How Much Does N_e Vary Among Species?

Nicolas Galtier*,[1] and Marjolaine Rousselle*,[†]
*Institute of Evolution Sciences of Montpellier (ISEM), CNRS, University of Montpellier, IRD, EPHE, 34095 Montpellier, France and
[†]Bioinformatics Research Centre, Aarhus University, DK-8000, DK Aarhus, Denmark
ORCID ID: 0000-0002-0479-4878 (N.G.)

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 also inferring the distribution of fitness effects of deleterious mutations. To this aim, we analyze the nonsynonymous (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 distribution of fitness effects 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 nonsynonymous 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.

KEYWORDS genetic drift; population size; mutation load; distribution of fitness effects; site-frequency spectrum

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, thereby increasing the genetic load (Ohta 1972; Lynch *et al.* 2011). Drift affects patterns of genome variation and can mimic or hide traces of adaptation (Jensen *et al.* 2005; Klopfstein *et al.* 2006; 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 analyzing the dynamics of allele frequency across a controlled number of generations (Jónás *et al.* 2016; Nené *et al.* 2018). This is convenient for experimentally evolving populations, but trickier 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 also be phrased in terms of variation in N_e.

The amount of neutral genetic diversity, or heterozygosity, carried by a population, π, is expected to reflect the mutation/drift balance and at equilibrium be proportional to the $N_e.\mu$ product, where μ is the mutation rate. In principle, one could therefore assess the variation in N_e among species from the variation in π. Empirical evidence shows that heterozygosity is indeed correlated with abundance across species (Ellegren and 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 π is not only influenced by N_e but also by μ, which of course might vary between species and obscure the signal. Second, population size can vary in

doi: https://doi.org/10.1534/genetics.120.303622
Manuscript received May 26, 2020; accepted for publication August 20, 2020; published Early Online August 24, 2020.
Supplemental material available at figshare: https://doi.org/10.25386/genetics.12833708.
[1]Corresponding author: Institut des Sciences de l'Evolution Université de Montpellier, Place E. Bataillon, CC65, 34095 Montpellier, France. E-mail: nicolas.galtier@umontpellier.fr

time. In this case π is expected to reflect not the contemporary N_e, but rather the long-term N_e, which more precisely is the time-harmonic mean of Ne (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. Third, selection at linked sites, either positive (Gillespie 2001) or negative (Charlesworth *et al.* 1995), can substantially affect π and may dominate over the effects of drift in large populations (Corbett-Detig *et al.* 2015; Elyashiv *et al.* 2016). All these effects make the concept of N_e a rather complex one, difficult to define and measure (Ewens 1984; Sjödin *et al.* 2005; Karasov *et al.* 2010).

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 nonsynonymous (= 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 analyzing the synonymous (= amino-acid conservative) variation, which can be assumed to be neutral. The ratio of nonsynonymous to synonymous heterozygosity, π_N/π_S, is often used as a measure of the mutation load (Chen *et al.* 2017). The π_N/π_S ratio has a number of desirable properties. First, it is mutation rate-independent, as indicated above, which is convenient for comparisons among distantly related species. Second, π_N/π_S is expected to approach its equilibrium faster than π after a change in N_e (Pennings *et al.* 2014; Brandvain and Wright 2016; Gravel 2016). For this reason, π_N/π_S should be less sensitive than π to ancient bottlenecks and selective sweeps. Empirically, π_N/π_S was found to be negatively correlated to population size in *Drosophila* (Jensen and 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 nonsynonymous 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 nonsynonymous 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 $-1/N_{22}$. If the DFE was such that most nonsynonymous 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 nonsynonymous mutations had intermediate selection coefficients, then the contrast would be sharper. Welch *et al.* (2008) demonstrated that in a Wright–Fisher population, if the fitness effect of nonsynonymous mutations follows a Gamma distribution of mean $\bar{s}$ and shape parameter β, then the expected π_N/π_S is proportional to $(N_e s)^{-\beta}$. As β 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 nonsynonymous *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 nonsynonymous mutations from the observed frequencies of nonsynonymous 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 nonsynonymous SFS compared to the synonymous one. Assuming Gamma-distributed deleterious effects, expressions were derived for the expected nonsynonymous 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 reused since then, with modifications, mostly with the aim of estimating the adaptive amino-acid substitution rate (Keightley and 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.* 2018), linked selection (Messer and Petrov 2013; Uricchio *et al.* 2019), beneficial mutations (Galtier 2016; Tataru *et al.* 2017), and selfish processes such as GC-biased gene conversion, a meiotic distorter that favors G and C over A and T alleles irrespective of fitness effects. Three recent studies have demonstrated that GC-biased gene conversion can strongly affect inferences based on the nonsynonymous *vs.* synonymous contrast, and must be seriously taken into account (Corcoran *et al.* 2017; Bolívar *et al.* 2018; Rousselle *et al.* 2019).

Using this method, one can estimate the distribution of the $S = 4N_e s$ 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 nonsynonymous mutations similar among species?; and (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 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.

Materials and Methods

Data sets

We used the coding sequence polymorphism data set assembled by Rousselle *et al.* (2020). 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.* (2020). Six to 20 individuals per species were genotyped at 531,360 to 14,112,150 coding positions from 1261 to 8604 genes (Supplemental Material, 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 (Lewontin 1974; 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*, and *Macaca mulatta* – and five species of fruit flies – *D. melanogaster*, *D. simulans*, *D. teissieri*, *D. yakuba*, and *D. santomea*.

In each species, the synonymous and nonsynonymous 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 $<n/2$ individuals were discarded. When genotypes were available in exactly $n/2$ individuals, the minor allele count was simply recorded. When genotypes were available in $>n/2$ individuals, hyper-geometric 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 $(n-k)^{\text{th}}$ categories, for every k. Unfolded SFS (Rousselle *et al.* 2020) 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 nonsynonymous SFS by only retaining A—T and G—C SNPs, following Rousselle *et al.* (2019). 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 <100 GC-conservative SNPs and were removed from the data set.

We also reanalyzed 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 (10 mammals, 1 bird, 2 fish), 9 insects (7 - *Anopheles* mosquitoes, 1 fruit fly, 1 butterfly) and 1 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 5 or above, *i.e.*, 6 species of vertebrates, 6 insects, 5 molluscs, 3 crustaceans, 3 echinoderms, 2 tunicates, 1 annelid, 1 cnidarian, and 1 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 nonsynonymous 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_{lth} of the nonsynonymous mutations are lethal, *i.e.*, cannot contribute any observable polymorphism, whereas a fraction $1 - p_{lth}$ 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 (https://github.com/BioPP/grapes). The amount of neutral polymorphism of each species was assessed using the π_S statistics, following Romiguier *et al.* (2014).

Table 1 Maximum likelihood estimates of $\bar{S}$ in primates and fruit flies under the Gamma + lethal DFE model

		β shared across species		one β per group	
	π_S (%)	P_{lth} = 0.96	P_{lth} = 0.65	P_{lth} = 0.6	P_{lth} = 0.65
Species					
H. sapiens	0.06	−3.36	−1.39	−2.75	−1.53
G. gorilla	0.11	−4.38	−2.08	−4.18	−2.22
P. troglodytes	0.11	−7.56	−3.30	−7.03	−3.45
P. abelii	0.16	−9.16	−4.64	−9.64	−4.59
M. mulatta	0.18	−20.0	−10.4	−24.5	−9.64
D. santomea	0.11	−401	−232	−395	−234
D. melanogaster	1.07	−502	−295	−493	−297
D. teissieri	1.52	−1860	−1080	−1820	−1090
D. yakuba	1.57	−946	−553	−928	−558
D. simulans	3.18	−1660	−970	−1630	−979
Max/min N_e	53.0	555	779	662	714
Bootstrap		[324–1050]	[438–1660]	[353–1430]	[422–1350]
Group median N_e ratio	13.8	125	168	132	162
Bootstrap		[84–203]	[102–296]	[87–237]	[103–281]

Simulations

Following Rousselle *et al.* (2018), we performed simulations using SLIM V2 (Haller and Messer 2016) to assess how quickly π_S and the estimated $\bar{S}$ (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 effective 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 2.2×10^{-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 nonsynonymous (if the fitness effect was different from zero), allowing us to compute π_N, π_S, and π_N/π_S at any time point throughout the simulation. To make simulations tractable, we used small effective 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 $4N_e\bar{s} = -0.025$, are relevant. Simulations were replicated 50 times.

Data availability

All the analyzed data sets are freely available from https://zenodo.org/record/3829893 (sequence alignments) and https://zenodo.org/record/3818299 (SFS). Supplemental material available at figshare: https://doi.org/10.25386/genetics.12833708.

Results

Synonymous genetic diversity

We first analyzed synonymous and nonsynonymous SFS in five species of primates and five species of fruit flies. The estimated synonymous diversity, π_S, was much higher in fruit flies than in primates (Table 1), consistent with the literature (Leffler *et al.* 2012). The estimated π_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 π_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 β, the shape parameter of the Gamma DFE: the maximum likelihood estimate of β varied from 0.09 to 0.36 among species (Figure 1, *y*-axis). When we instead fitted a model assuming a common β across species, the likelihood dropped severely (one β per species, log-likelihood = −1100.8; shared β, 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; $P < 10^{-20}$; 9 degrees of freedom). There was a trend for species from the same group to provide similar estimates of β (Figure 1). In primates, the range of estimated β was narrow ([0.09;0.16]) and the hypothesis of a common DFE shape was not rejected by LRT (optimal β for primates: 0.11; $P = 0.45$; 4 degrees of freedom). In fruit flies, the optimal β was close to 0.3 in four species, but much lower in *D. santomea* (Figure 1). The optimal fruit fly β was firmly rejected by primates as a group (LRT; $P < 10^{-20}$; 1 degree of freedom), and reciprocally (LRT; $P < 10^{-20}$; 1 degree of freedom; the test includes *D. santomea*).

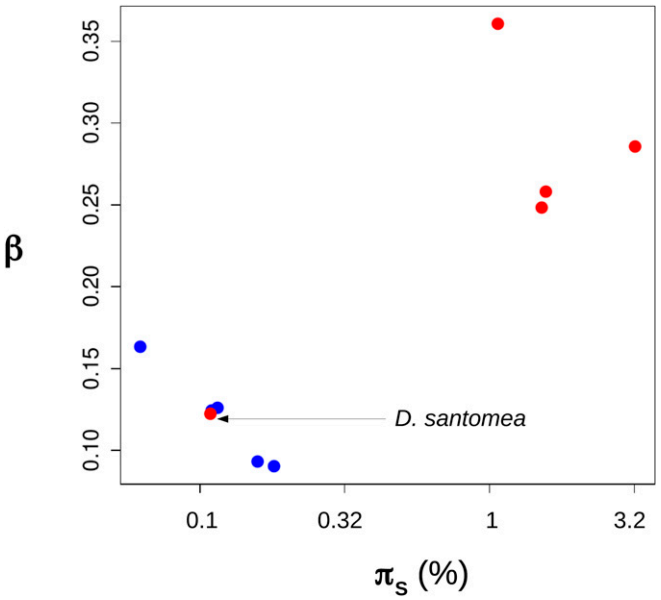

Figure 1 Estimated Gamma shape differs between primates and fruit flies. Blue dots: primates. Red dots: fruit flies. *x*-axis: synonymous genetic diversity (log_{10} scale). *y*-axis: estimated shape parameter assuming a Gamma DFE of nonsynonymous mutations.

Estimates of $\bar{S}$, the population-scaled mean selection coefficient, varied greatly among species in this analysis, but this was largely explained by variations in β. 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 β 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 β 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 nonsynonymous 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 β 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 of being observed at the 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_{lth} of lethal mutations, and a proportion $1 - p_{lth}$ of mutations of Gamma-distributed effects. Lethal mutations are here defined as mutations having a probability of being observed at polymorphic stage equal to zero. Figure 2 shows how the likelihood responds to p_{lth} and β in primates and fruit flies. In fruit flies, the optimal β was close to 0.3 irrespective of p_{lth}, and the likelihood was maximal when p_{lth} was close to 0.75. In primates, the optimal β varied more visibly with p_{lth}. It was close to 0.1 when p_{lth} was low, as indicated above, but increased toward higher values when p_{lth} increased. The maximal likelihood in primates was still obtained when p_{lth} was close to zero and β close to 0.1, but importantly, areas of the parameter space close to the fruit fly optimum (*e.g.*, p_{lth} 0.65 and β 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-β-per-species and shared-β models was substantially decreased under the Gamma + lethal model (191.0) compared to the Gamma model (281.9). Still, the one-β-per-species model significantly rejected the shared-β model in both cases, indicating that the inclusion of the p_{lth} parameter was not sufficient to erase every perceptible difference in DFE among species.

We jointly analyzed the 10 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_{lth} was in the range 0.6–0.65 (Figure S2). We considered these two values as plausible estimates of p_{lth}. When p_{lth} was set to 0.6, the maximum likelihood estimate of β 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_{lth} 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_{lth} be different from zero increased the log-likelihood by 85 units, which was highly significant ($P < 10^{-20}$, 1 degree of freedom). We also estimated $\bar{S}$ under the Gamma + lethal model assuming that the β 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

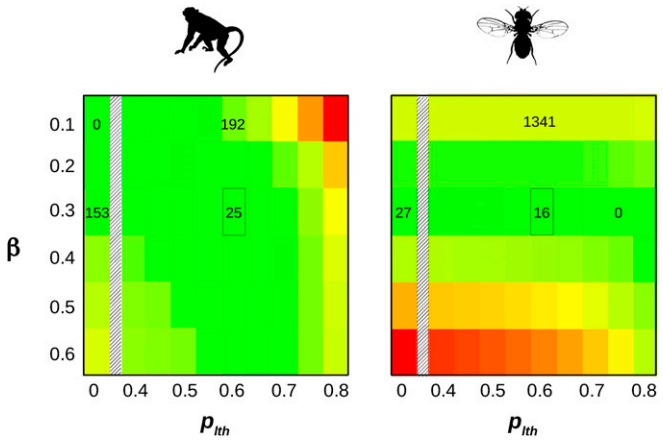

Figure 2 + Lethal model, likelihood surface. *x*-axis: proportion of lethal nonsynonymous mutations. *y*-axis: Gamma shape parameter for nonlethal, nonsynonymous mutations. The color scale is indicative of the log-likelihood: green = high, yellow = intermediate, red = low. The difference between local 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).

assuming shared β across all species (Table 1, right-most two columns), 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 β were such that the mean π_N/π_S 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 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," and "subsampled fruit flies" – we tried eight values of p_{lth}, from 0.4 to 0.75. We optimized β and the other parameters conditional on p_{lth} and recorded the likelihood. Table 2 shows the results for values of p_{lth} 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. One 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_{lth} 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_{lth}.

The GC-conservative data set included roughly 10 times less SNPs than the main one. The likelihood surface was flatter and close to its maximum at three values of p_{lth} – 0.5, 0.55, and 0.6 – while the estimated β 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 π_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 β was lower than in the main analysis, and the ratio 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 β 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 analyzed an additional 37 species from eight diverse groups of animals (Rousselle *et al.* 2020). In this 47-species data set, the estimated synonymous diversity π_S varied from 0.062% in *H. sapiens* to 4.4% in mussel *Mytilus trossulus*. The within-group median π_S varied from 0.12% 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 β (Figure S4). Butterflies and ants, for instance, tended to yield relatively high estimates of β, whereas the best fit in primates and birds was reached at relatively low β. 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_{lth} was 0.5 (with p_{lth} = 0.55 being close) and the estimated β was 0.26. We calculated the within-group median $\bar{S}$, which varied by a factor of 110 among groups (Table 3, first column). 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 ($r^2 > 0.97$ 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

Table 2 Robustness analysis, primates *vs.* fruit flies

Label	Analysis	SNP flies	SNP prim.	p_{lth}	β	Max/min $\bar{S}$	Median $\bar{S}$ ratio
a	Main	1.14×10^6	5.68×10^4	0.6	0.28	555	125
b_1	GC-conserv.	1.84×10^5	3.88×10^3	0.5	0.29	251	68
b_2	GC-conserv.	1.84×10^5	3.88×10^3	0.55	0.29	302	74
b_3	GC-conserv.	1.84×10^5	3.88×10^3	0.6	0.30	398	84
c	Refl. Gamma	1.14×10^6	5.68×10^4	0.6	0.30	341	88
d_1	Unfolded	5.9×10^6	5.68×10^4	0.55	0.23	1210	227
d_2	Unfolded	5.9×10^6	5.68×10^4	0.6	0.23	1480	274
e_1	Subsampled	5.68×10^4	5.68×10^4	0.55	0.23 [0.20–0.25]	1300 [752–3520]	240 [155–546]
e_2	Subsampled	5.68×10^4	5.68×104	0.6	0.25 [0.22–0.28]	900 [577–2020]	183 [128–356]

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 10 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: $r^2 = 0.27$, $P = 3.10^{-4}$; main analysis: $r^2 = 0.11$, $P = 3.10^{-2}$; log-transformed variables). Figure 3 shows the relationship between π_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: π_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_{lth} for this data set was 0.5, the estimated β was 0.29, and the ratio of maximal to minimal $\bar{S}$ was 581 – to be compared to a max/min ratio of π_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 to 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_{lth}: it varied by <2 log units over the 0–0.6 range for p_{lth}. The estimated β was close to 0.275 irrespective of p_{lth} (in the 0–0.6 range), and the ratio of maximal to minimal estimated $\bar{S}$ was 7300–7900, *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).

We conducted a goodness-of-fit analysis by comparing, for each species of the three data sets, the likelihood of the Gamma + lethal model to that of a saturated model, in which every class of the synonymous and nonsynonymous SFS has its own frequency parameter. Regarding the Rousselle *et al.* (2020) and Galtier (2016) data sets, we found that in a majority of species (37 out of 47 and 25 out of 28, respectively) the Gamma + lethal model was not rejected by an LRT (Table S2). The Chen *et al.* (2017) data set behaved a bit differently: the Gamma + lethal model was rejected in 12 species out of 23 for this data set. Please note that in these analyses the p_{lth} parameter was fixed to 0.6 in all species, while being considered as a free parameter of the Gamma + lethal model.

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 1–3), before moving to the interpretation and implications of our results (subsection 4–5).

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 nonsynonymous SFS. This model includes three parameters of interest: population mutation rate θ, average deleterious effect $\bar{S}$, and DFE shape β. In addition, the model has $n/2$ nuisance parameters, 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 nonsynonymous SFS. The r_i's are intended to capture any effect that similarly affects the fate of synonymous and nonsynonymous mutations – such as linked selection, population substructure, and departure from demographic equilibrium. In practice r_1 is set to 1, 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,

Table 3 Analysis of 47 species in 10 groups of animals

	Main	GC-cons.	Reflected	Unfolded	
Plth	0.5	0.55	0.5	0.45	
β	0.26	0.28	0.29	0.22	
Relative median $\bar{S}$:					Color:
Primates	1	1.6	1	1	Blue
Ants	1.6	1	1.6	1.6	Yellow
Butterflies	1.8	1.6	1.7	1.8	Orange
Passerines	3.1	4	2.9	3.1	Magenta
Earthworms	4.8	6.4	4.3	5.1	Cyan
Rodents	5.2	7.6	4.6	6.5	Gray
Fowls	21	7.8	17	25	Green
Ribbonworms	38	120	29	55	Black
Mussels	40	180	30	60	Brown
Fruit flies	110	130	75	200	Red

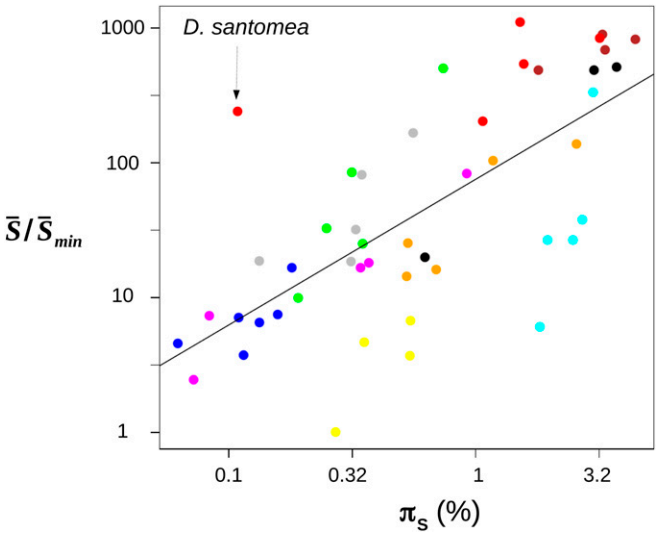

Figure 3 Variation in population-scaled mean selection coefficient *vs.* neutral genetic diversity across 47 species of animals. Each dot is for a species of animal. *x*-axis: relative synonymous heterozygosity. *y*-axis: relative estimated $\bar{S}$ Colors indicate groups (see Table 3). GC-conservative SNPs were used. Estimates of $\bar{S}$ were divided by its minimal value, which was obtained in ant *Formica pratensis*.

from −1297.7 to −71,404.7 (primates + fruit flies data set, Gamma + lethal, $p_{lth} = 0.6$). 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 and 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 (Galtier 2016; Tataru *et al.* 2017; Moutinho *et al.* 2019; Rousselle *et al.* 2020). The r_i-based approach is particularly appropriate when a relatively large number of species are analyzed, like here, when exploring many demographic scenarios for each species would be particularly costly.

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 θ and $\bar{S}$ (Kim *et al.* 2017). Consider for instance a data set yielding some estimate of θ, in which the estimated r_i's (i > 1) are all well below 1, meaning that singletons are in excess compared to the standard coalescent expectation. Such a pattern is expected for a population having experienced a recent, strong bottleneck. Now consider another data set yielding the same estimate of θ, but with estimated r_i's (i > 1) all well above 1, as expected under gradual population decline. 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 θ via the maximum likelihood method, but rather used π_S as our estimate of θ. We did, however, use the maximum likelihood estimate of $\bar{S}$. One should keep in mind that the meaning of these estimates is somehow dependent on the estimated r_i's, to an extent currently difficult to quantify.

To investigate this issue a bit deeper, we plotted the estimated r_i's for all SNP categories 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_{lth} 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. Although a more detailed analysis would be worthwhile, Figure S7 does not suggest to us that the r_i's pose a major problem of comparability in this analysis.

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 fruit flies, in contrast, we found that the value of β that best fitted the data were 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 β in *D. melanogaster*, taking special care of the problem of SNP mis-orientation. In butterflies, our median estimated β,

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 b 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 β.

The variation in estimated β we and Chen *et al.* (2017) report is 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 β is set to 0.1, as estimated in primates, then 53% of mutations are associated with a selection coefficient smaller than 0.001. If β was rather equal to 0.3, as in fruit flies, then this percentage would be 19%, and down to 12% if β=0.4, as in butterflies. Given the scarcity of experimental data on the nonsynonymous DFE in animals, one cannot firmly argue that such differences are implausible. We note, however, that if the estimates of β 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 β 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 β, 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. It is worth noting that this argument is based on a single species, though. Second, 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 β 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 distribution of the deleterious effect of nonsynonymous mutations among species.

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 nonsynonymous and a synonymous SFS implies accommodating 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 π_N/π_S ratio is relatively high, as in primates, there is a tendency for the Gamma model to converge toward low values of β, thus ensuring a large proportion of small effect nonsynonymous mutations, while a higher β could fit the difference in shape between the two spectra equally well, or maybe better. The additional p_{lth} parameter of the Gamma + lethal model, we suspect, somehow releases this constraint by controlling to a large extent the predicted π_N/π_S ratio.

This interpretation seems to fit reasonably well our analysis of the primates + fruit flies data set, and the report by Chen *et al.* (2017) of a lower estimate of β in small-N_e species. This rule, however, does not always apply. Our estimate of β 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 β 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_{lth} parameters did not change the picture much, compared to the Gamma model. This contrasts with our analysis and highlights the sensitiveness of this kind of analysis to the specificities of the data – number of genes, number of individuals, SNP calling procedure. Elyashiv *et al.* (2010) applied the Gamma + lethal model to yeast data and found that the shape parameters converged toward 0.35 under this model, which is very close to our joint estimate, while the Gamma model supported a higher value for β.

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 of size 10 to 20, as we are having here. This means that we lack information on the tail of the distribution we are trying to model. This also presumably explains why our p_{lth} estimate is substantially larger than existing estimates of the fraction of lethal mutations (Dickinson *et al.* 2016; Kim *et al.* 2017): we expect the parameter to capture not only lethal mutations, but also strongly deleterious mutations maintained at low frequency in the population.

Estimating N_e: mutation load vs. diversity

Our analysis of the load of deleterious mutations uncovered three [based on the Rousselle *et al.* (2020) and Chen *et al.* (2017) data sets] or four [based on the Galtier (2016)] 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 nonsynonymous to synonymous contrast, the genetic diversity of a species is influenced by the mutation rate. If the mutation rate was negatively correlated with the effective 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 μ, 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. 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 π_S during the recovery phase (Figure 4, generations 0–10,000). In these simulations, diversity-based estimates of N_e would be biased downward during a substantial period of time after the bottleneck, whereas the deleterious variation 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 (Kaiser and Charlesworth 2009; 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 deleterious variation 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 confirmation via specific theoretical developments, which ideally should also account for the effect of background selection.

The discrepancy between diversity-based and mutation load-based estimates of N_e was central in two recent studies of the DFE in animals. Huber *et al.* (2017) estimated the distribution of S for nonsynonymous mutations in *H. sapiens* and *D. melanogaster* using an approach similar to ours. Dividing by an estimate of N_e, they found that the distribution of s differed substantially between these two species, the average fitness effect of mutations being higher in *H. sapiens* than in *D. melanogaster*. This result seemingly contradicts our assumption of a constant DFE among species. It should be noted, however, that their estimate of N_e was obtained from a measure of the neutral genetic diversity, *i.e.*, was potentially affected by the caveats discussed above. The Huber *et al.* (2017) and this study actually agree in showing that estimates of $4N_es$ and $4Ne\mu$ do not scale proportionally across species. The two studies differ in their interpretation: Huber *et al.* (2017) invoked a difference in DFE, while we highlight linked selection and demographic fluctuations as potential causes of the discrepancy.

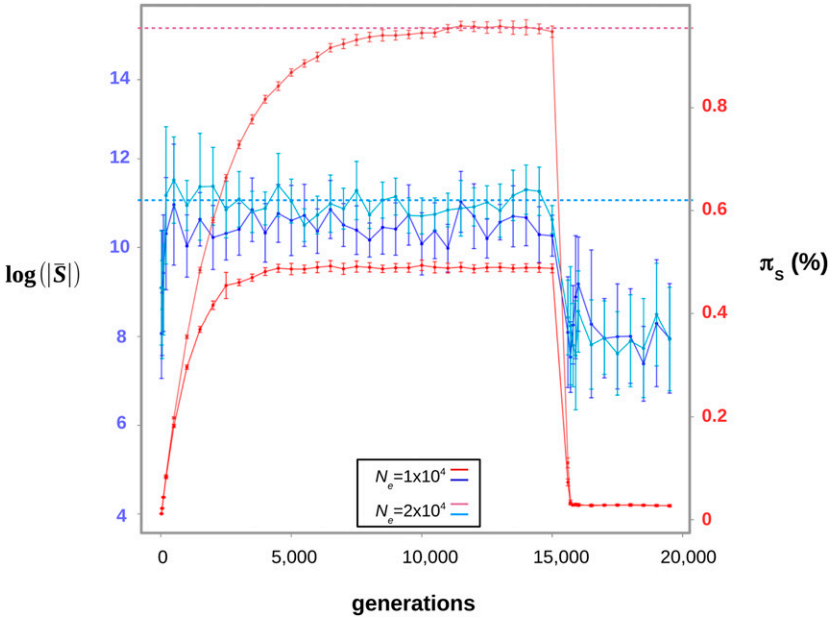

Figure 4 Evolution of the estimated $\bar{S}$ and π_S as N_e fluctuates. The simulation starts at time $t = 0$ in a population devoid of polymorphism. Two population sizes were used: N_e = 10,000 and 20,000. The estimated $\bar{S}$ converges toward its equilibrium value faster than π_S during the recovery phase. Then a bottleneck is simulated at t = 15,000 generations, where the population size drops to N_e = 500. The two statistics quickly reach their new equilibrium. Error bars represent the variance across 50 replicates. Simulated data sets in which the estimated shape parameter was <0.1 or >0.6 were discarded – they yield extreme estimates of $\bar{S}$. Error bars represent the variance across 50 replicates. Horizontal dotted lines represent the equilibrium values of π_S and $\bar{S}$ for N_e = 20,000.

Finally, Castellano *et al.* (2019) recently introduced a new hypothesis. These authors reported a wider range of variation in $\bar{S}$ than in π 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 could be 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 the variation in $\bar{S}$ would reflect a positive correlation between N_e and $\bar{s}$. This interesting hypothesis offers yet another potential explanation to the discrepancy between diversity-based and mutation load-based estimators of drift power.

Both Huber *et al.* (2017) and Castellano *et al.* (2019) challenge our assumption of a common DFE across species in invoking an effect of N_e on the average fitness effect of deleterious nonsynonymous mutations. They differ in their predictions, though, with Huber *et al.* (2017) suggesting a stronger average effect, and Castellano *et al.* (2019) a weaker average effect, in small-N_e species. How and how much the distribution of deleterious effects varies among species is as yet an unresolved issue, which clearly is key to the interpretation of the selected *vs.* neutral variation and the measurement of drift intensity.

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 toward reconciling genetic with ecological estimates of population size – but how big is this step?

On the 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 the effective 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 N_e generations becomes less and less plausible (Gillespie 2000). So maybe we should 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, and to Martin Lascoux and four anonymous reviewers

for constructive feedback and a positive recommendation. This work was supported by Agence Nationale de la Recherche grants no. ANR-15-CE12-0010 and ANR-19-CE12-0019-01.

Literature Cited

Bachtrog, D., K. Thornton, A. Clark, and P. Andolfatto, 2006 Extensive introgression of mitochondrial DNA relative to nuclear genes in the *Drosophila yakuba* species group. Evolution 60: 292–302. https://doi.org/10.1111/j.0014-3820.2006.tb01107.x

Barton, N. H., 2010 Genetic linkage and natural selection. Philos. Trans. R. Soc. Lond. B Biol. Sci. 365: 2559–2569. https://doi.org/10.1098/rstb.2010.0106

Bolívar, P., C. F. Mugal, M. Rossi, A. Nater, M. Wang *et al.*, 2018 Biased inference of selection due to GC-biased gene conversion and the rate of protein evolution in flycatchers when accounting for it. Mol. Biol. Evol. 35: 2475–2486. https://doi.org/10.1093/molbev/msy149

Böndel, K. B., S. A. Kraemer, T. Samuels, D. McClean, J. Lachapelle *et al.*, 2019 Inferring the distribution of fitness effects of spontaneous mutations in *Chlamydomonas reinhardtii*. PLoS Biol. 17: e3000192. https://doi.org/10.1371/journal.pbio.3000192

Boyko, A. R., S. H. Williamson, A. R. Indap, J. D. Degenhardt, R. D. Hernandez *et al.*, 2008 Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet. 4: e1000083. https://doi.org/10.1371/journal.pgen.1000083

Brandvain, Y., and S. I. Wright, 2016 The limits of natural selection in a nonequilibrium world. Trends Genet. 32: 201–210. https://doi.org/10.1016/j.tig.2016.01.004

Castellano, D., M. C. Macià, P. Tataru, T. Bataillon, and K. Munch, 2019 Comparison of the full distribution of fitness effects of new amino acid mutations across great apes. Genetics 213: 953–966. https://doi.org/10.1534/genetics.119.302494

Charlesworth, D., B. Charlesworth, and M. T. Morgan, 1995 The pattern of neutral molecular variation under the background selection model. Genetics 141: 1619–1632.

Chen, J., S. Glémin, and M. Lascoux, 2017 Genetic diversity and the efficacy of purifying selection across plant and animal species. Mol. Biol. Evol. 34: 1417–1428. https://doi.org/10.1093/molbev/msx088

Coop, G., 2016 Does linked selection explain the narrow range of genetic diversity across species? bioRxiv doi: 10.1101/042598 (Preprint posted March 7, 2016).https://doi.org/10.1101/042598

Coop, G., and P. Ralph, 2012 Patterns of neutral diversity under general models of selective sweeps. Genetics 192: 205–224. https://doi.org/10.1534/genetics.112.141861

Corbett-Detig, R. B., D. L. Hartl, and T. B. Sackton, 2015 Natural selection constrains neutral diversity across a wide range of species. PLoS Biol. 13: e1002112. https://doi.org/10.1371/journal.pbio.1002112

Corcoran, P., T. I. Gossmann, H. J. Barton, J. Slate, K. Zeng; Great Tit HapMap Consortium, 2017 Determinants of the efficacy of natural selection on coding and noncoding variability in two passerine species. Genome Biol. Evol. 9: 2987–3007 [corrigenda: Genome Biol. Evol. 10: 1062 (2018)]. https://doi.org/10.1093/gbe/evx213

Der, R., C. Epstein, and J. B. Plotkin, 2012 Dynamics of neutral and selected alleles when the offspring distribution is skewed. Genetics 191: 1331–1344. https://doi.org/10.1534/genetics.112.140038

Dickinson, M. E., A. M. Flenniken, X. Ji, L. Teboul, M. D. Wong *et al.*, 2016 High-throughput discovery of novel developmental phenotypes. Nature 537: 508–514. https://doi.org/10.1038/nature19356

Durrett, R., and J. Schweinsberg, 2005 A coalescent model for the effect of advantageous mutations on the genealogy of a population. Stochastic Process. Appl. 115: 1628–1657. https://doi.org/10.1016/j.spa.2005.04.009

Eldon, B., and J. Wakeley, 2006 Coalescent processes when the distribution of offspring number among individuals is highly skewed. Genetics 172: 2621–2633. https://doi.org/10.1534/genetics.105.052175

Ellegren, H., and N. Galtier, 2016 Determinants of genetic diversity. Nat. Rev. Genet. 17: 422–433. https://doi.org/10.1038/nrg.2016.58

Elyashiv, E., K. Bullaughey, S. Sattath, Y. Rinott, M. Przeworski *et al.*, 2010 Shifts in the intensity of purifying selection: an analysis of genome-wide polymorphism data from two closely related yeast species. Genome Res. 20: 1558–1573. https://doi.org/10.1101/gr.108993.110

Elyashiv, E., S. Sattath, T. T. Hu, A. Strutsovsky, G. McVicker *et al.*, 2016 A genomic map of the effects of linked selection in *Drosophila*. PLoS Genet. 12: e1006130. https://doi.org/10.1371/journal.pgen.1006130

Evans, S. N., Y. Shvets, and M. Slatkin, 2007 Non-equilibrium theory of the allele frequency spectrum. Theor. Popul. Biol. 71: 109–119. https://doi.org/10.1016/j.tpb.2006.06.005

Ewens, W. J., 1984 *Mathematical Population Genetics 1 Theoretical Introduction*. Springer, New York.

Eyre-Walker, A., 2002 Changing effective population size and the Mcdonald-Kreitman test. Genetics 162: 2017–2024.

Eyre-Walker, A., and P. D. Keightley, 2007 The distribution of fitness effects of new mutations. Nat. Rev. Genet. 8: 610–618. https://doi.org/10.1038/nrg2146

Eyre-Walker, A., and P. D. Keightley, 2009 Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. Mol. Biol. Evol. 26: 2097–2108. https://doi.org/10.1093/molbev/msp119

Eyre-Walker, A., M. Woolfit, and T. Phelps, 2006 The distribution of fitness effects of new deleterious amino acid mutations in humans. Genetics 173: 891–900. https://doi.org/10.1534/genetics.106.057570

Figuet, E., B. Nabholz, M. Bonneau, E. Mas Carrio, K. Nadachowska-Brzyska *et al.*, 2016 Life history traits, protein evolution, and the nearly neutral theory in amniotes. Mol. Biol. Evol. 33: 1517–1527. https://doi.org/10.1093/molbev/msw033

Filatov, D. A., 2019 Extreme Lewontin's paradox in ubiquitous marine phytoplankton species. Mol. Biol. Evol. 36: 4–14. https://doi.org/10.1093/molbev/msy195

Galtier, N., 2016 Adaptive protein evolution in animals and the effective population size hypothesis. PLoS Genet. 12: e1005774. https://doi.org/10.1371/journal.pgen.1005774

Gayral, P., J. Melo-Ferreira, S. Glémin, N. Bierne, M. Carneiro *et al.*, 2013 Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap. PLoS Genet. 9: e1003457. https://doi.org/10.1371/journal.pgen.1003457

Gillespie, J. H., 2000 The neutral theory in an infinite population. Gene 261: 11–18. https://doi.org/10.1016/S0378-1119(00)00485-6

Gillespie, J. H., 2001 Is the population size of a species relevant to its evolution? Evolution 55: 2161–2169. https://doi.org/10.1111/j.0014-3820.2001.tb00732.x

Good, B. H., A. M. Walczak, R. A. Neher, and M. M. Desai, 2014 Genetic diversity in the interference selection limit. PLoS Genet. 10: e1004222. https://doi.org/10.1371/journal.pgen.1004222

Gravel, S., 2016 When is selection effective? Genetics 203: 451–462. https://doi.org/10.1534/genetics.115.184630

Haller, B. C., and P. W. Messer, 2016 Slim 2: Flexible, interactive forward genetic simulations. Mol. Biol. Evol. 34: 230–240. https://doi.org/10.1093/molbev/msw211

Hartfield, M., and S. P. Otto, 2011 Recombination and hitchhiking of deleterious alleles. Evolution 65: 2421–2434. https://doi.org/10.1111/j.1558-5646.2011.01311.x

Hernandez, R. D., S. H. Williamson, L. Zhu, and C. D. Bustamante, 2007 Context-dependent mutation rates may cause spurious signatures of a fixation bias favoring higher GC-content in humans. Mol. Biol. Evol. 24: 2196–2202. https://doi.org/10.1093/molbev/msm149

Huber, C. D., B. Y. Kim, C. D. Marsden, and K. E. Lohmueller, 2017 Determining the factors driving selective effects of new nonsynonymous mutations. Proc. Natl. Acad. Sci. USA 114: 4465–4470. https://doi.org/10.1073/pnas.1619508114

Jensen, J. D., and D. Bachtrog, 2011 Characterizing the influence of effective population size on the rate of adaptation: Gillespie's Darwin domain. Genome Biol. Evol. 3: 687–701. https://doi.org/10.1093/gbe/evr063

Jensen, J. D., Y. Kim, V. B. DuMont, C. F. Aquadro, and C. D. Bustamante, 2005 Distinguishing between selective sweeps and demography using DNA polymorphism data. Genetics 170: 1401–1410. https://doi.org/10.1534/genetics.104.038224

Jónás, A., T. Taus, C. Kosiol, C. Schlötterer, and A. Futschik, 2016 Estimating the effective population size from temporal allele frequency changes in experimental evolution. Genetics 204: 723–735. https://doi.org/10.1534/genetics.116.191197

Kaiser, V. B., and B. Charlesworth, 2009 The effects of deleterious mutations on evolution in non-recombining genomes. Trends Genet. 25: 9–12. https://doi.org/10.1016/j.tig.2008.10.009

Karasov, T., P. W. Messer, and D. A. Petrov, 2010 Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. PLoS Genet. 6: e1000924. https://doi.org/10.1371/journal.pgen.1000924

Keightley, P. D., and A. Eyre-Walker, 2007 Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. Genetics 177: 2251–2261. https://doi.org/10.1534/genetics.107.080663

Keightley, P. D., J. L. Campos, T. R. Booker, and B. Charlesworth, 2016 Inferring the frequency spectrum of derived variants to quantify adaptive molecular evolution in protein-coding genes of *Drosophila melanogaster*. Genetics 203: 975–984. https://doi.org/10.1534/genetics.116.188102

Keightley, P. D., U. Trivedi, M. Thomson, F. Oliver, S. Kumar *et al.*, 2009 Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. Genome Res. 19: 1195–1201. https://doi.org/10.1101/gr.091231.109

Kim, B. Y., C. D. Huber, and K. E. Lohmueller, 2017 Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. Genetics 206: 345–361. https://doi.org/10.1534/genetics.116.197145

Klopfstein, S., M. Currat, and L. Excoffier, 2006 The fate of mutations surfing on the wave of a range expansion. Mol. Biol. Evol. 23: 482–490. https://doi.org/10.1093/molbev/msj057

Kong, A., M. L. Frigge, G. Masson, S. Besenbacher, P. Sulem *et al.*, 2012 Rate of de novo mutations and the importance of father's age to disease risk. Nature 488: 471–475. https://doi.org/10.1038/nature11396

Leffler, E. M., K. Bullaughey, D. R. Matute, W. K. Meyer, L. Ségurel *et al.*, 2012 Revisiting an old riddle: what determines genetic diversity levels within species? PLoS Biol. 10: e1001388. https://doi.org/10.1371/journal.pbio.1001388

Lewontin, R. C., 1974 *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York, London.

Loewe, L., B. Charlesworth, C. Bartolomé, and V. Nöel, 2006 Estimating selection on nonsynonymous mutations. Genetics 172: 1079–1092. https://doi.org/10.1534/genetics.105.047217

Lynch, M., L. M. Bobay, F. Catania, J. F. Gout, and M. Rho, 2011 The repatterning of eukaryotic genomes by random genetic drift. Annu. Rev. Genomics Hum. Genet. 12: 347–366. https://doi.org/10.1146/annurev-genom-082410-101412

Mackintosh, A., D. R. Laetsch, A. Hayward, B. Charlesworth, M. Waterfall *et al.*, 2019 The determinants of genetic diversity in butterflies. Nat. Commun. 10: 3466. https://doi.org/10.1038/s41467-019-11308-4

Messer, P. W., and D. A. Petrov, 2013 Frequent adaptation and the Mcdonald-Kreitman test. Proc. Natl. Acad. Sci. USA 110: 8615–8620. https://doi.org/10.1073/pnas.1220835110

Moutinho, A. F., F. F. Trancoso, and J. Y. Dutheil, 2019 The impact of protein architecture on adaptive evolution. Mol. Biol. Evol. 36: 2013–2028. https://doi.org/10.1093/molbev/msz134

Nené, N. R., A. S. Dunham, and C. J. R. Illingworth, 2018 Inferring fitness effects from time-resolved sequence data with a delay-deterministic model. Genetics 209: 255–264. https://doi.org/10.1534/genetics.118.300790

Ohta, T., 1972 Population size and rate of evolution. J. Mol. Evol. 1: 305–314. https://doi.org/10.1007/BF01653959

Peischl, S., I. Dupanloup, A. Foucal, M. Jomphe, V. Bruat *et al.*, 2018 Relaxed selection during a recent human expansion. Genetics 208: 763–777. https://doi.org/10.1534/genetics.117.300551

Pennings, P. S., S. Kryazhimskiy, and J. Wakeley, 2014 Loss and recovery of genetic diversity in adapting populations of HIV. PLoS Genet. 10: e1004000. https://doi.org/10.1371/journal.pgen.1004000

Piganeau, G., and A. Eyre-Walker, 2003 Estimating the distribution of fitness effects from DNA sequence data: implications for the molecular clock. Proc. Natl. Acad. Sci. USA 100: 10335–10340. https://doi.org/10.1073/pnas.1833064100

Romiguier, J., P. Gayral, M. Ballenghien, A. Bernard, V. Cahais *et al.*, 2014 Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature 515: 261–263. https://doi.org/10.1038/nature13685

Rousselle, M., M. Mollion, B. Nabholz, T. Bataillon, and N. Galtier, 2018 Overestimation of the adaptive substitution rate in fluctuating populations. Biol. Lett. 14: 20180055. https://doi.org/10.1098/rsbl.2018.0055

Rousselle, M., A. Laverré, E. Figuet, B. Nabholz, and N. Galtier, 2019 Influence of recombination and GC-biased gene conversion on the adaptive and nonadaptive substitution rate in mammals versus birds. Mol. Biol. Evol. 36: 458–471. https://doi.org/10.1093/molbev/msy243

Rousselle, M., P. Simion, M. K. Tilak, E. Figuet, B. Nabholz *et al.*, 2020 Is adaptation limited by mutation? A timescale-dependent effect of genetic diversity on the adaptive substitution rate in animals. PLoS Genet. 16: e1008668. https://doi.org/10.1371/journal.pgen.1008668

Ryman, N., L. Laikre, and O. Hössjer, 2019 Do estimates of contemporary effective population size tell us what we want to know? Mol. Ecol. 28: 1904–1918. https://doi.org/10.1111/mec.15027

Schneider, A., B. Charlesworth, A. Eyre-Walker, and P. D. Keightley, 2011 A method for inferring the rate of occurrence and fitness effects of advantageous mutations. Genetics 189: 1427–1437. https://doi.org/10.1534/genetics.111.131730

Sjödin, P., I. Kaj, S. Krone, M. Lascoux, and M. Nordborg, 2005 On the meaning and existence of an effective population size. Genetics 169: 1061–1070. https://doi.org/10.1534/genetics.104.026799

Sung, W., M. S. Ackerman, S. F. Miller, T. G. Doak, and M. Lynch, 2012 Drift-barrier hypothesis and mutation-rate evolution. Proc. Natl. Acad. Sci. USA 109: 18488–18492. https://doi.org/10.1073/pnas.1216223109

Tataru, P., M. Mollion, S. Glémin, and T. Bataillon, 2017 Inference of distribution of fitness effects and proportion of adaptive substitutions from polymorphism data. Genetics 207: 1103–1119. https://doi.org/10.1534/genetics.117.300323

Turissini, D. A., and D. R. Matute, 2017 Fine scale mapping of genomic introgressions within the *Drosophila yakuba* clade. PLoS Genet. 13: e1006971. https://doi.org/10.1371/journal.pgen.1006971

Uricchio, L. H., D. A. Petrov, and D. Enard, 2019 Exploiting selection at linked sites to infer the rate and strength of adaptation. Nat. Ecol. Evol. 3: 977–984. https://doi.org/10.1038/s41559-019-0890-6

Welch, J. J., A. Eyre-Walker, and D. Waxman, 2008 Divergence and polymorphism under the nearly neutral theory of molecular evolution. J. Mol. Evol. 67: 418–426. https://doi.org/10.1007/s00239-008-9146-9

Wright, S., 1938 Size of population and breeding structure in relation to evolution. Science 87: 430–431.

Communicating editor: S. Wright