How much does $N_e$ vary among species?

Nicolas Galtier ¹, Marjolaine Rousselle ¹,²

¹ ISEM, CNRS, Univ. Montpellier, IRD, EPHE, Montpellier, France
² Bioinformatics Research Centre, Aarhus University, DK Aarhus, Denmark

Corresponding author: Nicolas Galtier, nicolas.galtier@umontpellier.fr
Abstract

Genetic drift is an important evolutionary force of strength inversely proportional to \( N_e \), the effective population size. The impact of drift on genome diversity and evolution is known to vary among species, but quantifying this effect is a difficult task. Here we assess the magnitude of variation in drift power among species of animals via its effect on the mutation load – which implies to also infer the distribution of fitness effects of deleterious mutations (DFE). To this aim, we analyze the non-synonymous (amino-acid changing) and synonymous (amino-acid conservative) allele frequency spectra in a large sample of metazoan species, with a focus on the primates vs. fruit flies contrast. We show that a Gamma model of the DFE is not suitable due to strong differences in estimated shape parameters among taxa, while adding a class of lethal mutations essentially solves the problem. Using the Gamma + lethal model and assuming that the mean deleterious effects of non-synonymous mutations is shared among species, we estimate that the power of drift varies by a factor of at least 500 between large-\( N_e \) and small-\( N_e \) species of animals, i.e., an order of magnitude more than the among-species variation in genetic diversity. Our results are relevant to Lewontin's paradox while further questioning the meaning of the \( N_e \) parameter in population genomics.

Key words: genetic drift, population size, mutation load, distribution of fitness effect, site frequency spectrum
**Introduction**

Genetic drift, the fluctuation of allele frequencies due to the randomness of reproduction, is one of the major evolutionary forces. Drift affects the fixation probability of selected mutations and is therefore a source of genetic load (Ohta 1972, Lynch et al. 2011). Drift impacts patterns of genome variation and can mimic or hide traces of adaptation (Jensen et al. 2005; Klopinstein et al. 2005; Peischl et al. 2018). Quantifying drift and its variation is obviously an important goal. The strength of genetic drift can be directly assessed from time series data, i.e., by analysing the dynamics of allele frequency across a controlled number of generations (Jónás et al. 2016, Néné et al. 2018). This is convenient for experimentally evolving populations, but more tricky in natural conditions, where populations are less easily defined and the effects of immigration difficult to control for (Ryman et al. 2019). For these reasons, the strength of genetic drift is often approached at species level via its long-term interaction with other evolutionary forces. In a Wright-Fisher population the power of drift – i.e., the across-generation variance in allele frequency due to random sampling of organisms – is inversely proportional to the effective population size, $N_e$, and issues related to the variation in drift intensity can as well be phrased in terms of variation in $N_e$.

The amount of neutral genetic diversity, or heterozygosity, carried by a population, $\pi$, is expected to reflect the mutation/drift balance and at equilibrium be proportional to the $N_e \mu$ product, where $\mu$ is the mutation rate. In principle, one could therefore assess the variation in $N_e$ among species from the variation in $\pi$. Empirical evidence shows that heterozygosity is indeed correlated with abundance across species (Ellegren & Galtier 2016). The magnitude of the observed variation, however, is smaller than intuitively expected, and moderate differences in heterozygosities are sometimes reported between species that vary markedly in census population size – an observation often called Lewontin's paradox (Lewontin 1974, Leffler et al. 2012, Romiguier et al. 2014). Three main reasons have been invoked to explain this conundrum. First, the equilibrium $\pi$ is not only influenced by $N_e$ but also by $\mu$, which of course might vary between species and obscure the signal. Secondly, population size can vary in time.
In this case $\pi$ is expected to reflect not the contemporary $N_e$, but rather the long-term time $N_e$, which more precisely is the time-harmonic mean of $N_e$ (Wright 1938) and is strongly influenced by small values. Said differently, current genetic diversity might be largely determined by ancient bottlenecks or founder effects, irrespective of the amount of drift normally experienced by the considered population. Thirdly, selection at linked sites, either positive (Gillespie 2001) or negative (Charlesworth et al. 1995), can substantially affect $\pi$ and may dominate over the effects of drift in large populations (Corbett-Detig et al. 2015, Elyashiv et al. 2016). For all these reasons, even though genomic data have confirmed the impact of drift on species genetic diversity, one cannot safely assume that $\pi$ varies among species proportionally to $N_e$, or in inverse proportion to the strength of drift.

Here we attempt to assess the variation in $N_e$ among species by exploiting another drift-dependent variable, which is the load of segregating deleterious mutations: small-$N_e$ species are expected to carry a higher load than large-$N_e$ ones at selection/drift equilibrium. The segregating mutation load can conveniently be approached via the number and population frequency of non-synonymous (= amino-acid changing) variants. This implies focusing on the coding fraction of the genome, which can be seen as a limitation. On the other hand, coding sequences offer a unique opportunity to neatly control for the effect of mutation rate and demography by jointly analysis the synonymous (= amino-acid conservative) variation, which can be assumed to be neutral. The ratio of non-synonymous to synonymous heterozygosity, $\pi_N/\pi_S$, is a typical measure of the mutation load (Chen et al. 2017). The $\pi_N/\pi_S$ ratio has a number of desirable properties. First, it is mutation rate-independent, as indicated above. Secondly, $\pi_N/\pi_S$ is expected to approach its equilibrium faster than $\pi$ after a change in $N_e$ (Pennings et al. 2014, Brandvain & Wright 2016, Gravel 2016). For this reason $\pi_N/\pi_S$ should be less sensitive than $\pi$ to ancient bottlenecks and selective sweeps. Empirically, $\pi_N/\pi_S$ was found to be negatively correlated to population size in Drosophila (Jensen & Bachtrog 2011), birds (Figuet et al. 2016), animals (Romiguier et al. 2014), plants (Chen et al. 2017) and yeasts (Elyashiv et al. 2010).

So, can one quantify $N_e$, or its among species variation, based on coding sequence polymorphism data? One major hurdle is that the expected amount and pattern of non-synonymous variation is determined not only by $N_e$ but also by the strength of selection, or more precisely, the distribution of fitness effects (DFE) of non-synonymous mutations, which is unknown (Eyre-Walker and Keightley 2007). Drift is only expected to detectably affect the population frequency of mutations with selection coefficient, $s$,
of the order of $1/N_e$ or smaller. Consider two populations of effective sizes $N_1$ and $N_2$. The expected difference in mutation load between the two essentially depends on the amount of deleterious mutations of effect intermediate between $-1/N_1$ and $-1/N_2$. If the DFE was such that most non-synonymous mutations are either much more or much less deleterious than these two values, then a small difference in load is to be expected between the two species. If however a large fraction of the non-synonymous mutations had intermediate selection coefficients, then the contrast would be higher. Welch et al. (2008) demonstrated that in a Wright-Fisher population, if the fitness effect of non-synonymous mutations follows a Gamma distribution of mean $\bar{s}$ and shape parameter $\beta$, then the expected $\pi_s/\pi_S$ is proportional to $(N_e \bar{s})^{-\beta}$. As $\beta$ decreases, the distribution gets more skewed, and the expected load becomes less strongly dependent on $N_e$. These considerations imply that the variation in $N_e$ among species can only be assessed from the non-synonymous vs. synonymous contrast via a joint estimation of the shape of the DFE.

Eyre-Walker et al. (2006) introduced a method for estimating the DFE of non-synonymous mutations from the observed frequencies of non-synonymous and synonymous variants in a population sample – the so-called site-frequency spectra, or SFS. The idea is that slightly deleterious mutations tend to segregate at lower frequency than neutral ones, so they are expected to distort the non-synonymous SFS compared to the synonymous one. Assuming Gamma-distributed deleterious effects, expressions were derived for the expected non-synonymous and synonymous SFS at mutation/selection/drift equilibrium as a function of the population mutation rate, the shape and mean of the DFE, and nuisance parameters aimed at capturing demographic effects (Eyre-Walker et al. 2006). The method has been widely re-used since then, with modifications, mostly with the aim of estimating the adaptive amino-acid substitution rate (e.g. Keightley & Eyre-Walker 2007, Boyko et al. 2008, Eyre-Walker and Keightley 2009, Schneider et al. 2011, Galtier 2016, Tataru et al. 2017, Moutinho et al. 2019, Uricchio et al. 2019). The distinct versions of the method mainly differ in how they account for departures from model assumptions. These include ancient changes in $N_e$ (Eyre-Walker 2002, Tataru et al. 2017, Rousselle et al. 2018a), linked selection (Messer & Petrov 2013, Uricchio et al. 2019), beneficial mutations (Galtier 2016, Tataru et al. 2017), and selfish processes such as GC-biased gene conversion, a meioticdistorter that favours G and C over A and T alleles irrespective of fitness effects. Three recent studies (Corcoran et al. 2017, Bolivar et al. 2018, Rousselle et al. 2018b) have demonstrated that GC-
biased gene conversion can strongly affect inferences based on the non-synonymous vs. synonymous contrast, and must be seriously taken into account.

Using this method, one can get an estimate of the distribution of the $S=4N_e \bar{s}$ product, and particularly its mean $\bar{S}=4N_e \bar{s}$. Under the assumption that distinct species share a common DFE, and therefore a common average selection coefficient $\bar{s}$ (Loewe et al. 2006), one can estimate the between-species ratio of $N_e$ from the between-species ratio of $\bar{S}$. Here we analyze a recently generated population genomic data set in animals, with a focus on the primates vs. fruit flies comparison. We ask two questions: (i) is the DFE of non-synonymous mutations similar among species? (ii) if yes, by how much does $\bar{S}$, and therefore drift power, vary across species? Assuming a classical Gamma model for the DFE did not allow us to reliably assess the variation in $N_e$ due to important differences in estimated shape parameter between taxa. This problem, however, was alleviated by including in the model a $N_e$-independent fraction of lethal mutations. Controlling for the effect of GC-biased gene conversion and segregating beneficial mutations, we estimated that $N_e$ varies by a factor of $\sim 500$ between primates and fruit flies, i.e., an order of magnitude more than predicted from the variation in genetic diversity. An even wider range of variation was uncovered when we sampled more broadly across the metazoan phylogeny.

Material and Methods

Data sets

We used the coding sequence polymorphism data set assembled by Rousselle et al (2019). This includes 50 species from 10 diverse taxa of animals (hereafter called "groups"), namely primates, rodents, passerines, fowls, fruit flies, butterflies, ants, mussels, earthworms and ribbon worms. Data in the former five groups (vertebrates + fruit flies) were taken from public databases. In the other five groups, which are non-model invertebrates, exon capture data were newly generated by Rousselle et al. (2019). Six to twenty individuals per species were genotyped at 531,360 to 14,112,150 coding positions.
from 1261 to 8604 genes (Table S1). The data set was built by selecting in each group a unique set of genes common to all species. Distinct groups, however, have distinct gene sets.

The primate and fruit fly data sets are of particularly high quality in terms of genome annotation and sample size. The two groups, furthermore, have contrasted levels of genetic diversity, with primates being among the least polymorphic, and fruit flies among the most polymorphic, taxa of animals (Leffler et al. 2012, Romiguier et al. 2014). We therefore focused on these two groups in most of the analysis. In primates, the *Papio anubis* data set had a relatively low sample size of five diploid individuals and was not analyzed here. In fruit flies, the *Drosophila sechellia* data set contained a relatively small number of SNPs and was also removed. Our main data set therefore includes five species of catarrhine primates – *Homo sapiens, Pan troglodytes, Gorilla gorilla, Pongo abelii, Macaca mulatta* – and five species of fruit flies – *D. melanogaster, D. simulans, D. teissieri, D. yakuba, D. santomea*.

In each species, the synonymous and non-synonymous SFS were generated by counting, at each biallelic position (SNPs), the number of copies of the two alleles, tri- or quadri-allelic positions being ignored. To account for missing data, a specific sample size, $n$, was chosen for each species, this number being lower than twice the number of sampled individuals. Biallelic SNPs at which a genotype had been called in less than $n/2$ individuals were discarded. When genotypes were available in exactly $n/2$ individuals, the minor allele count was simply recorded. When genotypes were available in more than $n/2$ individuals, hypergeometric projection to the $\{1, n\}$ set was performed (Hernandez et al. 2007, Gayral et al. 2013). We used the so-called "folded" SFS in our main analysis, i.e., did not rely on SNP polarization, but rather merged counts from the $k^{th}$ and $(n-k)^{th}$ categories, for every $k$. Unfolded SFS (Rousselle et al. 2019) were also used in a control analysis.

To account for the confounding effect of GC-biased gene conversion, we also generated folded GC-conservative synonymous and non-synonymous SFS by only retaining A|T and G|C SNPs, following Rousselle et al. (2018b). Estimates based on GC-conservative SFS are expected to be unaffected by any bias due to GC-biased gene conversion, but this comes at the cost of a much smaller number of SNPs. Two species, ribbonworm *Lineus longissimus* and fowl *Pavo cristatus*, had less than 100 GC-conservative SNPs and were removed from the data set.
We also re-analysed two previously-published coding sequence population genomic data sets. Chen et al. (2017) gathered SFS data from published genome-wide analyses in 34 species of animals. We focused on the 23 species in which at least five diploid individuals were sampled. These include 13 vertebrates (ten mammals, one bird, two fish), nine insects (seven Anopheles mosquitoes, one fruit fly, one butterfly) and one nematode. Galtier (2016) analyzed a data set of 44 species from eight distinct metazoan phyla. We selected the 28 species in which sample size was five or above, i.e., six species of vertebrates, six insects, five molluscs, three crustaceans, three echinoderms, two tunicates, one annelid, one cnidarian, and one nematode.

Inference methods

For each species, the population-scaled mean selection coefficient of deleterious mutations, \( \bar{S} \), was estimated using the maximum likelihood method introduced by Eyre-Walker et al. (2006), in which a model assuming Gamma distributed deleterious effects of amino-acid changing mutations is fitted to the synonymous and non-synonymous SFS. Following up on Galtier (2016), we developed a multi-species version of this model, where the shape parameter of the Gamma distribution can be shared among species. We also implemented distinct models for the DFE, namely the Gamma+lethal, partially reflected Gamma and their combination. The Gamma+lethal model simply assumes that a fraction \( p_{lh} \) of the non-synonymous mutations are lethal, i.e., cannot contribute any observable polymorphism, whereas a fraction \( 1-p_{lh} \) has Gamma-distributed effects (Eyre-Walker et al. 2006, Elyashiv et al. 2010). The partially reflected Gamma model of the DFE was introduced by Piganeau and Eyre-Walker (2003) and considers the existence of back mutations from the deleterious state to the wild type. This model entails no additional parameter compared to the Gamma model, and can be easily combined with the "+ lethal" option. These methods were here newly implemented in a multi-SFS version of the grapes program (Galtier 2016, https://github.com/BioPP/grapes). The amount of neutral polymorphism of each species was assessed using the \( \pi_s \) statistics, following Romiguier et al. (2014).

Simulations
Following Rousselle et al. (2018a), we performed simulations using SLIM V2 (Haller & Messer 2016) in order to assess how quickly $\pi_s$ and the estimated $S$ (Gamma model) reach their equilibrium when $N_e$ varies in time. We simulated the evolution of coding sequences in a single population evolving forward in time and undergoing first an increase and then a decrease in population size. We considered a panmictic population of $10^4$ or $2.10^4$ diploid individuals whose genomes consisted of 1500 coding sequences, each of 999 base pairs. The mutation rate was set to $2.2 \times 10^{-7}$ per base pair per generation and the recombination rate to $10^7$ per base pair per generation. The assumed distribution of the fitness effect of mutations comprised 50% of neutral mutations and 50% of mutations following a negative gamma distribution of mean -2.5 and shape 0.3. Each mutation that arose during a simulation was categorized as either synonymous (if the fitness effect was zero) or non-synonymous (if the fitness effect was different from zero), allowing us to compute $\pi_n$, $\pi_s$, and $\pi_n/\pi_s$ at any time point throughout the simulation. In order to make simulations tractable, we used small population sizes and high mutation rates and selection coefficients, knowing that only the products of these quantities, i.e., $4N_e\mu=0.88.10^{-2}$ and $4N_es=-0.025$, are relevant. Simulations were replicated 50 times.

**Results**

*Synonymous genetic diversity*

We first analyzed synonymous and non-synonymous SFS in five species of primates and five species of fruit flies. The estimated synonymous diversity, $\pi_s$, was much higher in fruit flies than in primates (Table 1), consistent with the literature (Leffler et al. 2012). The estimated $\pi_s$ varied from 0.062% in H. sapiens to 3.2% in D. yakuba, i.e., a factor of 51 (figure 1, X-axis). The per group median $\pi_s$ was 0.11% in primates and 1.5% in fruit flies, i.e., varied by a factor of 13.

*Gamma DFE*

A model assuming Gamma-distributed population scaled selection coefficient was fitted to SFS data separately in the 10 species. We uncovered substantial variation in $\beta$, the shape parameter of the Gamma DFE: the maximum likelihood estimate of $\beta$ varied from 0.09 to 0.36 among species (figure 1,
Y-axis). When we rather fitted a model assuming a common $\beta$ across species, the likelihood dropped severely (one $\beta$ per species, log-likelihood=-1100.8; shared $\beta$, log-likelihood=-1382.7), and the hypothesis of a common DFE shape across the 10 species was strongly rejected by a likelihood ratio test (LRT; $p$-val<10^{-20}; nine degrees of freedom). There was a trend for species from the same group to provide similar estimates of $\beta$ (figure 1). In primates, the range of estimated $\beta$ was narrow ([0.09;0.16]) and the hypothesis of a common DFE shape was not rejected by LRT (optimal $\beta$ for primates: 0.11; $p$-val=0.45; four degrees of freedom). In fruit flies, the optimal $\beta$ was close to 0.3 in four species, but much lower in *D. santomea* (figure 1). The optimal fruit fly $\beta$ was firmly rejected by primates as a group (LRT; $p$-val<10^{-20}; one degree of freedom), and reciprocally (LRT; $p$-val<10^{-20}; one degree of freedom; the test includes *D. santomea*).

Estimates of $\bar{S}=4N_e\bar{s}$, the population scaled mean selection coefficient, varied greatly among species in this analysis but this was largely explained by variations in $\beta$. Gamma shape and mean are known to be correlated parameters, and this was verified here: the across species correlation coefficient between log-transformed $\bar{S}$ and log-transformed $\beta$ was -0.94 in primates and -0.99 in fruit flies (figure S1). The among-species variation in estimated $\bar{S}$ can therefore not be taken as a reliable indicator of the variation in $N_e$ when $\beta$ differs much among species.

The observed variation in estimated DFE shape might in principle reflect biological differences among species and groups, e.g., differences in composition/structure of the proteome/interactome, maybe strengthened by our gene sampling strategy – here, species from the same group share the same genes, whereas distinct groups have distinct gene sets. The *D. santomea* behavior, however, appears difficult to reconcile with this hypothesis. *D. santomea* is closely related to *D. yakuba* and *D. teissieri* (Turissini and Matute 2017), and shares the same gene set as other species of fruit flies. We see no obvious reason why the DFE of deleterious non-synonymous mutations in *D. santomea* would differ strongly from other fruit flies, and resemble the primate DFE. Interestingly, *D. santomea* shares with primates a relatively low genetic diversity (figure 1), perhaps as a consequence of its restricted geographic distribution (Bachtrog et al. 2006). For this reason, we hypothesized that the among-species variation in estimated $\beta$ we report could reflect a failure of the Gamma model to capture the details of the DFE at all values of $N_e$, rather than genuine differences in DFE among species.
Gamma + lethal DFE

It should be recalled that very highly deleterious alleles have essentially zero probability to be observed at polymorphic stage with the sample size we used here. We reasoned that, if the DFE included a proportion of mutations of effects essentially independent of \( N_e \), this could lead to undesired effects when fitting a Gamma distribution of \( S \), a variable proportional to \( N_e \).

To investigate this, we fitted to SFS data a model assuming a proportion \( p_{lh} \) of lethal mutations, and a proportion \((1-p_{lh})\) of mutations of Gamma distributed effects. Lethal mutations are here defined as mutations having a probability to be observed at polymorphic stage equal to zero. Figure 2 shows how the likelihood responds to \( p_{lh} \) and \( \beta \) in primates and fruit flies. In fruit flies, the optimal \( \beta \) was close to 0.3 irrespective of \( p_{lh} \), and the likelihood was maximal when \( p_{lh} \) was close to 0.75. In primates, the optimal \( \beta \) varied more visibly with \( p_{lh} \). It was close to 0.1 when \( p_{lh} \) was low, as indicated above, but increased towards higher values when \( p_{lh} \) increased. The maximal likelihood in primates was still obtained when \( p_{lh} \) was close to zero and \( \beta \) close to 0.1, but importantly, areas of the parameter space close to the fruit fly optimum (e.g., \( p_{lh} \sim 0.65 \) and \( \beta \sim 0.3 \)) provided a reasonably good fit to the data (figure 2). This suggests that the DFE perhaps does not differ so dramatically between primates and fruit flies, offering the opportunity to compare estimates of \( \bar{S} \) across species under the Gamma+lethal model. The difference in log-likelihood between the one-\( \beta \)-per-species and shared-\( \beta \) models was substantially decreased under the Gamma+lethal model (191.0) compared to the Gamma model (281.9). Still, the one-\( \beta \)-per-species model significantly rejected the shared-\( \beta \) model in both cases, indicating that the inclusion of the \( p_{lh} \) parameter was not sufficient to erase every perceptible difference in DFE among species.

We jointly analyzed the ten species from the two groups under the Gamma+lethal model assuming a shared shape parameter among species, and found that the likelihood was maximal when \( p_{lh} \) was in the range 0.6 – 0.65 (figure S2). We considered these two values as plausible estimates of \( p_{lh} \). When \( p_{lh} \) was set to 0.6, the maximum likelihood estimate of \( \beta \) was 0.278. In this analysis the maximum likelihood estimate of \( \bar{S} \) varied by a factor of 550 among species, and the median fruit fly \( \bar{S} \) was
125 times as large as the median primate \( \bar{S} \) (Table 1). Of note, the estimated \( \bar{S} \) in *D. santomea* was similar to that of other species of fruit flies, and much larger than in primates. Setting \( p_{lh} \) set to 0.65 instead of 0.6 yielded similar results (Table 1). Confidence intervals were obtained by bootstrapping SNPs (100 replicates). The Gamma+lethal model outperformed the Gamma model in this analysis: letting \( p_{lh} \) be different from zero increased the log-likelihood by 85 units, which was highly significant \((p\text{-val}<10^{-20}, \text{one degree of freedom})\). We also estimated \( \bar{S} \) under the Gamma + lethal model assuming that the \( \beta \) parameter was shared by species within a group, but could vary between the two groups. Estimates of \( \bar{S} \) and \( N_e \) ratios were close to the ones obtained assuming shared \( \beta \) across all species (Table 1, right-most two colons), confirming that the inferred DFEs of the two groups tend to converge when a fraction of lethal mutations are included. We checked that our estimate of \( \bar{S} \) and \( \beta \) were such that the mean \( \pi_s/\pi \) was linearly related to \((\bar{S})^{-\beta}\) across species, as expected at equilibrium with Gamma distributed fitness effects (Welch et al. 2008). This was the case under both the Gamma and Gamma+lethal models (figure S3).

**Robustness of the estimates**

We performed a number of additional analyses in order to investigate the robustness of the above results to various methodological settings. The results are summarized in Table 2. In each of the four control analyses – "GC-conservative", "reflected Gamma", "unfolded", "subsampled fruit flies" – we tried eight values of \( p_{lh} \), from 0.4 to 0.75. We optimized \( \beta \) and the other parameters conditional on \( p_{lh} \) and recorded the likelihood. Table 2 shows the results for values of \( p_{lh} \) within 2 log-likelihood units of its maximum.

The first line of Table 2 recalls some of the results presented in Table 1. In the "GC-conservative" control, we applied the same method as in the main analysis but only including C/G and A/T SNPs, which are supposedly immune from GC-biased gene conversion. The "reflected Gamma" analysis used the same data set as the main one, but assumed a reflected Gamma instead of a Gamma DFE, thus accounting for the presence of slightly beneficial mutations – in addition to a proportion of lethal mutations. The "unfolded" control used unfolded instead of folded SFS. Finally, in the "subsampled fruit flies" control, we reduced the size of the fruit fly data set to the size of the primate one, such that
the two groups have equal weights in the analysis. A hundred data sets were generated by randomly subsampling in *D. melanogaster* the number of SNPs available in *G. gorilla*, and similarly for *D. santomea/H. sapiens*, *D. simulans/M. mulatta*, *D. teissieri/P. troglodytes*, and *D. yakuba/P. abelii*. The most likely value for $p_{th}$ was 0.6 in 77 subsampled data sets and 0.55 in the other 23. Table 2 reports the median estimates and 95% confidence intervals across the 100 subsampled data sets for these two values of $p_{th}$.

The GC-conservative data set included roughly ten times less SNPs than the main one. The likelihood surface was flatter and close to its maximum at three values of $p_{th}$ – 0.5, 0.55, 0.6 – while the estimated $\beta$ was always close to 0.3. The ratio of $\bar{S}$ between large-$N_e$ fruit flies and small-$N_e$ primates was roughly twice as low as in the main analysis, either using extreme estimates or within-group medians. The reflected Gamma analysis also yielded ratios of $\bar{S}$ a bit lower than the main analysis – but still substantially higher than ratios of $\pi_s$, see Table 1. Of note, the maximum log-likelihood under the reflected Gamma + lethal model (-1301.6) was slightly decreased compared to Gamma + lethal (-1297.7). When unfolded SFS were analyzed, the estimated $\beta$ was lower than in the main analysis, and the ratios of $\bar{S}$ was twice as large. A similar pattern was obtained when we subsampled the fruit fly data set so as to match the primate sample size – a slightly lower estimate of $\beta$, and a max/min $N_e$ ratio of the order of $10^3$. In all these control analyses the Gamma+lethal model fitted the data significantly better than the Gamma model.

**Additional metazoan taxa**

We similarly analyzed an additional 37 species from eight diverse groups of animals (Rousselle et al. 2019). In this 47-species data set, the estimated synonymous diversity $\pi_s$ varied from 0.062% in *H. sapiens* to 4.4% in mussel *Mytilus trossulus*. The within-group median $\pi_s$ varied from 0.012 in primates to 3.3% in mussels, *i.e.*, by a factor of 27.

The results were largely consistent with the primates vs. fruit flies comparison. First applying the Gamma model, we detected a strong group effect on the estimated $\beta$ (figure S4). Butterflies and ants, for instance, tended to yield relatively high estimates of $\beta$, whereas the best fit in primates and birds was reached at relatively low $\beta$. This was mitigated by moving to the Gamma+lethal model, even
though the specificity of particular groups was still apparent (figure S5). When we jointly analyzed all
species assuming a Gamma+lethal DFE using folded SFS and all mutations, the maximal likelihood
was reached when \( p_{lh} \) was 0.5 (with \( p_{lh}=0.55 \) being close) and the estimated \( \beta \) was 0.26. We calculated
the within-group median \( \bar{S} \), which varied by a factor of 110 among groups (Table 3, first colon). We
performed the same GC-conservative, reflected Gamma and unfolded analyses as described above
(Table 3). The estimated \( \bar{S} \) in the reflected Gamma, unfolded and main analyses were strongly
correlated \( (r^2>0.97 \text{ in all three pairwise comparisons}) \), while the ratio of max/min median \( \bar{S} \) varied a
bit: it was 73 in the Gamma reflected analysis, 200 in the unfolded analysis. In all three analyses, fruit
flies were the group with the highest median \( \bar{S} \), and primates the lowest.

The GC-conservative analysis differed a bit and did not rank the ten groups in the same order (Table 3).
In this analysis, the lowest median \( \bar{S} \) was found in ants and the highest in mussels, the ratio between
these two numbers being 180. The GC-conservative estimates of \( \bar{S} \) were more strongly correlated
with species propagule size, a variable previously identified as a proxy for \( N_e \) in animals (Romiguier et
al. 2014), than were the other three estimates (GC-conservative: \( r^2=0.27, p\text{-val}=3.10^{-4} \); main analysis:
\( r^2=0.11, p\text{-val}=3.10^{-2} \); log-transformed variables). Figure 3 shows the relationship between \( \pi_S \) and the
relative \( \bar{S} \) as estimated from the GC-conservative data set. Note that the Y-axis encompasses three
orders of magnitude, vs. two on the X-axis. As mentioned above, \( D.\, santomea \) was an outlier: \( \pi_S \) in this
species was low, but the estimated \( \bar{S} \) was typical of fruit flies and other large-\( N_e \) species. The ratio of
max/min estimated \( \bar{S} \) across the 47 species was 491, 1100, 301 and 1076, respectively, in the main,
GC-conservative, reflected Gamma and unfolded analyses.

We similarly analyzed two additional, recently published data sets. The Chen et al. (2017) data set (23
species) yielded results similar to the Rousselle data set: the optimal \( p_{lh} \) for this data set was 0.5, the
estimated \( \beta \) was 0.29, and the ratio of maximal to minimal \( \bar{S} \) was 581 – to be compared to a max/min
ratio of \( \pi_S \) of 54. Of note, the estimated ratio of \( \bar{S} \) between \( D.\, melanogaster \) and \( H.\, sapiens \) was 139
in the Chen et al. (2017) data set, which is close the the 119 obtained in our main analysis (see Table 1).
The Galtier (2016) data set (28 species) differed from the other two in that the likelihood was largely
insensitive to \( p_{lh} \): it varied by less than 2 log units over the 0-0.6 range for \( p_{lh} \). The estimated \( \beta \) was
close to 0.275 irrespective of \( p_{lh} \) (in the 0-0.6 range), and the ratio of maximal to minimal estimated
\( \bar{S} \) was 7300-7900, \text{i.e.}, an additional order of magnitude compared to Rousselle's and Chen's data
sets. This result was mainly explained by a very high estimated $\bar{S}$ in the mosquito *Culex pipiens* and, particularly, the nematode *Caenorhabditis brenneri* (figure S6).

**Discussion**

Here we introduced a novel approach to compare the intensity of genetic drift among species based on coding sequence SFS data. Below we discuss the assumptions, merits and limitations of this approach (subsection 1 to 3), before moving to the interpretation and implications of our results (subsection 4 and 5). Subsections 1-3 are rather technical and can be skipped by readers mainly interested in biological aspects.

1. **Estimating $N_e$-related parameters from SFS data**

Here we used the approach introduced by Eyre-Walker et al. (2006) for fitting a population genetic model to a synonymous and a non-synonymous SFS. This model includes three parameters of interest: population mutation rate $\theta$, average deleterious effect $\bar{S}$ and DFE shape $\beta$. In addition, the model has $n/2$ parameters of nuisance, where $n$ is the sample size – the so-called $r_i$'s (see equations 1 and 2 in Eyre-Walker et al. 2006, here applied to folded SFS). Parameter $r_i$ multiplies the $i^{th}$ entry of the expected synonymous and non-synonymous SFS. The $r_i$'s are intended to capture any effect that similarly affects the fate of synonymous and non-synonymous mutations – such as linked selection, population substructure and departure from demographic equilibrium. In practice $r_1$ is set to one, and $r_i$, $i>1$, can be interpreted as the relative effective mutation rate of the $i^{th}$ frequency category, compared to the first category.

Including the $r_i$'s in the model is often necessary in terms of goodness of fit. This was the case here: when we set all the $r_i$'s to 1, i.e., assumed panmixy, no linked selection and demographic equilibrium, the likelihood dropped dramatically, from -1297.7 to -71,404.7 (primates + fruit flies dataset, Gamma+lethal, $p_{\text{ith}}=0.6$). Such a simplistic model is strongly rejected by the data and can hardly be used for inference purposes. Alternatively, one could try to explicitly model population substructure, linked selection (Good et al. 2014) and/or departure from demographic equilibrium (Evans et al. 2007,
Keightley & Eyre-Walker 2007). In practice, however, these effects are very difficult to disentangle. Messer and Petrov (2013), for instance, demonstrated that linked selection severely confounds inferences on the variation of \( N_e \) in a two-epoch model. This is why the flexible \( r_i \)-based parametrization has been used in numerous recent applications of the extended McDonald-Kreitman method (e.g. Galtier 2016, Tataru et al. 2017, Rousselle et al. 2018, Moutinho et al. 2019).

Introducing the \( r_i \)'s, however, has one drawback, which is that this tends to blur the interpretation of the estimates of parameters of interest, particularly \( \theta \) and \( \bar{S} \). Consider for instance a data set in which the estimated \( r_i \)'s (\( i>1 \)) are all well below one, meaning that singletons are in excess compared to the standard coalescent expectation. Such a pattern is expected for a population having experienced a strong bottleneck. What exactly \( \theta \) measures in this case is unclear, in absence of a formal model. Now consider another data set yielding the same estimate of \( \theta \), but with estimated \( r_i \)'s (\( i>1 \)) all well above one, as expected under population expansion. To conclude that the two considered populations have the same population mutation rate would appear somehow meaningless. A similar problem possibly applies to the interpretation and among-data sets comparisons of the estimated \( \bar{S} \). In this study we did not estimate \( \theta \) via the maximum likelihood method, but rather used \( \pi_S \) as our estimate of \( \theta \). We did, however, use the maximum likelihood of \( \bar{S} \). One should keep in mind that the meaning of these estimates are somehow dependent on the estimated \( r_i \)'s, to an extent currently difficult to quantify in the absence of a formal investigation of this issue.

To investigate this issue a bit deeper, we plotted the estimated \( r_i \)'s for all SNP categories \( i \) in primates and fruit flies (figure S7). In this analysis, the Gamma+lethal model was used, species were analyzed separately (no shared parameter), and \( p_{th} \) was set to 0.6. Figure S7 shows that the vast majority of the \( r_i \)'s belong to the [0.5, 1.5] interval, i.e., do not dramatically differ from 1, both in primates and fruit flies. The figure also shows that, although the median differs a bit between primates and fruit flies for SNP category 3 to 8, the distributions are largely overlapping among the two groups. A more detailed analysis would be worthwhile, but figure S7 does not suggest to us that the \( r_i \)'s pose a major problem of comparability in this analysis.

2. Strong variation in Gamma-DFE shape across species
When a Gamma DFE was assumed, our analysis revealed significant among-taxa variation in the estimated shape parameter. In primates, the best fit was achieved when the shape parameter was of the order of 0.1 – 0.15, and consistently so in five different species. These values are close to those obtained by Castellano et al. (2019), who similarly analyzed coding sequence SFS data in nine populations of great apes. Assuming a Gamma DFE, these authors found that the best model was one in which the shape parameter was shared among species and equal to 0.16. This estimate was only slightly increased when a fraction of beneficial mutations was modeled (Castellano et al. 2019). In *Drosophila*, in contrast, we found that the value of $\beta$ that best fitted the data was close to 0.3 – with the exception of *D. santomea*. This also is consistent with previous reports. Keightley et al. (2016), for instance, obtained a point estimate of 0.35 for $\beta$ in *D. melanogaster*, taking special care of the problem of SNP mis-orientation. In butterflies, our median estimated $\beta$, 0.37 (min=0.29, max=0.52), is close to the 0.44 reported by Mackintosh et al. (2019) based on a different approach. Our results corroborate those of Chen et al (2017), who provided a detailed analysis of the variation in estimated $\beta$ across various taxa of animals and plants, and like us reported a tendency for species with a low genetic diversity to exhibit low values of $\beta$.

The variation in estimated $\beta$ we and Chen et al (2017) report are substantial. It should be recalled that the shape parameter has a strong effect on the Gamma distribution, as it is inversely proportional to its variance. Consider for example Gamma-distributed DFEs sharing the same mean $s$ of, say, 0.1, but with distinct shape parameters. If $\beta$ is set to 0.1, as estimated in primates, then 53% of mutations are associated with a selection coefficient smaller than 0.001. If $\beta$ was rather equal to 0.3, as in fruit flies, then this percentage would be 19%, and down to 12% if $\beta$=0.4, as in butterflies. Given the scarcity of experimental data on the non-synonymous DFE in animals, one cannot firmly argue that such differences are implausible. We note, however, that if the estimates of $\beta$ we obtained reflected a biological reality, this would entail considerable variation in the prevalence of small effect mutations across animal proteomes, for which an explanation would be needed.

Here we rather hypothesized that the among-taxa variation in estimated $\beta$ is, for its largest part, due to model mis-specification, that is, we suggest that the true DFE might not be Gamma-distributed.
Besides the above intuitive argument, this hypothesis is based on two observations. The first one is the behavior of the *D. santomea* data set, which carries much less diversity than other species of fruit flies, and yielded a distinctively lower estimate of \( \beta \), suggesting that the Gamma model fails to correctly capture the shape of the DFE at all values of \( N_e \). The alternative explanation, *i.e.*, that the DFE in *D. santomea* truly differs from that of other fruit flies, appears dubious in this case given the low between-species divergence. Secondly, we found that the Gamma + lethal model provided a significantly better fit to the data than the Gamma model, while predicting DFE shapes that were more similar across taxa.

In this study, therefore, we implicitly attributed the observed variation in estimated Gamma shape to a methodological artifact. This is obviously questionable. It could be that the among-taxa variation in estimated \( \beta \) reported here and in Chen et al. (2017) has some biological relevance. This would deserve to be investigated experimentally, *e.g.* following the approach of Böndel et al. (2019), who obtained an estimate of \( \beta=0.3 \) in green algae *Chlamydomonas reinhardtii* based on crosses and fitness measurements in mutation accumulation lines. **It should be kept in mind that our discussion of the variation in \( N_e \) is conditional on the assumption of a common DFE among species.**

### 3. Modeling lethal mutations

The Gamma + lethal model, which considers an \( N_e \)-independent fraction of large-effect deleterious mutations, provided a significantly better fit to the data than the Gamma model. Here is a possible interpretation of this result. Fitting a DFE model to a non-synonymous and a synonymous SFS implies to accommodate both the difference in shape (the relative frequencies of singletons, doubletons, etc...) and in size (the total number of SNPs) between the two spectra. The former is determined by the balance between neutral and slightly deleterious mutations, whereas the latter mainly reflects the proportion of strongly deleterious mutations. We suggest that the Gamma distribution struggles to accommodate these two aspects at the same time. When the \( \pi_{\text{S}}/\pi_{\text{S}} \) ratio is relatively high, as in primates, there is a tendency for the Gamma model to converge towards low values of \( \beta \), thus ensuring a large proportion of small effect non-synonymous mutations, while a higher \( \beta \) could fit the difference in shape between the two spectra equally well, or maybe better. The additional \( p_{\text{lab}} \) parameter of the
Gamma + lethal model, we suspect, somehow releases this constraint by controlling to a large extent the predicted $\pi_N/\pi_S$ ratio.

This interpretation seems to fit reasonably well our analysis of the primates + fruit flies dataset, and the report by Chen et al (2017) of a lower estimate of $\beta$ in small-$N_e$ species. This rule, however, does not always apply. Our estimate of $\beta$ in *Formica* ants, for instance, was close to 0.4 under the Gamma model, despite a low genetic diversity in this taxon. More work would appear needed to determine whether the relatively high estimate of $\beta$ we obtained in ants reveals a real peculiarity of this group, or is due to the specific gene set we analyzed here, or can be explained by the relatively small SNP sample size of our ant data set, compared to primates and fruit flies (Table S1). To our knowledge, the Gamma + lethal model has been tried in two studies before this one. Eyre-Walker et al. (2006) analyzed a data set of 320 genes in 90 human individuals, and found that adding the $p_l$ parameters did not change much the picture, compared to the Gamma model. This contrasts with our analysis and highlights the sensitiveness of this kind of analyses to the specificities of the data – number of genes, of individuals, SNP calling procedure. Elyashiv et al. (2010) applied the Gamma + lethal model to yeast data and found that the shape parameters converged towards 0.35 under this model, which is very close to our joint estimate, while the Gamma model supported a higher value for $\beta$.

The Gamma + lethal model is a simple modification of the Gamma model, which was sufficient to significantly improve the fit in this analysis. The model, however, is not entirely satisfactory. In particular, our analysis makes the assumption of a common proportion of very strongly deleterious mutations among species with different $N_e$, which appears awkward knowing that the probability for a mutation to segregate at observable frequency is determined by the $N_e.s$ product. There might be more efficient, continuous ways to model the DFE and solve the problem posed by the Gamma distribution with this data set. One intrinsic difficulty with SFS model fitting is that deleterious mutations of sufficiently large effects will be equally unobservable irrespective of their precise selection coefficient – for instance, mutations of selection coefficient $s=-100/N_e$ or $s=-1000/N_e$ just have essentially zero probability to be observed in a population sample. This means that we lack information on the tail of the distribution we are trying to model.

4. Estimating $N_e$: mutation load vs. diversity
Our analysis of the load of deleterious mutations uncovered three (based on the Rousselle et al. 2019 and Chen et al. 2017 data sets) or four (based on the Galtier 2016 data set) orders of magnitude of variation in drift power among species of animals, when the neutral genetic diversity of the very same species varied by a factor of 100 or less. Below we discuss potential reasons for this discrepancy.

First, it should be recalled that, unlike the non-synonymous to synonymous contrast, the genetic diversity of a species is influenced by the mutation rate. If the mutation rate was negatively correlated with population size, and differed by one or two orders of magnitude between species of animals, then our results could be explained very simply. Empirical estimates in humans (Kong et al. 2012) and Drosophila (Keightley et al. 2009) indeed seem to point to an order of magnitude of difference in per base, per generation mutation rate between these two taxa. We lack, however, a reliable estimate of the mutation rate in the vast majority of the species of our data set. The existence of a negative relationship between \( N_e \) and \( \mu \), although somehow expected theoretically (Sung et al. 2012), is so far hypothetical.

Demographic fluctuations are another potential cause of discrepancy between genetic diversity-based and mutation load-based estimates of \( N_e \). Brandvain and Wright (2016) recalled that the mutation/selection/drift equilibrium is reached more quickly when selection is strong, with neutral mutations being the slowest to converge. This suggests that the mutation load might be less strongly influenced by ancient bottlenecks than the neutral genetic diversity, and therefore yield more reliable estimates of present-day drift power. In order to further investigate this hypothesis, we simulated coding sequence evolution in a population after a strong bottleneck. We found that the estimated \( \bar{S} \) indeed equilibrates faster than \( \pi_S \) during the recovery phase (figure 4, generations 0-10,000). In these simulations, diversity-based estimates of \( N_e \) would be biased downwards during a substantial period of time after the bottleneck, whereas the mutation load would more quickly provide a reliable estimate. Ancient bottlenecks, therefore, might explain why genetic diversity-based and mutation load-based estimates of drift power sometimes disagree – e.g. see above our discussion of the D. santomea case. Can this effect account for the increased between species variance in drift power we report, compared to genetic diversity-based estimates? This would require additional assumptions, such as, e.g., that large-\( N_e \) species tend to fluctuate more than small-\( N_e \) ones, or that the minimal value reached by \( N_e \) as
populations fluctuate varies less among species than the maximal one. Such hypotheses have already been proposed (Romiguier et al. 2014) but so far lack any empirical support.

A third and major factor potentially affecting the estimation of \( N_e \) is linked selection. The mutation load results from stochastic variations in allele frequency, which we have so far interpreted in terms of genetic drift. Linked selection – \( i.e. \), selective sweeps and background selection – is another source of stochasticity, which like drift is expected to result in a decreased genetic diversity and an increased mutation load (Barton 2010, Hartfield and Otto 2011). Corbett-Detig et al. (2015) demonstrated that the reduction in genetic diversity due to linked selection is stronger in large than in small population-sized species of plants and animals. Linked selection, therefore, tends to homogenize the genetic diversity among species. The impact of linked selection on the mutation load has been less thoroughly investigated, either theoretically or empirically. What we know is that recurrent selective sweeps result in patterns of neutral variation at linked loci that are best represented by a form of multiple-merger coalescent (Durrett and Schweinsberg 2005, Coop and Ralph 2012). Multiple-merger coalescents, on the other hand, are known to predict patterns of neutral and selected variation that depart the predictions of just drift (Eldon and Wakeley 2006, Der et al. 2012). So it might be that the respective effects of linked selection on the genetic diversity and the mutation load do not scale similarly with population size, perhaps explaining our results. This, again, is entirely hypothetical and would require to be confirmed via specific theoretical developments, which ideally should also account for the effect of background selection.

Finally, Castellano et al. (2019) recently introduced a new hypothesis. These authors reported a wider range of variation in \( \bar{S} \) than in \( \pi \) among species of great apes, which they interpreted in terms of positive epistasis. Castellano et al. (2019) suggested that the average effect of a new deleterious mutation is negatively related to the existing load, because of the interaction between deleterious mutations. According to their interpretation, species genetic diversity would scale linearly with \( N_e \), whereas variations in \( \bar{S} \) would reflect a positive correlation between \( N_e \) and \( \bar{S} \) – in contrast, we here assumed that \( \bar{S} \) is constant across species. This interesting hypothesis offers yet another potential explanation to the discrepancy between diversity-based and mutation load-based estimators of drift power.
5. How variable among species is $N_e$?

Lewontin's paradox has been a recurrent cause of concern/excitement over the last decade (Leffler et al. 2012, Romiguier et al. 2014, Corbett-Detig et al. 2015, Coop 2016, Filatov 2019, Mackintosh et al. 2019). In animals, the within-species genetic diversity roughly spans two orders of magnitude, whereas population density and geographic range vary considerably more across species. Our analysis of the mutation load rather suggests that $N_e$ varies by a factor of $10^3$, or maybe $10^4$, among species of animals. This is a step towards reconciling genetic with ecological estimates of population size – but how big is this step?

On one hand, one or two additional orders of magnitude can be seen as a moderate improvement, far from reconciling the effective and census population sizes of animal populations. Small insects or nematodes presumably outnumber large vertebrates by much more than a factor of 500 or 5000. On the other hand, Lewontin's paradox may appear somewhat naive in suggesting that the genetic diversity should be proportional to population size. Clearly, very large populations can just not follow the $\pi = 4N_e \mu$ prediction. This equation assumes mutation-drift equilibrium, which is only reached after a number of generations of the order of $N_e$. As $N_e$ increases, the assumption that the considered population has been devoid of bottlenecks and sweeps during the last $\sim N_e$ generations becomes less and less plausible (Gillespie 2000). So maybe should we be satisfied, after all, by an estimated ratio of $10^3$ or $10^4$ of long-term $N_e$ among species of animals? We suggest that Lewontin's "paradox" in part reflects the varying definition/usage of the $N_e$ parameter in the molecular evolutionary literature. Assessing the amount of stochasticity in allele frequency evolution, the prevalence of linked selection vs. drift, and their impact on genome evolution are key goals of current population genomics that perhaps do not need to be phrased in terms of a paradox, and probably cannot be reduced to just the issue of estimating one "$N_e$" per species.

Acknowledgments
We are grateful to Guillaume Achaz, Thomas Bataillon, Sylvain Glémin and Benoît Nabholz for insightful discussions. This work was supported by Agence Nationale de la Recherche grant no. ANR-15-CE12-0010.

References


Wright S. Size of population and breeding structure in relation to evolution. Science. 1938 87:430–431
Table 1. Maximum likelihood estimates of $\bar{S}$ in primates and fruit flies under the Gamma+lethal DFE model.

<table>
<thead>
<tr>
<th>species</th>
<th>$\pi_S$ (%)</th>
<th>$\beta$ shared across species</th>
<th>one $\beta$ per group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$p_{th}=0.6$</td>
<td>$p_{th}=0.65$</td>
</tr>
<tr>
<td>$H. sapiens$</td>
<td>0.06</td>
<td>-3.36</td>
<td>-1.39</td>
</tr>
<tr>
<td>$G. gorilla$</td>
<td>0.11</td>
<td>-4.38</td>
<td>-2.08</td>
</tr>
<tr>
<td>$P. troglodytes$</td>
<td>0.11</td>
<td>-7.56</td>
<td>-3.30</td>
</tr>
<tr>
<td>$P. abelii$</td>
<td>0.16</td>
<td>-9.16</td>
<td>-4.64</td>
</tr>
<tr>
<td>$M. mulatta$</td>
<td>0.18</td>
<td>-20.0</td>
<td>-10.4</td>
</tr>
<tr>
<td>$D. santomea$</td>
<td>0.11</td>
<td>-401</td>
<td>-232</td>
</tr>
<tr>
<td>$D. melanogaster$</td>
<td>1.07</td>
<td>-502</td>
<td>-295</td>
</tr>
<tr>
<td>$D. teissieri$</td>
<td>1.52</td>
<td>-1860</td>
<td>-1080</td>
</tr>
<tr>
<td>$D. yakuba$</td>
<td>1.57</td>
<td>-946</td>
<td>-553</td>
</tr>
<tr>
<td>$D. simulans$</td>
<td>3.18</td>
<td>-1660</td>
<td>-970</td>
</tr>
<tr>
<td>max/min $N_e$</td>
<td>53.0</td>
<td>555</td>
<td>779</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[324-1050]</td>
<td>[438-1660]</td>
</tr>
<tr>
<td>group-median $N_e$ ratio</td>
<td>13.8</td>
<td>125</td>
<td>168</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[84-203]</td>
<td>[102-296]</td>
</tr>
</tbody>
</table>
Table 2. Robustness analysis, primates vs. fruit flies.

<table>
<thead>
<tr>
<th>label</th>
<th>analysis</th>
<th>#SNP flies</th>
<th>#SNP prim.</th>
<th>$p_{th}$</th>
<th>$\beta$</th>
<th>max/min $\bar{S}$</th>
<th>median $\bar{S}$</th>
<th>ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>main</td>
<td>1.14 $10^6$</td>
<td>5.68 $10^4$</td>
<td>0.6</td>
<td>0.28</td>
<td>555</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>b₁</td>
<td>GC-conservative</td>
<td>1.84 $10^5$</td>
<td>3.88 $10^3$</td>
<td>0.5</td>
<td>0.29</td>
<td>251</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>b₂</td>
<td>GC-conservative</td>
<td>1.84 $10^5$</td>
<td>3.88 $10^3$</td>
<td>0.55</td>
<td>0.29</td>
<td>302</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>b₃</td>
<td>GC-conservative</td>
<td>1.84 $10^5$</td>
<td>3.88 $10^3$</td>
<td>0.6</td>
<td>0.30</td>
<td>398</td>
<td>84</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>reflected Gamma</td>
<td>1.14 $10^6$</td>
<td>5.68 $10^4$</td>
<td>0.6</td>
<td>0.30</td>
<td>341</td>
<td>88</td>
<td></td>
</tr>
<tr>
<td>d₁</td>
<td>unfolded</td>
<td>5.9 $10^6$</td>
<td>5.68 $10^4$</td>
<td>0.55</td>
<td>0.23</td>
<td>1210</td>
<td>227</td>
<td></td>
</tr>
<tr>
<td>d₂</td>
<td>unfolded</td>
<td>5.9 $10^6$</td>
<td>5.68 $10^4$</td>
<td>0.6</td>
<td>0.23</td>
<td>1480</td>
<td>274</td>
<td></td>
</tr>
<tr>
<td>e₁</td>
<td>subsampled fruit flies</td>
<td>5.68 $10^4$</td>
<td>5.68 $10^4$</td>
<td>0.55</td>
<td>0.23</td>
<td>1300</td>
<td>[752-3520]</td>
<td>240</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e₂</td>
<td>subsampled fruit flies</td>
<td>5.68 $10^4$</td>
<td>5.68 $10^4$</td>
<td>0.6</td>
<td>0.25</td>
<td>900</td>
<td>[577-2020]</td>
<td>183</td>
</tr>
</tbody>
</table>
Table 3. Analysis of 47 species in 10 groups of animals.

<table>
<thead>
<tr>
<th></th>
<th>main</th>
<th>GC-cons.</th>
<th>reflected</th>
<th>unfolded</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p_{th})</td>
<td>0.5</td>
<td>0.55</td>
<td>0.5</td>
<td>0.45</td>
</tr>
<tr>
<td>(\beta)</td>
<td>0.26</td>
<td>0.28</td>
<td>0.29</td>
<td>0.22</td>
</tr>
<tr>
<td>relative median (\bar{S}) :</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>primates</td>
<td>1</td>
<td>1.6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ants</td>
<td>1.6</td>
<td>1</td>
<td>1.6</td>
<td>1.6</td>
</tr>
<tr>
<td>butterflies</td>
<td>1.8</td>
<td>6.1</td>
<td>1.7</td>
<td>1.8</td>
</tr>
<tr>
<td>passerines</td>
<td>3.1</td>
<td>4</td>
<td>2.9</td>
<td>3.1</td>
</tr>
<tr>
<td>earthworms</td>
<td>4.8</td>
<td>6.4</td>
<td>4.3</td>
<td>5.1</td>
</tr>
<tr>
<td>rodents</td>
<td>5.2</td>
<td>7.6</td>
<td>4.6</td>
<td>6.5</td>
</tr>
<tr>
<td>fowls</td>
<td>21</td>
<td>7.8</td>
<td>17</td>
<td>25</td>
</tr>
<tr>
<td>ribbonworms</td>
<td>38</td>
<td>120</td>
<td>29</td>
<td>55</td>
</tr>
<tr>
<td>mussels</td>
<td>40</td>
<td>180</td>
<td>30</td>
<td>60</td>
</tr>
<tr>
<td>fruit flies</td>
<td>110</td>
<td>130</td>
<td>75</td>
<td>200</td>
</tr>
</tbody>
</table>
Figure legends

**Figure 1: The estimated Gamma shape differs between primates and fruit flies.**

**Figure 2: Gamma+lethal model, likelihood surface.**
X-axis: proportion of lethal non-synonymous mutations. Y-axis: Gamma shape parameter for non-lethal, non-synonymous mutations. The color scale is indicative of the log-likelihood: green=high, yellow=intermediate, red=low. The difference between local and maximal log-likelihood is indicated by numbers within a couple of relevant cells – 0 means maximal. The marked cell corresponds to the maximum of the likelihood when both data sets are jointly analyzed (see Figure S2). X-axis continuously covers the 0.4-0.8 range, and also shows the $p_{lh}=0$ case (Gamma model).

**Figure 3: Magnitude of variation in population scaled mean selection coefficient vs. neutral genetic diversity across 47 species of animals.**
Each dot is for a species of animals. X-axis: relative synonymous heterozygosity. Y-axis: relative estimated $\bar{S}$. Colors indicate groups, see Table 3. Estimates of $\bar{S}$ were divided by its minimal value, which was obtained in *Formica pratensis*.

**Figure 4: Simulation of the evolution of $\pi_s$ and $\pi_N/\pi_s$ as $N_e$ fluctuates.**
The simulation starts at time $t=0$ in a population devoid of polymorphism. The $\pi_N/\pi_s$ ratio converges towards its equilibrium value faster than $\pi_s$ during the recovery phase. Then a bottleneck is simulated at $t=15,000$ generations. The two statistics quickly reach their new equilibrium. Two population sizes were used: $N_e=10,000$ and 20,000. Error bars represent the variance across 50 replicates.
Figure 1: The estimated Gamma shape differs between primates and fruit flies.

Figure 2: Gamma+lethal model, likelihood surface.

X-axis: proportion of lethal non-synonymous mutations. Y-axis: Gamma shape parameter for non-lethal, non-synonymous mutations. The color scale is indicative of the log-likelihood: green=high, yellow=intermediate, red=low. The difference between local and maximal log-likelihood is indicated by numbers within a couple of relevant cells – 0 means maximal. The marked cell corresponds to the maximum of the likelihood when both data sets are jointly analyzed (see Figure S2). X-axis continuously covers the 0.4-0.8 range, and also shows the $p_{lth}=0$ case (Gamma model).
Figure 3: Magnitude of variation in population scaled mean selection coefficient vs. neutral genetic diversity across 47 species of animals.

Each dot is for a species of animals. X-axis: relative synonymous heterozygosity. Y-axis: relative estimated $\bar{S}$. Colors indicate groups, see Table 3. Estimates of $\bar{S}$ were divided by its minimal value, which was obtained in *Formica pratensis*. 
Figure 4: Simulation of the evolution of the estimated $\bar{S}$ and $\pi_s$ as $N_e$ fluctuates.

The simulation starts at time $t=0$ in a population devoid of polymorphism. The estimated $\bar{S}$ converges towards its equilibrium value faster than $\pi_s$ during the recovery phase. Then a bottleneck is simulated at $t=15,000$ generations. The two statistics quickly reach their new equilibrium. Two population sizes were used: $N_e=10,000$ and 20,000. Error bars represent the variance across 50 replicates. Simulated datasets in which the estimated shape parameter was smaller than 0.1 or greater than 0.6 were discarded – they yield extreme estimates of $\bar{S}$. 

39
**Figure S1:** Gamma model, separate analysis, primates vs. fruit flies, mean/shape correlation

Blue: primates. Red: fruit flies. Both axes are in log scale.

**Figure S2:** Gamma+lethal model, likelihood surface, primates + fruit flies.

Legend: see figure 2.

**Figure S3:** Relationship between $\frac{\pi_n}{\pi_s}$ and $(\bar{S})^{-\beta}$, primates + fruit flies.

A linear relationship between the two plotted variables is theoretically expected at equilibrium with Gamma distributed effects (Welch et al. 2008).

**Figure S4:** Gamma model, 47 species, shape parameter.

Each dot is for a species. Y-axis: separate analysis: each species has its own shape parameter. X-axis: species from the same group share a common shape parameter.

**Figure S5:** Gamma+lethal model, likelihood surface, 10 groups of animals


**Figure S6:** Magnitude of variation in population scaled mean selection coefficient vs. neutral genetic diversity, Galtier 2016 data set.

Each dot is for a species of animals. X-axis: relative synonymous heterozygosity. Y-axis: relative estimated $\bar{S}$. The highest two estimates of $\bar{S}$ were obtained in nematode *Caenorhabditis brenneri* and mosquito *Culex pipiens*.

**Figure S7:** Distribution of $r_i$'s across species of primates (blue) and fruit flies (red)

Model fitting was realized separately – no shared parameter among species. The Gamma+lethal model was applied to folded SFS, with parameter $p_{lh}$ set to 0.6. Only one species of fruit flies has more than 9 categories of SNPs.