Empirical validation of the nearly neutral theory at divergence and population genomic scale using 144 placental mammals genomes

M. Bastian[1], D. Enard[2], N. Lartillot[1]

[1]Laboratoire de Biométrie et Biologie Evolutive, UMR5558, Université Lyon 1, Villeurbanne, France

[2]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85719, USA

Corresponding author : nicolas.lartillot@univ-lyon1.fr

January 30, 2026

Abstract

By limiting the efficacy of selection, random drift is expected to play a major role in genome evolution. Formalizing this idea, the nearly-neutral theory predicts that the ratio of non-synonymous over synonymous polymorphism (π_N/π_S) within populations, and divergence (d_N/d_S) between species, should both correlate negatively with N_e. This has previously been tested in mammals and other groups. However, most studies have focused on either d_N/d_S or on π_N/π_S, thus not addressing the problem across evolutionary scales. In addition, many studies at the macro scale have used life-history traits (LHT) as a proxy of N_e, assuming that large-bodied organisms have lower N_e than small-bodied species. However, this assumption itself has rarely been validated against more objective measures of N_e, such as genetic diversity $\pi_S = 4N_\mathrm{e}\mu$, in part because π_S estimates are scarce. Here we propose an integrative test of the nearly-neutral predictions on 150 mammalian species, using 6000 orthologous genes, spanning the macro and the micro-evolutionary scale, using for the latter a measure of heterozygosity on each of the assembled diploid genomes. At the micro scale, we observe, for the first time in mammalian nuclear genomes, a relationship between π_N/π_S and π_S. At the macro scale, we confirm the positive correlation between d_N/d_S and LHT but, more importantly, establish that LHT and d_N/d_S are correlated with π_S, although weakly so. Together, these results provide the first global test of the nearly-neutral theory in mammals across time scales, suggesting all variables are correlated with a single hidden variable: N_e.

***K*eywords** population genomic, effective population size, selection efficacy, mammals, phylogenomic, micro-macro evolution, genetic drift

Significance Statement

Natural selection eliminates large-effect deleterious mutations but has limited efficacy for purging mutations of small effects. The effective population size (called N_e) is what determines the cutoff below which selection is inefficient. Since N_e varies between species, the prediction is that more mutations of intermediate effect should segregate and reach fixation in species with small populations. This has been widely studied both between species (macro-evolution) and within species (micro-evolution), although never in a fully integrated manner. Here, we analysed 144 genomes of placental mammals and 6002 orthologous genes, using the original sequence reads produced for genome assembly to estimate within-species genetic diversity based on the heterozygosity of the diploid individual whose genome has been sequenced. We show consistent relationships between genetic diversity within species, life-history traits such as body mass or longevity and selection efficacy simultaneously at the micro- and macro-evolutionary levels. Our analysis provides a globally consistent picture confirming the role of effective population size in tuning selection efficacy in placental mammals.

1 Introduction

Phylogenomics and population genomics have been the subject of extensive research over the years. However, they are rarely used in tandem or even compared. While macro-evolution is widely viewed as the long-term outcome of micro-evolutionary processes, direct integration of empirical divergence and polymorphism data over a broad set of species is still quite uncommon. Yet, a more decisive integration along those lines would benefit from the complementary strengths of polymorphism and divergence in investigating important aspects of evolutionary processes and their long-term impact on genomes.

One particularly important point on which this comparison could bring important insights is the question of the role of effective population size (N_e), and its variation, in genome evolution. Random drift limits the efficacy of selection, implying that many features of genomes might in fact be neutral or even maladaptive (Lynch and Gabriel, 1990; Lynch and Conery, 2003; Nei *et al.*, 2010). The tradeoff between selection and drift was first formalized by Ohta and Kimura (Ohta and Kimura, 1971; Ohta, 1973a), leading to the nearly-neutral theory. According to this theory, the majority of mutations are deleterious, with a broad distribution for their fitness effects (DFE), ranging from slightly to strongly deleterious. In this context, the inverse of the effective population size acts as a threshold on the DFE, such that the fate of mutations with selective effects smaller than $1/N_e$ is typically determined by genetic drift and may thus segregate and reach fixation, while the mutations with larger effects are efficiently purified by natural selection (Ohta, 1973a).

In the case of protein coding genes, synonymous mutations are often considered as neutral, at least in groups such as mammals (Ohta, 1995), thus providing a useful neutral background against which to evaluate the fate of non-synonymous mutations, which on the other hand have typically a broad range of fitness effects (Eyre-Walker *et al.*, 2006; Eyre-Walker and Keightley, 2007). Contrasting the non-synonymous and

synonymous compartments thus yields two measurable quantities: at the population (or micro-evolutionary) scale, the ratio of non-synonymous over synonymous polymorphism (or π_N/π_S) and at the phylogenetic (or macro-evolutionary) scale, the ratio of non-synonymous over synonymous divergence (or d_N/d_S). These two metrics measure the combined effect of selection and random drift. In both cases, lower values indicate that non-synonymous deleterious mutations are more efficiently purged by natural selection and are thus segregating at lower frequency (in the case of π_N/π_S) or less prone to reaching fixation (in the case of d_N/d_S). According to the nearly neutral theory, both are thus expected to decrease as a function of N_e, with a quantitative response that depends on the exact shape of the DFE (Welch *et al.*, 2008) or on the structure of the fitness landscape (Lourenço *et al.*, 2011; Goldstein, 2013; Latrille and Lartillot, 2021).

Testing these predictions, however, raises the question of how to estimate N_e and more fundamentally, how to define it (reviewed in Waples, 2022). In particular, some define it from a demographic point of view, when others invoke N_e as a more inclusive measure of genetic drift (thus including non-demographic contributions, such as linked selection). Some authors argue for a more complex definition of N_e and derive the concept at different time scales (Nadachowska-Brzyska *et al.*, 2022; Müller *et al.*, 2022). In the following, we opt for the inclusive definition of N_e, as a net measure of the intensity of drift, as it is more directly relevant for what determines the processes of population genetics and molecular evolution.

Concerning the estimation of N_e, the cases of macro- and micro-evolution are quite contrasted. At the micro-evolutionary scale, genetic diversity provides direct quantitative access to N_e, through the relation $\pi_S = 4N_\mathrm{e}\mu$ where μ is the mutation rate per site, per generation (Charlesworth, 2009; Wright *et al.*, 1939). Of note, π_S depends on both N_e and μ. In principle, μ can be estimated, which in turns gives the possibility of directly estimating N_e based on π_S (Eyre-Walker *et al.*, 2002; Brevet and Lartillot, 2021; Lynch *et al.*, 2023). In practice, it appears that most of the variation in π_S is contributed by the N_e component (Lynch *et al.*, 2023), making π_S itself a reliable proxy of N_e. Translating the nearly-neutral prediction in this context, π_N/π_S is predicted to correlate negatively with π_S, which was indeed observed in mammals, although thus far only for the mitochondrial genome (James *et al.*, 2017).

At the scale of divergence between species, molecular evolutionary theory does not provide any simple quantity directly related to N_e. To address this problem, much of previous comparative work has relied on the assumption of a correlation between N_e and life-history traits (named "LHT" in the following). The main argument is that large and long-lived organisms are generally characterized by a lower abundance than small-bodied organisms characterized by a shorter lifespan (Damuth, 1981; White *et al.*, 2007). In turn, N_e is expected to be directly influenced by true population sizes, at least to some extent. By this argument, random drift is expected to be stronger in large and long-living species (e.g. apes) than in short-lived organisms with small body size (e.g. murid, rodents). In this context, the nearly-neutral prediction is that of a positive correlation between d_N/d_S and LHT.

Empirically, a positive correlation between d_N/d_S and LHT has repeatedly been observed (Nabholz *et al.*, 2013; Popadin *et al.*, 2007; Figuet *et al.*, 2016; Romiguier *et al.*, 2014) and generally interpreted as a confirmation of the nearly-neutral theory. Actually, the literature on this subject is somewhat contrasted, with both positive and negative results, depending on the exact traits and the clade under scrutiny. More

fundamentally, in these analyses, the negative relation between LHT and N_e is invoked for explaining the observed positive correlation between LHT and d_N/d_S, but it is most often not itself independently validated against empirical data. A more robust test of the nearly-neutral theory would certainly require to empirically validate this independent assumption, by testing the correlation between LHT and an independent measure of N_e. Alternatively, one could measure the correlation of d_N/d_S more directly with this independent measure of N_e, rather than with the indirect proxy represented by LHT.

Since no clear quantitative measure of N_e is available at the macro-evolutionary scale, an independent validation of the correlation between LHT and N_e could rely on the micro-evolutionary proxy π_S. This has been surprisingly rarely investigated, essentially in metazoans (Romiguier *et al.*, 2014) and in birds (Figuet *et al.*, 2016) – in the latter case, precisely because the positive correlation between LHT and d_N/d_S was not observed, such that a deeper investigation was felt needed. Correlation analyses between π_S and LHT have never been reported in the case of mammals. Similarly, there have been few attempts of a direct confrontation of d_N/d_S with π_S (Eyre-Walker *et al.*, 2002; Brevet and Lartillot, 2021), which have also been limited in statistical power by the use of a small number of data points for π_S. Altogether, the current picture of the validity of the nearly-neutral theory across time scales is still incomplete.

The globally incomplete and undecisive picture offered by previous work is further affected by the use of disparate methods for estimating the molecular quantities and testing for their correlations. In this respect, one important aspect is to properly account for phylogenetic non-independence (Felsenstein, 1985), and this, in a context where some of the quantities of interest (in particular d_N/d_S) are not, strictly speaking, observed in extant species (as is the case for life history traits or π_S and π_N/π_S), but are only indirectly inferred along the branches of the phylogeny.

To deal with this issue, in the previous literature, two main avenues have been explored: a sequential approach, and an integrative one. The sequential approach consists in first estimating d_N/d_S on the terminal branches of the phylogeny and tabulating those estimates along with the other relevant variables. Then this table and a phylogeny are analysed using a standard methods based on independent contrasts (PGLS or Phylogenetic Generalized Least Squares, Pagel, 1997). In brief, these methods assume that the traits of interest jointly evolve according to a multivariate Brownian process over the tree and estimate the variance-covariance matrix of the process by maximum likelihood, based on the observed values at the tips. The integrative approach, on the other hand, proposes to directly integrate the comparative test (the phylogenetic regression) into the phylogenetic codon model used for inferring the patterns of synonymous and non-synonymous substitutions over the tree (Lartillot and Poujol, 2011). In practice, this amounts to condition the codon model on branch-specific rates of synonymous and non-synonymous substitutions that are derived from the underlying multivariate Brownian process, which is otherwise identical to that assumed by the PGLS approach. These two methods have their strengths and weaknesses. However, they have rarely been confronted on the same data (but see Figuet *et al.*, 2017).

Finally, in addition to relying on a robust methodology as clarified above, a proper testing of the nearly neutral predictions bridging the two times scales also requires adequate data at the genomic scale and globally across species of a large clade. In this direction, the currently available data from population genomics projects

are still relatively scarce, making it difficult to gather data on intra-specific polymorphism for all species across a clade, that would be matched with the genes included in the multiple sequence alignment used for reconstructing the patterns of d_N/d_S over the phylogeny.

On the other hand, the increasing availability of whole genomes sequences in some taxa (Zoonomia, 2020; Feng *et al.*, 2020) suggests an alternative approach not depending on population genomics projects. At least for taxonomic groups such as mammals, one can rely on the fact that all published genomes are from diploid individuals. As such, they contain information about the heterozygosity of those individuals, which itself is informative about the genetic diversity of the population (Li and Durbin, 2011). Although published genome sequences typically do not provide this information, it is nevertheless possible to obtain it by returning to the original sequencing reads (Figuet *et al.*, 2016; Zoonomia, 2020). Compared to a true population genomics project, this way to proceed provides more limited information, potentially affected by the fact that some individuals may happen to be inbred. On the other hand, it gives access to intra-specific diversity for all species for which complete genomes are available. Thus, it makes it possible to obtain a nearly complete data matrix, genome wide and over an entire clade, with which intra and inter-specific information can be globally confronted in the context of a comparative analysis.

In this article, we propose to combine whole-genome orthologous genes and heterozygosity data across a wide range of mammals. For this purpose, we annotated the complete one-to-one orthologous genes from 144 whole genomes. We then did variant calling on them to obtain the heterozygosity on the annotated genes for all species that are present in the phylogeny. This heterozygosity can then be confronted with the rich information currently available for mammals about life history traits (De Magalhaes *et al.*, 2009). We performed our correlation study using both an integrative and sequential approach, to do a global test of the nearly-neutral theory.

2 Results

Starting from 197 complete mammalian genomes and after several steps of data processing and filtering, we gathered a dataset of 6002 single copy orthologous genes for 144 placental species. Species and genes were chosen to ensure sufficient assembly quality, and maximise completeness of the data matrix. The genes were aligned and trimmed, and a phylogeny was reconstructed using a sample of 1000 of these genes. Based on this reference dataset, we considered an additional species subsampling scheme, meant to reduce the noise contributed by non-phylogenetic variation between closely related species. Specifically, we trimmed the dataset to keep only species that are separated by at least 10 My of divergence, resulting in a dataset of 89 species (referred to as the sparse dataset in the following, as opposed to the original dense dataset of 144 species).

This provides the macro-evolutionary backbone of our analysis, which was complemented with information about intra-specific polymorphism. Specifically, for each species represented in the dataset, the original reads that were used for genome assembly were mapped onto the genome to detect heterozygous sites. We focussed on the SNPs that were called in the coding sequences of the 6002 genes. This yields a set of synonymous and non-synonymous SNPs from which a π_S and π_N/π_S were inferred.

Out of the 144 species of the dense phylogenetic dataset, 6 had an abnormal allele frequency distribution and were removed for all subsequent analyses, thus giving 138 species with data about synonymous and non-synonymous heterozygosity. Visual inspection of the relation between π_S and π_N/π_S (Figure S1) suggests that 6 additional species, all of which have low coverage, may have insufficient data for reliable heterozygosity estimation. For this reason, we decided to implement a refined version of the dataset in which the polymorphism of these six species was excluded from the analysis. Here we analyze this adjusted version of the data (thus with 132 species with data about synonymous and non-synonymous heterozygosity). Results that include the six outlier species are presented in the supplementary material (Figure S12). In the context of the sparse phylogenetic backbone, this translates into 82 (or 86 including the 6 outliers just mentioned) species with data about heterozygosity, out of the 89 species represented in the phylogeny. All species filtering and subsampling steps are summarized in Supplementary Material, Table 5.11.

The relation between π_S, π_N/π_S, d_N/d_S and LHT were then examined using both an integrative method, FastCoevol, adapted from the Coevol approach Lartillot and Poujol (2011) and a sequential approach based on PGLS.

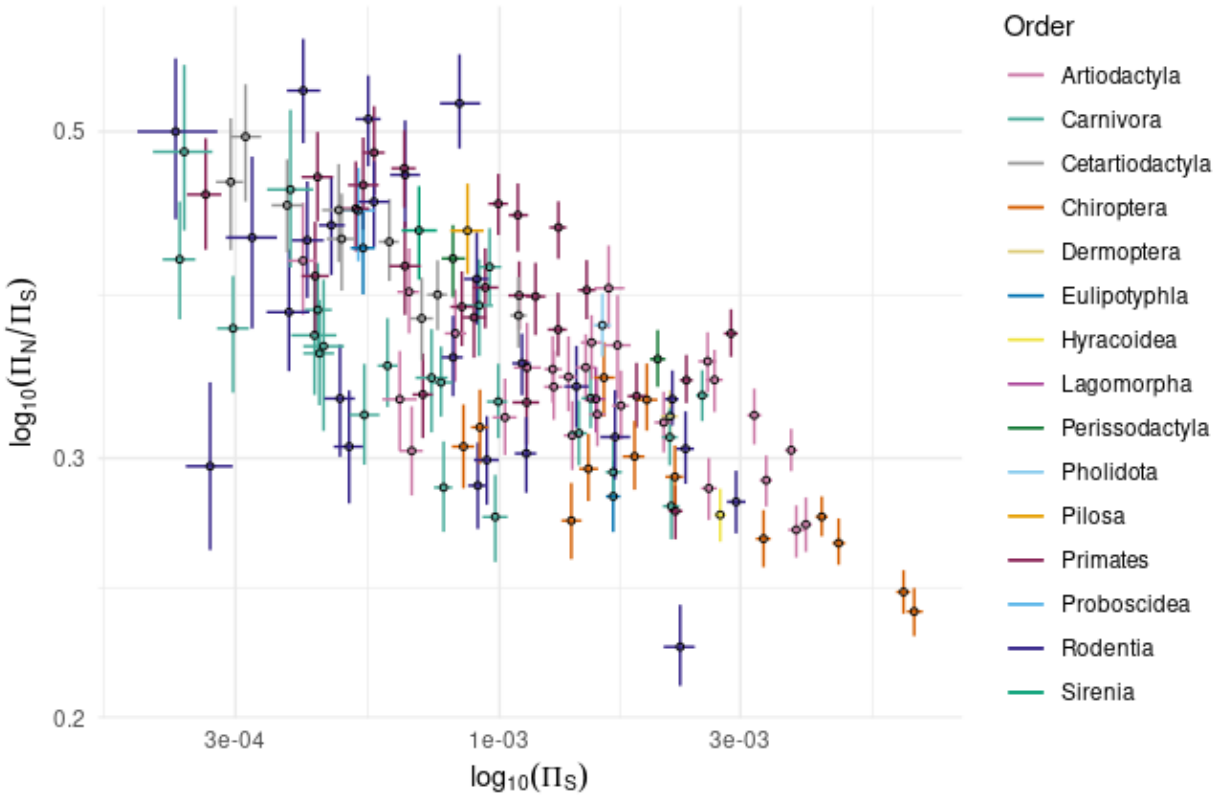

Figure 1: **Plot of π_N/π_S against π_S (in log-log scale)** for all species of the analysis except the six species with less than 1000 SNPs (bars: 95% CI computed by non-parametric bootstrap). Colours correspond to species families.

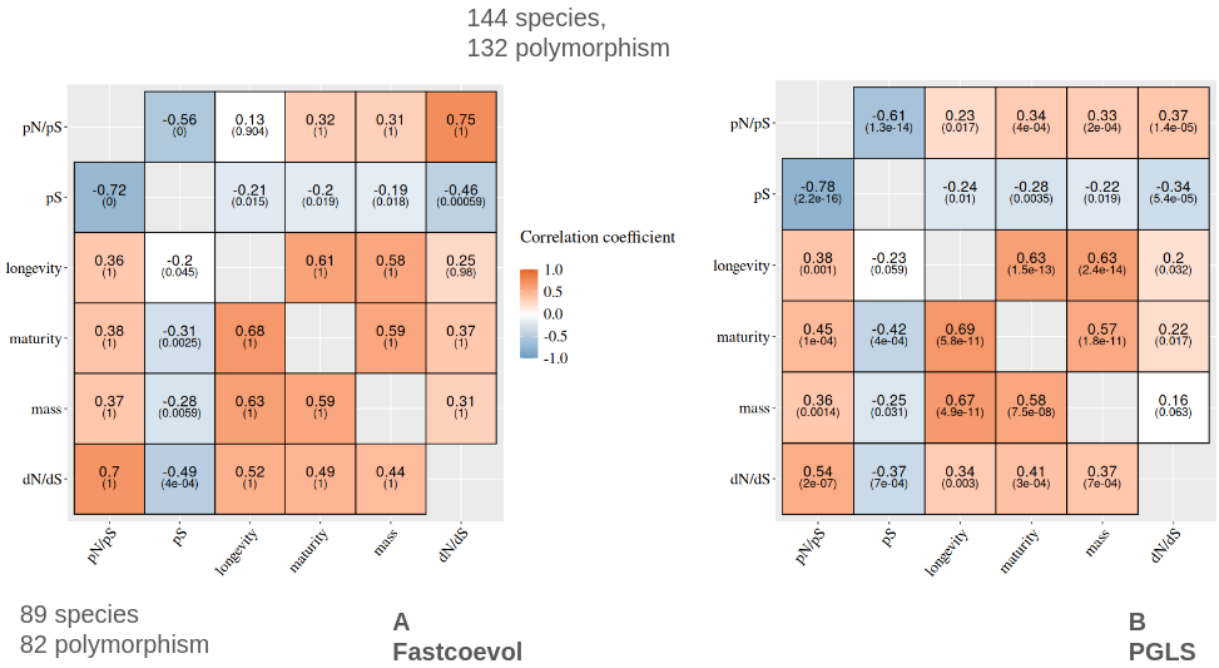

Figure 2: **Correlation coefficients from the FastCoevol and PGLS analyses**. **A**: FastCoevol analysis. **B**: PGLS analysis. **Upper diagonal**: using the 144-species phylogeny. **Bottom diagonal**: using the 89-species phylogeny. Colours correspond to the magnitude of the correlation coefficients (r). Numbers in brackets correspond to the posterior probability for FastCoevol analyses (which must be higher than 0.975 or lower than 0.025 for the correlation to be considered significant) or p-value for PGLS analyses (which must be lower than 0.05 for the correlation to be considered as significant). Non-coloured estimates correspond to non-significant correlations.

2.1 Micro-evolutionary scale : π_S versus π_N/π_S

At the micro-evolutionary scale, the efficacy of selection is reflected by π_N/π_S, while π_S provides a direct proxy of N_e. The nearly-neutral theory predicts a negative correlation between these two measures. Plotting π_N/π_S against π_S (Figure S1), a clear negative correlation is indeed observed between these two variables, except for six species (the hazel dormouse *Muscardinus avellanarius*, the hirola *Beatragus hunteri*, the cotton rat *Sigmodon hispidus*, the black rhinoceros *Diceros bicornis*, the humpback dolphin *Sousa chinensis* and the mantler howler *Alouatta palliata*) showing a abnormally low π_S, given their π_N/π_S. This could be the sign of recent inbreeding. However, these six species also tend to show a low genomic coverage and, for some of them, a poor genome assembly (see Material and Methods), which suggests more a data quality issue for these outlying points rather than a low real π_S.

Excluding these six low π_S species, a clear log-linear negative relation between π_S and π_N/π_S is observed (Figure 1), which is confirmed by the phylogenetic correlation analysis (Figure 2). The correlation is also significant and present similar correlation coefficient when these six species are included (Figure S12). This

is compatible with a more efficient purging of slightly deleterious mutations (lower π_N/π_S) in species with large N_e (larger π_S), in agreement with the nearly-neutral prediction.

Of note, we observe a stronger correlation between π_S and π_N/π_S with the sparse (89 species) phylogenetic dataset than with the original dense version with 144 species ($r = -0.72$ versus $r = -0.56$ with FastCoevol, $r = -0.78$ versus $r = -0.61$ with PGLS, Figure 2, compare above and below the diagonal in each panel). It may seem paradoxical that removing data points increases the strength of the inferred correlations. A reasonable explanation is that the independent contrasts between closely related species are dominated by estimations errors or short-term effects and that reducing the number of species reduces this source of noise.

The slope of $\log \pi_N/\pi_S$ as a function of $\log \pi_S$ is estimated at -0.16, with 95% credibility interval (-0.20, -0.13). This is smaller in absolute value than what was found in mammalian mitochondrial genomes (of the order of -0.63, James *et al.* (2017)), or in primates (-0.29, Brevet and Lartillot (2021)).

2.2 Macro-evolutionary scale : life history traits versus d_N/d_S

At the macro-evolutionary scale, life history traits are known to be strongly correlated with each other across mammals. This is recovered in our analysis using both PGLS and FastCoevol (Figure 2). As for d_N/d_S, its history over the phylogeny such as inferred by FastCoevol is shown in Figure 3 which presents the d_N/d_S reconstruction along the 89-species tree. From this figure, we observe a d_N/d_S ranging from 0.13 to 0.28 over the tips. Taxonomic groups with the lowest d_N/d_S are Rodentia (i.e. *Peromyscus maniculatus*, the eastern deer mouse, $d_N/d_S = 0.139$) and Lagomorpha (i.e. *Ochotona princeps*, the american pika, $d_N/d_S = 0.13$). At the other end of the range, taxonomic groups with the highest d_N/d_S correspond to Cetacea (i.e. *Eschritius robustus*, the gray whale, $d_N/d_S = 0.285$) and Primates (i.e. *Alouata palliata*, the mantled howler $d_N/d_S = 0.222$). This general observations from Figure 3 is globally suggestive of a pattern of a high d_N/d_S in large and long lived mammals (also recovered using the 144 species tree S5).

On the dense (144 species) tree, the FastCoevol analysis shows a significant positive correlation between d_N/d_S and all life history traits, with correlation coefficients varying from $r = 0.25$ to 0.37 depending on the trait (Figure 2 A, upper diagonal). Similarly, the PGLS analysis shows positive correlations ($r \simeq 0.21$), all significant except for the correlation between d_N/d_S and body mass (Figure 2 B, upper diagonal). The correlations are slightly stronger with FastCoevol, which can be explained by the fact that the additional variance contributed by the short-term overdispersion of d_N/d_S, which is modelled by FastCoevol, has been discounted. Using the sparse (89 species) tree, we observe stronger correlations under both methods. With the sparse dataset, correlations between d_N/d_S and mass also become significant in the PGLS analysis. This effect is also presents in the FastCoevol analysis, which suggests that the short-term over-dispersion of d_N/d_S along branches is not totally captured by the overdispersion component of the integrative approach.

Altogether, both PGLS and FastCoevol analyses agree on a significant positive correlation between d_N/d_S and LHT at the macro-evolutionary scale. These results are generally interpreted as an indirect correlation of both variable with N_e and are thus compatible with a nearly-neutral interpretation (Ohta, 1973b, 1992) suggesting less efficient selection in long-lived, large-bodied mammals, presumably characterised by lower N_e

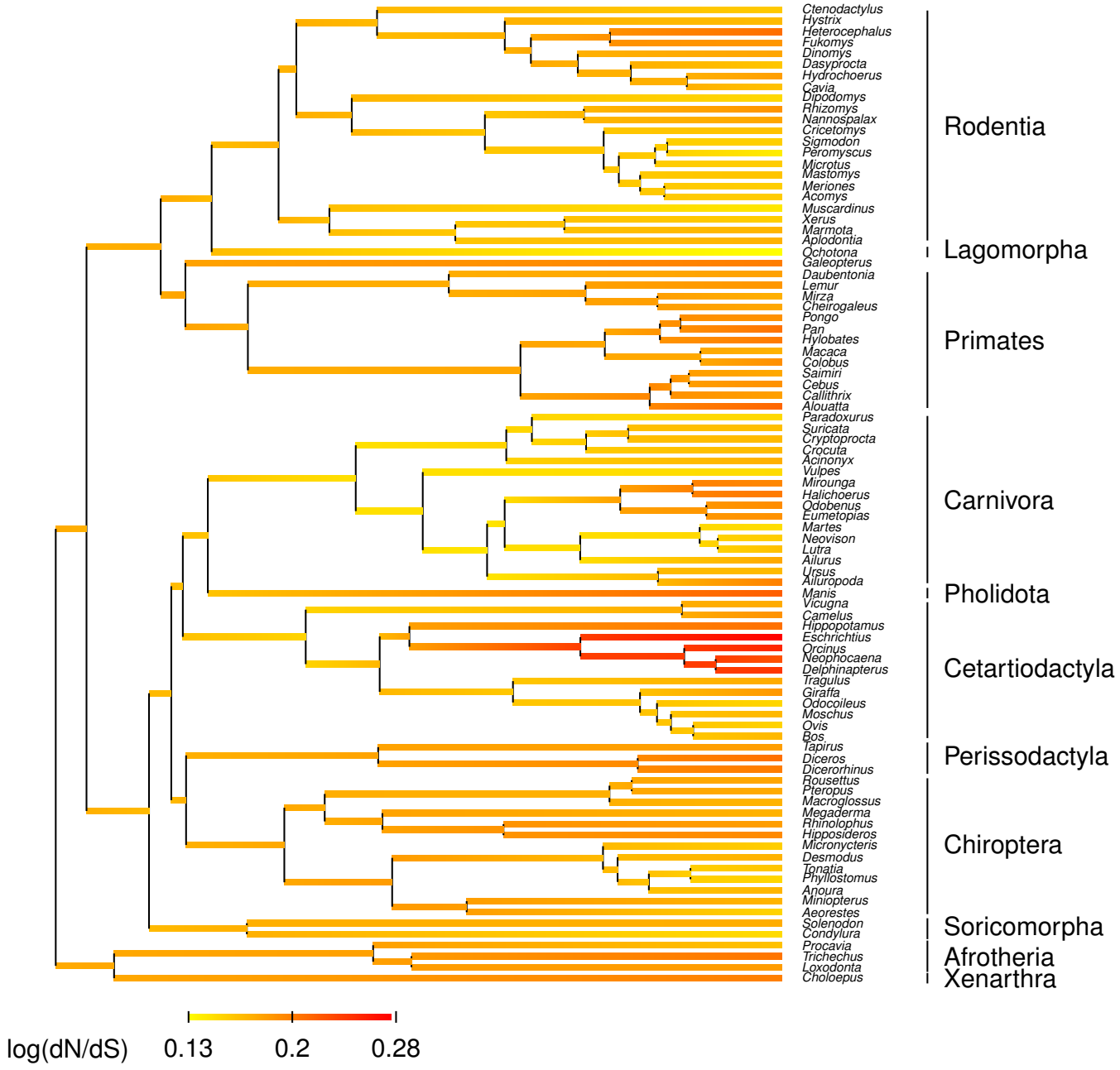

Figure 3: Reconstruction of $\log(d_N/d_S)$ along the phylogeny of placental mammals (89 species) by FastCoevol.

over the tree. As it stands, however, this interpretation is reliant on the additional and thus far untested assumption that LHT indeed correlate with N_e, a point which we now investigate.

2.3 Confronting the micro- and macro-evolutionary scales

If macro-evolutionary patterns are the long term result of micro-evolutionary processes, we expect to observe correlations of our metrics of interest across timescales. However, this depends on the extent to which short-term N_e (such as measured by π_S) reflects long-term N_e, and to what extent long-term N_e is indeed reflected in life history traits.

In this direction, with FastCoevol, we found significant negative correlations between π_S and life history traits, although only in the analysis without the six low-π_S species. Even then, the correlation coefficients are weak and below the significant threshold for longevity in the 89 species analysis (Figure 2A, bottom diagonal). The PGLS analysis also shows weak correlations and the statistical support vary depending on the life history traits and the taxon sampling (Figure 2B).

On the other hand, π_N/π_S shows a robust negative correlation with life history traits, for the sparse and dense datasets and with the two methods. The most straightforward explanation for this correlation is that both variables are indirectly correlated with N_e. The weaker correlation between π_S and LHT, compared to that between π_N/π_S and LHT, could be due to normalization errors in π_S and π_N estimates, which would cancel in their ratio (see discussion).

Here also, as above for d_N/d_S versus life history traits, the correlations are stronger and more consistent using FastCoevol rather than PGLS and when the dataset is thinned to remove closely-related species. Again, this could be due to the account for short-term overdispersion of d_N/d_S performed by FastCoevol and to the negative impact of non-phylogenetic variation. In the present case, short-term fluctuation in N_e may contribute. Taken together, these correlation patterns of π_S and π_N/π_S with life history traits, which are congruent in their sign, but contrasted in their strength and significance, are consistent with the hypothesis of a connection between short- and long-term Ne.

Conversely, and finally, d_N/d_S correlates negatively with π_S, thus providing a more direct evidence of an effect of N_e on the efficacy of selection at the macro-evolutionary scale. The slope of $\log d_N/d_S$ as a function of $\log \pi_S$ is equal to $-0.07(-0.12, -0.04)$. It is thus significantly shallower than that of $\log \pi_N/\pi_S$ as a function of $\log \pi_S$ (with non-overlapping credible intervals). The d_N/d_S also correlates positively with π_N/π_S. Overall, the correlation coefficient are globally stable across the different analyses for d_N/d_S versus π_S, but more variable for d_N/d_S versus π_N/π_S.

3 Discussion

The nearly neutral theory states that selection is less efficient at purging mildly deleterious mutations in smaller populations. Accordingly, it predicts higher levels of segregating non-synonymous polymorphisms within species and higher rates of non-synonymous substitutions between species, relative to the synonymous compartment, when N_e is low (Ohta, 1973b). In this study, we aim to confirm this hypothesis at two time scales : micro-evolutionary (using π_S and π_N/π_S) and macro-evolutionary (using life history traits and d_N/d_S), based on a large dataset of 6002 genes in 144 placental mammals.

The correlations observed in our study are recapitulated in Figure 4. Overall, all correlations are globally compatible with a nearly neutral interpretation. More specifically, all could be explained by all variables being correlated with a single hidden variable, which is N_e. The results are robust to alternative subsampling schemes, except for the correlation between π_S and life history traits, which is perhaps the weakest and least convincing among all observed correlations.

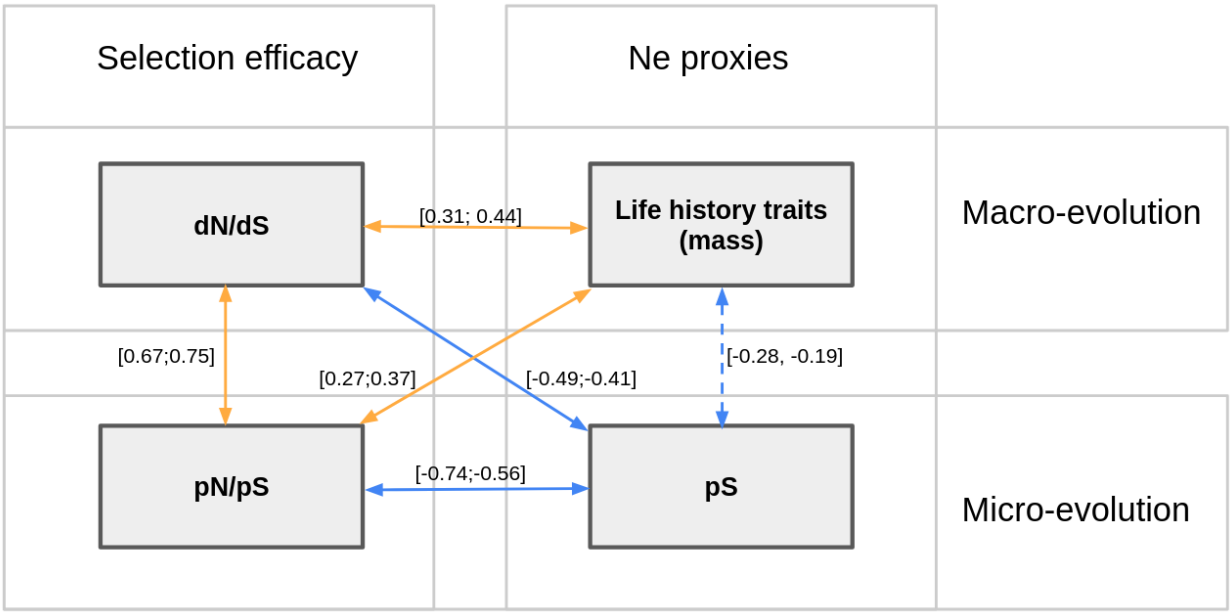

Figure 4: **Schematic summary of the correlations observed in the present study using FastCoevol.** Negative correlations are in blue, positive correlations in orange. Values within brackets correspond to the range of correlation coefficients obtained under the four different analyses using FastCoevol (sparse and dense dataset, with and without low π_S species). Mass is used here to represent life history traits. The dotted line between life history traits and π_S represents the instability of the relationship, which depends on the dataset used.

As presented in the introduction, several prior studies have already examined some of these correlations. In Table 5, we present a summary from the published literature concerning both the significant correlations and those sought but not found to be significant (labelled as "NS"). These different projects are not easy to compare as they are not consistent in data and methods used. Our results are also reported in this table, which thus highlights the correlations that we found and that were missing until now. Among them, the most important ones are: first, the correlation between LHT and π_S, which, even if it is weak, gives more objective support to the belief that LHT reflect N_e in mammals; second, the correlation between d_N/d_S and π_S, which provides a more direct test of a nearly-neutral effect on long-term protein evolution than the now well-established correlation between d_N/d_S and LHT; and finally, the correlation between π_N/π_S and π_S, which provides an equivalent test of the nearly-neutral predictions at the micro-evoluationary scale, and which was previously established only in the case of mitochondrial genes in mammals (James *et al.*, 2017) and was still missing for the nuclear compartment.

		dN/dS	LHT			pS	pN/pS
			mass	maturity	longevity		
dN/dS			Figuet (2016) r=0,38* Nabholz (2013) r=0,47** Lartillot Poujol (2011) r=0,58** Lartillot (2013) r=NS Lartillot Delsuc (2012) r=NS Popadin (2007) r=0,4** r=[0,29**;0,44**] 2/8 NS	r=0,71** r=0,38** r=NS r=0,27** r=0,36** r=[0,22*;0,49**]	r=0,69** r=0,51** R=0,13** r=0,27** r=0,27* r=[0,2*;0,53**]	r=[-0,49** ;-0,37**]	r=[0,37**;0,75**]
LHT	mass	Brevet (2021) primates r=NS Lartillot Poujol (2011) carnivores r=0,9** Figuet (2016) birds and reptiles r=NS Nabholz (2013) birds r=NS				r=[-0,28** ;-0,19**] 2/8 NS	r=[0,27**;0,41**]
	maturity	Brevet (2021) primates r=NS Lartillot Poujol (2011) carnivores r=NS Figuet (2016) birds r=NS ; reptiles r=0,52* Nabholz (2013) birds r=NS				r=[-0,42** ;-0,19*] 2/8 NS	r=[0,26**;0,46**]
	longevity	Brevet (2021) primates r=NS Lartillot Poujol (2011) carnivores r=NS Figuet (2016) birds r=NS ; reptiles r=0,68** Nabholz (2013) birds r=0,63**				r=[-0,24* ;-0,21**] 5/8 NS	r=[0,21*;0,38**] 2/8 NS
pS		Brevet (2021) primates r=NS	Brevet (2021) primates r=NS				James (2017) r=-0,98**
			Romiguier (2014) metazoan r=-0,32*		r=-0,53*		Piganeau Eyre-Walker (2009) r= -0,28*
			Figuet (2016) birds r=-0,39*	r=-0,52**	r=-0,45*		r=[-0,72** ;-0,56**]
pN/pS		Brevet (2021) primates r=NS	Brevet (2021) primates r=NS			Brevet (2021) primates r=-0,78**	
						Romiguier (2014) metazoan r= -0,71*	
			Romiguier (2014) metazoan r=0,28*		r=0,62*	Piganeau Eyre-Walker (2009)	
			Figuet (2016) birds r=0,46*	r=0,5**	r=0,45*	metazoan r=-0,11*, birds r=NS	

Figure 5: **Table summarizing previously published results.** Upper diagonal : Studies on mammals. Lower diagonal : Studies on other taxonomic groups. "*" mean a significant correlation (p-value < 0.05, pp> 0.95 or pp< 0.05), "**" strongly significant correlation (p-value < 0.001, pp> 0.975 or pp< 0.025); "NS": non-significant correlation. Results in bold blue correspond to our study. The values are within brackets as they represent the min and max of the correlation coefficient from the eight alternative sampling and method scheme. When a correlation is not significant in one of the eight modalities, we emphasize it by giving the number of non statistically significant observations over the eight. Results in pink correspond to analysis using mitochondrial data.

3.1 Limitations and potential sources of error

All the successive steps in our pipeline, from genome annotation to the alignment of gene sequences and the reconstruction of a phylogeny, are subject to different types of errors which could accumulate and could impact our conclusions (Simion *et al.*, 2020). Moreover, genomes of poor quality could be particularly problematic. Given the variable quality of the genome assemblies used here, a lot of work has been done to control for data errors (See Material and Methods).

Nevertheless, some errors may still be present in some amount. It thus raises the question of whether they could drive or compromise some of the correlations observed here. Different sources of errors at micro and macro-evolutionary scale, which are susceptible to bias the correlation analyses, can be identified.

Errors in alignment and SNP calling

At macro-evolutionary scale, alignment errors tends to inflate the apparent d_N/d_S. This effect is stronger for short branches (measured in synonymous evolutionary divergence) which could create an artefactual correlation between d_N/d_S and d_S. If long-living species tend to have a low d_S, this could induce an artefactual correlation between d_N/d_S and life history traits. Similarly, at micro-evolutionary scale, SNP calling errors will tend to inflate the apparent π_N/π_S. This effect will be stronger in species with a low true π_S and can create an artefactual correlation between π_S and π_N/π_S. By extension, if long-living mammals tend to have a low π_S, this will induce an artefactual correlation between π_N/π_S and life history traits. It is still difficult, as it stands, to completely rule out the possibility that some of the correlation that we observe here are driven by these potential biases. Of note, concerning π_N/π_S, this would imply a strong correlation between π_S and life-history traits.

Estimation of the number of mutational targets and its impact on π_S and π_N

To compute the synonymous and non-synonymous rates at the two scales, we need to normalise the counts by the number of mutational targets. In the case of polymorphism, this in turn depends on the estimation of the number of callable positions. Here, the number of callable positions is not readjusted after the filtering of low quality SNPs. This is particularly an issue concerning lower quality genomes because they present less raw SNPs detected by variant calling and lower SNP quality overall, inducing a stronger reduction of the total number of final SNPs, without removing them from the callable set. Of note, this bias should affect π_S and π_N but should cancel in their ratio, and thus is not expected to impact π_N/π_S. Such normalisation problems could explain the fact that π_S has much weaker correlation than π_N/π_S with the other traits.

Segregating polymorphism and its impact on d_N/d_S

Segregating polymorphism implies that the apparent d_N/d_S on the terminal branches of the tree contains a certain proportion of π_N/π_S in its estimation. This could contribute to making the apparent correlation between π_N/π_S and d_N/d_S stronger than it actually is. Indeed, this correlation is among the strongest observed in our analysis. More generally, this suggests care in interpreting the correlation of d_N/d_S with other variables, as it cannot be excluded that they entail a hidden correlation with π_N/π_S. On the other hand, this effect cannot explain everything – in particular, it cannot explain the fact that the correlation between d_N/d_S and LHT is stronger than that between π_N/π_S and LHT.

Use of heterozygosity of an individual genome as a proxy for population diversity

The use of the heterozygosity of a single individual as a proxy for population-level diversity has several conceptual and practical limitations. First, even if the heterozygosity computed on a whole genome is supposed to represent the population polymorphism, it can still deviate substantially from the population mean for a lot of reasons (inbreeding, population structure, bottleneck *etc*). Second, individual heterozygosity is influenced by genome-wide stochasticity. Even in a large, panmictic population, the realized heterozygosity of one diploid genome can differ from the expectation simply by chance, and this variance remains unquantified. Third, technical factors such as sequencing depth, variant calling errors, and assembly fragmentation can disproportionately affect heterozygosity estimates when only one genome per species is available. These limitations imply that heterozygosity from a single individual may not exactly capture the polymorphism of the population. However, in our study, both π_S and π_N/π_S are derived from the same individual, so their correlation is robust to these sources of variance even if their absolute values may not precisely reflect the real π_S and π_N/π_S.

Altogether, our results are still subject to several limitations and remain to be confirmed, in particular by implementing a more robust normalisation of π_S and by quantifying the contribution of the polymorphism to the d_N/d_S estimates. Nevertheless, as it stands, at worse the relation between π_S and LHT is underestimated, and the relation between π_N/π_S and LHT is perhaps the most robust result reported here pertaining to the relationship between micro and macro evolution.

3.2 Quantitative scaling of π_N/π_S and d_N/d_S as a function of π_S and N_e

The slope of the relationship between π_N/π_S and π_S, or that between d_N/d_S and π_S (on a log-log scale), have been of some interest in the previous literature, owing to their bearing on the shape of the underlying distribution of fitness effects. Specifically, if the DFE is gamma, of shape parameter β, and in the limit where β is small, then the slope of log π_N/π_S or d_N/d_S as a function of log Ne is expected to be equal to $-\beta$ (Welch *et al.*, 2008). Of note, this assumes that the DFE is constant across species, a point which has been questioned, both empirically (Castellano *et al.*, 2019) and theoretically (Goldstein, 2013; Latrille *et al.*, 2021). Also, as discussed in Brevet and Lartillot (2021), the variation of μ between species, which tends to be opposite to that of N_e, can result in the slope of log π_N/π_S as a function to log π_S being smaller that that of log π_N/π_S as a function of log N_e.

Here we see that the slopes as a function of π_S are globally shallower than previously reported results for mitochondria across mammals for π_N/π_S (James *et al.*, 2017) or nuclear genomes in primates for both π_N/π_S and d_N/d_S (Brevet and Lartillot, 2021). This could be the result of different behaviors in different genomic compartments or of clade-specific effects. Alternatively, incorrect estimation of π_S (see above) could also contribute to making apparent slopes shallower than true slopes.

We also observe a shallower slope for d_N/d_S than for π_N/π_S (as in Brevet and Lartillot (2021), but now with non-overlapping credible intervals). As discussed in Brevet and Lartillot (2021), this could be due to a non-nearly neutral contribution to the d_N/d_S (e.g. positive selection), unrelated to N_e. Alternatively, a

random short-term effect on N_e, on the top of its long-term phylogenetic trends, is expected to push the slope of d_N/d_S (but not π_N/π_S) towards smaller values.

3.3 Perspectives on micro versus macro-evolution

Our analysis provides a decisive entry point for further exploration of the micro-macro connections. About the methods, in the end, the use of an integrative (FastCoevol) rather than a sequential (PGLS) method doesn't seem to make a big difference. They both gives similar results. However, integrative methods can be furthered developed to reconstruct N_e. In particular, in this article, we make the assumption that π_S represents N_e, using the relation $\pi_S = 4N_e\mu$. However, μ itself varies between species, and is known to correlate with life history traits (Lanfear *et al.*, 2014; Lynch *et al.*, 2023). Correcting for μ would therefore be important in order to have a direct access to N_e. A classical approach to correct for μ is to estimate it by using the synonymous divergence between close species and fossil calibration with a generation time (Eyre-Walker *et al.*, 2002). In the integrative method, we can incorporate a decomposition of π_S into N_e and μ components by the reconstruction of μ along the topology. This was already implemented in Brevet and Lartillot (2021) and could be recruited and applied to the dataset obtained here for mammals.

Using this approach would make it possible to estimate slopes directly with respect to N_e, not π_S. However, this would first require explicitly modelling the mismatch between short-term and long-term N_e. Such an improvement in the model would provide a more quantitative idea about the intensity of short term fluctuations of N_e and see in particular whether it explains the shallower slope for d_N/d_S, compared to π_N/π_S.

Finally, here we focused on nearly neutral evolution, but our dataset could be used for investigating other aspects, as it presents the opportunity to compare two genomic compartments which respond differently to certain evolutionary processes such as positive selection (McDonald and Kreitman, 1991) or mutational bias and genetic biased conversion toward GC (Duret and Galtier, 2009).

4 Materials and Methods

Figure 6 summarizes the pipeline developed and used here, from the acquisition of the data to the final analyses. Below, we describe in detail the different steps of this pipeline.

4.1 Data acquisition

Four types of input data were used (depicted in light blue in Figure 6): complete genomes in fasta format, their associated reads, a BUSCO database and life-history traits. The reference genome assemblies were downloaded from NCBI Genomes in July 2021. At this time, we selected the assemblies of eutherian mammals using different criteria. First, we selected assemblies with median contig size (N50) of at least 30kb to avoid overly truncated genes. We then select assemblies with at least one individual with sequencing depth higher than 20x according to NCBI Genomes statistics. In addition, we excluded the assemblies from

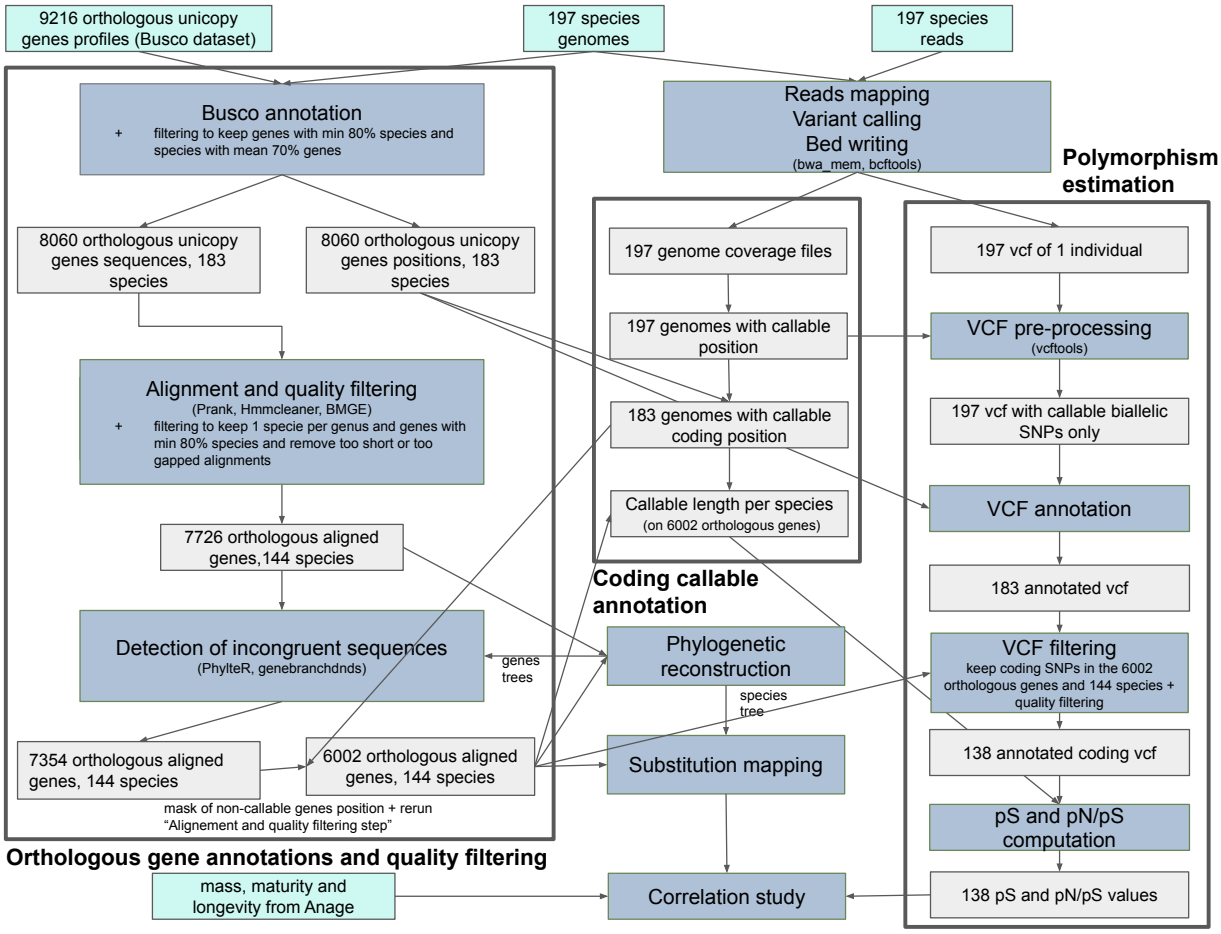

Figure 6: **Simplified summary of the pipeline from whole genomes acquisition and vcf writing to the final correlation analysis**. Light blue boxes correspond to the input data, grey boxes to the intermediate data and blue boxes to different steps of the pipeline. A more complete and more detailed desciption of the pipeline is available at https://github.com/MeloBastian/Mam_Ne

domestic species whose genetic variation is well-known to be reduced by the domestication process and the lab individuals that are usually highly inbred.

For each of the 197 assemblies that met all these requirements, we selected the latest reference assembly available in July 2021. Because the entire analysis relies on heterozygosity in single individuals, when multiple individuals were sequenced for a genome assembly project, we only selected and downloaded the reads from the individual with the highest sequencing depth. Sequencing reads were then mapped on their corresponding assembly with bwa_mem2 v2.2.1 used with default parameters (Vasimuddin *et al.*, 2019). We then conducted a variant calling to identify the heterozygotes positions and create for each species a whole-genome sequencing depth files, which was very useful for the sequences filtering.

4.2 Orthologous gene annotation and filtering

We developed a pipeline starting from the 197 assembled complete genomes, producing in output a data matrix of aligned and filtered single copy orthologous genes. At different steps of the process, quality controls and filters were applied (detailed below), leading to the removal of some species or genes from the analysis. Additionally, closely related species can share ancestral polymorphism. This shared polymorphism can affect the estimation of the d_N/d_S. To mitigate this issue, we chose to retain only one species per genus. The reduced dataset at the end of the pipeline contain 144 species and 6002 single-genes alignments (genes number per species and other metrics are resumed in Supplementary material, table 5.11).

4.2.1 Identification of orthologs

Not all genomes are annotated, and those that are were annotated using different procedures. Since our focus is on orthologous coding sequences, and to ensure a homogeneous treatment across the entire species set, we annotated all genomes using BUSCO (Simão *et al.*, 2015). BUSCO is a commonly used bioinformatics tool for assessing genome assembly quality based on the identification of orthologous genes assumed universal across the taxonomic group specified. Here we relied on the mammalian database (orthodb.10), defining a set of 9226 orthologous unicopy genes across this clade. The program provides a percentage of recovered orthologous genes, which is meant as an indicator of the genome quality. BUSCO also provides intermediate output files specifying the coordinates of the orthologous genes identified in the genome being analysed. These intermediate files were processed so as to generate, for each genome, a fasta file containing all coding sequences and a bed file with their position along the genome. From our initial dataset of 197 species, we only kept species with BUSCO complete single-copy score higher than 70%, meaning that at least 70% of the searched genes were found. In a second step, we removed genes present in less than 80% of the species to ensure their universality across placental mammals. After this step, we are left with 183 mammalian species and 8060 BUSCO one-to-one orthologous complete genes.

4.2.2 Alignment and quality filtering

The sequences of each orthologous gene were aligned using PRANK (Löytynoja, 2014). To remove poorly aligned regions that could bias the phylogenetic inference, we filtered them using HMMcleaner (Di Franco *et al.*, 2019) and BMGE (Criscuolo and Gribaldo, 2010) with default parameters. These two methods are complementary and improve the global quality of the alignments. BMGE trims unreliably aligned columns based on gap patterns, while HMMCleaner identifies and removes short, species-specific segments that deviate from the profile HMM of the alignment (Ranwez and Chantret, 2020).

Upon filtering, some alignments ended up empty or with a lot of gaps. We removed orthologous genes for which the alignment has less than 10 nucleotides or more than 75% gaps, resulting in 8056 aligned fasta files. Next, our dataset was reduced to only one species per genus leading to a set of 144 species. When multiple species occur for a genus, the one with the higher whole-genome coverage was chosen. We then removed the genes with less than 80% occurrence in this species set, resulting in a list of 7726 genes.

4.2.3 Detection of incongruent gene using gene tree

Although BUSCO aims at identifying single-copy orthologs, some alignments may nevertheless contain sequences with incorrect orthology assignments. Gene-level alignments for which the true evolutionary history differs from that of the species may introduce noise (Boussau and Scornavacca, 2020) and have a strong adverse input on the study and in particular on the species-level d_N/d_S estimates. Other sources of errors may also be present, such as sequencing errors resulting in compensated frameshifts, incorrect exon identification, etc.

To identify such deviant sequences and genes, we used a combination of two approaches. First, we reconstructed the 7726 gene trees using Iqtree2 with defaults parameters (Minh *et al.*, 2020). The trees where then given as an input to PhylteR (Comte *et al.*, 2023), which analyses all gene trees and provides a list of sequences and genes exhibiting outlier behaviour. Second, we developed a Bayesian shrinkage model (see Supplementary Material section 5.12) to identify sequences showing large deviations from the global pattern of synonymous branch lengths across the tree, which are expected to be very similar across genes.

PhylteR detected 20 genes that required complete removal and 1565 genes that contain at least one sequence to be removed. The Bayesian shrinkage model identified 1695 genes with at least one sequence displaying a deviant synonymous branch length. In total, 2015 genes out of the initial 7726 were flagged for at least one sequence, by either PhylteR or the Bayesian shrinkage model, with 1074 genes identified by both methods. We removed all the flagged sequences from the alignments and also genes containing less than 80% of the species after this filtering step, resulting in a dataset set of 7354 gene alignments.

4.2.4 Callable fraction of the genome

Using genomes and their associated reads mapped onto them, we computed bam files containing the sequencing depth at each position of the genomes using bcftools v1.13 (Danecek *et al.*, 2021). Given the large size of the resulting files, we converted the sequencing depth per position into a sequencing depth per window of 100 bp. We then computed for each species a mean genome sequencing depth and identified the 100 bp windows for which sequencing depth is two times higher or lower than this mean. These sites were considered as non-callable and were masked for the rest of the analysis. One species, *Neotragus schauinslandi*, for which more than 80% of the genome was masked, was removed from the analysis. Without *Neotragus schauinslandi*, the mean non-callable fraction of the genome is around 12% with a maximum of 40%.

Of note, some of the genomic regions considered as non-callable according to the criteria just described may overlap the coding sequence contained in the multiple sequence alignment across placentals (see above). To address this potential issue, we came back to the 7354 genes obtained after the use of PhylteR and the Bayesian shrinkage model, and replaced the non-callable coding sites by gaps directly in their Prank output. We then reran the filtering pipeline described above from HMMCleaner to the exclusion of the too short or gapped sequences and of the genes represented in less than 80% of species. This resulted in a final high-quality dataset of 6002 genes with few missing data.

4.3 Alternative species subsampling schemes

To investigate the impact of closely related species in our correlation analysis, we subsampled the 144 species to 89 species as follows. Using the dated tree reconstructed by the integrative analysis on the 144 species, we identified all the subclades whose most recent common ancestor is younger than 10% of the total tree depth (corresponding to approximatively 10 My, assuming a date of 100 My for the most recent common ancestor of placental mammals) and chose only one species per subclade, prioritizing species with higher sequencing depth and with all three life history traits informed. This leads to a reduced dataset of 89 species (listed in Supplementary Material). The 6002 single-genes alignments where restricted to these 89 species without realigning. At this step, we cannot remove genes present in less than 80% of species without removing too many genes.

4.4 Phylogenetic reconstruction

For the 89 and 144 species dataset, the species tree was estimated using Iqtree2 (Minh *et al.*, 2020) with the GTR+G4 model, on a concatenation of 1000 genes. The trees are in Supplementary Material (Figure S6 and S7). For gene trees, required for Phylter (Comte *et al.*, 2023), we use Iqtree2 too with default parameters and model search options.

4.5 VCF annotation and polymorphism

Variant calling was performed using default parameters of bcftools v1.13 (Danecek *et al.*, 2021) (https://samtools.github.io/bcftools/howtos/variant-calling.html). We obtained VCF files containing all variant sites detected. We used vcftools (Danecek *et al.*, 2011) to keep only bi-allelic SNPs and remove the SNPs located in the non-callable regions such as previously defined.

SNPs were annotated as synonymous, non-synonymous or non-coding, using a custom Python script that maps the SNPs and the coding orthologous annotation on the genomes. The script also labels each SNP by the name of the gene it belongs to. We then removed the non-coding SNPs and the ones not in the restricted list of 6002 orthologous genes. Furthermore, SNPs with GQ<150 and QUAL < 125 were removed as well as SNPs with an allelic frequency higher than 0.8 or lower than 0.2. After this step, we decided not to further consider the VCF of six species because they present a global abnormal allelic frequency distribution: *Acomys cahirinus* (Cairo spiny mouse), *Przewalskium albirostris* (Thorold's deer), *Mastomys coucha* (Southern multimammate mouse), *Litocranius walleri* (Gerenuk) ,*Cheirogaleus medius* (Fat-tailed dwarf lemur) and *Cephalophus harveyi* (Harvey's duiker) (see Supplementary Material, Figure S4). A summary of the number of SNPs per species at each step is provided in Supplementary Material table 5.11.

After all these filters, six species ended up with less than 1000 SNPs: *Muscardinus avellanarius* (Hazel dormouse), *Beatragus hunteri* (Hirola), *Sigmodon hispidus* (Hispid cotton rat), *Diceros bicornis* (Black rhinoceros), *Sousa chinensis* (Indo-Pacific humpback dolphin) and *Alouatta palliata* (Mantled howler). These species also have a low genome quality based either on the N50 and L50 metrics or on the genome coverage. Below, we will show correlation analyses with and without these 6 species to evaluate their impact on the

analysis. As we observe difference in the results using these two dataset, we systematically report the results with and without these six particular species (see Supplementary Material, section 5.9)

Estimates of π_S and π_N were obtained by counting the total number of synonymous (S) and non-synonymous SNPs (N) and then normalizing by a rough estimate of the number of synonymous and non-synonymous mutational opportunities (see Supplementary Material 5.2 for a more refine estimate). Thus, denoting by L the total number of coding and callable position across all genes, K_S and K_N the number of heterozygote synonymous and non-synonymous positions, and assuming that point mutations in coding sequences are twice more likely to be non-synonymous than synonymous:

$$\pi_S = \frac{K_S}{L}$$

$$\pi_N = \frac{K_N}{2L}$$

We then compute π_N/π_S using our π_S and π_N measures (reported in Supplementary Material table 5.11).

The sampling variance of our π_S and π_N/π_S estimates was estimated by bootstrap at the gene level (i.e. drawing 6002 genes from the list with replacement and recomputing π_S and π_N/π_S). A 95% confidence interval was computed.

We also question the independence between π_S and π_N/π_S witch is itself constructed using a π_S and π_N from a same set of genes (i.e. the independence between A and $\frac{B}{A}$). Results are presented in Supplementary Material, Section 5.7.

4.6 Life history traits

Data about adult body mass, longevity and age at female sexual maturity (called "maturity" for short in the following) across mammals were obtained from the Anage database (De Magalhaes *et al.*, 2009). To mitigate lack of data for some species, we systematically compute the mean of the trait per genus, even if the trait exists for the species (discussed in Supplementary material 5.8). We obtain respectively for the 144 and 89 species datasets, 127 and 81 data points for mass, 118 and 73 for maturity and 118 and 74 for longevity.

4.7 Phylogenetic correlation analysis

Phylogenetic correlation analyses between life history traits and molecular variables such as π_N/π_S or d_N/d_S (here after, collectively referred to as "traits"), were conducted using two alternative methods, both of which account for phylogenetic inertia.

4.7.1 Bayesian integrative approach: FastCoevol

Coevol (Lartillot and Poujol, 2011) is an integrative Bayesian approach for analysing the joint evolutionary patterns over the phylogeny of observable traits (such as body mass) and substitution rates (here, d_S, d_N/d_S). The underlying model recruits the conceptual basis of the independent contrasts approach (Felsenstein, 1985), by modeling the evolution of continuous characters (traits) via a multivariate Brownian diffusion process, while coupling it with a Brownian molecular clock for d_S and d_N/d_S (rates). The coupling implied by the joint

Brownian process for traits and rates enables gene alignments to be incorporated into the study, along with life-history traits and data about heterozygosity in extant species. It provides the opportunity to estimate their joint correlation patterns, while automatically accounting for phylogenetic inertia. Fitting the model to the data produces as an output a dated tree, an estimate of the covariance matrix between traits, as well as a reconstruction of traits along the branches of the tree (see Figure 7).

For the need of the present project, we implement FastCoevol, a faster version of the Coevol model. This version entails an additional level of branch-specific gamma deviates, meant to account for short-term deviations in d_S and d_N/d_S from the long-term trends implied by the Brownian process. This two-level model is a direct generalization of the mixed relaxed clock model introduced in Lartillot *et al.* (2016) in the context of a codon model.

In its original version, Coevol is computationally intensive, precluding applications on empirical data sets with more than a few tens of genes. To overcome this limitation, FastCoevol is based on substitution mapping approximation (Romiguier *et al.*, 2012; Lemey *et al.*, 2012). A detailed presentation of the approach is given in the Supplementary Methods (section Bayesian methodology). Specifically, the detailed substitution history for all genes is inferred under a simpler reference model (assuming a constant gene-level d_N/d_S and shared synonymous branch lengths across genes). This substitution history is then collapsed into posterior mean numbers of synonymous and non-synonymous substitutions, and posterior mean numbers of synonymous and non-synonymous mutational opportunities, and this, separately for each branch of the tree. Those numbers are then used as an approximate sufficient statistic for the molecular component of the multivariate Brownian model of Coevol. The use of this computation reduces the computational to 1-2 days for the substitution mapping and less than an hour for the fitting of the Brownian model on the data over the 6002 genes alignments. For computational reasons, we perform the substitution mapping under the reference model on lists of 1000 genes. We then summed statistics across all gene sets, on a per-branch basis, and used this as an input for the FastCoevol model. With FastCoevol, we produced five independent MCMC runs per datasets, for a total of 30000 points. We used a burn-in of 10000 points and computed the posterior mean estimates for the parameters of interest using 1 every 5 points after the burn-in. The values of the five resulting estimates of the covariance matrix were averaged, even though they are very similar to each other. We considered a positive or negative correlation as significant when the posterior probability of the correlation is respectively higher than 0.975 or lower than 0.025, as in Lartillot and Poujol (2011).

4.7.2 Sequential approach: PGLS

As an alternative to the integrative approach of FastCoevol, we used a sequential approach based on the standard PGLS method (Pagel, 1997). Specifically, the mapping statistics just described were summed over all genes and then used to compute an empirical d_N/d_S for each branch. Then, a linear regression analysis with correction for phylogenetic inertia was conducted using a phylogenetic covariance matrix derived from the evolutionary relationships among the species such as inferred by IQTree. For the PGLS analysis, we consider a correlation significant when the p-value is lower than 0.05.

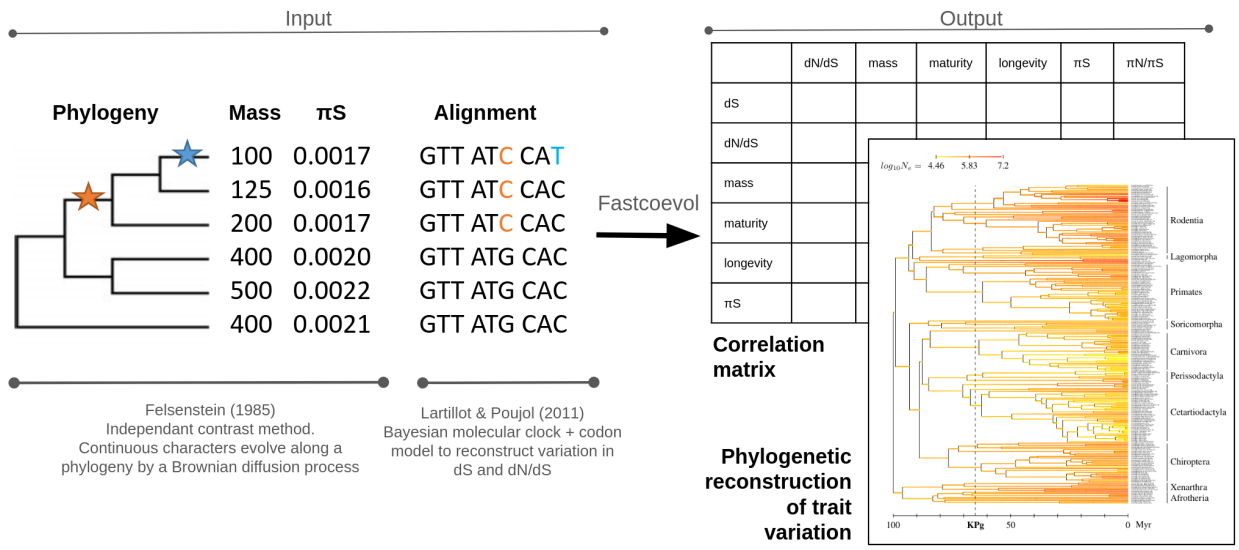

Figure 7: **FastCoevol pipeline**. The pipeline takes as an input a rooted phylogeny, a table containing the continuous traits of interest (here mass and π_S) and a multi-gene alignment. The output is a correlation matrix, with correlation coefficients and posterior probabilities for each pair of traits, including d_S and d_N/d_S, and a phylogenetic reconstruction of the variation of each trait along the phylogeny.

Acknowledgements

Damien de Vienne (phylter), Laurent Duret (discussions), Julien Joseph (vcf annotation script and discussion), Thibault Latrille (discussion). **Funding:** Agence Nationale de la Recherche (ANR-20-CE02-0008-01 "NeGA")

David Enard is funded by NIH MIGMS grant 5R35GM142677

Author contributions: Original idea: N.L M.B; Model conception: N.L; Code: M.B, N.L; Data analyses: M.B, D.E ; Interpretation: M.B, N.L; First draft: M.B; Editing and revisions: M.B, N.L, D.E; Project management and funding: NL.

Competing interests: The authors declare no conflicts of interest.

Data and materials availability: For scripts : https://github.com/MeloBastian/Mam$_N eFordata$: $https$: $//doi.org/10.5281/zenodo.15283024$

References

Boussau, B. and Scornavacca, C. 2020. Reconciling gene trees with species trees. *Phylogenetics in the genomic era*, pages 3–2.

Brevet, M. and Lartillot, N. 2021. Reconstructing the history of variation in effective population size along phylogenies. *Genome Biology and Evolution*, 13(8): evab150.

Castellano, D., Macià, M. C., Tataru, P., Bataillon, T., and Munch, K. 2019. Comparison of the full distribution of fitness effects of new amino acid mutations across great apes. *Genetics*, 213(3): 953–966.

Charlesworth, B. 2009. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10(3): 195–205.

Comte, A., Tricou, T., Tannier, E., Joseph, J., Siberchicot, A., Penel, S., Allio, R., Delsuc, F., Dray, S., and de Vienne, D. M. 2023. Phylter: efficient identification of outlier sequences in phylogenomic datasets. *bioRxiv*, pages 2023–02.

Criscuolo, A. and Gribaldo, S. 2010. Bmge (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC evolutionary biology*, 10: 1–21.

Damuth, J. 1981. Population density and body size in mammals. *Nature*, 290: 699–700.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., *et al.* 2011. The variant call format and vcftools. *Bioinformatics*, 27(15): 2156–2158.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., *et al.* 2021. Twelve years of samtools and bcftools. *Gigascience*, 10(2): giab008.

De Magalhaes, J., Costa, and J 2009. A database of vertebrate longevity records and their relation to other life-history traits. *Journal of evolutionary biology*, 22(8): 1770–1774.

Di Franco, A., Poujol, R., Baurain, D., and Philippe, H. 2019. Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evolutionary Biology*, 19: 1–17.

Duret, L. and Galtier, N. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annual review of genomics and human genetics*, 10: 285–311.

Eyre-Walker, A. and Keightley, P. D. 2007. The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, 8(8): 610–618.

Eyre-Walker, A., Keightley, P. D., Smith, N. G., and Gaffney, D. 2002. Quantifying the slightly deleterious mutation model of molecular evolution. *Molecular Biology and Evolution*, 19(12): 2142–2149.

Eyre-Walker, A., Woolfit, M., and Phelps, T. 2006. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics*, 173(2): 891–900.

Felsenstein, J. 1985. Phylogenies and the comparative method. *The American Naturalist*, 125(1): 1–15.

Feng, S., Stiller, J., Deng, Y., Armstrong, J., Fang, Q., Reeve, A. H., Xie, D., Chen, G., Guo, C., Faircloth, B. C., *et al.* 2020. Dense sampling of bird diversity increases power of comparative genomics. *Nature*, 587(7833): 252–257.

Figuet, E., Nabholz, B., Bonneau, M., Mas Carrio, E., Nadachowska-Brzyska, K., Ellegren, H., and Galtier, N. 2016. Life history traits, protein evolution, and the nearly neutral theory in amniotes. *Molecular biology and evolution*, 33(6): 1517–1527.

Figuet, E., Ballenghein, M., Lartillot, N., and Galtier, N. 2017. Reconstruction of body mass evolution in the Cetartiodactyla and mammals using phylogenomic data. *PCI Evol. Biol.*

Goldstein, R. A. 2013. Population size dependence of fitness effect distribution and substitution rate probed by biophysical model of protein thermostability. *Genome Biol. Evol.*, 5(9): 1584–1593.

James, J., Castellano, D., and Eyre-Walker, A. 2017. Dna sequence diversity and the efficiency of natural selection in animal mitochondrial dna. *Heredity*, 118(1): 88–95.

Lanfear, R., Kokko, H., and Eyre-Walker, A. 2014. Population size and the rate of evolution. *Trends in ecology & evolution*, 29(1): 33–41.

Lartillot, N. and Poujol, R. 2011. A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. *Molecular biology and evolution*, 28(1): 729–744.

Lartillot, N., Phillips, M. J., and Ronquist, F. 2016. A mixed relaxed clock model. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 371(1699).

Latrille, T. and Lartillot, N. 2021. Quantifying the impact of changes in effective population size and expression level on the rate of coding sequence evolution. *Theor Popul Biol*, 142: 57–66.

Latrille, T., Lanore, V., and Lartillot, N. 2021. Inferring Long-Term Effective Population Size with Mutation-Selection Models. *Mol. Biol. Evol.*, 38(10): 4573–4587.

Lemey, P., Minin, V. N., Bielejec, F., Kosakovsky Pond, S. L., and Suchard, M. A. 2012. A counting renaissance: Combining stochastic mapping and empirical Bayes to quickly detect amino acid sites under positive selection. *Bioinformatics*.

Leroy, T. and Nabholz, B. 2022. Response to kratochvíl and rovatsos. *Current Biology*, 32(1): R30–R31.

Li, H. and Durbin, R. 2011. Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357): 493–496.

Lourenço, J., Galtier, N., and Glémin, S. 2011. Complexity, pleiotropy, and the fitness effect of mutations. *Evolution*, 65(6): 1559–1571.

Löytynoja, A. 2014. Phylogeny-aware alignment with prank. *Multiple sequence alignment methods*, pages 155–170.

Lynch, M. and Conery, J. S. 2003. The origins of genome complexity. *science*, 302(5649): 1401–1404.

Lynch, M. and Gabriel, W. 1990. Mutation load and the survival of small populations. *Evolution*, 44(7): 1725–1737.

Lynch, M., Ali, F., Lin, T., Wang, Y., Ni, J., and Long, H. 2023. The divergence of mutation rates and spectra across the tree of life. *EMBO reports*, 24(10): e57561.

McDonald, J. H. and Kreitman, M. 1991. Adaptive protein evolution at the adh locus in drosophila. *Nature*, 351(6328): 652–654.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., and Lanfear, R. 2020. Iq-tree 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular biology and evolution*, 37(5): 1530–1534.

Müller, R., Kaj, I., and Mugal, C. F. 2022. A nearly neutral model of molecular signatures of natural selection after change in population size. *Genome Biology and Evolution*, 14(5): evac058.

Nabholz, B., Uwimana, N., and 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 biology and evolution*, 5(7): 1273–1290.

Nadachowska-Brzyska, K., Konczal, M., and Babik, W. 2022. Navigating the temporal continuum of effective population size. *Methods in Ecology and Evolution*, 13(1): 22–41.

Nei, M., Suzuki, Y., and Nozawa, M. 2010. The neutral theory of molecular evolution in the genomic era. *Annual review of genomics and human genetics*, 11(1): 265–289.

Ohta, T. 1973a. Slightly deleterious mutant substitutions in evolution. *Nature*, 246(5428): 96–98.

Ohta, T. 1973b. Slightly deleterious mutant substitutions in evolution. *Nature*, 246(5428): 96–98.

Ohta, T. 1992. The nearly neutral theory of molecular evolution. *Annual review of ecology and systematics*, 23(1): 263–286.

Ohta, T. 1995. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *Journal of molecular evolution*, 40: 56–63.

Ohta, T. and Kimura, M. 1971. On the constancy of the evolutionary rate of cistrons. *Journal of molecular evolution*, 1(1): 18–25.

Pagel, M. 1997. Inferring evolutionary processes from phylogenies. *Zoologica Scripta*, 26(4): 331–348.

Popadin, K., Polishchuk, L. V., Mamirova, L., Knorre, D., and Gunbin, K. 2007. Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals. *Proceedings of the National Academy of Sciences*, 104(33): 13390–13395.

Ranwez, V. and Chantret, N. N. 2020. Strengths and limits of multiple sequence alignment and filtering methods.

Romiguier, J., Figuet, E., Galtier, N., Douzery, E. J. P., Boussau, B., Dutheil, J. Y., and Ranwez, V. 2012. Fast and robust characterization of time-heterogeneous sequence evolutionary processes using substitution mapping. *PLoS One*, 7(3): e33852.

Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari, Y., Dernat, R., Duret, L., Faivre, N., *et al.* 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, 515(7526): 261–263.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. 2015. Busco: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19): 3210–3212.

Simion, P., Delsuc, F., and Philippe, H. 2020. To what extent current limits of phylogenomics can be overcome?

Vasimuddin, M., Misra, S., Li, H., and Aluru, S. 2019. Efficient architecture-aware acceleration of bwa-mem for multicore systems. In *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pages 314–324.

Waples, R. S. 2022. What is n e, anyway? *Journal of Heredity*, 113(4): 371–379.

Welch, J. J., Eyre-Walker, A., and Waxman, D. 2008. Divergence and polymorphism under the nearly neutral theory of molecular evolution. *Journal of molecular evolution*, 67: 418–426.

White, E. P., Ernest, S. K. M., Kerkhoff, A. J., and Enquist, B. J. 2007. Relationships between body size and abundance in ecology. *Trends Evol Evol*, 22: 323–330.

Wright, S., Teissier, G., *et al.* 1939. Statistical genetics in relation to evolution. *(No Title)*.

Zoonomia 2020. A comparative genomics multitool for scientific discovery and conservation. *Nature*, 587(7833): 240–245.