Empirical validation of the nearly neutral theory at divergence and population genomic scale across 150 placental mammal 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 : melodie.bastian@yahoo.fr ; nicolas.lartillot@univ-lyon1.fr

April 25, 2025

Abstract

By limiting the efficacy of selection on mutations of small effects, random drift is expected to play a major role in the evolution of genomes. Formalizing this idea, the nearly-neutral theory predicts that the proportion of non-synonymous mutations that are effectively neutral depends negatively on the effective population size (N_e). As a result, 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. Some of these predictions have previously been tested in mammals. However, most studies have either focused on d_N/d_S or on π_N/π_S and have not addressed the problem globally across evolutionary scales. Here we propose to test the nearly-neutral prediction in an integrative manner using the genomes of 150 mammals species and around 6000 orthologous genes. Our investigation spans the macro and the micro-evolutionary scale, using for the latter a measure of heterozygosity in the annotated orthologous coding sequences of the assembled diploid genomes. We confirm the positive correlation between d_N/d_S and life history traits. We also observe, for the first time in mammalian nuclear genomes, a relationship between π_N/π_S and π_S. Across time scales, the correlation patterns between micro and macro-evolutionnary quantities are in agreement with the nearly-neutral predictions, although the correlation between π_S and life-history traits is weaker and less robust than all other correlations. Together, these observed correlations are globally consistent with the nearly-neutral expectations and can be explained by invoking that all variables are in fact correlated with a single hidden variable: N_e.

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

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 confronted. Seeing macroevolution as the long-term result of micro-evolutionary processes is generally recognized and is sometimes even explicitly used, for instance to derive phylogenetic models directly from first principles of population genetics (Yang and Nielsen, 2008; Rodrigue *et al.*, 2010). On the other hand, direct confrontation, or integration, of the empirical data available both within and between species, over a broad set of species and within a single statistical framework, is still quite uncommon. Yet, a more decisive integration of comparative analyses both within and between species could benefit from their complementary strengths in revealing important aspects of evolutionary processes and their long-term impact on genomes.

One particularly important point on which this confrontation could bring important insights is the question of the role of effective population size (N_e), and its variation between species, in genome evolution. The role of N_e in genome evolution has been the subject of a long-lasting debate. 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 (Kimura, 1968; Ohta, 1973a), leading to the nearly-neutral theory. According to this theory, the majority of mutations are deleterious, although 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 $1/N_e$ are typically determined by genetic drift and may thus segregate and reach fixation, while the mutations with larger effects are efficiently purified by natural selection.

In the case of protein coding genes, mutations can be distinguished as either synonymous or non-synonymous. Synonymous mutations are often considered as neutral (or nearly-neutral) in 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, because they change the encoded amino acid (Eyre-Walker *et al.*, 2006; Eyre-Walker and Keightley, 2007). Using this contrast between the non-synonymous and the synonymous compartments leads to two key molecular quantities that can be measured or inferred based on coding sequences: 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 represent key measures of the efficacy of selection against random drift (or for short, in the following, *selection efficacy*). According to the nearly neutral theory, they are expected to decrease (i.e. more efficient selection against random drift) as a function of N_e at their respective scale.

Testing these predictions, however, raises the question of how to estimate N_e and more fundamentally, how to define it, depending on the time scale considered. There are different ways to define N_e (recently reviewed on Waples (2022)). Some will define it from a demographic point of view, when others invoke the effective population size as a measure of genetic drift. 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 our work, we will consider two types of N_e proxies, each defined at a different evolutionary scale. At micro-evolutionary scale, we use π_S as proxy of N_e based on the formula from coalescent theory : $\pi_S = 4N_e\mu$ 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, although it appears that most of the variation in π_S in contributed by the N_e component (Lynch *et al.*, 2023). At the macro-evolutionary scale, the molecular evolutionary theory doesn't provide a direct proxy of N_e. Furthermore, the exact meaning of a long-term N_e, which would stand for a measure of random drift at the macro-evolutionary scale, is not so well formalized. Nevertheless, it is generally assumed that large long-lived mammals present lower population sizes than small-bodied species (Popadin *et al.*, 2007; Figuet *et al.*, 2016; Nabholz *et al.*, 2013). Accordingly, and following previous work, we opt for the use of life history traits as a proxy of long-term N_e, acknowledging that its relation with N_e is somewhat indirect and less well theoretically formalized compared to its short term equivalent, π_S.

These two proxies of N_e now defined, we want to interrogate the nearly-neutral predictions at the macro- and micro-evolutionary scales, by empirically examining the relations between π_S and π_N/π_S, which is expected to be negative, on the one hand, and the relation between life-history traits and d_N/d_S, which is expected to be positive, on the other hand. Moreover, considering that macroevolution is the result of the accumulation of microevolutionary processes over longer timescale, we want to assess the congruence between these two scales, namely that the patterns of π_N/π_S and d_N/d_S are congruent with each other and match the nearly-neutral expectations. As an important aside, confronting the two time scales would also offer the opportunity to verify that life-history traits are indeed correlated with N_e.

A number of previous study have already attempted to test those predictions. At the macro-evolutionary scale, a positive correlation between life history traits and d_N/d_S as repeatedly been observed (Nabholz *et al.*, 2013; Popadin *et al.*, 2007; Figuet *et al.*, 2016; Romiguier *et al.*, 2014) which is often interpreted as a confirmation of the indirect correlation of long-term N_e with life-history traits. Actually, the literature on this subject is somewhat contrasted, with both positive and negative results, depending on the exact traits and on the taxonomic scale (Lartillot and Delsuc, 2012; Nabholz *et al.*, 2013; Figuet *et al.*, 2016; Brevet and Lartillot, 2021). At the micro-evolutionary scale, some studies have reported an empirical relation between π_N/π_S and π_S (James *et al.*, 2017; Brevet and Lartillot, 2021). Across time-scales, only the relation between micro-evolutionary process metrics and life-history traits in birds and across metazoan has been reported (Figuet *et al.*, 2016; Romiguier *et al.*, 2014). Equivalent results have never been reported in the case of mammals. This incomplete picture of the relation between the different micro and macro-evolutionary metrics across mammals calls for further investigation.

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) (Pagel, 1997). 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). These two methods have their strengths and weaknesses. However, they have rarely been confronted on the same data.

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 macro-evolutionary patterns 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 Materials and Methods

Figure 1 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.

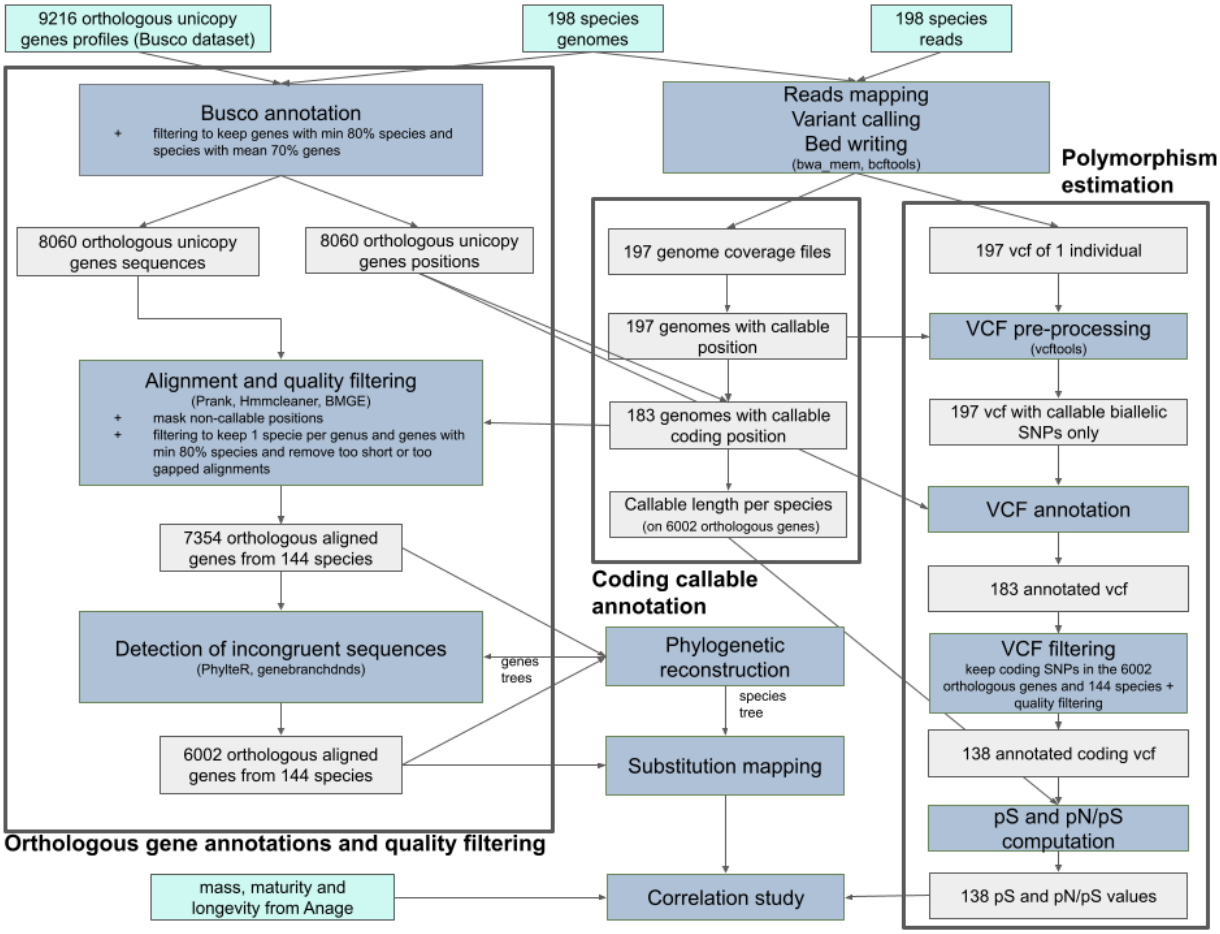

Figure 1: **Simplified summary of the pipeline from whole genomes acquisition and vcf writing to the final correlation analysis**. Light blue boxes correspond to the input data, grey boxes to the intermediate data and blue boxes to different steps of the pipeline. A more complete and more detailed desciption of the pipeline is available at https://github.com/MeloBastian/Mam_Ne

2.1 Data acquisition

Four types of input data were used (depicted in light blue in Figure 1): 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 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). Variant calling and the generation of whole-genome sequencing depth files were done with samtools bcftools v1.13, using default parameters (Danecek *et al.*, 2021) (https://samtools.github.io/bcftools/howtos/variant-calling.html).

2.2 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.

2.2.1 Identification of orthologs

Not all the 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 in the studied clade. After this step, we are left with 183 mammalian species and 8060 BUSCO one-to-one orthologous complete genes.

2.2.2 Alignment and quality filtering

The sequences of each orthologous gene were aligned using PRANK (Löytynoja, 2014) and then filtered using HMMcleaner (Di Franco *et al.*, 2019) and BMGE (Criscuolo and Gribaldo, 2010) with default parameters. These two filtering methods are complementary: BMGE filter alignments by columns, while HMMCleaner filters small segments of sequence on a per species basis (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.

2.2.3 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 Methods) 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.

2.2.4 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 (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.

2.3 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 (See results, tree in Figure 3), 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.

2.4 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 S3 and S4). For gene trees, required for Phylter (Comte *et al.*, 2023), we use Iqtree2 too with default parameters and model search options.

2.5 VCF annotation and polymorphism

Variant calling was performed using default parameters of samtools bcftools v1.13 (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 home-made 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*, *Przewalskium albirostris*, *Mastomys coucha*, *Litocranius walleri* ,*Cheirogaleus medius* and *Cephalophus harveyi*) (see Supplementary Material, section VCF filtering and Figure S1).

After all these filters, six species ended up with less than 1000 SNPs (*Muscardinus avellanarius*, *Beatragus hunteri*, *Sigmodon hispidus*, *Diceros bicornis*, *Sousa chinensis* and *Alouatta palliata*). These species also have a low genome quality based on the N50 and L50 metrics. Below, we will show correlation analyses with and without these 6 species to evaluate their impact on the analysis.

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. 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. To address concerns regarding the potential circularity in the correlation between π_N/π_S and π_S (due to the double use of π_S), we repeated the analysis using two non-overlapping sets of genes to compute π_S and π_N/π_S and found no substantial difference in the results (see Supplementary Material section Decoupled estimation of and π_N/π_S, and Figure S7).

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.

2.6 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. 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.

2.7 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.

2.7.1 Bayesian integrative approach: FastCoevol

We used a fast implementation of the Coevol model (Lartillot and Poujol, 2011), implemented for the need of the present project. Coevol 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 2).

Compared to the original Coevol model, the version used here, named FastCoevol, 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 time from one week, using the original Coevol version, 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).

2.7.2 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.

3 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. This

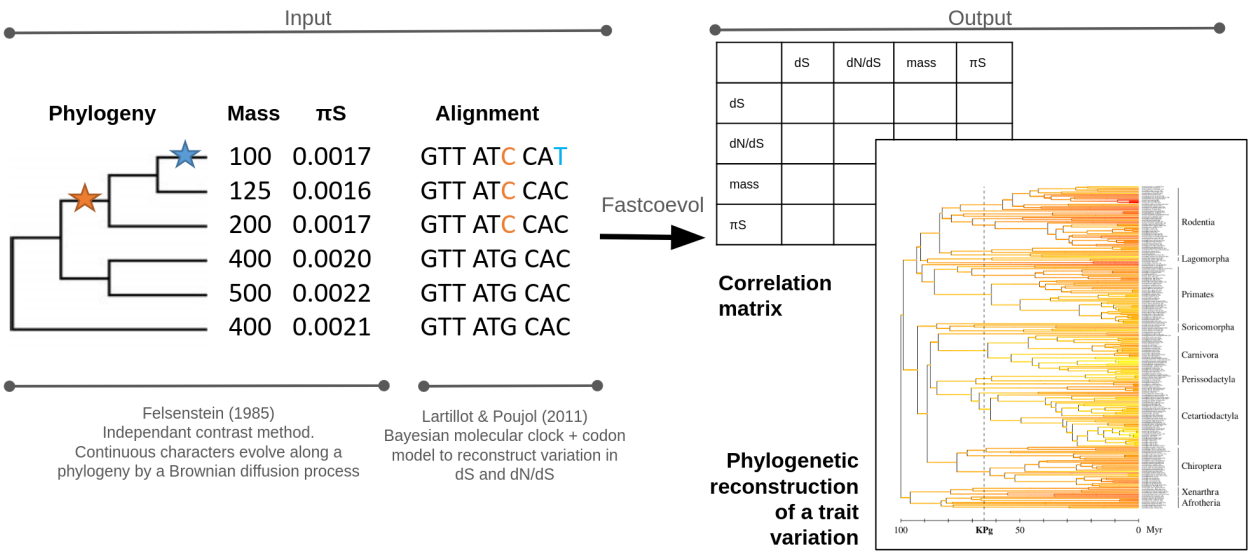

Figure 2: **FastCoevol pipeline**. The pipeline takes as an input a rooted phylogeny, a table containing the continuous traits of interest (here mass and π_S) and a multi-gene alignment. The output is a correlation matrix, with correlation coefficients and posterior probabilities for each pair of traits, including d_S and d_N/d_S, and a phylogenetic reconstruction of the variation of each trait along the phylogeny.

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. In the end, this yields a set of synonymous and non-synonymous SNPs from which a π_S and π_N/π_S were inferred.

Starting from this reference dataset, we considered two additional species subsampling schemes. First, 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. Second, we conducted the analysis without the polymorphism data for the six species for which less than 1000 SNPs were called (*Muscardinus avellanarius*, *Beatragus hunteri*, *Sigmodon hispidus*, *Diceros bicornis*, *Sousa chinensis* and *Alouatta palliata*). This leads to four version of the datasets, which were then systematically analysed using both standard (PGLS) and integrative (FastCoevol) approaches.

3.1 Visualisation of d_N/d_S variation along the tree

The history of the changes in d_N/d_S over the phylogeny, such as infered by FastCoevol, is shown in Figure 3 (the tree with the 89 species datated is presented in Suplementary Material, Figure S2). The d_N/d_S ranges from 0.15 to 0.34 over the tips. Taxonomic groups with the lowest d_N/d_S are Rodentia (i.e. *Peromyscus maniculatus*, $d_N/d_S = 0.148$), Lagomorpha (i.e. *Ochotona princeps*, $d_N/d_S = 0.151$), and Caniforms in the Carnivora groups (i.e. *Vulpes vulpes*, $d_N/d_S = 0.157$). At the other end of the range, taxonomic groups with the highest d_N/d_S correspond to Cetacea (i.e. *Monodon mococeros* $d_N/d_S = 0.296$) and Primates (i.e. *Homo sapiens* $d_N/d_S = 0.258$). 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. Of note, some isolated lineages show more unexpected patterns, such as *Mustela putorius*, a Caniform which presents the highest d_N/d_S (0.341) while some other species within the same group exhibit the lowest d_N/d_S. At this stage, we have no explanation for this particular observation.

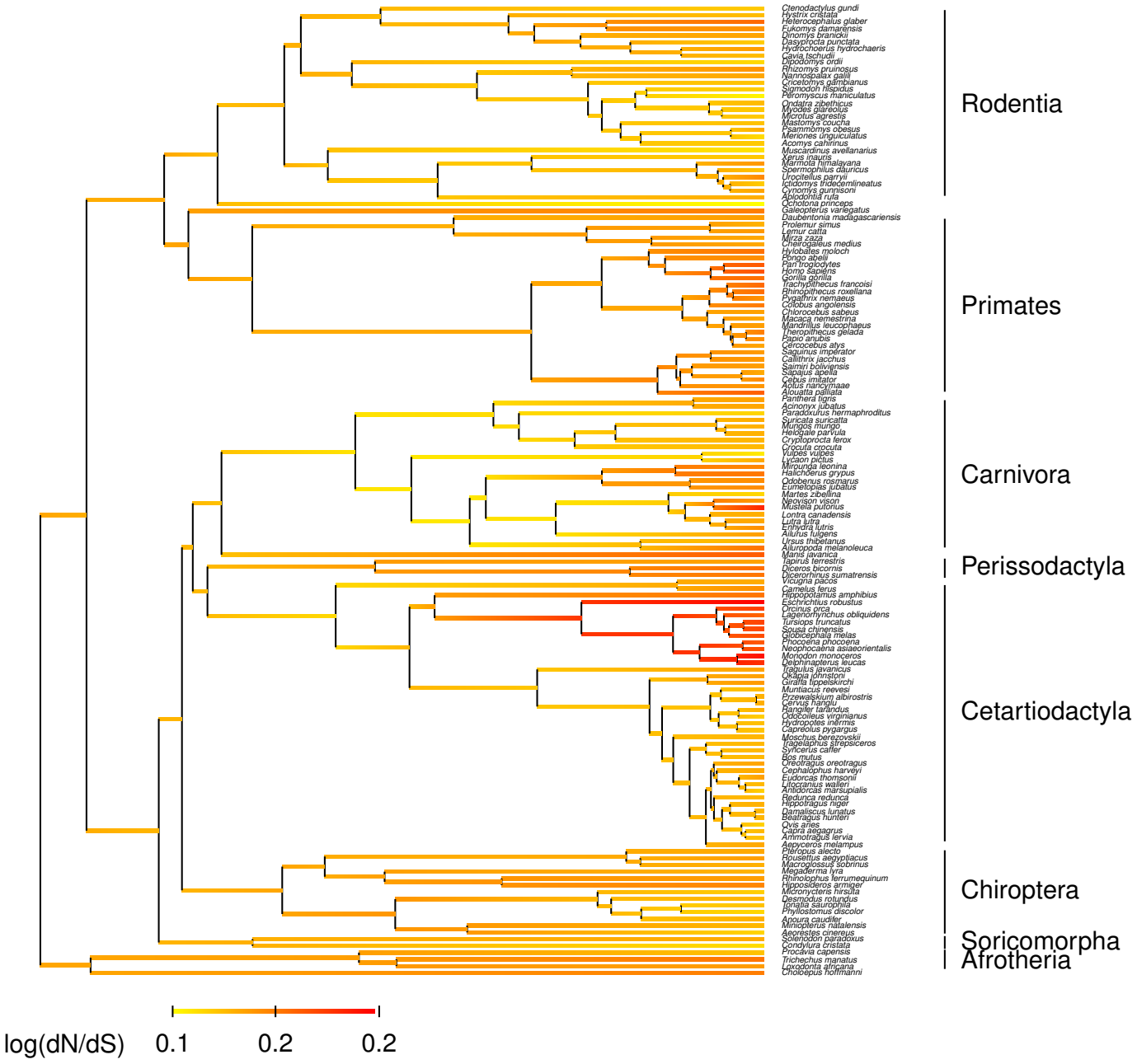

Figure 3: Reconstruction of $\log(d_N/d_S)$ along the phylogeny of placental mammals (144 species) by FastCoevol

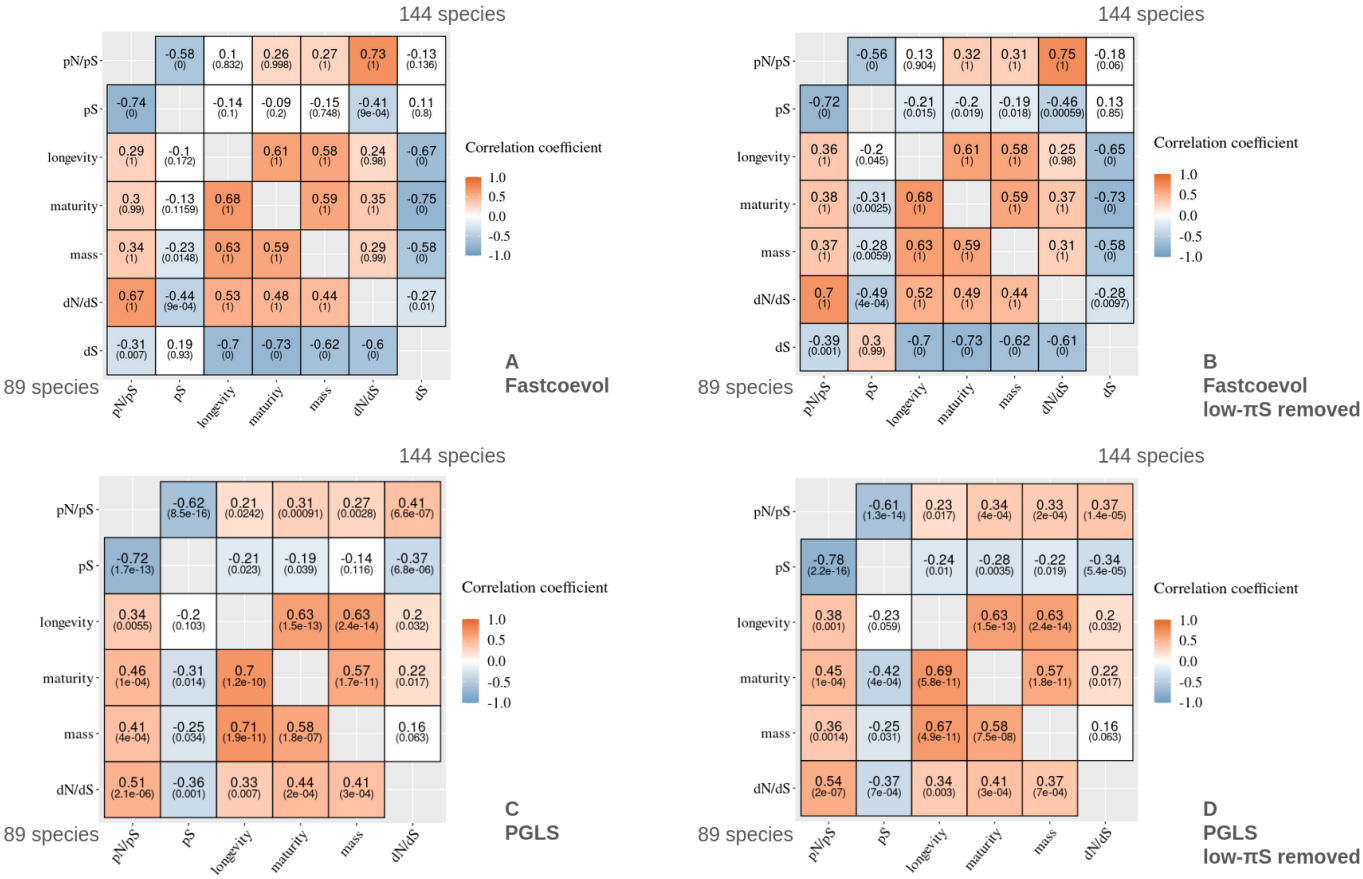

Figure 4: **Correlation coefficients from the FastCoevol and PGLS analyse**. **A**: FastCoevol analysis with the full dataset. **B**: FastCoevol analysis with the dataset excluding polymorphism data for the six low π_S species. **C**: PGLS analysis with the full dataset. **D**: PGLS analysis with the dataset excluding polymorphism data for the six low π_S species. **Upper diagonal**: the 144-species dataset. **Bottom diagonal**: the 89-species dataset excluding short branches. Colours correspond to the magnitude of the correlation coefficient (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.

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

At the macro-scale, life history traits are known to be strongly correlated with each other. This well known result is recovered in our analysis using both PGLS and FastCoevol (on the 144 and 89 species tree) (Figure 4 A-C). Moreover, a highly significant negative correlation is observed between d_S and life history traits (Figure 4 A). This negative relation is congruent with what is already known and generally interpreted as a generation time effect (Thomas *et al.*, 2010).

On the 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 $r = 0.24$ to 0.35 depending on the trait

(Figure 4 A). 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 4 C). 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. Interestingly, using the 89 species tree, we observe stronger correlations under both methods. With the 89 species, correlations between d_N/d_S and mass also became significant in the PGLS analysis.

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 This effect is also presents comparing the 89 versus 144 species correlations matrix from FastCoevol, which suggests that the short-term over-dispersion of d_N/d_S along branches is not totally captured by the method.

Altogether, both PGLS and FastCoevol analyses agree on a significant positive correlation between proxies of long-term N_e and selection efficacy 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 over the tree.

3.3 Micro-evolutionary scale : π_S versus π_N/π_S

At the micro-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 then predicts a negative correlation between these two measures. In the phylogenetic correlation analysis, we indeed observe a negative correlation using both methods (Figure 4 A-C). Again, we note stronger correlations in the 89 species analysis than in the one with 144 species ($r = -0.74$ versus $r = -0.58$ with FastCoevol, $r = -0.72$ versus $r = -0.62$ with PGLS).

However, plotting π_N/π_S against π_S (Figure 5) reveals some species showing a departure from the general pattern. Those species harbour a low π_S value given their π_N/π_S. This could be the sign of recent inbreeding, although these species don't seem to be outliers in their runs of homozygosity (see Supplementary Material). Low π_S could also reflect recent demographic bottlenecks, which would more strongly impact π_S than π_N/π_S. More probably, however, incorrect normalisation of the synonymous and non-synonymous counts, and more specifically systematic overestimation of the number of callable positions when coverage is low, could lead to a downward bias in the π_S (and π_N) estimates, without impacting π_N/π_S. In Figure 5, we coloured the points using the species genome depth (see Material and Methods) and we marked with squares (instead of dots) the six species already identified for their smaller number of high-quality coding SNPs. We note that these flagged species correspond to the outliers in the graph and present a lower sequencing depth than the other species in the core of the general pattern. Altogether, this suggests more a data quality issue for these outlying points rather than a low real π_S. Accordingly, in the following, we systematically analyse the correlation with (Figure 4 A-C) and without (Figure 4 B-D) polymorphism data for these six species with small SNP counts.

In the end, our analysis shows a robust correlation between selection efficacy and π_S as a proxy of N_e at the micro-evolutionary scale. The correlation at this scale appears stronger than at the macro scale. One possible reason for this is that π_S is a more direct proxy of N_e in comparison to life history traits.

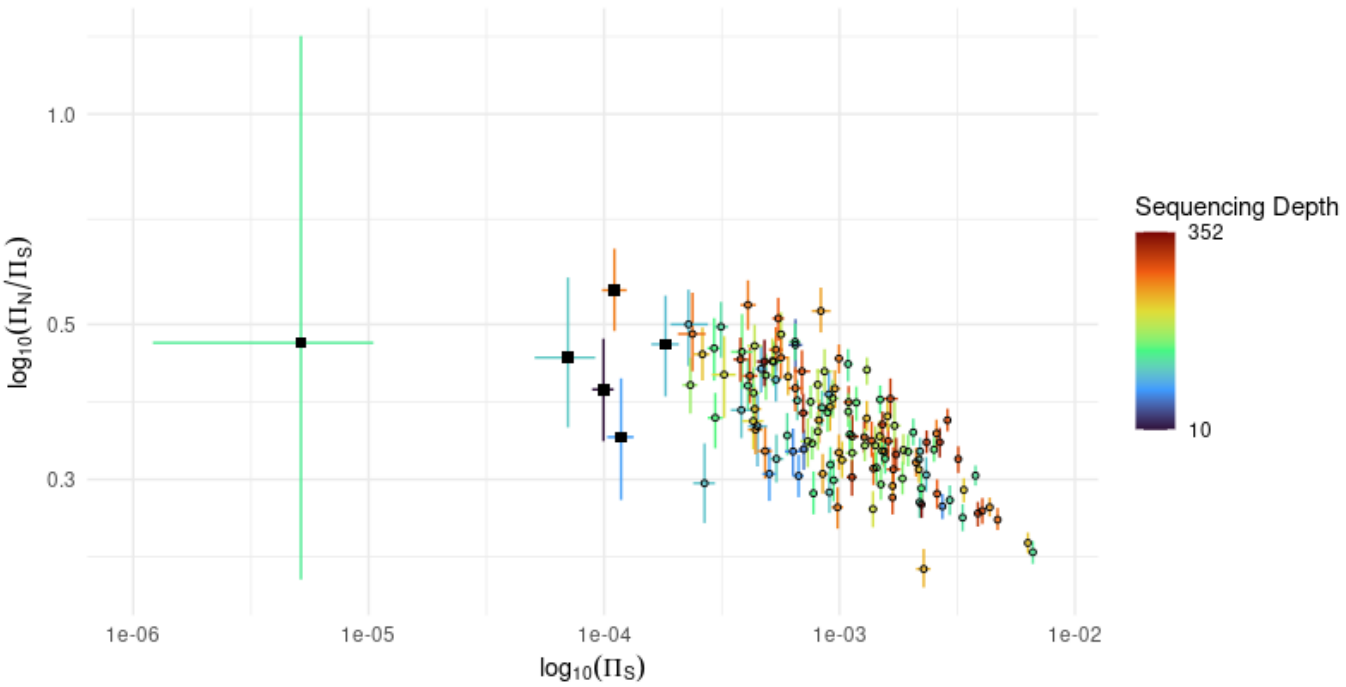

Figure 5: Plot of π_N/π_S againstπ_S (in log-log scale) for all species of the analysis (bars: 95% CI computed by non-parametric bootstrap). Colours correspond to genome sequencing depth. The square points correspond to the six species with less than 1000 coding SNPs after filtering.

3.4 Comparing 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 (such as presumably 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 mass in the 89 species analysis (Figure 4 A-B). The, PGLS analysis also shows weak correlations and the statistical support vary depending on the life history traits and the taxon sampling (Figure 4 C-D).

On the other hand, π_N/π_S shows a robust negative correlation with life history traits, under the four data settings and with the two methods. The most straightforward explanation for this correlation is a nearly-neutral effect. Here also, as above for d_N/d_S versus life history traits, the correlations are stronger and more consistent when the dataset is thinned to remove closely-related species. Again, this could be due

to the negative impact of non-phylogenetic variation. In the present case, short-term fluctuation in N_e may represent an important contribution (see discussion).

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 both negatively with π_S and positively with π_N/π_S. 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.

4 Conclusion and Discussion

The nearly neutral theory states that, the higher the effective population size is, the more the selection is able to purge mildly deleterious mutations (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 6. 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.

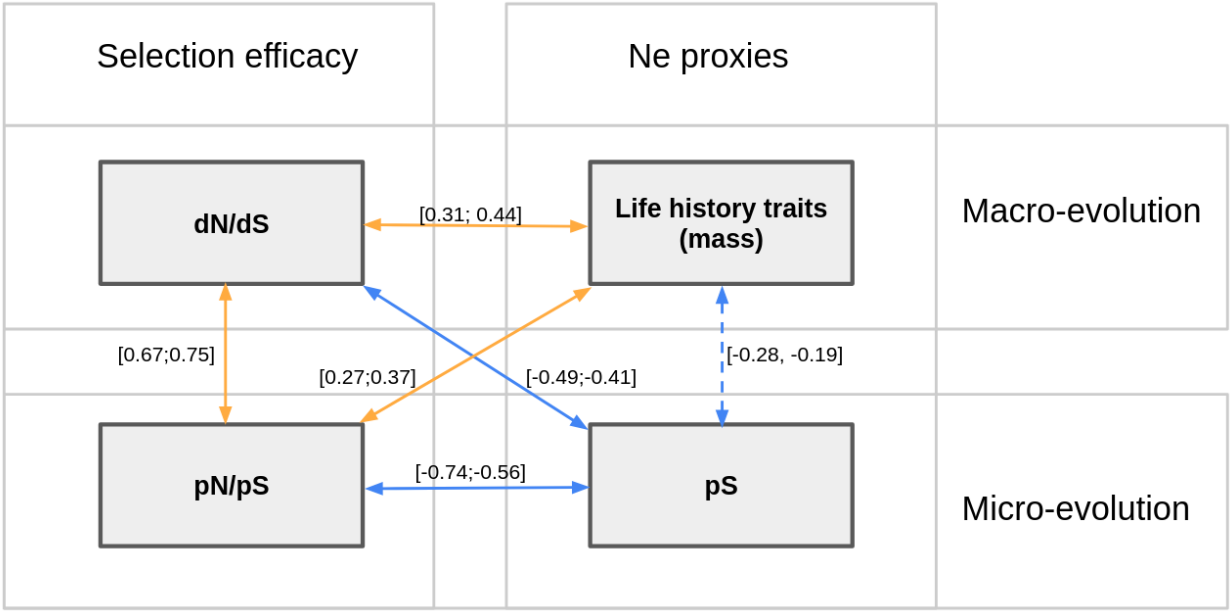

Figure 6: **Schematic summary of the correlations observed in our study.** Negative correlations are in blue, positive correlations in orange. Values within brackets correspond to the range of correlation coefficients obtained under the four different analysis using FastCoevol. Mass is used here to represent life history traits.

As presented in the introduction, several prior studies have already examined some of these correlations. In Table 1, we present a sample 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. Nevertheless, our results, also reported in this table, highlights correlations between micro and macro-evolutionary measures that have not been studied until now at mammals scale.

	dN/dS	LHT			pS	pN/pS
		mass	maturity	longevity		
dN/dS		Figuet *et al.* (2016) r=0.38*	r=0.71**	r=0.69**		
		Nabholz *et al.* (2013) r=0.47**	r=0.38**	r=0.51**		
		Lartillot and Poujol (2011) r=0.58**	r=NS	R=0.13**		
		Lartillot (2012) r=NS	r=0.27**	r=0.27**		
		Lartillot and Delsuc (2012) r=NS	r=0.36**	r=0.27*		
		Popadin *et al.* (2007) r=0.4**				
		r=[0.29;0.44**] 2/8 NS**	**r=[0.22**;0.49**]**	**r=[0.2**;0.53**]**	**r=[-0.49**;-0.37**]**	**r=[0.37**;0.75**]**
mass	Brevet and Lartillot (2021) primates r=NS					
	Lartillot and Poujol (2011) carnivores r=0.9**					
	Figuet *et al.* (2016) birds and reptiles r=NS				**r=[-0.28**;-0.19**] 2/8 NS**	**r=[0.27**;0.41**]**
	Nabholz *et al.* (2013) birds r=NS					
maturity	Brevet and Lartillot (2021) primates r=NS					
	Lartillot and Poujol (2011) carnivores r=NS					
	Figuet *et al.* (2016) birds r=NS; reptiles r=0.52*				**r=[-0.42**;-0.19*] 2/8 NS**	**r=[0.26**;0.46**]**
	Nabholz *et al.* (2013) birds r=NS					
longevity	Brevet and Lartillot (2021) primates r=NS					
	Lartillot and Poujol (2011) carnivores r=NS					
	Figuet *et al.* (2016) birds r=NS ; reptiles r=0.68**				**r=[-0.24*;-0.21**] 5/8 NS**	**r=[0.21*;0.38**] 2/8 NS**
	Nabholz *et al.* (2013) birds r=0.63**					
pS	Brevet and Lartillot (2021) primates r=NS	Brevet and Lartillot (2021) primates r=NS				James *et al.* (2017) r=-0.98**
		Romiguier *et al.* (2014) metazoan r=-0.32*		r=-0.53*		Piganeau and Eyre-Walker (2009) r=-0.28*
		Figuet *et al.* (2016) birds r=-0.39*	r=-0.52**	r=-0.45*		**r=[-0.72**;-0.56**]**
pN/pS	Brevet and Lartillot (2021) primates r=NS	Brevet and Lartillot (2021) primates r=NS			Brevet and Lartillot (2021) primates r=-0.78**	
		Romiguier *et al.* (2014) metazoan r=0.28*		r=0.62*	Piganeau and Eyre-Walker (2009) metazoan r=-0.11*, birds r=NS	
		Figuet *et al.* (2016) birds r=0.46*	r=0.5**	r=0.45**		

Table 1: Summary of previously published results on mammals (upper diagonal) and other metazoans (lower diagonal)

Note.- "*" 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 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.

4.1 Robustness to data errors

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.

Two main sources of errors, susceptible to bias the correlation analyses, can be identified. First, 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. Second, and similarly, SNP calling errors will tend to inflate the apparent π_N/π_S. This effect will be stronger in species with a low 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.

Additionally, we cannot exclude the possibility of improper normalisation of π_S (and π_N), due to incorrect 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 mismatch between π_S and π_N/π_S in the low-π_S species of Figure 4, but also the fact that π_S has much weaker correlation than π_N/π_S with life history traits.

4.2 Impact of species subsampling

To examine the impact of closely related species in the correlation analysis, we conducted in parallel the correlation analysis using the full 144 species set and a reduced set of 89 species separated by at least 10 My. We observed that the reduced dataset gives globally stronger correlations. As explained above, this can be due to the non-phylogenetic signal present in the variables being analysed. Several factors could contribute to this non-phylogenetic signal, like measurement errors (already discussed above), intra-specific variation and short-term fluctuation, in particular in N_e.

Concerning intra-specific variation, the use of the heterozygosity of a single individual as a proxy for population-level diversity can be problematic if the individual used does not reflect its population diversity (due to population structure, inbreeding, etc). To assess the potential impact of inbreeding, we investigated the length of runs of homozygosity (ROH), but did not observe any strong signal (see Supplementary Material,

section ROH study and Figure S5 and S6). In any case, using a single individual will always contribute some additional variance, that remains to be quantified. Of note, here, since π_S and π_N/π_S are calculated on the same individual, their correlations should not be subject to this additional variance.

The increased correlation coefficients between π_S and life history traits or d_N/d_S, when we shift from the 144-species to the 89-species tree, can be an indication for a possible short-term fluctuation in N_e and thus a mismatch between short-term and long-term N_e which is certainly an interesting aspect to further investigate and quantify.

In the case of d_N/d_S, in addition to short-term fluctuation (which are detected and are inferred to be substantial by the over-dispersed model of FastCoevol), another important factor could be at play, namely segregating polymorphism. Segregating polymorphism implies that the apparent d_N/d_S contains a certain proportion of π_N/π_S in its estimation. This proportion is higher in the terminal branches of the 144-species tree because of their shorter length. Thus, supposing that the real d_N/d_S correlates more strongly with life-history traits (both being long-term quantities) than does π_N/π_S, then the π_N/π_S part present in the observed d_N/d_S, would reduce the correlation coefficient between d_N/d_S and life history traits for the 144-species tree. This hypothesis is supported by a stronger correlation between π_N/π_S and d_N/d_S using the 144 species tree.

4.3 Perspectives on micro versus macro-evolution

In spite of these potential weaknesses, in the end, our analysis confirms and consolidates the correlations among long-term quantities (d_N/d_S, life history traits), and among short-term quantities (π_N/π_S, π_S). More importantly, it provides a more decisive entry point on micro-macro connections. Nevertheless, our results still remain to be confirmed in particularly by implementing a more robust normalisation of π_S and by quantifying the contribution of the polymorphism to the d_N/d_S estimates.

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 this article, we make the assumption that π_S represents N_e, using the relation $\pi_S = 4N_\mathrm{e}\mu$. However, μ itself varies between species, and is known to correlate with life history traits (Lanfear *et al.*, 2014; Lynch *et al.*, 2023). Mutations rate became a crucial factor which is important to take into account in our N_e reconstruction. 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. In the integrative method, we can incorporate a decomposition of π_S in is 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. However, this would first require explicitly modelling the mismatch between short-term and long-term N_e in the context of this integrative approach. Such improvement in the model would also provide a more quantitative idea about the intensity of the short term N_e fluctuations.

Finally, here we focused on nearly neutral evolution, but our dataset could be used to investigate other aspects, as it presents the opportunity to confront two genomic compartiments 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).

Acknowledgements

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: Available soon.

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.

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.

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., 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.

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.

Kimura, M. 1968. Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genetics research*, 11(3): 247–270.

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. 2012. Interaction between selection and biased gene conversion in mammalian protein-coding sequence evolution revealed by a phylogenetic covariance analysis. *Molecular biology and evolution*, 30(2): 356–368.

Lartillot, N. and Delsuc, F. 2012. Joint reconstruction of divergence times and life-history evolution in placental mammals using a phylogenetic covariance model. *Evolution*, 66(6): 1773–1787.

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).

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.

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.

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

Piganeau, G. and Eyre-Walker, A. 2009. Evidence for variation in the effective population size of animal mitochondrial dna. *PloS one*, 4(2): e4396.

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.

Rodrigue, N., Philippe, H., and Lartillot, N. 2010. Mutation-selection models of coding sequence evolution with site-heterogeneous amino acid fitness profiles. *Proceedings of the National Academy of Sciences*, 107(10): 4629–4634.

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?

Thomas, J. A., Welch, J. J., Lanfear, R., and Bromham, L. 2010. A generation time effect on the rate of molecular evolution in invertebrates. *Molecular biology and evolution*, 27(5): 1173–1180.

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.

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

Yang, Z. and Nielsen, R. 2008. Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. *Molecular biology and evolution*, 25(3): 568–579.

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