



**HAL**  
open science

## Life History Traits, Protein Evolution, and the Nearly Neutral Theory in Amniotes

Emeric Figuet, Benoit Nabholz, Manon Bonneau, Eduard Mas Carrio, Krystyna Nadachowska-Brzyska, Hans Ellegren, Nicolas Galtier

► **To cite this version:**

Emeric Figuet, Benoit Nabholz, Manon Bonneau, Eduard Mas Carrio, Krystyna Nadachowska-Brzyska, et al.. Life History Traits, Protein Evolution, and the Nearly Neutral Theory in Amniotes. *Molecular Biology and Evolution*, 2016, 33 (6), pp.1517 - 1527. 10.1093/molbev/msw033 . hal-01900644

**HAL Id: hal-01900644**

**<https://hal.science/hal-01900644v1>**

Submitted on 30 Oct 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Life History Traits, Protein Evolution, and the Nearly Neutral Theory in Amniotes

Emeric Figuet,<sup>1</sup> Benoît Nabholz,<sup>1</sup> Manon Bonneau,<sup>1</sup> Eduard Mas Carrio,<sup>1</sup> Krystyna Nadachowska-Brzyska,<sup>2</sup> Hans Ellegren,<sup>2</sup> and Nicolas Galtier<sup>\*,1</sup>

<sup>1</sup>UMR 5554 Institut des Sciences de l'Evolution, CNRS, Université de Montpellier, IRD, EPHE, Montpellier, France

<sup>2</sup>Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

\*Corresponding author: nicolas.galtier@univ-montp2.fr.

Associate editor: Jeffrey Thorne

## Abstract

The nearly neutral theory of molecular evolution predicts that small populations should accumulate deleterious mutations at a faster rate than large populations. The analysis of nonsynonymous ( $d_N$ ) versus synonymous ( $d_S$ ) substitution rates in birds versus mammals, however, has provided contradictory results, questioning the generality of the nearly neutral theory. Here we analyzed the impact of life history traits, taken as proxies of the effective population size, on molecular evolutionary and population genetic processes in amniotes, including the so far neglected reptiles. We report a strong effect of species body mass, longevity, and age of sexual maturity on genome-wide patterns of polymorphism and divergence across the major groups of amniotes, in agreement with the nearly neutral theory. Our results indicate that the rate of protein evolution in amniotes is determined in the first place by the efficiency of purifying selection against deleterious mutations—and this is true of both radical and conservative amino acid changes. Interestingly, the among-species distribution of  $d_N/d_S$  in birds did not follow this general trend:  $d_N/d_S$  was not higher in large, long-lived than in small, short-lived species of birds. We show that this unexpected pattern is not due to a more narrow range of life history traits, a lack of correlation between traits and  $N_e$ , or a peculiar distribution of fitness effects of mutations in birds. Our analysis therefore highlights the bird  $d_N/d_S$  ratio as a molecular evolutionary paradox and a challenge for future research.

**Key words:** substitution rate, synonymous, birds, mammals, reptiles, effective population size, longevity, body mass.

## Introduction

The effective population size ( $N_e$ ) is a central parameter of population genetic theory that determines the rate of genetic drift, that is, the strength of stochastic allele frequency fluctuations across generations.  $N_e$  is expected to influence not only neutral aspects of genetic diversity, such as within-species levels of genetic polymorphism, but also nonneutral processes. This is because genetic drift counteracts the action of natural selection: When stochastic effects dominate, weakly favorable mutations are less likely to be selected, whereas slightly deleterious mutations have a nonnegligible probability to reach fixation by chance. The nearly neutral theory therefore predicts that small populations should accumulate a heavier load of deleterious mutations than large ones (Ohta 1987; Lynch 2007; Akashi et al. 2012).

Amniotes are an ~320 My old group of vertebrates containing mammals and sauropsids, the latter represented by reptiles and birds. A number of studies have recently attempted to test the nearly neutral hypothesis in this group of organisms. Keightley et al. (2005) reported a higher conservation level of regulatory elements in rodents than in hominid primates and invoked a reduced  $N_e$  to explain the degradation of control regions in hominids. Eöry et al. (2010), however, detected a stronger level of constraint on 4-fold synonymous sites in primates than in murids, somewhat

countering the predicted role of  $N_e$  in modulating weakly selected sites. Analyzing exonic splice enhancers in mammals, Wu and Hurst (2015) suggested that selection for proper splicing could be stronger in small  $N_e$  than in large  $N_e$  species if the former have accumulated more numerous, longer introns.

Besides regulatory elements, a popular marker of selective constraint at the molecular level is the ratio of nonsynonymous ( $d_N$ ) to synonymous ( $d_S$ ) substitution rate in coding sequences. Synonymous mutations are often supposed to be mostly neutral in vertebrates (but see above), whereas a large number of nonsynonymous mutations are presumably deleterious (Eyre-Walker and Keightley 2007). Because the latter are expected to be more efficiently purged from large than small populations, theory predicts a reduced  $d_N/d_S$  ratio in species with large  $N_e$ . In mammals, several studies have reported a positive relationship between  $d_N/d_S$  and species body mass or maximal lifespan, both of which are expected to correlate negatively with  $N_e$  (Nikolaev et al. 2007; Popadin et al. 2007; Lartillot and Delsuc 2012; Romiguier et al. 2013; Figuet, Romiguier, et al. 2014). For example, based on a data set of ~1,000 orthologous nuclear genes, the average  $d_N/d_S$  ratio was reported to be above 0.17 in presumably small  $N_e$  apes, but below 0.13 in the presumably large  $N_e$  mouse, rat, and shrew (Romiguier et al. 2013), in agreement with theoretical expectations.

However, in birds, analysis of neither mitochondrial (Nabholz et al. 2013) nor nuclear (Lanfear et al. 2010; Weber, Nabholz, et al. 2014) coding sequences have revealed a positive relationship between  $d_N/d_S$  and species body mass, longevity, or sexual maturity; in fact, weak negative relationships were reported. These results, which were robust to various controls, cast doubts on the pertinence of the relationship previously reported in mammals. On the other hand, two studies have reported a significantly positive correlation between the  $K_R/K_C$  ratio, another marker of the selective pressure applying to protein sequences, and body mass in birds, adding to the confusion (Nabholz et al. 2013; Weber, Nabholz, et al. 2014).  $K_R/K_C$  measures the ratio of the rate of biochemically radical ( $K_R$ ) versus conservative ( $K_C$ ) amino acid changes (Zhang 2000; Sainudiin et al. 2005). Why this ratio would be affected by  $N_e$  is theoretically not so obvious and depends on the distribution of selection coefficients among radical and conservative mutations (Smith 2003).

There is thus only equivocal support for the nearly neutral theory in amniotes and several factors might be invoked to explain this unclear situation. First, it should be noted that  $N_e$  in the above-reviewed studies was approximated by species life history traits (LHTs), which are at best indirect indicators of the true  $N_e$ . Body mass is related to current population density in mammals (Damuth 1981; White et al. 2007) and birds (Nee et al. 1991), but whether it is relevant to the long-term  $N_e$  and similarly so in all groups of amniotes, remains to be demonstrated. Second, the  $d_N/d_S$  ratio is influenced not only by the efficiency of purifying selection against deleterious mutations, but also by the rate of fixation of adaptive amino acid changes, which might alter the relationship to LHT (Weber, Nabholz, et al. 2014). Specifically, the substitution rate for advantageous mutations (and therefore their contribution to  $d_N/d_S$ ) should be positively related to  $N_e$ . Third, GC-biased gene conversion (gBGC), a recombination-associated segregation distortion in favor of G and C alleles, is known to impact the  $d_N/d_S$  ratio (Galtier et al. 2009; Gossmann et al. 2014; Bolívar et al. 2016). Avian and mammalian genomes differ in many respects, and particularly regarding the distribution and conservation of local recombination rates (Kawakami et al. 2014), which might influence the efficacy of purifying and positive selection, and the role of gBGC. There is also the issue of the reliability of  $d_N$  and  $d_S$  estimates in the two groups, owing to substitutional saturation and compositional effects (Lartillot 2013). Finally, the above-reviewed studies have been heterogeneous in terms of analyzed gene sets and methods, which might have contributed to the discrepancy.

In this study, we took two complementary approaches to investigate further the among between LHTs,  $N_e$  and coding sequence evolution in amniotes. First, we extended the analysis to include nonavian sauropsids (“reptiles”) based on a data set we recently generated (Figuert, Ballenghien, et al. 2014). The paraphyletic reptiles comprise over 8,000 species of amniotes split into 4 orders: Squamata (lizard and snakes, ~7,900 species), Sphenodontia (tuatara, only 2 species), Testudines (turtles, around 300 species), and Crocodylia

(crocodiles and alligators, 25 species), the latter being the sister clade of birds (Janke and Arnason 1997). Nonavian sauropsids are critical for the interpretation of the previously reported relationships between  $d_N/d_S$  and LHTs in amniotes. If the pattern seen in birds was confirmed in reptiles, one should simply conclude that mammals and sauropsids behave differently. On the other hand, if analyses of nonavian sauropsids corroborated the mammalian response, then the nearly neutral interpretation of the  $d_N/d_S$  versus LHT relationship would be strengthened, and birds would appear as an interesting exception.

Second, we investigated the impact of LHTs on polymorphism levels. Genome-wide population data are only available in a relatively small number of amniote species, and mainly in mammals (Perry et al. 2012). But 45 recently published bird genomes were obtained by sequencing at high depth a single individual each (Zhang, Li, Li, et al. 2014), thus offering the opportunity to estimate heterozygosity. Provided that inbreeding levels are not too high, the genome-wide heterozygosity of a single individual informs on former population genetic processes and species-level diversity (Li and Durbin 2011; Nadachowska-Brzyska et al. 2015). If species LHTs were reliable predictors of  $N_e$ , they should affect diversity levels and  $d_N/d_S$ , as recently demonstrated at the Metazoa scale (Romiguier et al. 2014).

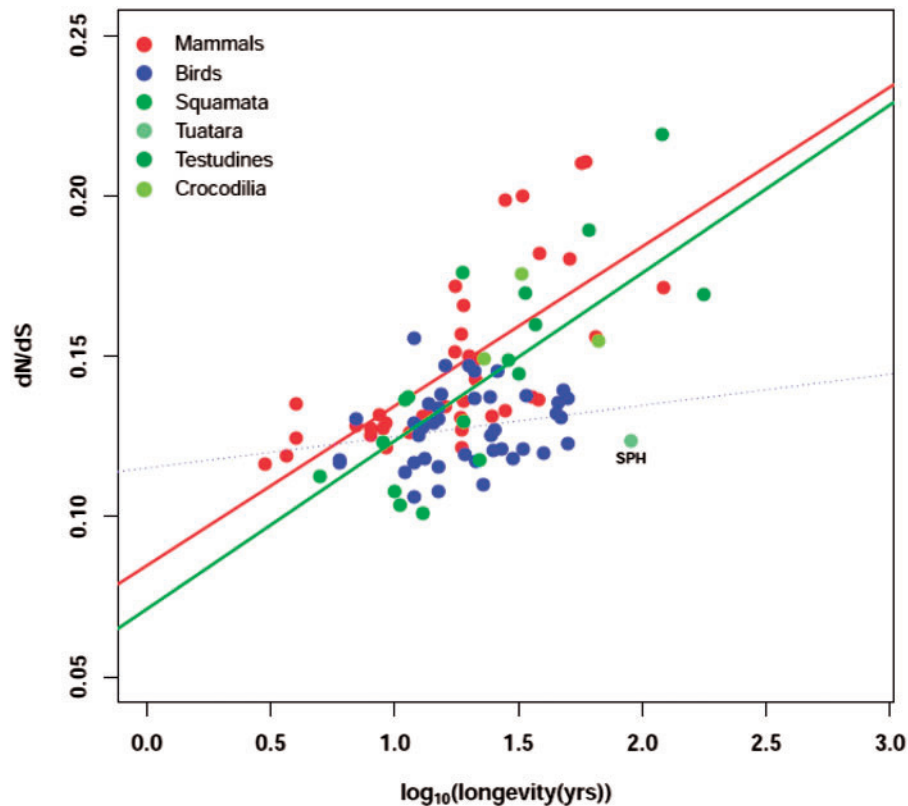
This battery of new analyses revealed a strong and consistent effect of LHTs on molecular variation patterns in amniotes, demonstrating that  $N_e$  deeply impacts genome evolution in this group of organisms in agreement with the nearly neutral theory. However, the lack of relationship between  $d_N/d_S$  and LHTs in birds remains a mysterious exception to this general trend.

## Results

### $d_N/d_S$ Versus Life History Traits in Amniotes

The relationship between  $d_N/d_S$  and LHTs was investigated in the major amniote lineages using a common set of 1,077 orthologous genes, in which synonymous and nonsynonymous substitutions were mapped to branches of the underlying phylogeny (Romiguier et al. 2012). Consistent with previous studies (Weber, Nabholz, et al. 2014), we failed to detect a positive relationship between  $d_N/d_S$  and any of the three analyzed LHTs in birds (fig. 1, table 1, and supplementary table S4, Supplementary Material online). Nonavian amniotes, in contrast, yielded strong positive relationships between  $d_N/d_S$  and longevity (fig. 1), body mass, and age of sexual maturity (table 1). The relationships were statistically significant within mammals, in agreement with previous reports, and within nonavian reptiles as far as longevity and age of sexual maturity were concerned (table 1).

The tuatara *Sphenodon punctatus* was an outlier (fig. 1), with a relatively low  $d_N/d_S$  ratio (0.123) despite an elevated maximal lifespan (90 years). This might be explained by the very long terminal branch for this genus, which is the single representative of a ~250 My old lineage (Mulcahy et al. 2012). Whether current LHTs representatively reflect long-term trait values are particularly uncertain for this species. Removing



**Fig. 1.** Relationship between  $\log_{10}(\text{longevity})$  and  $d_N/d_S$  in amniotes. Each dot corresponds to a terminal branch. Regression line for each group is drawn using the same color scheme. The tuatara (*Sphenodon punctatus*) is identified as SPH.

the tuatara resulted in a substantial increase in strength of correlations ( $d_N/d_S$  vs. longevity:  $r = 0.71$ ,  $P < 10^{-9}$  in nonavian amniotes;  $r = 0.78$ ,  $P < 10^{-4}$  in nonavian sauropsids).

Birds therefore appear as an exception among amniotes in showing no relationship between  $d_N/d_S$  and LHTs.  $d_N/d_S$  in short-lived birds was similar, on average, to that of short-lived mammals and reptiles, but no increase in  $d_N/d_S$  was detected in long-lived birds. Bird species displayed a narrower range of longevity than the other groups considered here. Still, the positive correlations between LHTs and  $d_N/d_S$  remained significant in mammals and nonavian sauropsids when we only selected species within these two groups for which traits were within the ranges observed in birds (supplementary table S7, Supplementary Material online).

The three LHTs used in this study are positively correlated to each other. A multiple regression analysis of  $d_N/d_S$  on all traits gave results in general agreement with the pairwise correlations reported in table 1. In mammals, the three traits had a significant marginal effect on  $d_N/d_S$  with a multiple  $r^2$  of 0.59. In nonavian sauropsids, the multiple  $r^2$  was still high (0.52), but only longevity kept a significant marginal effect in the multiple regression. In birds, no significant effect was detected, with the three traits collectively explaining no more than 4.5% of the between-species variance in  $d_N/d_S$ .

The reported relationships between  $d_N/d_S$  and LHTs were robust to alternative methodologies. Patterns very similar to the substitution mapping analysis were observed when  $d_N/d_S$  values were estimated using the codeml software from PAML (Yang 2007), or when the correlation between  $d_N/d_S$  and

LHTs was modeled as a multivariate Brownian motion in coEvol (Lartillot and Poujol 2011). These control analyses were performed on subsets of genes due to the elevated computational requirements of codeml and coEvol. The strength and statistical significance of the relationships were generally lower with these reduced data sets than with the complete one (supplementary methods, table S9, and fig. S4, Supplementary Material online). The integrated approach implemented in coEvol, in which phylogenetic inertia is directly modeled, represents the state-of-the-art but could not be applied to our entire data set due to computational limitations.

### Radical Versus Conservative Amino Acid Substitutions

We annotated nonsynonymous substitutions according to their biochemical impact and calculated the rate of radical ( $K_R$ ) and conservative ( $K_C$ ) amino acid substitutions. Using  $d_S$  as a neutral reference, we observed that the  $K_R/d_S$  and the  $K_C/d_S$  ratios behaved in a similar fashion to  $d_N/d_S$  in all groups of amniotes regarding their relationship to LHTs: Significant positive correlations were detected in nonavian amniotes, but no correlation was found in birds (table 1 and supplementary table S4, Supplementary Material online). These results suggest that the both radical and conservative mutations can result in mild deleterious effects and respond to variations in  $N_e$ , which complicates the interpretation of the  $K_R/K_C$  ratio under the nearly neutral theory. No significant correlation was detected between  $K_R/K_C$  and life history traits in

**Table 1.** Relationships between Molecular Substitution Rates and Life History Traits in Amniotes.

$d_N/d_S$	Body Mass		Longevity		Sexual Maturity	
	<i>r</i>	<i>P</i> value	<i>r</i>	<i>P</i> value	<i>r</i>	<i>P</i> value
Nonavian amniotes	0.37	0.0004	0.67	<10 <sup>-8</sup>	0.56	<10 <sup>-5</sup>
Mammals	0.38	0.017	0.69	<10 <sup>-5</sup>	0.71	<10 <sup>-6</sup>
Reptiles	0.34	0.13	0.68	<10 <sup>-3</sup>	0.52	0.019
Birds	0.13	0.32	0.21	0.20	0.08	0.62
$k_R/d_S$						
Nonavian amniotes	0.33	0.011	0.68	<10 <sup>-8</sup>	0.60	<10 <sup>-6</sup>
Mammals	0.39	0.015	0.70	<10 <sup>-6</sup>	0.72	<10 <sup>-6</sup>
Reptiles	0.31	0.17	0.65	0.001	0.47	0.036
Birds	0.28	0.06	0.26	0.10	0.21	0.16
$k_C/d_S$						
Nonavian amniotes	0.30	0.020	0.68	<10 <sup>-8</sup>	0.61	<10 <sup>-6</sup>
Mammals	0.40	0.012	0.73	<10 <sup>-6</sup>	0.73	<10 <sup>-6</sup>
Reptiles	0.32	0.16	0.65	0.001	0.48	0.032
Birds	0.01	0.96	0.15	0.36	-0.02	0.90
$k_R/k_C$						
Nonavian amniotes	0.26	0.051	0.23	0.09	0.17	0.20
Mammals	0.28	0.09	0.43	0.007	0.52	<10 <sup>-3</sup>
Reptiles	0.23	0.31	0.26	0.25	0.16	0.51
Birds	0.67	<10 <sup>-6</sup>	0.36	0.022	0.56	<10 <sup>-4</sup>

NOTE.—All life history trait values were log<sub>10</sub> transformed. Strength of correlations do not account for species relatedness. Significant correlations are indicated in bold.

nonavian amniotes (table 1). In birds, however, the  $K_R/K_C$  ratio showed strong positive correlations with LHTs, particularly with body mass (fig. 2). This is consistent with Weber, Boussau, et al. (2014), but still difficult to interpret knowing that neither  $K_R$  nor  $K_C$  did respond to LHTs in this group. Correlations between LHTs and substitution rates were globally robust to our control for phylogenetic dependence (supplementary methods and figs. S1 and S2, Supplementary Material online), with the exception of nonavian sauropsids for which fewer data points were available.

### Genome-Wide Patterns of Heterozygosity in Birds

A plausible explanation for the lack of relationship between  $d_N/d_S$  and LHTs in birds would be that LHTs are not, or only weakly, correlated with  $N_e$  in this group. To indirectly test this hypothesis, we analyzed genome-wide heterozygosity in 45 bird species and assumed that between-species variation in heterozygosity should in the first place be influenced by  $N_e$  (Romiguier et al. 2014). Single nucleotide polymorphism (SNP) density varied by an order of magnitude among species: from 500,000 to 7 million SNPs per genome were called, and heterozygosity levels varied between  $4.2 \times 10^{-4}$  and  $6.6 \times 10^{-3}$ . A significantly negative relationship was detected between heterozygosity and each of body mass ( $r = -0.39$ ,  $P < 10^{-2}$ ; fig. 3A and table 2), longevity ( $r = -0.45$ ,  $P < 10^{-2}$ ), and age of sexual maturity ( $r = -0.52$ ,  $P < 10^{-3}$ ). This result strongly suggests that LHTs are indeed correlated with  $N_e$  in birds.

Then we calculated for each bird species the average ratio of nonsynonymous to synonymous polymorphism,  $\pi_N/\pi_S$ ,

which is expected to be higher in small than in large populations under the nearly neutral theory, and deleterious nonsynonymous mutations reach higher frequencies when genetic drift is strong. We detected a significantly positive correlation between all three LHTs and  $\pi_N/\pi_S$ , in agreement with the hypothesis of a higher effective population size and more efficient purifying selection in small-sized, short-lived birds (fig. 3B and table 2). Correlation coefficients were similar for the three traits (between 0.41 and 0.50,  $P < 10^{-2}$ ) in this case. This result indicates that LHTs are reliable predictors of the strength of purifying selection in birds. No significant relationship was found between  $\pi_N/\pi_S$  and  $d_N/d_S$  in birds ( $r = 0.15$ ), which is an intriguing result. A similar analysis was conducted in mammals using a transcriptome-based 16 species data set. A significantly positive relationship between  $\pi_N/\pi_S$  and longevity was detected ( $r^2 = 0.47$ ,  $P = 0.0032$ ), consistent with the bird analysis. The correlation between  $\pi_S$  and longevity was negative, as expected, but not significant.

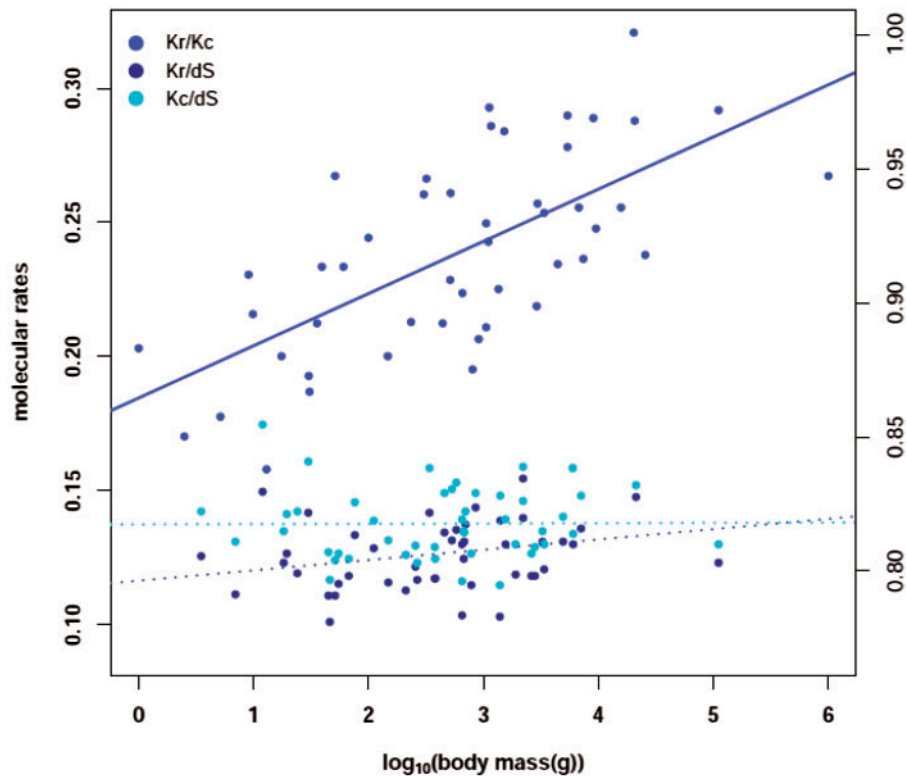
## Discussion

### Nearly Neutral Molecular Evolution in Amniotes

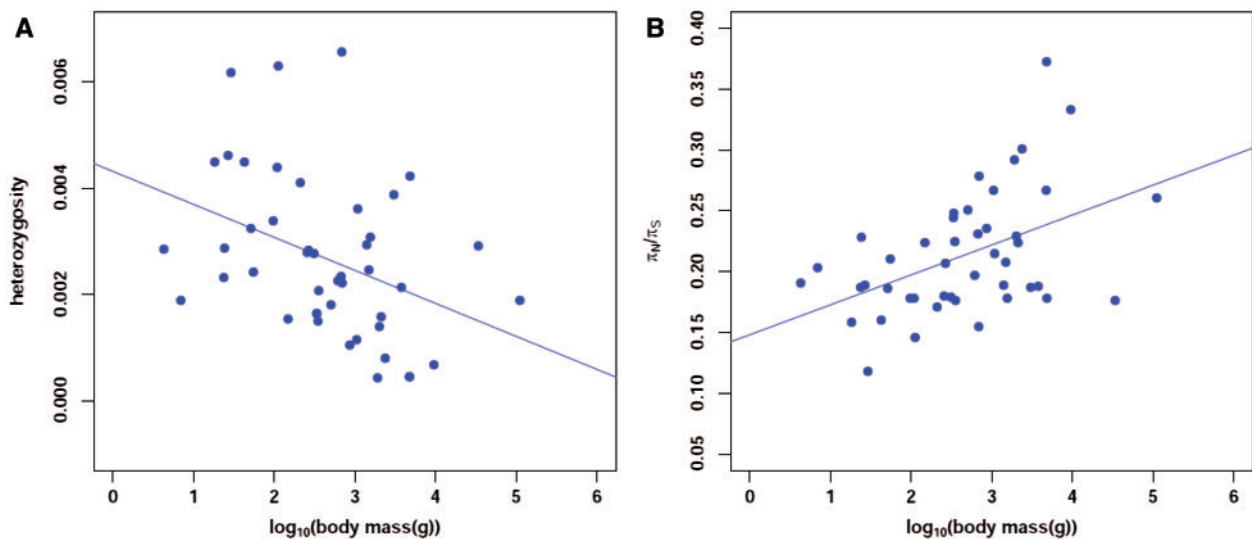
Earlier studies of the relationship between coding sequence substitution rates and LHTs in amniotes have been contradictory in reporting a positive relationship between  $d_N/d_S$  and longevity or body mass in mammals, but a negative relationship in birds, whereas  $K_R/K_C$  was responding positively in birds (Popadin et al. 2007; Lanfear et al. 2010; Romiguier et al. 2013; Nabholz et al. 2013; Weber, Nabholz, et al. 2014). These inconsistent results have cast doubts on the effect of  $N_e$  on protein evolution and on the pertinence of the nearly neutral theory (Ohta 1987) in amniotes. Our analysis adds three important results that help clarify the picture.

First, we observed that each of  $K_R/d_S$  and  $K_C/d_S$  responds to LHTs in a way similar to  $d_N/d_S$ , indicating that both radical and conservative nonsynonymous changes include slightly deleterious mutations. This complicates the interpretation of the  $K_R/K_C$  ratio, and its relationship to  $N_e$ , because of the unknown relative contribution of slightly deleterious mutations to the two categories of substitutions. Additional theoretical and empirical work is required to clarify the determinants and meaning of this variable. In mammals and reptiles, the  $K_R/K_C$  ratio was only a weak, if any, correlate of LHTs. The positive relationship detected between  $K_R/K_C$  and LHTs in birds is therefore intriguing, especially knowing that  $d_N/d_S$  does not respond to LHTs in this group (see below).

Second, using a common set of orthologs, we found a positive link between species longevity and  $d_N/d_S$  in both nonavian sauropsids and mammals, in agreement with the nearly neutral theory, but not in birds. The limited species sampling precluded a separate analysis of squamates, turtles, and crocodylians, so that we lack a formal proof that the relationship exists within each of these clades. However, figure 1 clearly illustrates the singularity of birds compared with other sauropsids, and particularly the relatively low level of  $d_N/d_S$  in long-lived birds.



**Fig. 2.** Relationship between  $\log_{10}(\text{body mass})$  and radical versus conservative amino acid substitution rates. Each dot corresponds to a terminal branch. Y-axis on the left is relevant to  $K_R/d_S$  and  $K_C/d_S$ . Y-axis on the right is relevant to  $K_R/K_C$ . Only the  $K_R/K_C$  relationship is statistically significant (plain regression line).



**Fig. 3.** Genome-wide patterns of heterozygosity in birds. (A) Relationship between  $\log_{10}(\text{body mass})$  and heterozygosity. (B) Relationship between  $\log_{10}(\text{body mass})$  and the ratio of nonsynonymous to synonymous SNPs in coding sequences.

Third, we found that heterozygosity and the  $\pi_N/\pi_S$  ratio correlate with species body mass, longevity, and age of sexual maturity in birds. This result demonstrates that LHTs do influence population genetic processes in birds. Large, long-lived birds carry a higher load of presumably deleterious mutations (i.e., nonsynonymous polymorphisms) than small, short-lived birds, presumably as a result of smaller population

sizes in the former and again in agreement with the nearly neutral theory.

Our analyses therefore indicate that species LHTs deeply influence patterns of genomic polymorphism and divergence in amniotes. The higher  $d_N/d_S$  and  $\pi_N/\pi_S$  ratio we report in large, long-lived organisms is in agreement with the hypothesis of a reduced efficiency of purifying selection in small

**Table 2.** Relationships between Genome-Wide Patterns of Heterozygosity and Life History Traits in Birds.

	Body Mass		Longevity		Sexual Maturity	
	<i>r</i>	<i>P</i> value	<i>r</i>	<i>P</i> value	<i>r</i>	<i>P</i> value
He	−0.39	0.008	−0.45	0.004	−0.52	<10 <sup>−3</sup>
π <sub>N</sub> /π <sub>S</sub>	0.46	0.001	0.41	0.009	0.50	<10 <sup>−3</sup>

NOTE.—All life-history trait values were log<sub>10</sub>-transformed. Strength of correlations do not account for species relatedness.

populations and that among-species variation in genome-average  $d_N/d_S$  is determined in the first place by nearly neutral processes. Amniote genomics thus provides empirical corroboration to Lynch’s (2007) hypothesis of an accumulating load of deleterious mutations in the genomes of small  $N_e$  species.

Available genome-wide estimates in other groups of animals suggest that the relationship we see is not restricted to amniotes (Romiguier et al. 2014). In the presumably very large  $N_e$  *Ciona intestinalis*, for instance, the genome-average  $\pi_N/\pi_S$  is  $\sim 0.05$  (Tsagkogeorga et al. 2012), that is, much lower than the values reported here in birds and mammals (typically above 0.12, and up to 0.37). This suggests that the vast majority of the nonsynonymous variation observed in large, long-lived amniote species, including humans, correspond to deleterious mutations that would be purged if  $N_e$  was larger in these taxa.

### Molecular Evolution and Life History Traits in Reptiles and Amniotes

The relationship between LHTs and  $d_N/d_S$  (or other measures of the efficiency of natural selection) in mammals has been interpreted in the first place as mediated by  $N_e$  through body mass, which is known to be inversely related to population density (Damuth 1981; Popadin et al. 2007). The link between longevity or age of sexual maturity and  $d_N/d_S$  that was subsequently uncovered (Romiguier et al. 2013; Nabholz et al. 2013) was merely thought as resulting from the positive relationship between these traits and body mass.

Here we show that, in contrast to mammals, body mass is not a good predictor of  $d_N/d_S$ ,  $K_R/d_S$ , or  $K_C/d_S$  in nonavian sauropsids. Maximal longevity and age of sexual maturity, on the other hand, are strongly correlated with molecular variables in reptiles (table 1). Why these two variables, but not body mass, would be related to  $N_e$  in reptiles is unclear. The three traits are expected to be more plastic—that is, dependent on the availability of resources—in ectotherms than in endotherms (Shine 2005), to the point that environmental variables such as latitude predicted rates of neutral evolution more accurately than did LHTs in turtles (Lourenço et al. 2013). Species of nonavian sauropsids showing a low  $d_N/d_S$  despite a relatively high body mass include golden tegu (*Tupinambis teguixin*) and king cobra (*Ophiophagus hannah*), whereas the opposite was seen in European pond turtle (*Emys orbicularis*) and cornsnake (*Pantherophis guttatus*).

The lack of a positive relationship between body mass and  $d_N/d_S$  in reptiles questions the dominant interpretation of body mass as the main driver of the relationship between LHTs and molecular evolution. In a comparative analysis of

the amount of genetic diversity across 90 species of animals, Romiguier et al. (2014) reported a significant but relatively weak effect of adult body mass, compared with species longevity and egg or juvenile size. The interpretation of these results are currently uncertain, but they suggest that the underlying ecological mechanisms might be more complex than previously thought: A dependence of population density to body mass mediated by metabolic requirements (Damuth 1981; Nee et al. 1991; White et al. 2007) is apparently not a sufficient explanation to explain the currently available data on LHTs versus molecular rates. Romiguier et al. (2014) speculated that long-lived species could be more resilient to fluctuations in density—for example, able to recover after severe bottlenecks—so that their minimal  $N_e$ , which determines the long-term  $N_e$ , would be smaller than that of short-lived species. Analyses of the relationships between LHTs and molecular rates in a variety of taxa would appear needed to confirm or infirm this suggestion, and clarify the underlying ecological processes.

### The Bird $d_N/d_S$ Exception

Our analysis of 1,077 genes did not corroborate the recent report of a negative relationship between  $d_N/d_S$  and body mass in birds (Weber, Nabholz, et al. 2014). The bird data set we used here differs from that of Weber, Nabholz, et al. (2014) in containing only genes for which orthologs are available in reptiles and mammals. It is therefore biased toward more highly constrained genes, as illustrated by the lower range of  $d_N/d_S$  we report (0.106–0.156), compared with Weber, Nabholz, et al. (2014) (0.13–0.175). This might contribute to explain the difference between the two analyses. At any rate, the two studies agree in not detecting a positive relationship between  $d_N/d_S$  and species body mass, longevity, or age of sexual maturity in birds, in spite of a species-rich data set with low amounts of missing data. This is particularly intriguing knowing that the nonavian sauropsid pattern is consistent with the prediction of the nearly neutral theory. A number of possible explanations to this unexpected result were investigated.

First, we showed that the narrower range of variation in LHTs in birds, compared with other amniotes, is not a sufficient explanation; the significant relationship recovered in both mammals and reptiles remained when we removed species showing extreme traits. Second, the hypothesis suggesting that LHTs would not be correlated to  $N_e$  in birds was contradicted by our analysis of genome-wide heterozygosity and  $\pi_N/\pi_S$ . The  $\pi_N/\pi_S$  analysis also demonstrates that the bird pattern is not due to a peculiar distribution of fitness effects of mutations in this group—there is apparently a substantial fraction of slightly deleterious mutations that occur and segregate in bird populations.

The distinctive response of  $\pi_N/\pi_S$  and  $d_N/d_S$  to LHTs might potentially result from a particularly fast rate of evolution of  $N_e$  in birds, such that the current  $N_e$ , which should determine  $\pi_N/\pi_S$ , would not reflect the long-term  $N_e$ , which determines  $d_N/d_S$ . Recent studies, however, have reported a negative relationship between coding sequence substitution rates and body mass in birds (Jarvis et al. 2014; Weber, Nabholz, et al.

2014). Dividing synonymous terminal branch lengths by estimated divergence times, we similarly detected a significantly faster per-year rate of evolution in small-sized than in large-sized birds ( $r = -0.71$ ,  $P < 1.10^{-7}$ ,  $n = 44$ ), presumably reflecting a higher per-year mutation rate in the former due to shorter generation time. Although indirect, these results at least indicate that LHTs are sufficiently conserved in time and across lineages to impact molecular evolution in the long run, questioning the hypothesis of fast evolving traits as an explanation to the bird  $d_N/d_S$  paradox. We also confirmed that the distribution of synonymous terminal branch length (supplementary fig. S3, Supplementary Material online), as well as the strength of correlations between the  $d_N$  and  $d_S$ , were similar in the three analyzed groups. This suggests that the bird data set is not more prone to saturation and estimation problems than in the other groups of amniotes.

One well-documented peculiarity of birds, among amniotes, is the high level of conservation of karyotype and recombination landscape (Backström et al. 2010; Kawakami et al. 2014). Recombination affects genome evolution not only by enhancing multilocus selection (Gossmann et al. 2014), but also through the action of gBGC, a meiotic segregation distortion process that favors G and C alleles against A and T ones (Galtier et al. 2001; Montoya-Burgos et al. 2003; Webster et al. 2005, 2006; Duret and Galtier 2009). gBGC has been documented in all groups of amniotes (Romiguier et al. 2010; Figuet, Ballenghien, et al. 2014; Weber, Boussau, et al. 2014), and is known to impact the  $d_N/d_S$  ratio (Webster and Smith 2004; Galtier and Duret 2007; Ratnakumar et al. 2010; Lartillot 2013; Bolívar et al. 2016), raising the possibility that gBGC affects the genome average  $d_N/d_S$  in a peculiar way in birds compared with other amniotes and confounds the effect of genetic drift.

We calculated in each species the among-genes average GC content at third codon positions (GC3). We found that GC3 is significantly negatively correlated with body mass ( $r = -0.61$ ,  $P < 1.10^{-5}$ ), longevity, and age of sexual maturity in birds, consistent with Nabholz et al. (2011), but not in other groups of amniotes (mammals:  $r = -0.20$ , NS for body mass; non-avian sauropsids:  $r = -0.18$ , NS; nonavian amniotes:  $r = -0.10$ , NS), in agreement with the idea that gBGC might modulate the impact of LHTs on molecular evolution in a specific way in birds. However, when we split the data sets in three bins of genes based on GC3 (thresholds of 40% and 50%), results largely similar to table 1 were obtained for each bin (supplementary table S8, Supplementary Material online). AT-rich genes, which have been largely immune from gBGC, responded similarly to GC-rich ones—although correlations were a bit lower in AT-rich genes. This preliminary analysis, however, does not reject the hypothesis of a possible role of gBGC in the unexpected bird pattern. Additional theoretical and empirical efforts are required to clarify the complex relationship between LHTs, genetic drift, gBGC, and coding sequence evolution. Finally, it should be recalled that mitochondrial genes are fully consistent with nuclear genes in supporting a positive relationship between  $d_N/d_S$  and

longevity in mammals (Popadin et al. 2007) but not in birds (Nabholz et al. 2013), even though gBGC is not documented, and very unlikely to be at work, in the mitochondrial genome.

The unexpected behavior of the avian  $d_N/d_S$  is therefore still largely mysterious. A number of hypotheses have been rejected in this study, but no satisfactory explanation to the observed pattern was identified. The impact of adaptive evolution, which can obviously influence the  $d_N/d_S$  ratio, still needs to be assessed. The effect of fluctuating  $N_e$  in relation to the shape of the distribution of fitness effects of mutations (Goldstein 2013) would also require more attention. Other biological models such as bats, which like birds experience peculiar longevity-related selective pressure compared with terrestrial mammals (de Magalhães et al. 2007; Munshi-South and Wilkinson 2010), could help exploring the effect of a flying lifestyle and its metabolic consequences on molecular evolution. Resolving the  $d_N/d_S$  paradox in birds would be of primary interest, as it might uncover a facet of the molecular evolutionary process that we currently do not understand well.

## Materials and Methods

### Coding Sequence Data and Alignments

#### Mammals

Coding sequence alignments of orthologous genes in 38 species were retrieved from the ORTHOMAM v8 database (Douzery et al. 2014), which is based on ENSEMBL v73 annotations. The highly divergent platypus (*Ornitorhynchus anatinus*) was not used owing to the risk of misalignment and substitutional saturation. Positions containing 30% of gaps/missing data or more were not considered. Alignments shorter than 200 bp or in which 12 species or more were missing were discarded, leading to a total of 12,939 coding sequences.

#### Birds

Coding sequence alignments of 8,295 orthologous genes in 47 species were obtained from Jarvis et al. (2014), in which orthology was predicted based on sequence similarity and synteny between chicken (*Gallus gallus*) and zebra finch (*Taeniopygia guttata*). The white-tailed eagle (*Haliaeetus albicilla*) was not used because of its close relatedness with another species of the same genus (bald eagle *Haliaeetus leucocephalus*). The same filtering criteria as in mammals were applied except that alignments in which six species or more were missing were discarded.

#### Non-Avian Sauropsids

Partial or complete coding sequences from 21 species from all major clades of reptiles (3 crocodylians, 6 turtles, 11 squamates, and 1 tuatara) were retrieved from annotated complete genomes (2 species) or transcriptomic studies (19 species; supplementary table S1, Supplementary Material online; Figuet, Ballenghien, et al. 2014). Orthology was predicted by first applying the OrthoMCL software (Li et al. 2003) with



default parameters to amino acid–translated sequences. Among the candidate clusters of orthologs returned by OrthoMCL, we kept those containing exactly one sequence from each of the two fully sequenced genomes (the green anole [*Anolis carolinensis*] and the Chinese softshell turtle [*Pelodiscus sinensis*]). Regarding the other 18 species, when more than one sequence per species per cluster was available, these were merged into a single sequence if they were either nonoverlapping or >95% similar to each other. Otherwise, the longest sequence was kept. Coding sequences were aligned using the frameshift-aware MACSE program (Ranwez et al. 2011). Obviously misaligned regions were semiautomatically cleaned. Positions containing 50% of gaps/missing data or more were not considered. Alignments shorter than 200 bp or in which 9 species or more were missing were discarded.

#### Shared Set of Genes

We used ENSEMBL v77 orthology prediction among green anole, chicken, and human genomes to restrict our data set to a common set of genes. This led to a total of 1,077 orthologous genes shared by mammals and sauropsids, representing around 1.7 Mb coding sequence length. The avian and mammalian alignments contained only 7% and 15% of gaps/missing data, respectively, but this proportion reached 46% in nonavian sauropsids.

#### Phylogenies and Divergence Dates

The topology of the trees used for phylogenetic analyses were adapted from recent publications (Man et al. 2011; Guillon et al. 2012; Pyron et al. 2013; Douzery et al. 2014; Jarvis et al. 2014; supplementary methods, Supplementary Material online). Estimated divergence dates for internal nodes of the bird phylogeny were obtained from the molecular dating analysis of Jarvis et al. (2014), in which a maximal constraint of 86.5 My was placed on Neornithes, and were used to estimate per-year substitution rates in terminal branches. We evaluated the robustness of our phylogenetic analyses to the choice of topology by employing the alternative trees from Prum et al. (2015) for birds and from Meredith et al. (2011) for mammals, and by switching the controversial (*Iguana*, (*Basiliscus*, *Anolis*)) node in nonavian sauropsids (Wiens et al. 2012; supplementary methods, Supplementary Material online).

#### $d_N/d_S$ and $K_R/K_C$ Estimation

The substitution mapping procedure was used to calculate the mean  $d_N/d_S$  of each species in the three groups (Dutheil et al. 2012; Romiguier et al. 2012). This consisted of first fitting the YN98 codon model, as implemented in the bppML program (Dutheil and Boussau 2008), to each alignment of protein-coding sequences. The estimated model parameters were then used to map synonymous and nonsynonymous substitutions to internal and terminal branches of the tree, respectively—that is, for each codon site and each branch, we calculated the expected number of synonymous and nonsynonymous substitutions conditional on initial and final state, summing across all possible pairs of initial and final states, weighting by their posterior probabilities (Romiguier et al.

2012). The number of synonymous and nonsynonymous sites were estimated by considering the mutational opportunities of each extant sequence following Nei and Gojori (1986), using the transition/transversion ratio estimated by the YN98 codon model. For each species, the number of synonymous and nonsynonymous sites and the number of synonymous and nonsynonymous substitutions in the corresponding terminal branch of the tree were summed across alignments and  $d_N/d_S$  was calculated by taking the ratio of the sums. For each species, only alignments in which both this species and at least one representative of its sister group (potentially a single species) were available were considered, thus ensuring a consistent definition of terminal branches across alignments despite incomplete taxon sampling. A similar procedure was adopted for the  $K_R/K_C$  ratio calculation. We defined radical substitutions as those substantially modifying amino acid volume or polarity according to the classification of Sainudiin et al. (2005). The WAG01 model was used to map radical and conservative substitutions, and the numbers of radical and conservative sites were estimated by considering mutational opportunities in the same way as for the  $d_N/d_S$  analysis.

Based on two subsets of 19 and 20 genes (~20 kb each), we also calculated  $d_N/d_S$  values using the codeml software from the PAML package (Yang 2007). We employed the YN98 codon model and assigned a different value of  $d_N/d_S$  to each terminal branch of the tree. As an additional control, we also applied the approach introduced in the coEvol program version 1.3 (Lartillot and Poujol 2011) where the correlated evolution of molecular variables with LHT is modeled using a multivariate Brownian diffusion process (supplementary methods, Supplementary Material online).

#### Life History Traits

Body mass, maximum longevity, and age of sexual maturity (in each case taking the mean of male and female trait values) were retrieved from the AnAge database build 13 (De Magalhães and Costa 2009). We used for each species the mean of all available species in its genus in order to get an estimate of the long-term trait average. In mammals, only genus *Nomascus* was missing from the database while there were 34 missing data points across the three traits in sauropsids. In the absence of information on AnAge, LHT data were obtained from the literature (supplementary tables S2 and S3, Supplementary Material online); six longevity and three maturity records were still missing in sauropsids after this. In birds, data on body mass was taken from the CRC Handbook of Avian body masses (Dunning 2007) (supplementary table S5, Supplementary Material online), and longevity and age of sexual maturity were obtained from the literature (supplementary tables S2 and S3, Supplementary Material online). All values were log<sub>10</sub> transformed before linear correlation analysis.

#### Bird Genome-Wide Heterozygosity

The reference genomes of 45 bird species were downloaded from gigaScience (Zhang, Li, Li, Gilbert, et al. 2014). Raw reads were downloaded from the Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/sra/>). For the 21 species with highest

coverage, we only used a fraction of the reads to ensure optimal comparability across species. Reads generated from libraries similar to those used in the less covered species (insert size between 500 and 800 bp) were kept in the first place (supplementary table S6, Supplementary Material online). We obtained a range of mean coverage from 22 to 55X among bird species.

We mapped reads onto the reference genome of each bird species with the BWA program (Li and Durbin 2009), using the paired-end algorithm “sampe.” Duplicate reads, which are supposed to result from polymerase chain reaction duplicate artifacts, were removed. Reads were realigned around potential indel positions using GATK IndelRealigner (McKenna et al. 2010). The mean coverage of each species was recalculated with GATK DepthOfCoverage.

We called SNPs using the GATK Unified Genotyper with default parameters. Only positions at which coverage was between one third and twice the mean coverage, mapping quality score above 30, base calling quality score above 30, and genotyping quality score above 35 were retained. Heterozygosity level was estimated by dividing the number of SNPs by the total number of sites meeting the coverage criteria. As a control, heterozygosity levels were also calculated following the procedure of Nadachowska-Brzyska et al. (2015), which led to very similar results.

We used the snpEff software (Cingolani et al. 2012) and coding sequence annotations from gigaScience (Zhang, Li, Li, Gilbert, et al. 2014) to annotate SNPs as noncoding, synonymous, or nonsynonymous. We calculated the  $\pi_N/\pi_S$  ratio of each species by dividing the number of nonsynonymous polymorphisms by three times the number of synonymous polymorphisms. Within-species coding sequence polymorphism was also investigated in 16 mammalian species using the  $\pi_N$  and  $\pi_S$  estimates of Romiguier et al. (2014).

### Regression Analyses

All statistical analyses were done with R. Correlations were approached parametrically using Pearson’s linear regression. In nonavian amniotes, the effect of phylogenetic inertia was tested through the method of phylogenetically independent contrasts (Felsenstein 1985) with the “ape” R package version 3.2 (Paradis et al. 2004) using the pic function. The bird tree contains very short internal branches near the root (Jarvis et al. 2014), for which contrasts are not expected to be particularly meaningful (Garland et al. 1992). Species relatedness in the bird phylogeny was therefore accounted for by keeping a single species per order (or other appropriate taxonomic level), thus making the tree essentially star-like and removing any expected covariance due to phylogenetic relatedness. No internal branch had a length above 0.005 in our modified tree. Detailed methodology and results for phylogenetic inertia analyses are given in supplementary methods and figs. S1 and S2, Supplementary Material online.

### Supplementary Material

Supplementary tables S1–S9, figures S1–S4, and methods are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

### Acknowledgments

We thank Jonathan Romiguier, Nicolas Lartillot, and three anonymous reviewers for helpful discussion. This work was supported by Agence Nationale de la Recherche grant ANR-10-BINF-01-01 (Ancestrôme), the Swedish Research Council, and Knut and Alice Wallenberg Foundation.

### References

- Akashi H, Osada N, Ohta T. 2012. Weak selection and protein evolution. *Genetics* 192:15–31.
- Backström N, Forstmeier W, Schielzeth H, Mellenius H, Nam K, Bolund E, Webster MT, Ost T, Schneider M, Kempnaers B, et al. 2010. The recombination landscape of the zebra finch *Taeniopygia guttata* genome. *Genome Res.* 20:485–495.
- Bolívar P, Mugal CF, Nater A, Ellegren H. 2016. Recombination rate variation modulates gene sequence evolution mainly via GC-biased gene conversion, not Hill-Robertson interference, in an avian system. *Mol Biol Evol.* 33:216–227.
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Ruden DM, Lu X. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118 F<sub>1</sub>; iso-2; iso-3. *Fly (Austin)* 6:80–92.
- Damuth J. 1981. Population density and body size in mammals. *Nature* 290:699–700.
- de Magalhães JP, Costa J. 2009. A database of vertebrate longevity records and their relation to other life-history traits. *J Evol Biol.* 22:1770–1774.
- de Magalhães JP, Costa J, Church GM. 2007. An analysis of the relationship between metabolism, developmental schedules, and longevity using phylogenetic independent contrasts. *J Gerontol A Biol Sci Med Sci.* 62:149–160.
- Douzery EJP, Scornavacca C, Romiguier J, Belkhir K, Galtier N, Delsuc F, Ranwez V. 2014. OrthoMaM v8: a database of orthologous exons and coding sequences for comparative genomics in mammals. *Mol Biol Evol.* 31:1923–1928.
- Dunning JB Jr. 2007. CRC handbook of avian body masses. 2nd ed. Boca Raton (FL): CRC Press.
- Duret L, Galtier N. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet.* 10:285–311.
- Dutheil J, Boussau B. 2008. Non-homogeneous models of sequence evolution in the Bio++ suite of libraries and programs. *BMC Evol Biol.* 8:255
- Dutheil JY, Galtier N, Romiguier J, Douzery EJ, Ranwez V, Boussau B. 2012. Efficient selection of branch-specific models of sequence evolution. *Mol Biol Evol.* 29:1861–1874.
- Eöry L, Halligan DL, Keightley PD. 2010. Distributions of selectively constrained sites and deleterious mutation rates in the hominid and murid genomes. *Mol Biol Evol.* 27:177–192.
- Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet.* 8:610–618.
- Felsenstein J. 1985. Phylogenies and the comparative method. *Am Nat.* 125:1–15.
- Figuet E, Ballenghien M, Romiguier J, Galtier N. 2014. Biased gene conversion and GC-content evolution in the coding sequences of reptiles and vertebrates. *Genome Biol Evol.* 7:240–250.
- Figuet E, Romiguier J, Dutheil JY, Galtier N. 2014a. Mitochondrial DNA as a tool for reconstructing past life-history traits in mammals. *J Evol Biol.* 27:899–910.
- Galtier N, Duret L. 2007. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. *Trends Genet.* 23:273–277.
- Galtier N, Duret L, Glémin S, Ranwez V. 2009. GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates. *Trends Genet.* 25:1–5.

- Galtier N, Piganeau G, Mouchiroud D, Duret L. 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics* 159:907–911.
- Garland T, Harvey P, Ives A. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Syst Biol* 41:18–32.
- Goldstein RA. 2013. Population size dependence of fitness effect distribution and substitution rate probed by biophysical model of protein thermostability. *Genome Biol Evol* 5:1584–1593.
- Gossmann TI, Santure AW, Sheldon BC, Slate J, Zeng K. 2014. Highly variable recombinational landscape modulates efficacy of natural selection in birds. *Genome Biol Evol* 6:2061–2075.
- Guillon J, Hulin V, Girondot M. 2012. A large phylogeny of turtles (Testudines) using molecular data. *Contrib Zool* 81:147–158.
- Janke A, Arnason U. 1997. The complete mitochondrial genome of *Alligator mississippiensis* and the separation between recent archosauria (birds and crocodiles). *Mol Biol Evol* 14:1266–1272.
- Jarvis ED, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT, Suh A, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346:1126–1138.
- Kawakami T, Smeds L, Backström N, Husby A, Qvarnström A, Mugal CF, Olason P, Ellegren H. 2014. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Mol Ecol* 23:4035–4058.
- Keightley PD, Lercher MJ, Eyre-Walker A. 2005. Evidence for widespread degradation of gene control regions in hominid genomes. *PLoS Biol* 3:e42
- Lanfear R, Ho SYW, Love D, Bromham L. 2010. Mutation rate is linked to diversification in birds. *Proc Natl Acad Sci U S A* 107:20423–20428.
- Lartillot N. 2013. Interaction between selection and biased gene conversion in mammalian protein-coding sequence evolution revealed by a phylogenetic covariance analysis. *Mol Biol Evol* 30:356–368.
- Lartillot N, Delsuc F. 2012. Joint reconstruction of divergence times and life-history evolution in placental mammals using a phylogenetic covariance model. *Evolution* 66:1773–1787.
- Lartillot N, Poujol R. 2011. A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. *Mol Biol Evol* 28:729–744.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
- Li H, Durbin R. 2011. Inference of human population history from whole genome sequence of a single individual. *Nature* 475:493–496.
- Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13:2178–2189.
- Lourenço JM, Glémin S, Chiari Y, Galtier N. 2013. The determinants of the molecular substitution process in turtles. *J Evol Biol* 26:38–50.
- Lynch M. 2007. The origins of genome architecture. Sunderland (MA): Sinauer Associates Inc.
- Man Z, Yishu W, Peng Y, Xiaobing W. 2011. Crocodylian phylogeny inferred from twelve mitochondrial protein-coding genes, with new complete mitochondrial genomic sequences for *Crocodylus acutus* and *Crocodylus novaeguineae*. *Mol Phylogenet Evol* 60:62–67.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303.
- Meredith RW, Janec JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A, Eizirik E, Stadler T, Rabosky DL, et al. 2011. Impacts of the cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* 334:521–524.
- Montoya-Burgos JI, Boursot P, Galtier N. 2003. Recombination explains isochores in mammalian genomes. *Trends Genet* 19:128–130.
- Mulcahy DG, Noonan BP, Moss T, Townsend TM, Reeder TW, Sites JW, Wiens JJ. 2012. Estimating divergence dates and evaluating dating methods using phylogenomic and mitochondrial data in squamate reptiles. *Mol Phylogenet Evol* 65:974–991.
- Munshi-South J, Wilkinson GS. 2010. Bats and birds: exceptional longevity despite high metabolic rates. *Ageing Res Rev* 9:12–19.
- Nabholz B, Künstner A, Wang R, Jarvis ED, Ellegren H. 2011. Dynamic evolution of base composition: causes and consequences in avian phylogenomics. *Mol Biol Evol* 28:2197–2210.
- Nabholz B, Uwimana N, Lartillot N. 2013. 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. *Genome Biol Evol* 5:1273–1290.
- Nadachowska-Brzyska K, Li C, Smeds L, Zhang G, Ellegren H. 2015. Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. *Curr Biol* 25:1375–1380.
- Nee S, Read A, Greenwood J, Harvey P. 1991. The relationship between abundance and body size in British birds. *Nature* 351:312–313.
- Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426.
- Nikolaev SI, Montoya-Burgos JI, Popadin K, Parand L, Margulies EH, Antonarakis SE. 2007. Life-history traits drive the evolutionary rates of mammalian coding and noncoding genomic elements. *Proc Natl Acad Sci U S A* 104:20443–20448.
- Ohta T. 1987. Very slightly deleterious mutations and the molecular clock. *J Mol Evol* 26:1–6.
- Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20:289–290.
- Perry GH, Melsted P, Marioni JC, Wang Y, Bainer R, Pickrell JK, Michelini K, Zehr S, Yoder AD, Stephens M, et al. 2012. Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Res* 22:602–610.
- Popadin K, Polishchuk LV, Mamirova L, Knorre D, Gunbin K. 2007. Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals. *Proc Natl Acad Sci U S A* 104:13390–13395.
- Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, Lemmon AR. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* 526:11–13.
- Pyron RA, Burbrink FT, Wiens JJ. 2013. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evol Biol* 13:1–53.
- Ranwez V, Harispe S, Delsuc F, Douzery EJP. 2011. MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS One* 6:e22594
- Ratnakumar A, Mousset S, Glémin S, Berglund J, Galtier N, Duret L, Webster MT. 2010. Detecting positive selection within genomes: the problem of biased gene conversion. *Philos Trans R Soc Lond B Biol Sci* 365:2571–2580.
- Romiguier J, Figuet E, Galtier N, Douzery EJP, Boussau B, Dutheil JY, Ranwez V. 2012. Fast and robust characterization of time-heterogeneous sequence evolutionary processes using substitution mapping. *PLoS One* 7:e33852
- Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, Chiari Y, Dornat R, Duret L, Faivre N, et al. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* 515:261–263.
- Romiguier J, Ranwez V, Douzery EJP, Galtier N. 2010. Contrasting GC-content dynamics across 33 mammalian genomes: relationship with life-history traits and chromosome sizes. *Genome Res* 20:1001–1009.
- Romiguier J, Ranwez V, Douzery EJP, Galtier N. 2013. Genomic evidence for large, long-lived ancestors to placental mammals. *Mol Biol Evol* 30:5–13.
- Sainudiin R, Wong WSW, Yogeewaran K, Nasrallah JB, Yang Z, Nielsen R. 2005. Detecting site-specific physicochemical selective pressures: applications to the Class I HLA of the human major histocompatibility complex and the SRK of the plant sporophytic self-incompatibility system. *J Mol Evol* 60:315–326.
- Shine R. 2005. Life-history evolution in reptiles. *Annu Rev Ecol Evol Syst* 36:23–46.
- Smith NGC. 2003. Are radical and conservative substitution rates useful statistics in molecular evolution? *J Mol Evol* 57:467–478.

- Tsagkogeorga G, Cahais V, Galtier N. 2012. The population genomics of a fast evolver: high levels of diversity, functional constraint, and molecular adaptation in the tunicate *Ciona intestinalis*. *Genome Biol Evol.* 4:740–749.
- Weber CC, Boussau B, Romiguier J, Jarvis ED, Ellegren H. 2014. Evidence for GC-biased gene conversion as a driver of between-lineage differences in avian base composition. *Genome Biol.* 15:1–16.
- Weber CC, Nabholz B, Romiguier J, Ellegren H. 2014.  $K_R/K_C$  but not  $d_N/d_S$  correlates positively with body mass in birds, raising implications for inferring lineage-specific selection. *Genome Biol.* 15:1–13.
- Webster MT, Axelsson E, Ellegren H. 2006. Strong regional biases in nucleotide substitution in the chicken genome. *Mol Biol Evol.* 23:1203–1216.
- Webster MT, Smith NGC. 2004. Fixation biases affecting human SNPs. *Trends Genet.* 20:116–122.
- Webster MT, Smith NGC, Hultin-Rosenberg L, Arndt PF, Ellegren H. 2005. Male-driven biased gene conversion governs the evolution of base composition in human alu repeats. *Mol Biol Evol.* 22:1468–1474.
- White EP, Ernest SKM, Kerkhoff AJ, Enquist BJ. 2007. Relationships between body size and abundance in ecology. *Trends Ecol Evol.* 22:323–330.
- Wiens JJ, Hutter CR, Mulcahy DG, Noonan BP, Townsend TM, Sites JW Jr, Reeder TW. 2012. Resolving the phylogeny of lizards and snakes (Squamata) with extensive sampling of genes and species. *Biol Lett.* 8:1043–1046.
- Wu X, Hurst LD. 2015. Why selection might be stronger when populations are small: intron size and density predict within and between-species usage of exonic splice associated *cis*-motifs. *Mol Biol Evol.* 32:1847–1861.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.
- Zhang G, Li B, Li C, Gilbert M, Jarvis ED, Wang J. 2014. Comparative genomic data of the Avian Phylogenomics Project. *Gigascience* 3:1–8.
- Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, et al. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* 346:1311–1320.
- Zhang J. 2000. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *J Mol Evol.* 50:56–68.