of ND6 (supplementary table S3, Supplementary Material online).

Interestingly, for birds, the relationship between $\pi_C$ and $\pi_T$ and LHT goes in the opposite direction, compared with what is observed in placental mammals: in birds, $\pi_C$ is negatively correlated with body mass although the correlation is not supported ($R = -0.15$, PP $= 0.08$) and $\pi_T$ is positively correlated with body mass, with marginal support ($R = 0.17$, PP $= 0.95$). As in mammals, controlling for $\pi_C$ and $\pi_T$ or for GC bias, AT skew and GC skew, with or without ND6, does not reduce the support for the correlation between body mass and Kr/Kc.

### Table 4

Partial Correlations between dN/dS and LHT Controlling for GC3 Content

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Body Mass</th>
<th>Longevity</th>
<th>Age at Sexual Maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placentalia ($N = 194$)</td>
<td>0.38 (−0.99)**</td>
<td>0.45 (0.91)</td>
<td>0.16 (0.82)</td>
</tr>
<tr>
<td>Birds</td>
<td>−0.21 (0.15)</td>
<td>−0.54 (0.01)**</td>
<td>−0.25 (0.20)</td>
</tr>
</tbody>
</table>

**PP < 0.025 or > 0.975.

### Table 5

Covariance between Kr/Kc and LHT Controlling for Base Composition

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Control</th>
<th>Body Mass</th>
<th>Longevity</th>
<th>Age at Sexual Maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placentalia (All)</td>
<td>$\pi_C$ and $\pi_T$</td>
<td>0.33 (−0.99)**</td>
<td>0.46 (−0.99)**</td>
<td>0.53 (−0.99)**</td>
</tr>
<tr>
<td>Placentalia (All)</td>
<td>GC bias, AT/GC skew</td>
<td>0.33 (−0.99)**</td>
<td>0.38 (0.98)**</td>
<td>0.54 (−0.99)**</td>
</tr>
<tr>
<td>Birds</td>
<td>$\pi_C$ and $\pi_T$</td>
<td>0.42 (0.99)**</td>
<td>0.29 (0.88)</td>
<td>0.19 (0.76)</td>
</tr>
<tr>
<td>Birds</td>
<td>GC bias, AT/GC skew</td>
<td>0.41 (0.99)**</td>
<td>0.27 (0.86)</td>
<td>0.20 (0.77)</td>
</tr>
</tbody>
</table>

* N0s.—Change in polarity and volume are considered as radical substitutions.
** PP < 0.025 or > 0.975.
Macroevolutionary Patterns of Kr/Kc and LHT

The reconstructions of $dN/dS$ and Kr/Kc are displayed in supplementary figures S3 and S4, Supplementary Material online, for $dN/dS$ and figures 2 and 3 for Kr/Kc. In the case of $dN/dS$ it confirms that compositional variation cannot entirely explain observed correlations between Kr/Kc and LHT.
branches with large values of $d_N/d_S$ are mostly ancient branches. This is further indicated by the strong positive correlation between divergence times and $d_N/d_S$ observed in placental (fig. 4A), suggesting that $d_N/d_S$ could indeed be overestimated because of an underestimation of $d_S$ in the more ancient branches of the phylogeny, itself due to the greater mutational saturation affecting synonymous substitutions. In the context of a nearly neutral interpretation of the variation in $d_N/d_S$, such differential saturation effects would result in artifactual inference of small ancestral effective population sizes. In addition, as the model uses the relationship between LHT and $d_N/d_S$ to reconstruct ancestral LHT, the overestimation of $d_N/d_S$ in ancient lineages may bias the estimation of ancestral LHT toward higher values.

In contrast, patterns of Kr/Kc (fig. 2) are more consistent within orders, and across the tree: high values of Kr/Kc in orders such as Cetartiodactyla (especially cetaceans), Carnivora (particularly Ursidae and Pinnipedia), Primates (especially Catarrhini), and more isolated clades like elephants or sloths. Conversely, Glires (i.e., Rodentia + Lagomorpha), Eulipotyphla (i.e., insectivores within Laurasiatheria), or Afrotheria have low Kr/Kc values. Chiroptera have relatively low values, with large Pteropodidae having higher Kr/Kc ratios than other smaller Chiroptera. These patterns are globally

**FIG. 3.**—Reconstruction of Kr/Kc (as defined by change in polarity or volume) along the phylogeny of birds. Color stands for the marginal posterior mean instant value of Kr/Kc at each node, with linear interpolation along branches.
consistent with Kr/Kc representing a more reliable molecular correlate of population size and LHT. Unlike what is observed with dN/dS, the early lineages of placental mammals do not have especially high values of Kr/Kc, and a negative, instead of a positive, correlation is observed between Kr/Kc and divergence times in placentals (Spearman’s rho = −0.67, \( P \) value < 0.01; fig. 4B), but note the increase of Kr/Kc toward the very base of the tree (fig. 4B). Nonetheless, the ancestor of placental mammals is estimated to have a bigger body mass using the MGdNdS model (between 2 and 40 kg) than under the RadConsAA model (between 0.6 and 7 kg). Similarly, the ancestors of each of the major groups of placental mammals are all inferred to be two to four times smaller using RadConsAA than using MGdNdS (table 6).

The saturation detected in nucleotide sequences may also explain why the relationship between LHT and dN/dS is weaker than the relationship between LHT and Kr/Kc. Based on all of these observations, we choose to favor the
reconstruction of the ancestral LHT using the RadConsAA model over MGdNdS, and we will only refer to the results obtained using the RadConsAA model hereafter.

The ancestor of extant placental mammals was estimated to weigh 2 kg (95% credibility interval [CI]: 0.6–7 kg), reach its sexual maturity after 440 days (CI: 277–696 days) and have a life span of 15 years (CI: 10–23 years). This estimation is compatible with estimates obtained with a nuclear data set (mean = 1.6 kg; CI: 0.2–7 kg; Lartillot and Delsuc 2012). Comparing estimates obtained under the covariant or the uncorrelated models suggests that Kr/Kc brings information in favor of a slight downward revision of the ancestral LHT of a certain number of ancestors. For example, all the ancestor of major orders such as Primates, Rodentia, Chiroptera, Cetartiodactyla, and Carnivora are inferred to be heavier using the uncorrelated model. The largest discrepancy was observed for Cetartiodactyla where the ancestor is estimated to be almost twice as large with the uncorrelated model (median = 33 kg; CI: 9–113 kg) compared with what is observed under the covariant model (median = 15 kg; CI: 4–53 kg). The estimation for the ancestor of placental mammals is, however, nearly identical between the uncorrelated (median = 1.7 kg; CI: 0.5–6 kg) and the uncorrelated models.

The negative correlation between Kr/Kc and divergence times alluded to above, combined with the positive correlation between Kr/Kc and LHT, would seem to suggest the existence of a global trend in increasing body size in placental mammals, akin to a Cope’s rule. On the other hand, early placental do not have especially low Kr/Kc ratios: Kr/Kc is larger in early branches connecting placental orders, than within Rodentia or Eulipotyphla (mean Kr/Kc = 0.85; CI: 0.69–1.1 and mean = 0.85; CI: 0.70–0.99 for Laurasiatheria and Euarchontoglossa ancestor vs. mean = 0.93; CI = 0.71 and 1.22 for terminal branches of Rodentia and mean = 0.72; CI: 0.63–0.96 for terminal branches of Eulipotyphla). Such an intermediate Kr/Kc ratio suggests that early placental were not especially small: indeed, even when taking into account the covariance between body mass and Kr/Kc (which leads to a smaller inferred ancestral body mass), the estimated body mass of the last common ancestor of placental is bigger than the mass of a majority of extant mammals. The median body mass of extant mammals is 100 g, and only 23% of the extant species weigh more than 2 kg (N = 3,913 species available in PanTheria database, Jones et al. 2009).

Therefore, our estimate for the ancestor represents a fairly large animal compared with a typical extant placental mammal, and this, fundamentally because species-rich group of mammals such as Rodentia, Chiroptera, and Eulipotyphla are all small animals.

To further explore the trend in body mass during placental evolution, we computed the difference between the median ancestral body mass estimated using the RadConsAA-PoVo model and the median of the body mass of extant placental mammals species (all in base 10 logarithm). A positive (respectively negative) value indicates a decrease (respectively an increase) in body mass. Taxonomy is as in PanTheria (Jones et al. 2009).

![Fig. 5.—Difference between the median ancestral body mass estimated using the RadConsAA-PoVo model and the median of the body mass of extant placental mammals species (all in base 10 logarithm). A positive (respectively negative) value indicates a decrease (respectively an increase) in body mass. Taxonomy is as in PanTheria (Jones et al. 2009).](image-url)
birds mainly occurred through size reduction. The ancestor of Neoaves (all the extant birds excluding Galloanserae and Palaeognathae) was already estimated to be smaller than the Neornithes ancestor (median = 341 g; CI: 142–820 g) and the great majority of modern birds (6,556 species out of 8,177 Neoaves) have today a body mass lower than 150 g (Dunning 2007). This trend is, however, not followed by Palaeognathae (Ratites + Tinamiformes), nor by Galloanserae. The ancestor of Palaeognathae was estimated to be heavier than the Neornithes ancestor (median = 2.1 kg and 95% CI: 530 g to 8.7 kg), after which body mass still increases toward the very large extant Ratites (but not in Tinamiformes). In Galloanserae, body mass remains fairly stable from approximately 800 g (95% CI: 210 g to 3.1 kg) for the ancestral species to a median of 705 g (95% CI: 102 g to 4.2 kg) for the extant species. Using the same method as for Placentalia, we computed the difference between the median ancestral body mass and the median body mass of extant species for the 11 orders (or clades) (fig. 6). In 7 out of the 11 clades, body mass has decreased during Avian evolution (fig. 6). This might indicate that, in contrast to placental mammals, life-history evolution preferentially occurred through size reduction in birds. Finally, most ancestral body mass estimates under the uncorrelated model are slightly larger than under the covariant model with Kr/Kc.

**Discussion**

Extending several recent methodological developments concerning the analysis of correlations between substitution patterns and LHT, and adapting them so as to deal with the specific features of the evolutionary regime of mitochondrial DNA (in particular the high rate of synonymous substitution), the present article proposes a synthetic view of what the comparative analysis of mitochondrial genomes can tell us about the evolution of LHT, and population genetics parameters in birds and mammals.

**Mutational Saturation and the Relation between Substitution Patterns and LHT**

Assuming that body mass represent a proxy of long-term effective population size (discussed later), the nearly neutral theory predicts a positive correlation of dNdS and Kr/Kc with body size. Globally, placental mitochondrial genomes agree with this expectation. Birds, on the other hand, display contradictory patterns: negative correlation between dNdS and body mass, but positive correlation between Kr/Kc and body mass (albeit not robust to controlling for other traits and dS). We argue that the conflicting results observed in the case of birds reflect the poor estimation of dNdS in mitochondrial data sets, rather than an absence of relationship between body mass and the efficacy of selection in birds. This interpretation relies on two points. First, synonymous substitutions appear to be highly saturated in mitochondrial sequences (see Results section and fig. 1). Despite the fact that the mitochondrial mutation rate is lower in birds than in mammals (Hickey 2008; Nabholz et al. 2009), saturation may be more pronounced in birds because, in our data sets, terminal branch lengths are often longer in the case of birds than for placentalas (median = 0.13 vs. 0.43 substitution per nucleotide sites, for placental mammals and birds, respectively, in both cases, using third-codon positions). This appears to be an indirect consequence of the taxonomic sampling of the bird data set, mainly designed to tackle the Neoaves phylogeny (Gibb et al. 2007; Pratt et al. 2009). Second, in placental mammals, the relationship between Kr/Kc and LHT is stronger that the relationship with dNdS (compare tables 1 and 3), further suggesting that the RadConsAA model, despite its increased overall parameterization, is better at estimating the rate of fixation of slightly deleterious mutations in mitochondrial sequences. Finally, it is expected that any relationship between the substitution process and LHT would be more difficult to pinpoint in birds, because they have a markedly narrower range of variation in LHT compared with mammals.

In the case of mammals, owing to a better taxonomic sampling and a wider range of LHT variation in this group, mutational saturation, although strong, does not qualitatively
change the correlation patterns of dN/dS with LHT. On the other hand, the reconstruction of dN/dS along the phylogeny (supplementary fig. S2, Supplementary Material online) clearly appears to be unreliable: the large values observed in ancient and fast evolving lineages are probably the result of a differential saturation of dS and dN, and are thus almost certainly artificial. Such biases in the estimation of ancestral dN/dS are likely to have a strong impact on the marginal reconstruction of LHT evolution along the tree. Typically, given the positive correlation inferred between dN/dS and LHT, this might result in an upward bias in the estimated ancestral body sizes. In contrast, as far as ancestral LHT inference is concerned, Kr/Kc seems to represent a more reliable molecular marker than dN/dS.

Variation in Kr/Kc among Lineages: A Nearly Neutral Effect?

The nearly neutral theory predicts that both dN/dS and, under certain conditions, Kr/Kc, should correlate negatively with effective population size (Nₑ). Here, and elsewhere (Nikolaev et al. 2007; Lartillot and Deluc 2012; Romiguier et al. 2013), what is more specifically observed is a correlation between dN/dS or Kr/Kc and body mass (and other LHT). Interpreting these observations in the light of the nearly neutral theory requires additional assumptions concerning the link between body mass and long-term Nₑ. In this direction, a negative correlation between population density and body size has been reported in a wide range of taxonomical groups (White et al. 2007). This correlation seems stronger in mammals than in birds (e.g., Silva et al. 1997). In turn, it is reasonable to assume that effective population size is correlated with population density. This correlation is expected to be weakened, although not suppressed, by additional covariates of Nₑ such as unequal sex ratio, population structure, inbreeding, and variance in the number of offspring could influence Nₑ (Frankham 1995; Charlesworth 2009). These additional factors may in fact explain the relatively weak correlation between body size and Kr/Kc observed here, as well as the weak correlations previously reported between dN/dS and LHT. Note that the fact that the level of mitochondrial neutral diversity is only weakly negatively correlated with body mass could be taken as an argument against the link between body mass and Nₑ (Berlin et al. 2007; Nabholz et al. 2008 but see Piganeau and Eyre-Walker 2009). Nucleotide diversity is, however, very sensitive to recent drop in population size and therefore could be a very poor estimate of long term Nₑ. This sensitivity to population fluctuation could explain why, in birds, no correlation was found between dN/dS and Nₑ (Berlin et al. 2007; Nabholz et al. 2008) but see Piganeau and Eyre-Walker 2009). Nucleotide diversity is, however, very sensitive to recent drop in population size and therefore could be a very poor estimate of long term Nₑ. By imposing an upper limit on the population size of a given species, body mass could influence long term Nₑ and therefore be a better predictor of variation in dN/dS (and Kr/Kc) than short term Nₑ such as estimated from genetic diversity or current census population size.

From a molecular evolutionary perspective, the most important question raised by the present analysis is whether, and to what extent, Kr/Kc reflects variation in effective population size. Although the correlation patterns observed in our analysis are globally compatible with a nearly neutral model, alternative interpretations should also be considered.

Among potentially confounding factors acting on Kr/Kc, positive selection represents one possible candidate. Positive selection has been suggested as a dominant force in the molecular evolutionary regime of mitochondrial genomes (Bazin et al. 2006), although mostly in invertebrates. In contrast, patterns of genetic diversity in mitochondrial genomes of birds and mammals suggest a regime dominated by purifying selection and random drift in these groups (Muligan et al. 2006; Nabholz et al. 2008, 2009). Isolated cases of positive selection in mitochondrial protein sequences have been reported in mammals, and particularly in anthropoid primates (Grossman et al. 2004). In the specific case of anthropoids, we have checked that the correlation between Kr/Kc and LHT was robust to the removal of all sequences of this clade (not shown).

Mitochondrial nucleotide composition is correlated with LHT and particularly with longevity (Jobson et al. 2010), a relationship confirmed by our results. This raises the possibility that changes in nucleotide composition, and more generally in the mutation spectrum could influence the estimation of variation in dN/dS or Kr/Kc and confound the correlation patterns of these two variables with LHT. Here, we have checked that the correlation between Kr/Kc and LHT remains significant upon controlling for compositional variation (table 5). Similarly, the correlation between dN/dS and LHT is robust when controlling for equilibrium GC3 (table 4). Finally, the fact that Kr/Kc positively correlates with LHT in birds and mammals, despite of opposite compositional trends in the two groups, provides an internal negative control suggesting that compositional variation at the nucleotide level cannot entirely explain the correlation patterns of Kr/Kc with LHT. This said, our controls do not entirely rule out potential biases due to more general changes in the mutation spectrum (e.g., the transition–transversion ratio, Belle et al. 2005). An in depth analysis of the question would require explicit models of the mutation-selection balance at the codon level, and of how this balance may shift as a result of changes in either the mutation process or the efficacy of selection.

Altogether some caution is therefore needed when interpreting variation in mitochondrial Kr/Kc among lineages. Explanations in terms of positive selection, or more subtle artifacts caused by changes in the mutation spectrum, cannot be entirely excluded. Nevertheless, when combining the observations made on mammals and on birds, a nearly neutral
interpretation remains the best explanation currently available. Accordingly, figures 2 and 3 would represent a tentative (albeit uncalibrated) reconstruction of the evolutionary trends in long-term effective population size in mammals and birds. In the case of mammals, this can be compared with a previous reconstruction based on the nuclear dN/dS (restricted to GC conservative transversions, which are immune from GC-biased gene conversion, Lartillot 2013). Although effective population size is not expected to be the same for nuclear and mitochondrial sequences, owing to differences in the mode of transmission, and in the recombination and selection regimes, nevertheless, at the scale of placental mammals, the qualitative patterns of variation should be relatively concordant between the two genetic compartments. Indeed, the two reconstructions agree in many respects. In particular, the long-term effective population size is inferred to be small in anthropoids, Cetacea, Panungulata (except Procavia) and Pilosa, large in Eulipotyphla, Rodentia, and Chiroptera, and intermediate in the ancestor of Placentalia. There are also differences between the two reconstructions, in particular for Carnivora, inferred to have a smaller long-term effective population size based on mitochondrial Kr/Kc, than based on the nuclear GC-conservative dN/dS (although in both cases, Canidae are inferred to have larger effective population sizes than Ursidae, Pinnipeds, or Felidae). Such discrepancies suggest the presence of confounding factors (among which compositional variation or local episodes of positive selection). Ultimately, these reconstructions could be calibrated using data about polymorphism in extant species, although this would require specific assumptions about the correlation between short- and long-term effective population size.

**Evolution of Body Mass in Birds and Mammals**

The correlation between Kr/Kc and LHT implies that information about ancestral LHT can be gained from reconstructions of patterns of Kr/Kc along phylogenies. In this respect, our analysis provides interesting insights into the evolution of life-history, and particularly the evolution of body-size, in placental mammals. Instead of a global trend across the entire group, our reconstruction integrating information about ancestral Kr/Kc suggests that the ancestors of some specific orders (e.g., Carnivora, Perissodactyla, and Cetartiodactyla) might have been slightly smaller than what would be inferred based on the body mass of extant taxa. This in turn suggests the existence of some local trends in increasing body size in these specific orders. On the other hand, other orders clearly show a pattern of stasis, or even of a decrease, in body size (e.g., Chiroptera, Rodentia, and Eulipotyphla).

Our analysis also suggests that the ancestors of both placental mammals and birds were not small, compared with extant species. This is particularly obvious in birds where most of the diversification seems to have been accompanied by reduction in size, but the same is true for placental mammals: the last common ancestor of Placentalia is inferred to be between 0.6 and 7 kg. This is larger than a majority of extant placental mammals. Comparison of this estimate with the fossil record is not an easy task. Nevertheless, current paleontological knowledge suggests that, although Eutherians known from the Cretaceous had apparently undergone a large ecological diversification (Luo 2007; Wilson et al. 2012) they were, on average, rather small (between 0.1 and 1 kg), with the exception of the 12–14 kg carnivorous *Repenomamus giganticus* (Hue et al. 2005). Our estimation is, therefore, in the upper side of the range of body masses inferred for cretaceous fossils, even if some isolated larger fossils are documented. Interestingly, our estimate is more compatible with a recent study using nuclear protein-coding gene GC evolution (Romiguier et al. 2013), although on the lower side of the confidence interval reported in this study.

Like mammals, birds were also ecologically and morphologically fairly diverse in the cretaceous (Chiappe and Dyke 2002). Body mass of these birds range from a few dozen grams to approximately 1 kg (body mass extrapolated using limbs size and scaling equation from Hone et al. 2008), which is compatible with our estimation (0.2 to 1.9 kg). Small individuals often belong to Enantiornithes, whereas larger birds were found in Ornithurae (e.g., Hesperornis and Ichthyornis)—a clade including the modern Neornithes (Chiappe and Dyke 2002). In the Cretaceous, Neornithes fossils are unfortunately very fragmentary and can be difficult to identify (Mayr 2009). Within the Paleogene however, most modern orders are represented in the fossil record (Mayr 2009), however, body mass has not yet been estimated for these taxa.

The present reconstruction assumes that the evolution of body mass and Kr/Kc is correctly described by a nondirectional Brownian motion. In a similar analysis of correlations between body size, LHT, and substitution rates in nuclear sequences in placental mammals (Lartillot and Delsuc 2012), it was noted that the assumptions behind the Brownian model are probably violated, and that this may have an impact on ancestral reconstructions, even taking into account correlations between body mass and substitution patterns. In the long term, further elaborating on the present model, integrating several molecular markers, and accounting for departures from the undirected Brownian assumption, could result in more contrasted and more informative patterns of LHT evolution.

**Supplementary Material**

Supplementary figures S1–S4 and tables S1–S3 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).
Acknowledgments
This work was supported by the ARCAD foundation project (http://www.arcad-project.org/) to B.N.; the Natural Science and Engineering Research Council of Canada and the Canadian Foundation for Innovation to N.U. and N.L.; and the French Agence Nationale pour La Recherche, Ancestrome project to N.L. This is ISEM publication number ISE-M 2013-068. Computational resources were provided by Calcul Québec, Compute Canada, and Montpellier MBB Platform.

Literature Cited