



**HAL**  
open science

# Large Variation in the Ratio of Mitochondrial to Nuclear Mutation Rate across Animals: Implications for Genetic Diversity and the Use of Mitochondrial DNA as a Molecular Marker

Remi Allio, Stefano Donega, Nicolas Galtier, Benoit Nabholz

► **To cite this version:**

Remi Allio, Stefano Donega, Nicolas Galtier, Benoit Nabholz. Large Variation in the Ratio of Mitochondrial to Nuclear Mutation Rate across Animals: Implications for Genetic Diversity and the Use of Mitochondrial DNA as a Molecular Marker. *Molecular Biology and Evolution*, 2017, 34 (11), pp.2762 - 2772. 10.1093/molbev/msx197 . hal-01815350

**HAL Id: hal-01815350**

**<https://hal.science/hal-01815350>**

Submitted on 15 Jun 2021

**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.



Distributed under a Creative Commons Attribution 4.0 International License

# Large Variation in the Ratio of Mitochondrial to Nuclear Mutation Rate across Animals: Implications for Genetic Diversity and the Use of Mitochondrial DNA as a Molecular Marker

Remi Allio, Stefano Donega, Nicolas Galtier, and Benoit Nabholz\*

ISEM, Univ. Montpellier, CNRS, IRD, EPHE, Montpellier, France

\*Corresponding author: E-mail: benoit.nabholz@umontpellier.fr.

Associate editor: Nadia D. Singh

## Abstract

It is commonly assumed that mitochondrial DNA (mtDNA) evolves at a faster rate than nuclear DNA (nuDNA) in animals. This has contributed to the popularity of mtDNA as a molecular marker in evolutionary studies. Analyzing 121 multilocus data sets and four phylogenomic data sets encompassing 4,676 species of animals, we demonstrate that the ratio of mitochondrial over nuclear mutation rate is highly variable among animal taxa. In nonvertebrates, such as insects and arachnids, the ratio of mtDNA over nuDNA mutation rate varies between 2 and 6, whereas it is above 20, on average, in vertebrates such as scaled reptiles and birds. Interestingly, this variation is sufficient to explain the previous report of a similar level of mitochondrial polymorphism, on average, between vertebrates and nonvertebrates, which was originally interpreted as reflecting the effect of pervasive positive selection. Our analysis rather indicates that the among-phyla homogeneity in within-species mtDNA diversity is due to a negative correlation between mtDNA per-generation mutation rate and effective population size, irrespective of the action of natural selection. Finally, we explore the variation in the absolute per-year mutation rate of both mtDNA and nuDNA using a reduced data set for which fossil calibration is available, and discuss the potential determinants of mutation rate variation across genomes and taxa. This study has important implications regarding DNA-based identification methods in predicting that mtDNA barcoding should be less reliable in nonvertebrates than in vertebrates.

**Key words:** mitochondria, mutation rate, substitution rate, polymorphism, natural selection.

## Introduction

In animals, mitochondrial DNA (mtDNA) is assumed to experience a higher mutation rate than nuclear DNA (nuDNA) (Ballard and Whitlock 2004). Surprisingly, despite the popularity of mtDNA as a marker in evolutionary studies, this assumption only relies on a handful of comparisons involving mostly vertebrates species (Brown et al. 1979; Miyata et al. 1982; Nabholz et al. 2009). The situation is much less clear in nonvertebrates. Depending on species, a mtDNA mutation rate much higher (Thomas and Wilson 1991; Blouin et al. 1998; Oliveira et al. 2008; Willett 2012), similar (Powell et al. 1986; Vawter and Brown 1986; Brower and DeSalle 1998), or lower (Shearer et al. 2002; Hellberg 2006; Huang et al. 2008) than the nuDNA mutation rate has been reported. Taking a comparative approach, Lynch et al. (2006) reported contrasted values for the ratio of mtDNA to nuDNA mutation rate ( $\mu_{mit}/\mu_{nuc}$ ) among animals. In this analysis,  $\mu_{mit}/\mu_{nuc}$  varied from 0.57 and 1.43 in corals and mosquitoes, respectively, to around 18 in crickets and aphids and up to 25 in vertebrates. More recently, Havird and Sloan (2016) also reported a large range of  $\mu_{mit}/\mu_{nuc}$  in animals and plants with flies (Diptera) showing a clearly lower  $\mu_{mit}/\mu_{nuc}$  than mammals. These analyses, however, were based on a limited number of data sets (less than ten nonvertebrate taxa), so

that the extent of the variation in  $\mu_{mit}/\mu_{nuc}$  and its distribution among taxonomic groups is currently unclear.

Understanding this variation has several important implications. mtDNA is by far the most frequently used marker for DNA-based species identification (the so-called “DNA barcoding” (Hebert et al. 2003; Hebert and Gregory 2005) and biodiversity monitoring using environmental DNA (Thomsen et al. 2012). The success rate of species identification using barcoding techniques is known to vary among taxonomic groups (Meier et al. 2006; Elias et al. 2007); whether this pertains to the variation in mtDNA mutation rate is worth investigating. Full mitochondrial genome data, typically obtained from next-generation sequencing, are also widely used in phylogenetic analyses (Botero-Castro et al. 2013; Barker 2014; Fabre et al. 2014; Tilak et al. 2014). Characterizing the taxonomic variation in mtDNA vs. nuDNA evolutionary rate should help assessing the performance of the mtDNA marker, compared to nuclear genes, for phylogenetic purposes.

Importantly, the  $\mu_{mit}/\mu_{nuc}$  ratio is also relevant to the interpretation of patterns of within-species polymorphism across genomes and species (Ellegren and Galtier 2016). In a meta-analysis of 1,683 species of animals, Bazin et al. (2006) found that the level of mtDNA polymorphism was surprisingly constant among taxonomic groups and uncorrelated to

nuDNA polymorphism. This was interpreted as reflecting the effect of positive selection on mtDNA evolution. Under neutrality the genetic polymorphism is governed by the mutation/drift balance, but the fixation of a newly arising, positively selected allele (that is, a selective sweep) leads to complete loss of genetic diversity at linked loci—a process known as genetic hitchhiking (Maynard-Smith and Haigh 1974). If pervasive, genetic hitchhiking is expected to homogenize the amount of polymorphism among species and erase the relationship between polymorphism and  $N_e$  (Gillespie 2001). This hypothesis appeared plausible because the animal mitochondrial genome, which is nonrecombining and gene dense, is expected to be more strongly affected by linked selection than nuclear loci (Hurst and Jiggins 2005; Bazin et al. 2006; Berlin et al. 2007). This interpretation, however, was debated because a correlation between mtDNA and nuDNA polymorphism was reported within specific groups such as mammals (Mulligan et al. 2006; Nabholz et al. 2008b; Piganeau and Eyre-Walker 2009).

Much of this discussion has focused on the effects of  $N_e$  and selection, in absence of large scale data on the relative mtDNA vs. nuDNA mutation rate. This is, we believe, in large part due to the fact that to explain the lack of relationship between mtDNA and nuDNA diversity, the  $\mu_{mit}/\mu_{nuc}$  ratio would have to exactly compensate for the difference in average  $N_e$  between taxa, a hypothesis deemed implausible by Bazin et al. (2006). However, more recently, Piganeau and Eyre-Walker (2009) have suggested the existence of a negative relationship between mitochondrial mutation rate per generation and  $N_e$  in mammals. Theoretically, the mutation rate is expected to evolve towards a lower limit determined by the power of random genetic drift ( $1/N_e$ ), and therefore could be negatively correlated to  $N_e$  (Lynch 2010). There is no obvious reason, however, why this relationship should apply specifically to the mitochondrial genome, and not the nuclear genome.

In this study, we take advantage of the availability of phylogenetic data sets in a large number of taxa to revisit the mutation rate hypothesis. Analyzing 121 multilocus and four phylogenomic data sets, we estimate the  $\mu_{mit}/\mu_{nuc}$  ratio in a variety of animal phyla and show that the variation in  $\mu_{mit}/\mu_{nuc}$  is indeed sufficient to explain the absence of correlation between mtDNA and nuDNA polymorphism. Taxa with a high average nuclear genetic diversity (such as insects or mollusks) experience a mtDNA mutation rate that is close to the nuDNA mutation rate, leading to similar level of polymorphism in the two genomic compartments. In contrast, taxa with a low average nuclear genetic diversity (such as birds, mammals, or scaled reptiles) have a mtDNA mutation rate markedly higher than the nuDNA mutation rate, leading to a mtDNA polymorphism much more elevated than the nuDNA polymorphism. Controlling for the variation in  $\mu_{mit}/\mu_{nuc}$  we were able to recover a positive correlation between nuDNA and mtDNA polymorphism. We conclude that there is no need to invoke positive selection to explain the results of Bazin et al. (2006), and discuss the causes and consequences of  $\mu_{mit}/\mu_{nuc}$  variation across animal phyla.

## Results

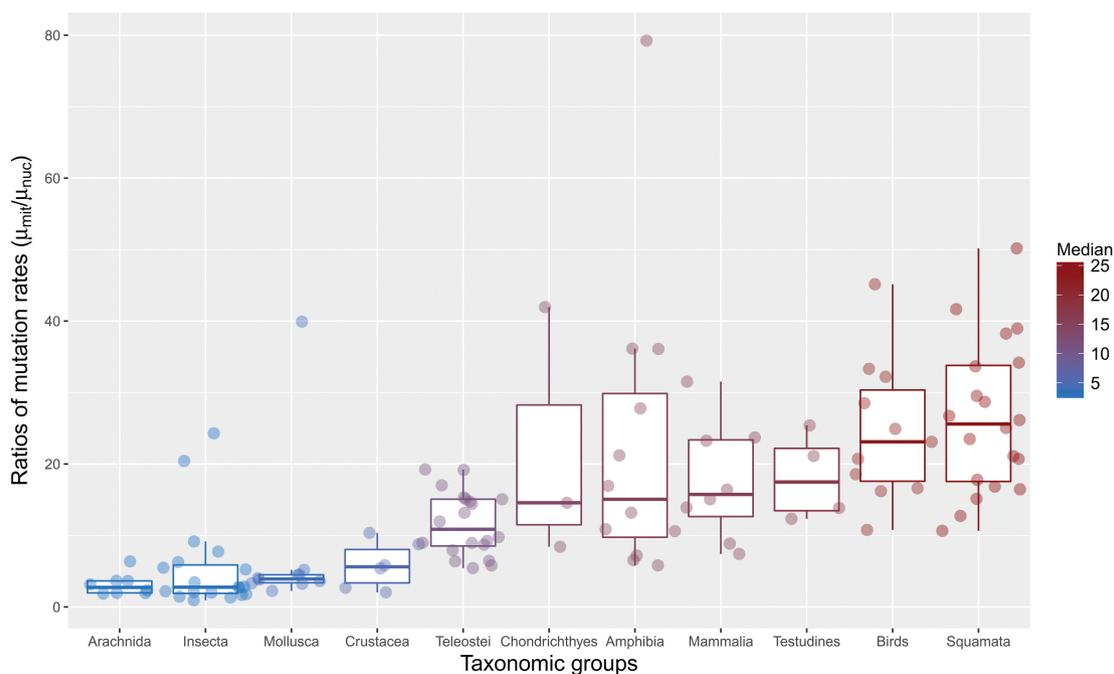
### Literature Review for Multilocus Data sets

We performed a systematic review of volumes 62–97 of the journal *Molecular Phylogenetics and Evolution* (January 2012–April 2016) and retrieved 122 data sets comprising at least one mitochondrial and one nuclear alignment covering a common set of at least five closely related species (supplementary table S1, Supplementary material online). We sorted these data sets according to 12 taxonomic groups, each of them represented by a variable number of data sets: Squamata  $N = 20$ ; Teleostei  $N = 20$ ; Insecta  $N = 19$ ; Amphibia  $N = 13$ ; Aves  $N = 11$ ; Mollusca  $N = 10$ ; Mammalia  $N = 8$ ; Arachnida  $N = 8$ ; Crustacea  $N = 6$ ; Testudines  $N = 4$ ; Chondrichthyes  $N = 3$ ; Porifera  $N = 1$ . These groups are similar to those used by Bazin et al. (2006) except that we split “Pisces” in Teleostei (bony fishes) and Chondrichthyes (cartilaginous fishes) and Sauropsida in Aves (birds), Squamata (scaled reptiles) and Testudines (turtles). We also retrieved data in Arachnida, but did not find any suitable Echinoderm data set. Our single sponge (Porifera) data set yielded a ratio of mitochondrial to nuclear mutation rate lower than 1 ( $\mu_{mit}/\mu_{nuc} = 0.52$ ), which is consistent with Huang et al. (2008). Porifera were excluded from further analyses due to insufficient sample size ( $N = 1$ ). The final data sets comprised an average 38.6 species, 1.4 mitochondrial genes and 2.1 nuclear genes.

### Extensive Variation in the Ratio of Mitochondrial to Nuclear Mutation Rate

The expected molecular divergence at neutrally evolving sites is equal to the product of mutation rate by divergence time (Kimura 1968; Birky and Walsh 1988). We selected the same set of species for mtDNA and nuDNA, such that the ratio of molecular divergences estimates the ratio of mutation rates, assuming that the times of divergences are similar for mtDNA and nuDNA (i.e., neglecting the variation in ancestral coalescence time among loci). To limit the impact of natural selection, we only analyzed third codon positions and estimated branch lengths using classical phylogenetic maximum likelihood methods. Molecular divergence was defined as the sum of branch lengths, and the  $\mu_{mit}/\mu_{nuc}$  ratio was estimated by dividing the mitochondrial molecular divergence by the nuclear one. The average number of substitutions per phylogeny was 3,377 and 443 for mitochondrial and nuclear loci, respectively.

The estimated ratio of mutation rates ( $\mu_{mit}/\mu_{nuc}$ ) was highly variable among taxonomic groups (fig. 1, ANOVA testing the effect of taxonomical groups on  $\mu_{mit}/\mu_{nuc}$  variation,  $R^2 = 0.63$ ,  $P < 2.2e-16$ ). In Arachnida, the average mtDNA mutation rate ( $\mu_{mit}$ ) was only 3.1 times higher than the average nuDNA mutation rate ( $\mu_{nuc}$ ), whereas in Squamata (scaled reptiles) mitochondrial genes had a mutation rate 26.4 times higher than nuclear genes, on average. Among the 22 pairwise comparisons revealing a significant difference, 21 involved a comparison between a vertebrate and a non-vertebrate taxon (Tukey Test, adjusted  $p < 0.01$ ). No statistical difference could be detected among nonvertebrate



**Fig. 1.** Across-taxa distribution of the ratios of mitochondrial to nuclear mutation rate in animals. Boxes give the quartiles; whiskers extend to 1.5 times the interquartile range. Dots indicate ratio of individual data sets. Colors vary according to the value of the median of the taxon.

groups, whereas only one comparison was significant among vertebrate groups, namely, the Teleostei–Squamata comparison (Tukey Test,  $P = 0.003$ , Teleostei have a ratio of 11.6 on average). Between vertebrates and nonvertebrates, 21 comparisons out of 28 showed statistically significant differences (Tukey Test, adjusted  $P < 0.01$ ).

A number of surprising outliers were uncovered in several groups. The data set with the highest  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  ratio in nonvertebrates was a mollusk ( $\mu_{\text{mit}}/\mu_{\text{nuc}} = 39.9$ ), whereas all the other mollusks showed a ratio close to 4 (fig. 1). This extreme data set is composed of 95 species of land snails (Pulmonata: Bradybaenidae, Hirano et al. 2014) and we have no strong reason to question the reliability of this estimate. Similarly, in insects, two data sets yielded a  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  ratio above 20, with the highest ratio obtained from a data set composed of 127 species of large Australian cicada ( $\mu_{\text{mit}}/\mu_{\text{nuc}} = 24.3$ , Hemiptera: Cicadidae; Owen et al. 2015). The data set with the largest ratio of all is composed of only ten species of Eurasian toads (Anura: Bufonidae;  $\mu_{\text{mit}}/\mu_{\text{nuc}} = 79.2$ ; Recuero et al. 2012) and therefore might be affected by sampling error.

### Extensive Variation of $\mu_{\text{Mit}}/\mu_{\text{Nuc}}$ Is Supported by Genome-Wide Data sets

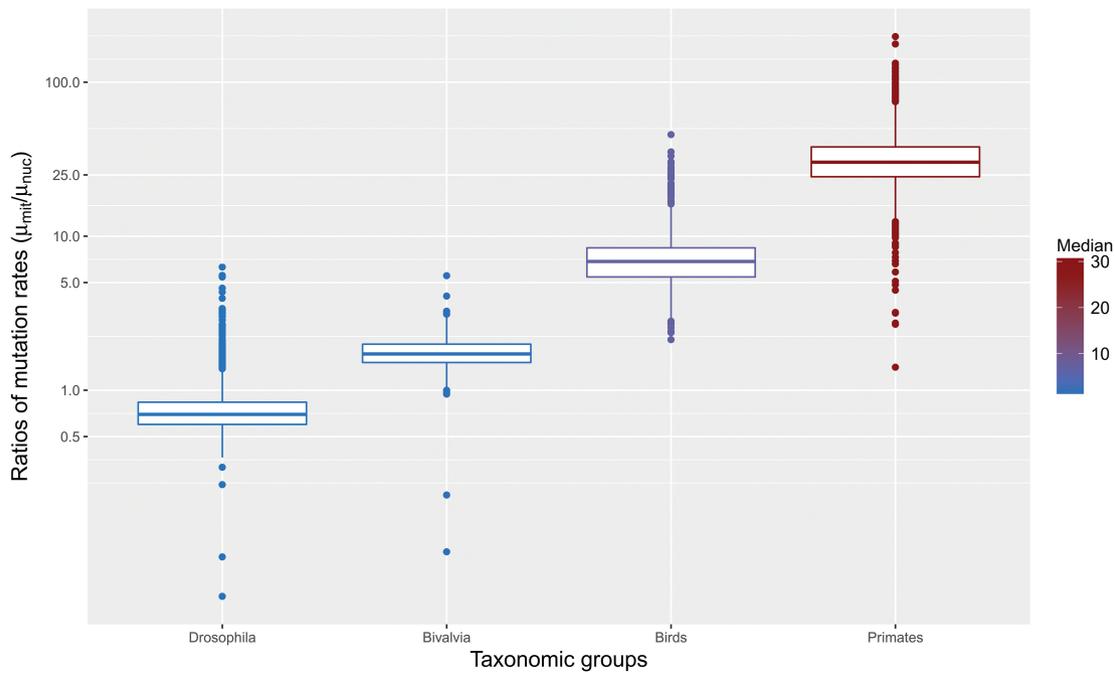
Each of the data sets analyzed above included just a handful of genes. To assess the genome-wide variation in  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  we used four phylogenomic data sets—two in vertebrates (primates and birds, 5,566 and 3,146 genes, respectively) and two in nonvertebrates (*Drosophila* and *Bivalvia*, 8,359 and 398 genes, respectively). In this analysis, we computed a  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  per nuclear gene by dividing the mitochondrial divergence (averaged across all mitochondrial genes) by the divergence of each individual nuclear gene. Some among-nuclear gene variation in  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  was detected in all four

taxonomic groups, but the distribution of  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  was clearly influenced in the first place by taxonomy (fig. 2, ANOVA,  $R^2 = 0.96$ ,  $P < 2.2e-16$ ). The mean  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  per gene varied from 0.8 in *Drosophila* and 1.8 in bivalves to 7.4 in birds and 32.5 in primates. This genome-wide analysis therefore confirms the existence of a conspicuous difference in  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  between vertebrate and nonvertebrate taxa, and demonstrates that the effect reported in figure 1 is not specific to the genes typically used in phylogenetic analyses.

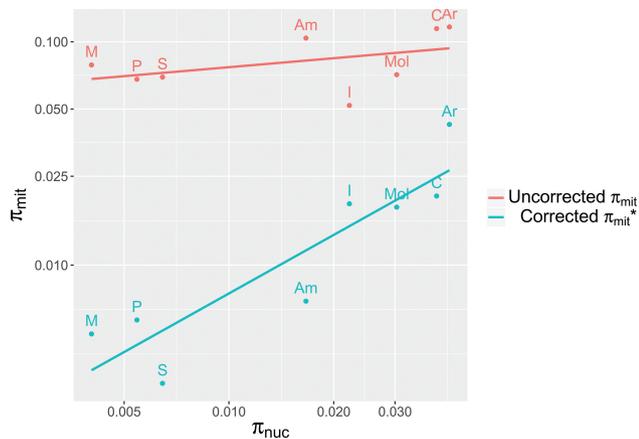
### Correcting for Mutation Rates Recovers the Correlation between mtDNA and nuDNA Coalescence Time

Our data clearly demonstrate that mtDNA and nuDNA mutation rates do not vary in a similar way across metazoan taxa. This result calls for a reappraisal of the relationship between mtDNA and nuDNA polymorphism (Bazin et al. 2006). To this aim, we computed the corrected mtDNA polymorphism,  $\pi_{\text{mit}}^*$ , defined as the mtDNA polymorphism  $\pi_{\text{mit}}$  divided by the ratio of mutation rates  $\mu_{\text{mit}}/\mu_{\text{nuc}}$ . Assuming selective neutrality and a balanced sex ratio, the corrected mtDNA polymorphism is expected to equal 1/4 of the nuDNA polymorphism ( $\pi_{\text{nuc}}$ ). We obtained the  $\pi_{\text{mit}}$  and  $\pi_{\text{nuc}}$  values used by Bazin et al. (2006) and use the  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  ratio estimated above.

As reported by Bazin et al. (2006), the average  $\pi_{\text{mit}}$  varied little among taxa and was not correlated to the average  $\pi_{\text{nuc}}$  (Spearman's rank correlation  $\rho = 0.52$ ,  $P = 0.20$ , fig. 3). Once corrected for  $\mu_{\text{mit}}/\mu_{\text{nuc}}$ , the correlation coefficient increased and became statistically significant (Spearman's rank correlation  $\rho = 0.90$ ,  $P = 0.004$ , fig. 3), revealing a higher average mtDNA coalescence time in species with an elevated  $\pi_{\text{nuc}}$ . Our results therefore suggest that the absence of correlation



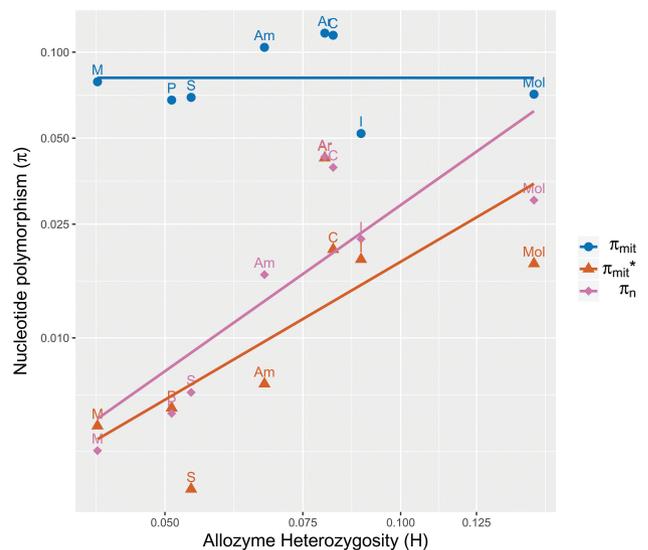
**FIG. 2.** Across-genes distribution of the ratio of mitochondrial to nuclear mutation rate in four phylogenomic data sets. Boxes give the quartiles; whiskers extend to 1.5 times the interquartile range. Dots indicate ratio of individual genes above or below the whiskers limit. Colors vary according to the value of the median of the taxon. Y-axis is log-transformed.



**FIG. 3.** Relationship between mitochondrial and nuclear average synonymous diversity across metazoan taxa. Y-axis: uncorrected (red) and mutation rate-corrected (blue) average level of synonymous polymorphism in mtDNA X-axis: average level of synonymous polymorphism in nuclear DNA. Y-axis and X-axis are log-transformed. Data from Bazin et al. (2006). Am: Amphibia, Ar: Arachnida, C: Crustacea, I: insecta, M: Mammalia, Mol: Molluska, P: Pisces, S: Sauropsida.

between  $\pi_{mit}$  and  $\pi_{nuc}$  can be explained by the variation in  $\mu_{mit}/\mu_{nuc}$  among animal taxa.

We also reproduced the main figure of Bazin et al. (2006, their fig. 1) by plotting the average  $\pi_{mit}^*$  and  $\pi_{nuc}$  against the average allozyme heterozygosity (fig. 4).  $\pi_{mit}^*$  was not significantly correlated to the allozyme heterozygosity (Spearman's rank correlation  $\rho = 0.69$ ,  $P = 0.07$ ) but the correlation coefficient increased markedly compared to the uncorrected  $\pi_{mit}$  ( $\pi_{mit}$  vs. allozymic heterozygosity, Spearman's rank



**FIG. 4.** Relationship between mitochondrial synonymous diversity and allozyme heterozygosity across metazoan taxa. Y-axis: uncorrected (blue/circles) and mutation rate-corrected (red/triangles) average level of synonymous polymorphism in mtDNA and nuclear DNA (purple/squares) X-axis: average level of allozyme heterozygosity Y-axis and X-axis are log-transformed. Abbreviations: see legend to figure 3.

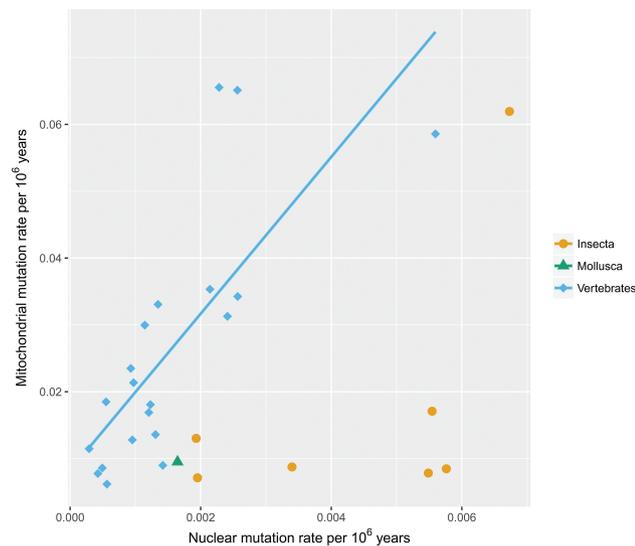
correlation  $\rho = 0.02$ ,  $P = 0.98$ ) and became close to the correlation between allozymic and  $\pi_{nuc}$ .

In our data sets,  $\pi_{mit}^*$  is on average 1.5 times lower than  $\pi_{nuc}$ . This is substantially higher than the neutral expectation ( $\pi_{nuc} = 4 \pi_{mit}^*$ ). This could be explained by (i) an underestimation of the  $\mu_{mit}/\mu_{nuc}$  ratio, (ii) and unbalanced sex ratio with an excess of females.

## Absolute Mutation Rates

The large variation in  $\mu_{mit}/\mu_{nuc}$  among metazoan taxa could result from a variable  $\mu_{mit}$ , a variable  $\mu_{nuc}$  or variation in both  $\mu_{mit}$  and  $\mu_{nuc}$ . To investigate these possibilities, we estimated the per year mutation rate for 28 data sets in which at least one fossil calibration point was available (Sauropsida  $N = 8$ ; Insecta  $N = 7$ ; Mammalia  $N = 5$ ; Amphibia  $N = 2$ ; Testudines  $N = 2$ , Aves  $N = 1$ ; Chondrichthyes  $N = 1$ ; Mollusca  $N = 1$ ; Teleostei  $N = 1$ ).

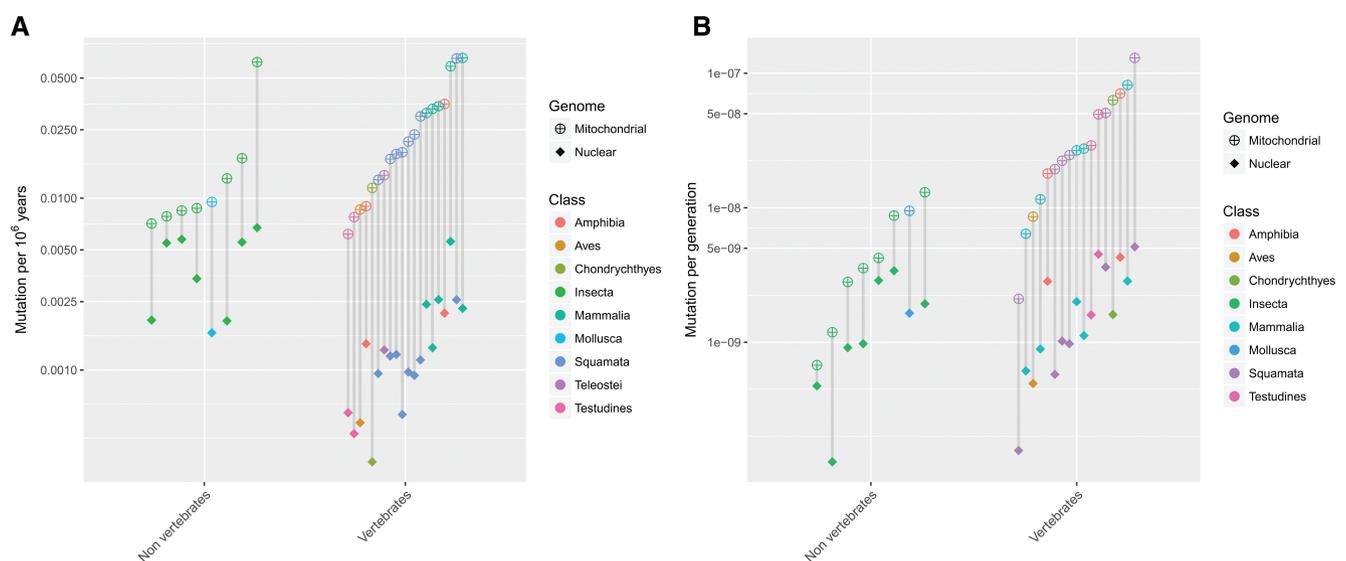
Including all data sets,  $\mu_{mit}$  was not correlated to  $\mu_{nuc}$  (Spearman's rank correlation  $\rho = 0.31$ ,  $P = 0.11$ ; fig. 5), but



**FIG. 5.** Relationship between the per year nuclear and mitochondrial mutation rates. Vertebrate data sets are in blue (squares). Nonvertebrate are in orange (Insects, circles) and green (Mollusk  $N = 1$ , triangle). The blue line is the regression line computed using only vertebrate data.

a significant correlation was detected within vertebrates, that is, excluding insects and mollusks (Spearman's rank correlation  $\rho = 0.71$ ,  $P < 0.001$ ; fig. 5). The absolute  $\mu_{mit}$  was higher than  $\mu_{nuc}$  in all taxa but the difference was much more pronounced in vertebrates (fig. 6A). The per year nuclear mutation rate was significantly lower in vertebrates than in insects and mollusks (median  $\mu_{nuc} = 0.0012$  per site per  $10^6$  years in vertebrates, 0.0044 in nonvertebrates, Wilcoxon' test  $P = 0.002$ ), whereas no significant difference was detected between vertebrates and nonvertebrates as far as the mitochondrial mutation rate was concerned (median  $\mu_{mit} = 0.0199$  per site per  $10^6$  years in vertebrates, 0.009 in nonvertebrates, Wilcoxon' test  $P = 0.07$ ).

Polymorphism levels are determined by mutation rate per generation, not per year. We obtained rough estimates of the average per generation mutation rates using age at sexual maturity or number of generation per year as a proxy for generation time. This considerably altered the distribution of mutation rate across vertebrates and nonvertebrates (fig. 6B). The average estimated per generation  $\mu_{mit}$  was higher in vertebrates than in nonvertebrates (median  $\mu_{mit} = 26.75 \times 10^{-9}$  per site per generation in vertebrates,  $3.90 \times 10^{-9}$  in nonvertebrates, Wilcoxon' test  $P < 0.001$ ) whereas the average per generation  $\mu_{nuc}$  was not significantly different between nonvertebrates and vertebrates (median  $\mu_{nuc} = 1.60 \times 10^{-9}$  per site per generation in vertebrates,  $1.31 \times 10^{-9}$  in nonvertebrates, Wilcoxon' test  $P = 0.58$ , fig. 6B). Therefore, the shorter generation time of nonvertebrate species seems to result in a lower mitochondrial mutation rate per generation. The lower per generation mutation rate of nonvertebrate species apparently compensates for their higher  $N_e$  leading to a similar  $\pi_{mit}$  between vertebrates and nonvertebrates and explains the homogeneity of mitochondrial polymorphism levels across animals.



**FIG. 6.** Distribution of the per year and per generation mutation rates across taxa and genomes. (A) Per year mitochondrial (circles with crosses) and nuclear (diamonds) mutation rate in vertebrates and nonvertebrates. (B) Per generation mitochondrial (circles with crosses) and nuclear (diamonds) mutation rate in vertebrates and nonvertebrates. Vertical grey lines joint mitochondrial and nuclear estimates obtained from the same data set. Y-axis is log-transformed.

## Estimation of Neutral Divergence and the Effect of GC Content

In order to estimate mutation rate variation from divergence data it is important to limit the impact of natural selection. The divergence of neutral sites is strictly proportional to mutation rate (Kimura 1968). This is true even for neutral sites genetically linked to selected sites (e.g., third codon position) (Birky and Walsh 1988). In this respect, a common practice is to use synonymous sites as proxy of neutral sites (but see Chamary et al. 2006). However, the estimation of the number of synonymous sites could be an issue. Specifically, the method of Goldman and Yang (1994) (GY94) appears to be influenced by GC content and non-stationary base composition (Bierne and Eyre-Walker 2003; Guéguen and Duret 2017). In our data set, we detected a strong relationship between the proportion of synonymous sites (i.e., number of synonymous sites/total number of sites) and GC content (supplementary fig. S5, Supplementary Material online). In the GY94 model, synonymous and nonsynonymous sites are counted as mutational opportunities, that is, the relative number of synonymous and nonsynonymous sites per codon depends on the estimated rates of substitution to its neighbor codons. We use the F3x4 model, where the equilibrium codon frequencies are empirically estimated from observed nucleotides frequencies. Equilibrium codon frequencies influence substitution rates between codons under GY94 and, consequently, the relative number of synonymous and nonsynonymous positions. In GC-rich alignments, the substitution rate between G- and C-ending codons, which are often nonsynonymous, are assumed to be higher than in alignments with balanced GC content—and similarly for substitutions between A- and T-ending codons in AT-rich alignments. This leads to low estimated numbers of synonymous sites in alignments with extreme GC content. This is illustrated by the fact that the proportion of synonymous sites greatly increases if we assume a common equilibrium frequency for all codons ( $f = 1/61$ ) instead of  $F3 \times 4$ . For example, the most AT-rich mitochondrial alignment (COI in *Melitta*, Hymenoptera; GC = 7%) has an estimated 9% of synonymous sites with the F3x4 parameter but this proportion increases to 30% when equal equilibrium codon frequencies are assumed. Similarly, in the most GC-rich nuclear alignment (*shaw* in Aleyrodidae, Hemiptera; GC = 63%), the proportion of synonymous sites increase from 9% to 38%.

It is unclear whether the  $F3 \times 4$  parameter reflects a true biological feature, that is, whether in AT-rich alignments, transversion between AT-rich codons are truly more frequent than in alignments with a more balanced GC content. Nevertheless, we replicated all our analyses using either the synonymous divergence (estimated under GY94) or 4-fold degenerated sites, instead of all third-codon positions in our main analysis. The 4-fold degenerated sites have the advantage to be both synonymous and straightforward to count. However, the number of 4-fold degenerated sites can be limiting especially when the divergence is high—we only included sites that were 4-fold degenerated in all the analyzed species. With both type of sites, the main results remained unchanged. All the results obtained using the

synonymous divergences and 4-fold degenerated sites are presented as supplementary material text S1 and supplementary figures S1–S4, Supplementary Material online.

Finally, the distribution of GC content greatly varies between genomes and taxa. Particularly, the most AT-rich loci are nonvertebrate mitochondrial loci and the most GC-rich loci are nuclear loci. It is possible that GC content is biologically linked to substitution rate via GC dependent mutation rate and/or GC-biased gene conversion (Duret and Galtier 2009). However, the average mitochondrial GC content and nuclear GC content have not effect on the variation of  $\mu_{mit}/\mu_{nuc}$  as tested in a multiple-factor type II ANOVA accounting for the effect of taxonomy ( $P > 0.10$  for GC content effect and  $P = 7.7e-08$  for taxonomy).

## Discussion

### High $\mu_{mit}/\mu_{nuc}$ in Vertebrates

Our analysis of 121 data sets encompassing 4,676 species of animals revealed extensive variation in  $\mu_{mit}/\mu_{nuc}$  among phyla of animals. Although somewhat continuous, this variation was mostly partitioned between vertebrates (with  $\mu_{mit}/\mu_{nuc}$  typically above 10) and nonvertebrates (with  $\mu_{mit}/\mu_{nuc}$  typically below 5). Our results confirm early reports, based on very small samples, of a similar order of magnitude for  $\mu_{mit}$  and  $\mu_{nuc}$  in nonvertebrate taxa (Powell et al. 1986; Vawter and Brown 1986; Brower and DeSalle 1998). We also detected a few non-vertebrate taxa with a high  $\mu_{mit}/\mu_{nuc}$  similar to that of vertebrates, but these are exceptions.

Analyzing 28 data sets for which fossil calibrations were available, we estimated the per year mutation rate separately for mtDNA and nuDNA. We found that the per year  $\mu_{nuc}$  is higher in nonvertebrates than in vertebrates, perhaps reflecting an effect of generation time. The per year  $\mu_{mit}$  in contrast, did not differ significantly between vertebrates and nonvertebrates. This is surprising, knowing the extensive variation in per year  $\mu_{mit}$  that was previously reported, and here confirmed, within mammals (Nabholz et al. 2008a; Welch et al. 2008) and within birds (Nabholz et al. 2009, 2016; Nguyen and Ho 2016). In particular, the strong effect of life-history traits, such as body mass, longevity, and generation time, on the per year  $\mu_{mit}$  that was documented in mammals and birds (Nabholz et al. 2008a; Welch et al. 2008; Nabholz et al. 2009, 2016) does not seem to hold at the Metazoa scale. Using a proxy of generation time, we observed that the average per generation  $\mu_{mit}$  which matters as far as polymorphism is concerned, is much higher in vertebrates than in nonvertebrates.

### Robustness to Gene Sampling

Many of the  $\mu_{mit}/\mu_{nuc}$  ratios were estimated based on data sets comprising just one locus per genome (supplementary table S1, Supplementary Material online), which might not be representative of the genome-wide ratio. To test whether the variation in  $\mu_{mit}/\mu_{nuc}$  is robust to gene sampling we gathered four genome-wide data sets of vertebrates and non-vertebrates species. In all cases, we identified substantial variation in  $\mu_{mit}/\mu_{nuc}$  among genes. This could be due to a

variable nuclear mutation rate across the genome (Ellegren et al. 2003; Hodgkinson and Eyre-Walker 2011), but part of the variation could also come from stochastic effects (although we excluded alignments shorter than 200 bp) and alignment errors. Despite this large variation, a strong taxonomic effect was detected, confirming that the variation in  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  among taxonomic groups is not an artifact of gene sampling. Many of our Arachnida and Mollusca data sets all comprise a single nuclear locus, namely Histone H3. This might have influenced our result if third codon positions were particularly fast-evolving in this gene. However, in the bivalves phylogenomic data set, the  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  ratio for histone H3 is 3.6 which is higher than the mean  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  (1.8). This demonstrates that histone H3 is rather a slow evolving gene (leading an high  $\mu_{\text{mit}}/\mu_{\text{nuc}}$ ) compared to other nuclear protein-coding genes and that our results, if anything, underestimate the true effect.

### Reconsidering the Genetic Draft Hypothesis

Our analysis demonstrates that the variation in  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  among animal clades explains the absence of correlation between  $\pi_{\text{mit}}$  and  $\pi_{\text{nuc}}$  reported by Bazin et al. (2006). The  $\mu_{\text{mit}}/\mu_{\text{nuc}}$ -corrected  $\pi_{\text{mit}}^*$  was well correlated to  $\pi_{\text{nuc}}$  indicating that the variation in effective population size across animals affects the nuclear and mitochondrial genomes in a comparable way, as previously reported within some vertebrate clades (Mulligan et al. 2006; Popadin et al. 2007; Nabholz et al. 2008a; Piganeau and Eyre-Walker 2009). Bazin et al. (2006) invoked a more pronounced effect of genetic draft (Gillespie 2001) on mtDNA in large- $N_e$  taxa to interpret the lack of correlation between  $\pi_{\text{mit}}$  and  $\pi_{\text{nuc}}$ . Our analysis provides an alternative explanation to the results of Bazin et al. (2006). It shows that the constancy of mean  $\pi_{\text{mit}}$  among animals groups could be explained by a negative correlation between  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  and effective population size.

Our results imply that the action of natural selection is not pervasive enough to erase the effect of genetic drift on mitochondrial diversity. They do not, however, exclude that positive selection sometimes occurs in mtDNA evolution (da Fonseca et al. 2008). Recently, James et al. (2016) have estimated that  $\sim 26\%$  (5.7–45%) of amino-acid substitutions affecting mitochondrial proteins are adaptive in animals, and some of the reported instances of mitochondrial/nuclear phylogenetic incongruence have been interpreted as reflecting adaptive mitochondrial introgression (Toews and Brelsford 2012). In our analysis, we report a significant correlation between  $\pi_{\text{mit}}^*$  and  $\pi_{\text{nuc}}$  but the correlation between  $\pi_{\text{mit}}^*$  and allozyme heterozygosity was not statistically significant. It should be noted that here the values of  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  (this study) and polymorphism (Bazin et al. 2006) were estimated from distinct data sets, including different loci and species. As a consequence, the correlation coefficients we report are likely to be underestimated.

### Why Is $\mu_{\text{Mit}}/\mu_{\text{Nuc}}$ Correlated to $N_e$ ?

There is no obvious theoretical reason to expect a relationship between  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  and effective population size. This partly explains why Bazin et al. (2006) did not consider

mutation rate as a plausible confounding factor, in absence of large-scale data on  $\mu_{\text{mit}}/\mu_{\text{nuc}}$ . To gain insight into the causes of  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  variation, we estimated absolute mutation rates per year. Our results indicate that both  $\mu_{\text{mit}}$  and  $\mu_{\text{nuc}}$  are variable among species and higher level taxa. The variation between vertebrates and nonvertebrates seems more pronounced for the nuclear genome than for the mitochondrial genome. Finally,  $\mu_{\text{mit}}$  appears correlated to  $\mu_{\text{nuc}}$  in vertebrates, suggesting that the per year mitochondrial and nuclear mutation rates are influenced by common determinants. This is not surprising as several life-histories traits such as body-mass and sexual maturity are known to influence both mitochondrial and nuclear mutation rates (Bromham 2009). The absence of correlation between the two rates in insects could simply result from insufficient power due to the limited number of data sets available ( $N = 7$ ). We were able to estimate absolute mutation rates in only 28 data sets in total, which limits our ability to test alternative hypotheses that could explain the variation in  $\mu_{\text{mit}}/\mu_{\text{nuc}}$ . Below, we discuss four nonexclusive hypotheses that could potentially be tested with an extended data set.

First, the metabolic rate could have an impact on the mtDNA mutation rate (Martin et al. 1992; Martin and Palumbi 1993; Rand 1994). Because of its location, the mitochondrial genome is expected to be impacted by mutagenic byproducts of aerobic respiration such as reactive oxygen species (ROS) (Turrens 2003). Insects have a lower mass-specific metabolic rate than endothermic vertebrates (Makarieva et al. 2008) and therefore, insects could have a reduced production rate of ROS that could limit their mtDNA mutation rate. However, many ectothermic vertebrates have a mass-specific metabolic rate similar or lower than that of many insects (Makarieva et al. 2008). Moreover, scaled reptiles have a similar  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  than birds and mammals despite their lower mass-specific metabolic rate (White et al. 2006). Finally, the relationship between mass-specific metabolic rate and ROS is currently unclear and may not be linear (Galtier et al. 2009).

Second, the number of genome replications per generation could differ between the mitochondrial and nuclear genomes, and among taxa of animals. A high  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  could be explained by a higher ratio of mtDNA/nuDNA replication cycles per generation in vertebrates than in nonvertebrates. In mammals, it has been shown that oocyte maturation involves a dramatic increase in the number of mitochondrial genomes (Mishra and Chan 2014). A large increase in mtDNA amounts, however, was also reported in the in oocytes of some nematodes (Tsang and Lemire 2002). Whether the number of mtDNA replications per generation generally differs between vertebrates and nonvertebrates is therefore currently unclear.

Third, the mitochondrial genome is replicated by a specific DNA polymerase, namely DNA polymerase  $\gamma$ . A single event involving a reduction in the fidelity of DNA polymerase  $\gamma$  in the ancestor of vertebrates would be sufficient to explain most of the  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  variation we report.

Fourth, Lynch (2010) provided some theoretical justification and empirical evidence for a negative correlation

between per generation mutation rate and effective population size. Natural selection is expected to push down the mutation rates to a value at which further decrease would imply a negligible fitness increase, compared to the strength of genetic drift ( $1/N_e$ ) (Lynch 2010). The relationship between  $N_e$  and the fitness effect of the germline mutation rate might be different for the mitochondrial and the nuclear genomes therefore leading to a variable  $\mu_{mit}/\mu_{nuc}$  scaling with  $N_e$ . Direct estimation of mutation rates through sequencing of genomes in mutation accumulation or pedigree experiments are still rare (Lynch 2010; Smeds et al. 2016) and a formal test of this hypothesis may not be possible in the short term. However, indirect evidence for or against the above hypotheses could be obtained by confronting phylogenetic estimation of  $\mu_{mit}/\mu_{nuc}$  with species' life-history traits taken as proxies of metabolic rate (Makarieva et al. 2008), generation time (Bromham et al. 1996), and effective population size (Romiguier et al. 2014).

### Implications for DNA Barcoding

Finally, our results have implications for DNA-based identification methods. An ideal barcoding marker should form robust monophyletic clades reflecting species boundaries. This ability depends on the mutation rate of the considered marker, which determines the expected number of mutations diagnostic of a given species, and the strength of genetic drift ( $1/N_e$ ), which determines the age of the most recent common ancestor of individuals in a species. If species boundaries were predictable from nuDNA divergence (Roux et al. 2016), our results would imply that mtDNA barcoding is more powerful in vertebrates than in non-vertebrates species (fig. 7). For a given level of nuDNA divergence, one should sequence 3–10 times more mtDNA sites in a nonvertebrate than in a vertebrate to obtain the same level of phylogenetic resolution. To some extent, this is also true of bony fishes and amphibians compared to birds, mammals and scaled reptiles. This is in line with reports of a relatively low success of the barcoding technique in some non-vertebrate taxa (Meier et al. 2006; Elias et al. 2007). Finally, direct comparison of species delimitation methods and biodiversity monitoring using the same mtDNA markers across distinct phyla of animals may be biased knowing the large variation in  $\mu_{mit}/\mu_{nuc}$  we report.

## Materials and Methods

### Multilocus Data sets

We extracted all data sets from « *Molecular Phylogenetics and Evolution* » by systematically scanning the table of content between volume 62 (January 2012) and 97 (April 2016). All selected data sets have at least one mitochondrial gene and one nuclear gene. Any species represented by only nuclear or mitochondrial sequence were excluded. We also selected only one sequence per species, following the taxonomy adopted in the publications. This step aims at excluding polymorphism data from the data sets. The data sets including less than five species were excluded. The data sets were split in 12 taxonomical groups: Amphibia; Arachnida; Aves; Chondrichthyes; Crustacea; Insecta; Mammalia; Molluska; Porifera; Squamata;

Teleostei; Testudines. Because the literature review was very time consuming, we limited the research to 20 data sets per taxonomical group.

Sequences were extracted from NCBI/GenBank (<http://www.ncbi.nlm.nih.gov/nucleotide/>; last accessed April 2016) using the accession numbers available from the publications. Details of the data sets are presented in supplementary table S1, Supplementary Material online.

### Molecular Divergence Estimation

Coding DNA Sequences (CDS) were extracted using custom made BioPython scripts (Cock et al. 2009). Sequences were aligned gene by gene using MACSE (Ranwez et al. 2011). Phylogenetic trees were inferred using RAxML with the substitution model GTRCAT (Stamatakis 2006). One phylogeny per locus was inferred. This will avoid potential discordance between gene trees that could lead to a spurious excess of inferred substitutions produced by incomplete lineage sorting (Mendes and Hahn 2016).

Next, we extracted third codon positions in order to limit the effect of natural selection. Sites containing more than 20% of missing data were excluded. Molecular divergence was estimated using the BASEML program (Yang 2007) and the F84 substitution model without rate variation among sites. Total third codon positions Divergence (TD) was estimated gene by gene by computing the sum of branch lengths.

Assuming that all mutations are neutral at third codon positions, we have:  $TD = \mu \times t$ ; where  $\mu$  is mutation rate and  $t$  the divergence time. For every data set, the total divergence time is the same for the nuclear and mitochondrial phylogenies because the same set of species was analyzed for both markers. As a result, we are able to estimate the ratio of mitochondrial and nuclear mutation rate  $\mu_{mit}/\mu_{nuc}$  by computing the ratio of TD. When several mtDNA and/or nuDNA markers were available, we took the mean ratio across all mitochondrial versus nuclear pairs of markers.

Besides third codon position analysis, we also calculated branch lengths in unit of per synonymous site synonymous substitutions using CODEML (Yang 2007) and using 4-fold degenerated sites. Four-fold degenerated sites were extracted from the alignment using custom C++ program based on the Bio++ libraries (Guéguen et al. 2013). These analyses were conducted the same way as explained above.

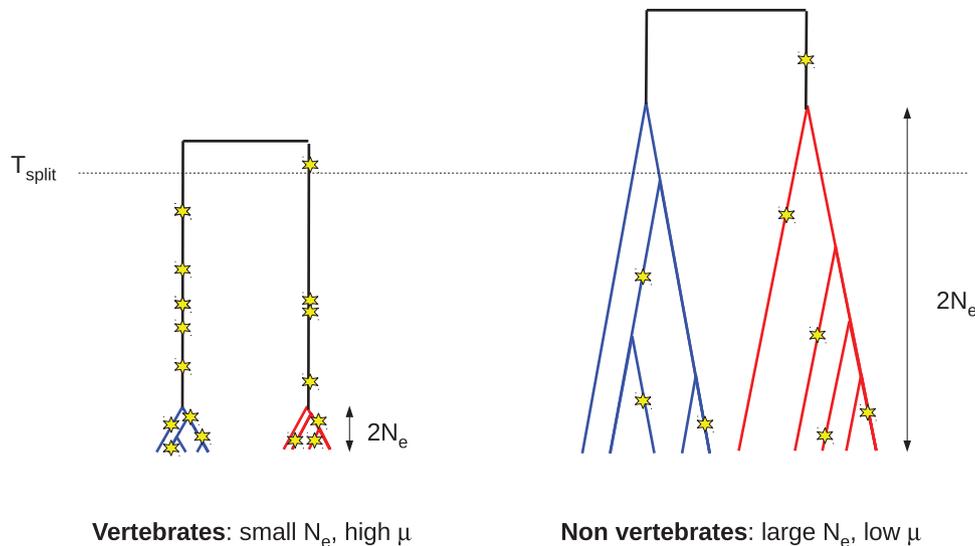
Alignments are available from Figshare repository (URL: <https://doi.org/10.6084/m9.figshare.5117467.v1>).

### Polymorphism Data

Mitochondrial and nuclear synonymous polymorphism ( $\pi_{mit}$  and  $\pi_{nuc}$ , respectively) were obtained from Bazin et al. (2006) (Eric Bazin, *pers. commun.*). We defined the mutation-corrected polymorphism,  $\pi_{mit}^*$ , as:

$$\pi_{mit}^* = \pi_{mit} \mu_{nuc} / \mu_{mit}$$

$\mu_{nuc}/\mu_{mit}$  being defined as the median  $\mu_{nuc}/\mu_{mit}$  of each taxonomical group of Bazin et al. (2006). For 'Pisces' group, we used the median  $\mu_{nuc}/\mu_{mit}$  of the Teleostei and Chondrichthyes.



**Fig. 7.** Schematic representation of mitochondrial DNA genealogies in two fictive pairs of vertebrate and non-vertebrate species.  $T_{\text{split}}$  is the age of population split (without gene flow) between the two species in number of generations. The blue and red parts of the genealogies correspond to within-species polymorphism.  $2N_e$  is the average time to the most recent common ancestor in a demographically stable population. Stars represent mutations.  $N_e$  is assumed to be ten times higher, and  $\mu$  ten times lower, in nonvertebrates than in vertebrates. In vertebrates mtDNA, a small  $N_e$  and a high mutation rate facilitate the identification of monophyletic clades that match species boundaries. In nonvertebrates, a high  $N_e$  and a low mutation rate decrease the expected number of diagnostic mutations.

Assuming neutrality and a balanced sex-ratio,  $\pi_{\text{mit}}^*$  is expected to be proportional to  $\pi_{\text{nuc}}$ .

$$\pi_{\text{mit}}^* = N_e \mu_{\text{mit}} \mu_{\text{nuc}} / \mu_{\text{mit}} = N_e \mu_{\text{nuc}} = \pi_{\text{nuc}} / 4$$

### Phylogenomic Data sets

Four phylogenomic data sets were analyzed: two vertebrates (Birds and Primates) and two nonvertebrates (*Drosophila* and Bivalvia). In birds, we used the alignment of Jarvis et al. (2014). We focused on six species of passerines + parrots: ((*Melopsittacus*, *Nestor*), *Acanthisitta* (*Corvus*, *Geospiza*, *Taeniopygia*)). In primates, alignments were obtained from the ORTHOMAM database (version 9, <http://www.orthomam.univ-montp2.fr/>; last accessed April 2016; Douzery et al. 2014). The data set included 12 species: (((((((*Homo*, *Pan*), *Gorilla*), *Pongo*), *Nomascus*), ((*Macaca*, *Papio*), *Chlorocebus*)), *Callithrix*, *Tarsius*, (*Otolemur*, *Microcebus*)). The Bivalvia data sets were obtained from González et al. (2015). We focused on the 11 species of Neoheterodonte: (((((((*Cycladicama*, *Diplodonta*), ((*Mercenaria*, (*Glossus*, *Arctica*)), (*Corbicula*, *Cyrenoida*)))))). Finally, the *Drosophila* data sets were obtained from Romiguier et al. (2014). The phylogenetic relationship between the 12 species is (((((((*D. melanogaster*, (*D. Simulans*, *D. sechellia*)), (*D. erecta*, *D. yakuba*)), *D. ananassae*), (*D. pseudoobscura*, *D. persimilis*)), *D. willistoni*), ((*D. mojavensis*, *D. virilis*), *D. grimshawi*)).

Nucleotide alignments were available for all the data sets except for the Bivalvia in which only amino-acid alignments were produced. For these data sets, we extracted the nucleotide sequences from assembled transcriptomes (Vanessa L. Gonzalez, *pers. commun.*). Sequences were aligned gene by gene using MACSE (Ranwez et al. 2011).

Full mitochondrial genomes were available for birds, primates and *Drosophila*. For Bivalvia, we used only the

Cytochrome Oxidase 1 gene. Mitochondrial sequences were aligned gene by gene using MACSE (Ranwez et al. 2011).

For all the data sets, alignments were cleaning using Gblocks option:  $-t=c$   $-b4=5$   $-b5=h$  (Talavera and Castresana 2007) and sites containing more than 20% of missing data were excluded. Alignments including less than 200 sites were excluded. Molecular divergence was estimated using BASEML program (Yang 2007) and the F84 substitution model without rate variation among sites. Total third codon positions Divergence (TD) was estimated gene by gene by computing the sum of branch lengths.  $\mu_{\text{mit}}/\mu_{\text{nuc}}$  was estimated by dividing the TD of a mitochondrial gene by the mean TD of the nuclear genes.

### Absolute Mutation Rates and Generation Times

When they were available, fossil calibration points were extracted from the publications. We re-estimated a single phylogeny based on mitochondrial and nuclear sequences using RAXML with the substitution model GTRCAT partitioned by gene (Stamatakis 2006). Molecular dating was performed gene by gene using MCMCTREE (Yang 2007) with the HKY85 + Gamma model. The absolute mutation rate was computed as the average rate across genes.

Generation times were approximated using age at sexual maturity or number of generations per year. For vertebrates, we mostly relied on the AnAge database (version 13, Tacutu et al. 2012). For nonvertebrates, generation times were obtained from the literature and, in rare cases, from websites (see supplementary table S2, Supplementary Material online for details). In most cases, no information was available for the very species of our data sets and we used the median value of a taxonomical group including our species (mostly at the family level). For example, for a data set composed of

*Trioceros* Chameleo species (Ceccarelli et al. 2014), we used the median value of all Chameleoniae in AnAge.

### Statistical Analysis

All the statistical analyses were performed using R (R Core Team 2013). The effect of taxonomy on  $\mu_{mit}/\mu_{nuc}$  was tested with an Analysis of variance (ANOVA). The effect of the average GC content of mitochondrial (GCmito) and nuclear (GCnucl) GC content was tested with a multiple-factor type II ANOVA in a model including taxonomy ( $\log(\mu_{mit}/\mu_{nuc}) \sim \text{taxonomy} + \text{GCmito} + \text{GCnucl}$ ). Variables were log transformed. Correlations were assessed using Spearman's non-parametric method.

### Acknowledgments

This work is dedicated to the memory of Eric Bazin. We are grateful of Vanessa L. Gonzalez for provided the transcriptomes of *Bivalvia* species. We thank Sylvain Glémin and the anonymous reviewers for helpful discussion. This work was supported by Agence Nationale de la Recherche grants ANR-15-CE12-0010 "DarkSideOfRecombination" to B.N. and N.G, ANR-14-CE02-0002-01 "BirdIslandGenomic" to B. N and ANR-10-BINF-01-01 "Ancestrome" to N.G. This is ISEM publication number ISEM 2017-141.

### Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

### References

Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. *Mol Ecol*. 13:729–744.

Barker FK. 2014. Mitogenomic data resolve basal relationships among passeriform and passeridan birds. *Mol Phylogenet Evol*. 79:313–324.

Bazin E, Glémin S, Galtier N. 2006. Population size does not influence mitochondrial genetic diversity in animals. *Science* 312:570–572.

Berlin S, Tomaras D, Charlesworth B. 2007. Low mitochondrial variability in birds may indicate Hill–Robertson effects on the W chromosome. *Heredity* 99:389–396.

Bierne N, Eyre-Walker A. 2003. The problem of counting sites in the estimation of the synonymous and nonsynonymous substitution rates: implications for the correlation between the synonymous substitution rate and codon usage bias. *Genetics* 165:1587–1597.

Birky CW, Jr, Walsh JB. 1988. Effects of linkage on rates of molecular evolution. *Proc Natl Acad Sci U S A*. 85:6414–6418.

Blouin MS, Yowell CA, Courtney CH, Dame JB. 1998. Substitution bias, rapid saturation, and the use of mtDNA for nematode systematics. *Mol Biol Evol*. 15:1719–1727.

Botero-Castro F, Tilak M, Justy F, Catzeflis F, Delsuc F, Douzery EJP. 2013. Next-generation sequencing and phylogenetic signal of complete mitochondrial genomes for resolving the evolutionary history of leaf-nosed bats (Phyllostomidae). *Mol Phylogenet Evol*. 69:728–739.

Bromham L. 2009. Why do species vary in their rate of molecular evolution? *Biol Lett*. 5:401–404.

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

Brower AVZ, DeSalle R. 1998. Patterns of mitochondrial versus nuclear DNA sequence divergence among nymphalid butterflies: the utility of wingless as a source of characters for phylogenetic inference. *Insect Mol Biol*. 7:73–82.

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

Ceccarelli FS, Menegon M, Tolley KA, Tilbury CR, Gower DJ, Laserna MH, Kasahun R, Rodriguez-Prieto A, Haggmann R, Loader SP. 2014. Evolutionary relationships, species delimitation and biogeography of Eastern Afrotropical horned chameleons (Chamaeleonidae: Trioceros). *Mol Phylogenet Evol*. 80:125–136.

Chamary JV, Parmley JL, Hurst LD. 2006. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat Rev Genet*. 7:98–108.

Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, et al. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25:1422–1423.

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.

Duret L, Galtier N. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet*. 10:285–311.

Elias M, Hill RI, Willmott KR, Dasmahapatra KK, Brower AVZ, Mallet J, Jiggins CD. 2007. Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proc R Soc Lond B Biol Sci*. 274:2881–2889.

Ellegren H, Galtier N. 2016. Determinants of genetic diversity. *Nat Rev Genet*. 17:422–433.

Ellegren H, Smith NG, Webster MT. 2003. Mutation rate variation in the mammalian genome. *Curr Opin Genet Dev*. 13:562–568.

Fabre P-H, Vilstrup JT, Raghavan M, Der Sarkissian C, Willerslev E, Douzery EJ, Orlando L. 2014. Rodents of the Caribbean: origin and diversification of hutias unravelled by next-generation museomics. *Biol Lett*. 10:20140266.

da Fonseca RR, Johnson WE, O'Brien SJ, Ramos MJ, Antunes A. 2008. The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics* 9:1.

Galtier N, Nabholz B, Glémin S, Hurst GDD. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol Ecol*. 18:4541–4550.

Gillespie JH. 2001. Is the population size of a species relevant to its evolution? *Evolution* 55:2161–2169.

Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol*. 11:725–736.

González VL, Andrade SC, Bieler R, Collins TM, Dunn CW, Mikkelsen PM, Taylor JD, Giribet G. 2015. A phylogenetic backbone for *Bivalvia*: an RNA-seq approach. *Proc R Soc B*. 282: 20142332.

Guéguen L, Duret L. 2017. Unbiased estimate of synonymous and non-synonymous substitution rates with non-stationary base composition. *bioRxiv* 124925.

Guéguen L, Gaillard S, Boussau B, Gouy M, Groussin M, Rochette NC, Bigot T, Fournier D, Pouyet F, Cahais V, et al. 2013. Bio ++: efficient extensible libraries and tools for computational molecular evolution. *Mol Biol Evol*. 30:1745–1750.

Havird JC, Sloan DB. 2016. The roles of mutation, selection, and expression in determining relative rates of evolution in mitochondrial versus nuclear genomes. *Mol Biol Evol*. 33:3042–3053.

Hebert PD, Ratnasingham S, de Waard JR. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc R Soc Lond B Biol Sci*. 270:596–599.

Hebert PDN, Gregory TR. 2005. The promise of DNA barcoding for taxonomy. *Syst Biol*. 54:852–859.

Hellberg ME. 2006. No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. *BMC Evol Biol*. 6:24.

Hirano T, Kameda Y, Kimura K, Chiba S. 2014. Substantial incongruence among the morphology, taxonomy, and molecular phylogeny of the land snails *Aegista*, *Landouria*, *Trishoplita*, and *Pseudobuliminus* (Pulmonata: Bradybaenidae) occurring in East Asia. *Mol Phylogenet Evol*. 70:171–181.

Hodgkinson A, Eyre-Walker A. 2011. Variation in the mutation rate across mammalian genomes. *Nat Rev Genet*. 12:756–766.

Huang D, Meier R, Todd PA, Chou LM. 2008. Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding. *J Mol Evol*. 66:167–174.

- Hurst GDD, Jiggins FM. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proc Biol Sci.* 272:1525–1534.
- James JE, Piganeau G, Eyre-Walker A. 2016. The rate of adaptive evolution in animal mitochondria. *Mol Ecol.* 25:67–78.
- Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SY, Faircloth BC, Nabholz B, Howard JT, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346:1320–1331.
- Kimura M. 1968. Evolutionary rate at the molecular level. *Nature* 217:624–626.
- Lynch M. 2010. Evolution of the mutation rate. *Trends Genet.* 26:345–352.
- Lynch M, Koskella B, Schaack S. 2006. Mutation pressure and the evolution of organelle genomic architecture. *Science* 311:1727–1730.
- Makarieva AM, Gorshkov VG, Li B-L, Chown SL, Reich PB, Gavrillov VM. 2008. Mean mass-specific metabolic rates are strikingly similar across life's major domains: evidence for life's metabolic optimum. *Proc Natl Acad Sci U S A.* 105:16994–16999.
- Martin AP, Naylor GJ, Palumbi SR. 1992. Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. *Nature* 357:153–155.
- Martin AP, Palumbi SR. 1993. Protein evolution in different cellular environments: cytochrome b in sharks and mammals. *Mol Biol Evol.* 10:873–891.
- Maynard-Smith J, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res.* 23:23–35.
- Meier R, Shiyang K, Vaidya G, Ng PKL. 2006. DNA barcoding and taxonomy in diptera: a tale of high intraspecific variability and low identification success. *Syst Biol.* 55:715–728.
- Mendes FK, Hahn MW. 2016. Gene tree discordance causes apparent substitution rate variation. *Syst Biol.* 65:711–721.
- Mishra P, Chan DC. 2014. Mitochondrial dynamics and inheritance during cell division, development and disease. *Nat Rev Mol Cell Biol.* 15:634–646.
- Miyata T, Hayashida H, Kikuno R, Hasegawa M, Kobayashi M, Koike K. 1982. Molecular clock of silent substitution: at least six-fold preponderance of silent changes in mitochondrial genes over those in nuclear genes. *J Mol Evol.* 19:28–35.
- Mulligan CJ, Kitchen A, Miyamoto MM. 2006. Comment on “Population size does not influence mitochondrial genetic diversity in animals”. *Science* 314:1390.
- Nabholz B, Glémin S, Galtier N. 2008a. Strong variations of mitochondrial mutation rate across mammals—the longevity hypothesis. *Mol Biol Evol.* 25:120–130.
- Nabholz B, Glémin S, Galtier N. 2009. The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. *BMC Evol Biol.* 9:54.
- Nabholz B, Lanfear R, Fuchs J. 2016. Body mass-corrected molecular rate for bird mitochondrial DNA. *Mol Ecol.* 25: 4438–4449.
- Nabholz B, Mauffrey J-F, Bazin E, Galtier N, Glémin S. 2008b. Determination of mitochondrial genetic diversity in mammals. *Genetics* 178:351–361.
- Nguyen JMT, Ho SYW. 2016. Mitochondrial rate variation among lineages of passerine birds. *J Avian Biol.* 47:690–696.
- Oliveira DC, Raychoudhury R, Lavrov DV, Werren JH. 2008. Rapidly evolving mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp *Nasonia* (Hymenoptera: Pteromalidae). *Mol Biol Evol.* 25:2167–2180.
- Owen CL, Marshall DC, Hill KB, Simon C. 2015. The phylogenetic utility of acetyltransferase (ARD1) and glutaminyl tRNA synthetase (QtRNA) for reconstructing Cenozoic relationships as exemplified by the large Australian cicada *Pauropsalta* generic complex. *Mol Phylogenet. Evol.* 83:258–277.
- Piganeau G, Eyre-Walker A. 2009. Evidence for variation in the effective population size of animal mitochondrial DNA. *PLoS One* 4:e4396.
- 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.
- Powell JR, Caccone A, Amato GD, Yoon C. 1986. Rates of nucleotide substitution in *Drosophila* mitochondrial DNA and nuclear DNA are similar. *Proc Natl Acad Sci U S A.* 83:9090–9093.
- R Core Team. 2013. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Core Team. Available from: <http://www.R-project.org/>.
- Rand DM. 1994. Thermal habit, metabolic rate and the evolution of mitochondrial DNA. *Trends Ecol Evol.* 9:125–131.
- Ranwez V, Harispe S, Delsuc F, Douzery EJ. 2011. MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS One* 6:e22594.
- Recuero E, Canestrelli D, Vörös J, Szabó K, Poyarkov NA, Arntzen JW, Crnobrnja-Isailovic J, Kidov AA, Coghlaniceanu D, Caputo FP, et al. 2012. Multilocus species tree analyses resolve the radiation of the widespread *Bufo bufo* species group (Anura, Bufonidae). *Mol Phylogenet Evol.* 62:71–86.
- Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, Chiari Y, Derrat R, Duret L, Favre N, et al. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* 515:261–263.
- Roux C, Fraise C, Romiguier J, Anciaux Y, Galtier N, Bierne N. 2016. Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biol.* 14:e2000234.
- Shearer TL, Van Oppen MJH, Romano SL, Wörheide G. 2002. Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). *Mol Ecol.* 11:2475–2487.
- Smeds L, Qvarnstrom A, Ellegren H. 2016. Direct estimate of the rate of germline mutation in a bird. *Genome Res.* 26:1211–1218.
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- Tacutu R, Craig T, Budovsky A, Wuttke D, Lehmann G, Taranukha D, Costa J, Fraifeld VE, de Magalhães JP. 2012. Human ageing genomic resources: integrated databases and tools for the biology and genetics of ageing. *Nucleic Acids Res.* 41:D1027–D1033.
- Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 56:564–577.
- Thomas WK, Wilson AC. 1991. Mode and tempo of molecular evolution in the nematode *Caenorhabditis*: cytochrome oxidase II and calmodulin sequences. *Genetics* 128:269–279.
- Thomsen PF, Kielgast J, Iversen LL, Wiuf C, Rasmussen M, Gilbert MTP, Orlando L, Willerslev E. 2012. Monitoring endangered freshwater biodiversity using environmental DNA. *Mol Ecol.* 21:2565–2573.
- Tilak M-K, Justy F, Debais-Thibaud M, Botero-Castro F, Delsuc F, Douzery EJP. 2014. A cost-effective straightforward protocol for shotgun Illumina libraries designed to assemble complete mitogenomes from non-model species. *Conserv Genet Resour.* 7:37–40.
- Toews DP, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Mol Ecol.* 21:3907–3930.
- Tsang WY, Lemire BD. 2002. Mitochondrial genome content is regulated during nematode development. *Biochem Biophys Res Commun.* 291:8–16.
- Turrens JF. 2003. Mitochondrial formation of reactive oxygen species. *J Physiol.* 552:335–344.
- Wawter L, Brown WM. 1986. Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. *Science* 234:194–196.
- Welch J, Bininda-Emonds O, Bromham L. 2008. Correlates of substitution rate variation in mammalian protein-coding sequences. *BMC Evol Biol.* 8:53.
- White CR, Phillips NF, Seymour RS. 2006. The scaling and temperature dependence of vertebrate metabolism. *Biol Lett.* 2:125–127.
- Willett CS. 2012. Quantifying the elevation of mitochondrial DNA evolutionary substitution rates over nuclear rates in the intertidal copepod *Tigriopus californicus*. *J Mol Evol.* 74:310–318.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.