Genetic Diversity and the Efficacy of Purifying Selection across Plant and Animal Species

Jun Chen, Sylvain Glémin, Martin Lascoux

Genetic Diversity and the Efficacy of Purifying Selection across Plant and Animal Species

Jun Chen,[1] Sylvain Glémin,[1,2] and Martin Lascoux*,[1]

[1]Department of Ecology & Genetics, Evolutionary Biology Centre, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

[2]Institut des Sciences de l'Evolution UMR 5554, Université Montpellier-CNRS-IRD-EPHE, Montpellier, France

***Corresponding author:** E-mail: martin.lascoux@ebc.uu.se.

Associate editor: John Parsch

Abstract

A central question in evolutionary biology is why some species have more genetic diversity than others and a no less important question is why selection efficacy varies among species. Although these questions have started to be tackled in animals, they have not been addressed to the same extent in plants. Here, we estimated nucleotide diversity at synonymous, π_S, and nonsynonymous sites, π_N, and a measure of the efficacy of selection, the ratio π_N/π_S, in 34 animal and 28 plant species using full genome data. We then evaluated the relationship of nucleotide diversity and selection efficacy with effective population size, the distribution of fitness effect and life history traits. In animals, our data confirm that longevity and propagule size are the variables that best explain the variation in π_S among species. In plants longevity also plays a major role as well as mating system. As predicted by the nearly neutral theory of molecular evolution, the log of π_N/π_S decreased linearly with the log of π_S but the slope was weaker in plants than in animals. This appears to be due to a higher mutation rate in long lived plants, and the difference disappears when π_S is rescaled by the mutation rate. Differences in the distribution of fitness effect of new mutations also contributed to variation in π_N/π_S among species.

Key words: **effective population size, distribution of fitness effects, purifying selection, life history traits, nearly neutral theory.**

Introduction

Genetic diversity within species depends on the number of mutations entering the population each generation and their effect on fitness, which determine what happens to them if, by any chance, they do not vanish right away. Under the neutral theory *sensu stricto* (Kimura 1983), most mutations are supposed to be either strongly deleterious or neutral so that the amount of genetic polymorphism primarily reflects a balance between random genetic drift and mutations: it is predicted to be proportional to the effective population size, N_e, a commonly used measure of the amount of genetic drift (Charlesworth 2009), and to the mutation rate towards neutral variants, which depends on the level of constraint on the genomic elements under consideration. Despite its conceptual utility, the neutral theory presents several limitations (Gillespie 2004). The nearly neutral theory was developed to address some of the shortcomings of the neutral theory by adopting a continuous distribution of fitness effect of mutations: strongly deleterious or neutral mutations are still common and advantageous ones are still rare but there is now a large class of mildly deleterious mutations (Akashi et al. 2012). Purifying selection against these mutations is weak enough for some of them to be able to establish themselves in the population and contribute to polymorphism, if selection is of the same order or lower than genetic drift. Under the nearly neutral theory, N_e thus plays a central role as it determines not only the level of polymorphism but also the global efficacy of selection, the rate of evolution (Lanfear et al. 2014) and potentially genome size and organization (Lynch 2007).

Within the neutral theory framework, the question of the determinants of genetic polymorphism across species has remained a long-standing puzzle (Leffler et al. 2012; Ellegren and Galtier 2016). Assuming that the variation in mutation rates is limited among species, searching for determinants of polymorphism comes down to explaining variation in N_e. Recently, a large comparative study showed that polymorphism among 76 animal species was better explained by propagule size (defined as the size of the organism—juveniles, eggs, or larva—at time of dispersal) rather than by geographical variables such as species distribution range (Romiguier et al. 2014). It suggests that life history traits (LHTs) linked to a "*r*/*K*" gradient could determine N_e by affecting how species respond to environmental perturbations: long-lived species with strong parental investment (roughly "*K*-strategists") could bear lower density population size, hence have lower long-term N_e than short-lived species with low parental investment (roughly "*r*-strategists"). It has also been proposed that linked selection constrains levels of diversity (Corbett-Detig et al. 2015), but likely not enough to explain the narrow range of nucleotide diversity observed across species, except for highly selfing organisms (Coop 2016). So far, no equivalent studies have been conducted in plants, and previous results rely on allozyme data (Hamrick and Godt 1996) or very few

Article

genes per species (Leffler et al. 2012; Glémin et al. 2006). Plants differ from animals in major ways that could have a significant impact on nucleotide diversity. For example, in plants, germen and soma are not separated and some tree species can also live much longer than most animal species. This could affect the mutation rate (Moorjani et al. 2016) and thereby alter the direct relationship between N_e and polymorphism. Also, LHTs are not necessarily comparable between plants and animals and, for instance, plants present a greater diversity of mating systems than animals, so that mating system could be a stronger determinant of genetic diversity than other traits in plants (see Hamrick and Godt 1996; Glémin et al. 2006).

A related question is how N_e, hence possibly some LHTs, affect the efficacy of selection at the genomic levels. Does it conform to the nearly neutral theory with higher accumulation of weakly deleterious mutations in species with low N_e? This prediction has been tested in several groups of animal species using the ratio of nonsynonymous (D_N) versus synonymous (D_S) substitution rates. Generally, the LHTs used as a proxy of N_e correlate well with the D_N/D_S ratio, in agreement with the nearly neutral theory (Popadin et al. 2007; Lartillot and Delsuc 2012; Lourenço et al. 2012; Figuet et al. 2016), birds being a notable exception (Figuet et al. 2016). The ratio of nonsynonymous versus synonymous nucleotide diversity (π_N/π_S) or, equivalently, the ratio of 0-fold and 4-fold degenerate positions in protein-coding sequences, π_0/π_4, are other common measures of selection efficacy within species (Harrang et al. 2013; Lohmueller 2014; Do et al. 2015; Henn et al. 2015).

π_0/π_4 depends on N_e and the distribution of fitness effect (DFE), $\phi(s)$, both being often combined in the distribution of the scaled fitness effect, $\Phi(S)$ where $S = 4N_e s$. In most studies, the test of the relationship between N_e and the efficacy of purifying selection thus implicitly assumes that $\phi(s)$ is constant and that changes in $\Phi(S)$ are fully explained by changes in N_e. The DFE is often assumed to be gamma-distributed with the shape parameter (β hereafter) determining the skewness of the distribution: the lower the β value the more skewed the distribution, with many weakly deleterious and only a few very strongly deleterious mutations. Interestingly, under these assumptions, the nearly neutral model predicts a simple relationship between π_N/π_S and N_e: the log of π_N/π_S is linearly related to the log of N_e, the slope being the shape parameter of the DFE (and similarly for D_N/D_S) (Kimura 1979; Welch et al. 2008). Consequently, a log–log linear relationship between π_S and π_N/π_S is also predicted and was observed in animals (Romiguier et al. 2014). However, shift in selective pressures may occur as found for some genes in two closely related yeast species (Elyashiv et al. 2010) and it is possible that the whole *unscaled* DFE, $\phi(s)$, also varies across species. For example, predictions based on the Fisher Geometric Model suggest that the DFE should depend on species characteristics such as organism "complexity," level of pleiotropy of mutations or population size (Martin and Lenormand 2006; Lourenço et al. 2011; Tenaillon 2014). It is thus not granted that a general relationship between N_e and measures of the efficacy of purifying selection holds for any group and at any taxonomic level. For instance, we do not know whether plants and animals share the same relationship between π_S and π_N/π_S.

In any case, the nearly neutral theory provides us with an ideal conceptual framework linking measures of the efficacy of selection to two crucial explanatory variables, the effective population size and the distribution of fitness effects, both of which in turn are related to the biology of the species, and in particular to its life history traits (Figure S1 in supplentary SI text, Supplementary Material online). Under this framework we asked the following questions. How does genetic diversity vary across plant and animal species? Which LHTs (if any) determine genetic diversity in plants and do ecological strategies play a central role in plants as proposed for animals (Romiguier et al. 2014)? How much does the efficacy of purifying selection vary and can it be explained by variation in N_e alone or does variation in DFE also play a role? How are these relationships affected by differences in LHTs?

To address these questions, we estimated π_0/π_4 in 34 animal and 28 plant species and assessed its relationship with neutral nucleotide diversity, π_4, N_e, the DFE, and life history traits. Overall, our study lends further support to the nearly neutral theory as π_0/π_4 is log–log linearly related to π_4. However, variation in N_e alone is not sufficient to explain all variations and the relationship depends on whether one considers animals or plants, in agreement with substantial variation in the shape of the DFE and mutation rates. It also highlights the important impact of life history traits on nucleotide diversity and selection efficacy and the relatively weak effect of population history on those parameters.

Results

Data Set

We gathered SNP data in 62 species whose genome has been recently resequenced at high coverage with the Illumina HiSeq platform. These included 34 animal species composed of 17 mammals, 12 insects, three fishes, one bird, and one nematode, and 28 plant species composed of 14 herbaceous and 14 woody/shrubby plants. Ten plant species were selfers and 18 were outcrossers. We also sampled domesticated counterparts of ten wild species, namely maize, sorghum, wheat, rice, cucumber, soybean, watermelon, tomato, cassava, and dog (see supplementary tables S1 and S2 in SI text, Supplementary Material online). Additionally, in humans, *D. melanogaster*, *A. thaliana*, *P. trichocarpa*, and a few other species, we also analyzed separately several populations. In total 145 species/populations were investigated in this study with sample size ranging from 2 to 50 chromosomes. After data were curated (see Materials and Methods), the number of analyzed genes per species/populations ranged from 1,280 to 23,189 (see supplementary S1 file, Supplementary Material online). For each species/population, we computed the diversity at 0-fold and 4-fold positions, π_0 and π_4, respectively. In our data set plants have higher neutral genetic diversity π_4 than animals (median $= 5.4 \times 10^{-3}$ and 2.5×10^{-3}, Kolmogorov–Smirnov test P value$= 5.6 \times 10^{-3}$, see also fig. 1A) but span a lower range (from 6.7×10^{-4} to 1.8×10^{-2} in plants vs. 6.28×10^{-4} to 3.68×10^{-2} in animals).

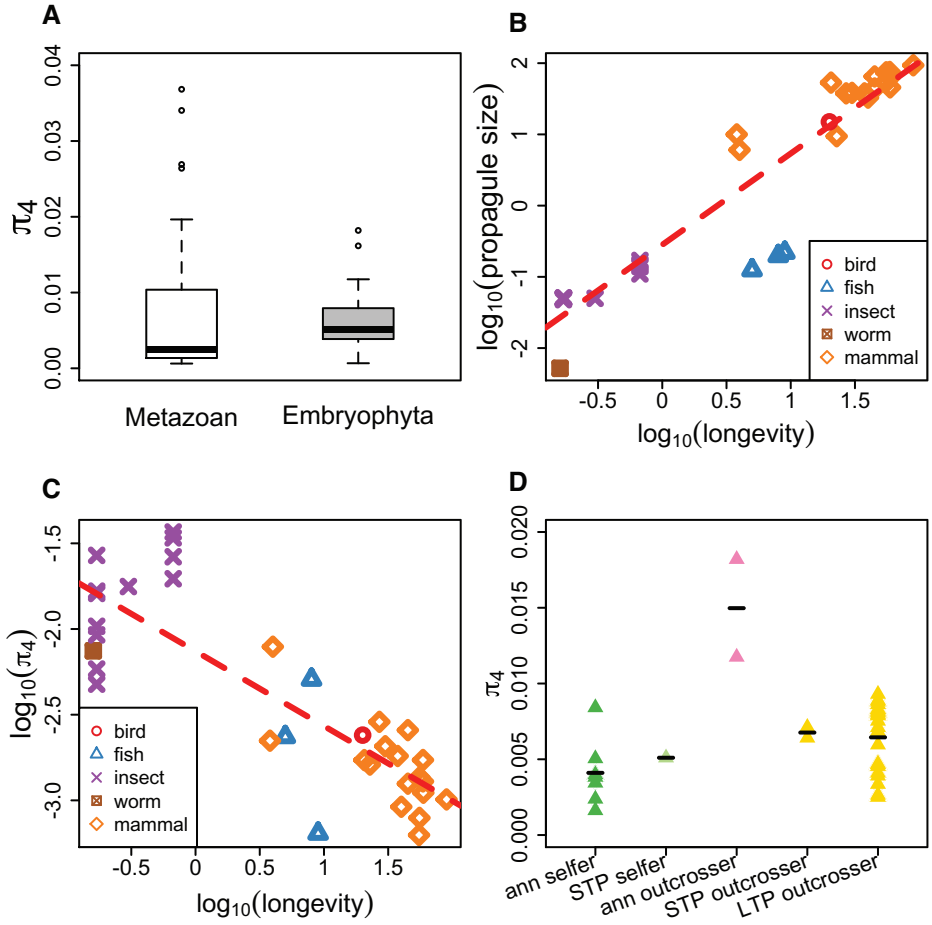

FIG. 1. The correlation of genetic diversity π_4 with life history traits. (*A*) π_4 difference between animals and plants. (*B*) Correlation of animal propagule size with longevity. (*C*) Correlation with longevity in animals. (*D*) Difference in mating system and lifespan in plants. STP, short-term perennial; LTP, long-term perennial.

The nucleotide diversities observed in animals in this study are of the same order of magnitude but with slightly lower values than those observed by Romiguier et al. (2014). To assess the quality of our results, we examined the potential bias of sample size, number of genes, read depth, genotype quality, and SNP calling software and no significant effect was identified (see supplementary SI text, Supplementary Material online).

Similar LHTs Predict Genetic Diversity in Plants and Animals

Romiguier et al. (2014) showed that life history traits are strong predictors of species genetic diversity, especially propagule size and longevity, both being strongly correlated and capturing the "*r*/*K*" gradient. We also observed a strong correlation between longevity and propagule size (Pearson's correlation coef. = 0.95, *P* value = $2.2e-16$, fig. 1*B*) but a covariance analysis showed that only longevity is significantly correlated with π_4 (*P* value = 8.8×10^{-4}, adjusted $R^2 = 0.7$, fig. 1*C*): short-lived insects have the highest genetic diversity and long-lived mammals the smallest whereas nematode, fishes, and bird have intermediate values.

In plants, following previous studies (Glémin et al. 2006; Hamrick and Godt 1996), we tested for the effect of a series of LHTs on genetic diversity, namely mating system, longevity, plant height, pollen and seed dispersal, and seed mass. Among those, mating system is the best predictor of genetic diversity (*P* value = 1.78×10^{-3}), whereas for outcrossing species lifespan is another important predictor (*P* value = 5.8×10^{-3}, fig. 1*D*). On average, genetic diversity is twice as high in outcrossing plants than in selfing ones ($\pi_4 = 7.4 \times 10^{-3}$ and 3.8×10^{-3}). Additionally, for outcrossing plant species the longer they live the lower their genetic diversity: π_4 decreases from 0.015 in annual outcrossers to 0.01 in short-term perennial (STP) outcrossers, and to 6×10^{-3} in long-term perennial (LTP) outcrossers (fig. 1*D*). No conclusion could be drawn on the influence of lifespan for selfing species due to lack of data. Other LHTs have no significant effect (see supplementary SI text, Supplementary Material online), in particular seed mass, which can be compared with propagule size in animals. However, it is worth noting that in plants, longevity and seed mass are not correlated (ANOVA F = 0.67, *P* value = 0.52) in contrast to what is observed in animals. In animals, the effect of mating system could not be tested because only one species, *Caenorhabditis briggsae*, is a selfer. To correct for a potential bias caused by overrepresentation of some species groups and for phylogeny we took the average values of species within the same genus. To ascertain that

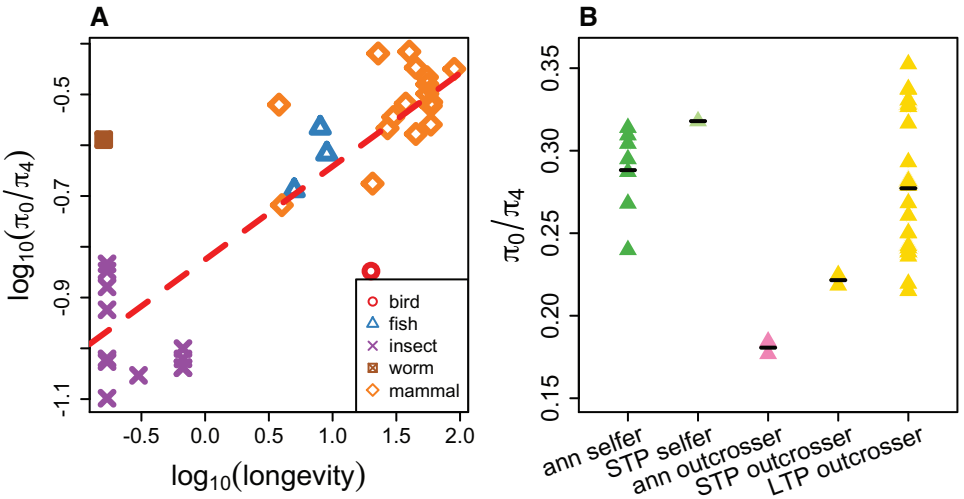

FIG. 2. The correlation of π_0/π_4 with life history traits. (*A*) For animals π_0/π_4 is positively correlated with longevity. (*B*) For plants π_0/π_4 differs between mating system and lifespan. STP, short-term perennial; LTP, long-term perennial.

phylogenetic relationships among species did not affect the results, we further used Felsenstein's phylogenetic independent contrast analysis (Felsenstein 1985) (see *SI Text*). In both cases, we recovered the same trends in genetic diversity. This confirms that the variation of genetic diversity among species reflects differences in life-history traits, mainly longevity and mating systems, as suggested earlier on by Romiguier et al. (2014).

The Relationship between π_0/π_4 and π_4 Differs between Plants and Animals

According to the nearly neutral theory, N_e controls the efficacy of purifying selection so that π_0/π_4 is directly and negatively related to N_e. We thus predict that (i) the LHTs affecting π_4 should also affect π_0/π_4 and that (ii) π_0/π_4 and π_4 should be negatively correlated. In agreement with the first prediction, longevity has the strongest impact on π_0/π_4 (slope = 0.18, P value = 7.5×10^{-10}, $R^2 = 0.69$), as for π_4 (fig. 2*A*). Similarly, in plants, both longevity and mating system affected π_0/π_4 significantly (P value = 0.019 and 1.1×10^{-4}, $R^2 = 0.51$). More specifically, selfers tend to have, on average, higher π_0/π_4 than outcrossers for species with similar lifespans whereas, for the same mating system, the longer a plant species lives the larger its π_0/π_4 is (fig. 2*B*).

As shown on figure 3, the second prediction is also globally verified: π_0/π_4 significantly decreases with π_4. However, the global pattern differs between plants and animals. First, although plants have a higher π_4 on average, they tended to also have higher π_0/π_4 ratios (median = 0.29) than animals (0.25, Kolmogorov–Smirnov test, P value = 0.015), whereas the reverse would be expected. Second, π_0/π_4 and π_4 were significantly and negatively correlated (P value $< 2 \times 10^{-16}$, $R^2 = 0.77$, fig. 3) in ANCOVA analysis, but the slope of animals and plants differed dramatically (slope = −0.38 vs. −0.1, P value = 7.35×10^{-5}). A separate analysis on plants also showed that, compared with animals, the correlation between π_0/π_4 and π_4 was much weaker but still significant if mating system and longevity were taken into account (P value = 0.02, 0.011, and 0.0018 for π_4, longevity, and mating system, respectively, $R^2 = 0.49$).

There are two nonexclusive explanations for this result. First, instead of being a direct estimate of effective population size, π_4 is actually the product of N_e and the mutation rate per generation, μ. Mutation rate variation among species could alter the relationship. A universal relationship between π_0/π_4 and π_4 only holds if mutation rates vary weakly among species. Second, a universal relationship would also imply a roughly constant DFE among species.

To remove the effect of μ, we rescaled the results by the nucleotide mutation rate (per site and per generation). This rescaling has a limited effect on the relative species position, except for several tree species (*Populus*, *Pinus*, and *Amborella*) that have the highest mutation rates per generation. After rescaling they shifted left of the herbaceous species on the N_e scale (left-bottom panel of fig. 3; see also supplementary file S3, Supplementary Material online, for an enlarged version). This result also indicates that the effect of longevity on N_e is stronger than predicted by π_4, especially for plants.

After rescaling, the negative correlation between π_0/π_4 and N_e is still significant (P value = 2×10^{-16}, $R^2 = 0.77$) but animals and plants now share the same regression line of π_0/π_4 and N_e if we only consider outcrossers (P value = 0.08 and 0.13 for difference of slope and intercept, respectively). All these results could be recovered after taking average values per genus or using phylogenetic independent contrasts (see supplementary SI text, Supplementary Material online). Thus, differences in mutation rates should be regarded as one of the main contributors to the difference in the $\pi_0/\pi_4 - \pi_4$ correlation between plants and animals.

The Characteristics of the DFE among Species

Although differences in mutation rates could be the main cause of the observed pattern on figure 3, variation in the distribution of fitness effect (DFE) could still play an important role. Moreover, how the DFE varies among species and whether it is affected by LHTs is still poorly known.

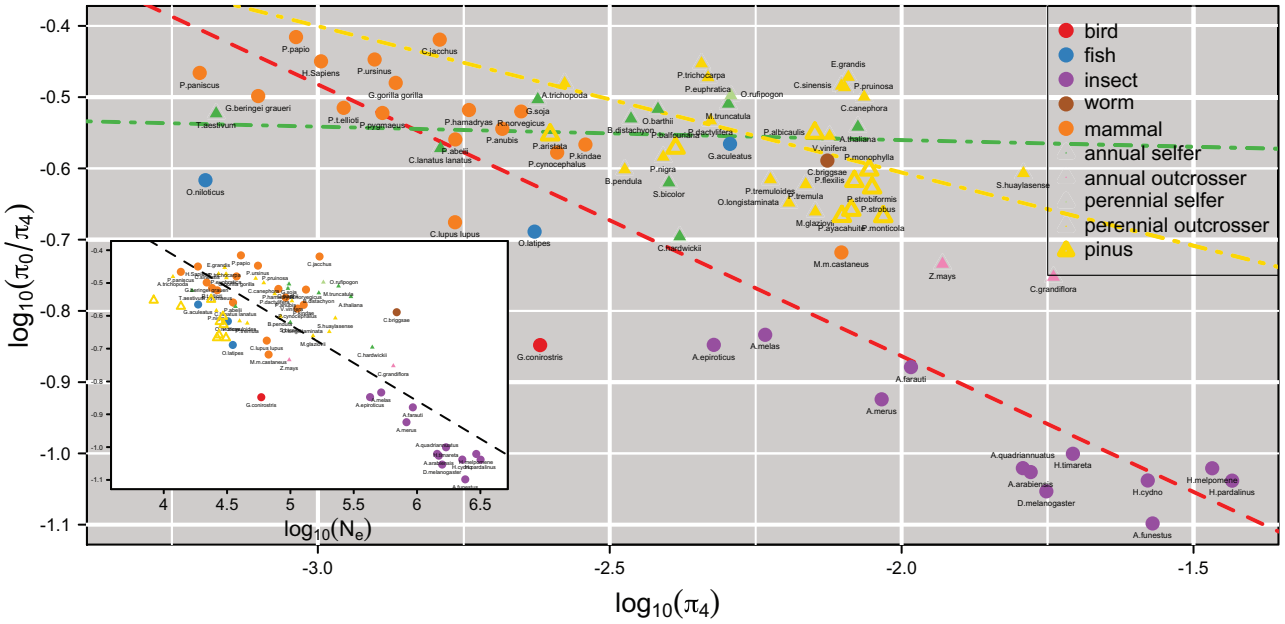

FIG. 3. The correlation of π_0/π_4 and π_4. The circles represent animal species and triangles plants species. Colors are used to distinguish further grouping. The yellow dashed line shows the regression of π_0/π_4 over π_4 for woody/shrubby plants and the green line is for herbaceous plants. The red line shows the regression of π_0/π_4 over π_4 for animal species. All domesticated/cultivated species were excluded from the regression analysis and the calculation of the correlation. The left-bottom panel shows the regression of π_0/π_4 over N_e. A larger version of the inlay is available in Supplementary File S3. Pinus data were taken from (Eckert et al. 2013) and were not used in the regression analysis.

We estimated DFE of nonsynonymous mutations using the model developed by Eyre-Walker et al. (2006) and Welch et al. (2008). We fitted a gamma distribution of deleterious mutations to the folded allele frequency spectrum of 4-folded and 0-folded sites. This model neglects beneficial mutations, and this can slightly bias the estimate of the shape of the DFE (Tataru et al. 2016). Unfolded frequency spectra are necessary to correctly fit a DFE with both deleterious and beneficial mutations, but outgroups to polarize mutations were not available in most data sets. The DFE for each species was summarized by either the shape parameter (β) and mean of a gamma distribution ($N_e\bar{s}$) or the proportion of mutations under four categories of selection strength: effectively neutral mutations, SS1 ($0 < N_es \leq 1$), mildly deleterious mutations, SS2 ($1 < N_es \leq 10$), deleterious mutations, SS3 ($10 < N_es \leq 100$), and highly deleterious mutations, SS4 ($100 < N_es$).

Under the nearly neutral theory, β is often assumed to be small (Welch et al. 2008; Kimura 1979) and this seems to hold in nature: indeed the DFE is generally characterized by a strongly L-shaped distribution (Eyre-Walker and Keightley 2007). The present study confirms this but also shows that there are major differences between groups of species: animals tend, on average, to have higher β values than plants (median = 0.13 vs. 0.082, P value = 0.041). In animals, β decreased slowly but significantly with the increase of longevity (slope = −0.066, P value = 6.2×10^{-4}, see supplementary fig. S2A, Supplementary Material online). Insects and nematodes have much higher β values ($0.24 \sim 0.27$) than mammals, birds, and fishes ($0.1 \sim 0.13$). Primates had slightly smaller β values (0.09) than other mammals. In plants, annual outcrossers had higher β values than other groups (0.22 vs. $0.74 \sim 0.11$, P value > 0.36, see supplementary fig. S2B, Supplementary Material online). β did not vary much among closely related species within the same genus (including wild and domesticated species), with a mean standard deviation of 0.04 (fig. 4A). β was even less variable among populations within a species (s.d. = 0.032, fig. 4B). An analysis of variance showed that the variance among genus and higher orders explained about 54% of the total variance in β, variance among species explained around 31% and variance among populations explained the rest, 15%. The *unscaled* mean s effect of the DFE, s_{mean}, also varies among species. In theory, we could obtain a raw estimation as $s_{mean} = \frac{\mu}{\pi_4} S_{mean}$. However, s_{mean} can take very high values (varying on a double exponential scale), is estimated with a very large variance and is highly correlated with β (Spearman's coef. = −0.95, P value $< 2 \times 10^{-16}$, see supplementary fig. S3, Supplementary Material online), which is estimated much more precisely. As a result, we did not find any clear association between s_{mean} and species groups.

We then tested whether the parameters of the DFE can also predict π_0/π_4 in addition to N_e. Despite the problems mentioned above we used both β and s_{mean} but the potential limits must be kept in mind. We fitted the following model to the combined data set of animal and plant species:

$$\log(\pi_0/\pi_4) \sim intercept + \log(N_e) + \log(\beta) + \log(\log(s_{mean})) + err$$

The three variables had a significant and negative effect on π_0/π_4 as predicted by theory (P values $< 2.9 \times 10^{-7}$, $R^2 = 0.82$). Significance remained after taking the average of each genus (P values < 0.01, $R^2 = 0.53$, see

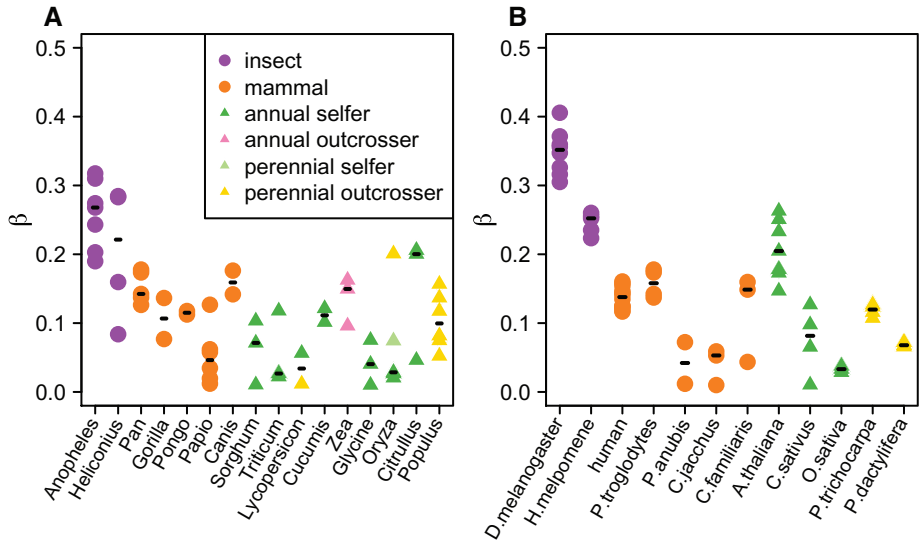

FIG. 4. The distribution of the DFE shape parameter β among (*A*) species of the 16 genera and (*B*) populations of the 13 species for which population level data were available.

supplementary SI text, Supplementary Material online) colorredor by using independent contrasts (P values < 0.05, $R^2 = 0.84$, see supplementary SI text, Supplementary Material online). Overall, the results are thus in agreement with the nearly neutral theory but suggest that, in addition to variation in N_e, the characteristics of the DFE must be taken into account to properly explain the variation in π_0/π_4 among species.

Effect of Population Structure and Demography

As noted in the introduction, Gravel and others (Gravel 2016; Brandvain and Wright 2016) have pointed out that it can be misleading to use π_N/π_S (or equivalently π_0/π_4) as a measure of the selection efficiency when comparing populations that have not reached equilibrium for allele frequency spectrum: π_N will reach equilibrium faster than π_S and populations that are far from equilibrium will tend to exhibit larger values of π_N/π_S.

To test the robustness of our results, we therefore used the proportion of effectively neutral mutations (category SS1) instead of π_0/π_4, since it is a more direct measure of the efficacy of selection against weakly deleterious mutations and its estimate is partly corrected for demography and populations structure (see Materials and Methods). SS1 was significantly correlated to π_0/π_4 (Spearman's coef.=0.92, P value $< 2.2 \times 10^{-16}$) and the conclusions were similar to those reached for π_0/π_4 (see supplementary figs. S4–S6, Supplementary Material online). For the species for which different populations were available we also assessed how π_0/π_4 varied among populations. Although π_4 can vary substantially, for example under the effect of domestication, within species variation in π_0/π_4 and in the whole DFE are small. This within species variation is much weaker than the variation among species. A variance component analysis also suggested that population differences only explained $\sim 0.8\%$ of the variance in π_0/π_4. For populations of humans, *D. melanogaster*, and *A. thaliana*, the π_0/π_4 ratio was negatively correlated to genetic diversity (slope $= -0.13$, -0.13, and -0.15, P value $= 6.3 \times 10^{-7}$, 0.03, and 3.8×10^{-6}, respectively).

The DFE between Africans and "out-of-Africa" (OOA) human populations differed significantly (P values < 0.02) though there was only a small shift in SS1 values (mean- $= 0.259$ vs. 0.267), the main difference being actually between larger selection coefficient categories (see supplementary fig. S7*A*, Supplementary Material online). Although African and OOA populations of *Drosophila* did not significantly differ for any of the four categories (see supplementary fig. S7*B*, Supplementary Material online), SS1 was dramatically smaller than the SS1 values of human or other animal species. In plants, we investigated populations of *Populus trichocarpa* and *A. thaliana*. *P. trichocarpa* populations shared the same value of SS1 = 0.31 which might reflect the weak population structure in this species (see supplementary fig. S7*C*, Supplementary Material online). On the other hand, SS1 in populations of *A. thaliana* followed a pattern similar to "isolation-by-distance," though with a very limited range of values from 0.19 to 0.25 (see supplementary fig. S7*D*, Supplementary Material online). For domesticated species, which went through recent and severe bottleneck(s), the SS1 category was generally slightly smaller in wild than in domesticated species (median = 0.24 vs. 0.26, see supplementary fig. S8, Supplementary Material online) and non significant (P value = 0.23). The difference in the three other categories was more pronounced but the pattern across species was not consistent. Overall, population structure and demography have a limited impact compared with among species differences.

Genome Size and the Efficacy and Selection

In agreement with the nearly neutral theory, the results presented above suggest that species with small N_e are less efficient to remove weakly deleterious mutations in coding sequences because of the reduced efficacy of purifying

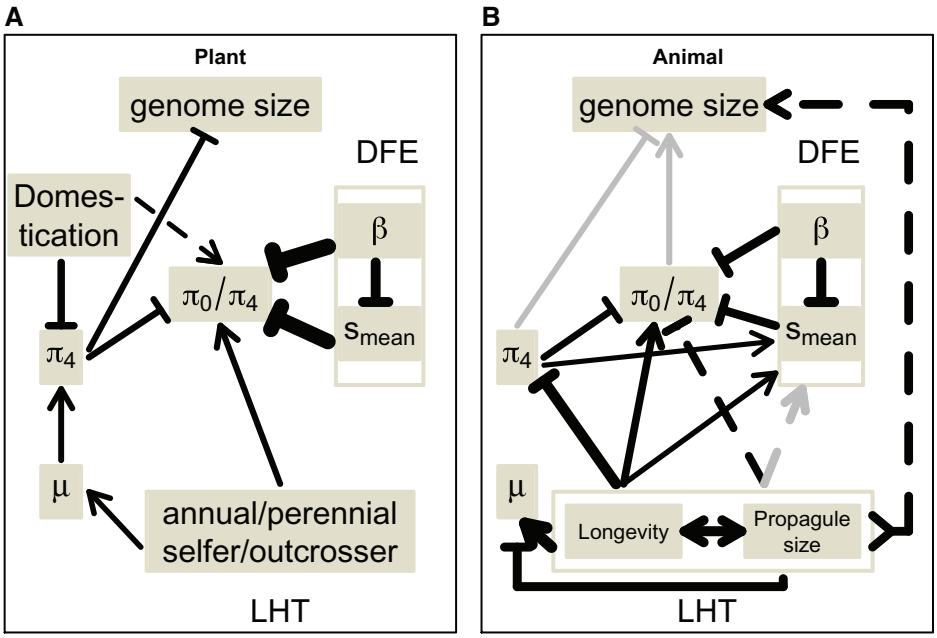

FIG. 5. The path analysis of predictors of π_0/π_4. (*A*) Plants and (*B*) animals. Direct effects are indicated with solid lines and indirect effects are represented by dashed lines. A positive effect is represented with an arrow and a negative ends with a bar. The width of the lines corresponds to the standardized coefficient. Black lines indicate significant correlation with *P* value $\leq$ 0.05 and gray lines show *P* value $\leq$ 0.1.

selection. Lynch (Lynch 2007) extended systematically this idea to all genome components and proposed that N_e is a main driver of genome size evolution. Genome size is correlated with N_e in eukaryote species (Lynch 2007) but this result has been criticized (Whitney et al. 2011) and was not necessarily retrieved at all taxonomic scales (Whitney et al. 2010). Using the current data set and genome sizes (see supplementary tables S1 and S2 in SI text, Supplementary Material online), we tested Lynch's hypothesis. Genome size was significantly correlated to N_e (*P* value $= 1.2 \times 10^{-10}$, $R^2 = 0.52$) but plants exhibited a shallower slope than animals (−0.38 vs. −1.1, *P* value = 0.01).

Disentangling Causal Relationships between LHT and Genomic Diversity

Above we showed that some LHTs affect both π_4 and π_0/π_4 and that demography only has a weak effect. Although LHTs can affect π_0/π_4 through N_e, the above results indicate that it could also affect it through the parameters of the DFE. To disentangle the causal relationships among these variables, we performed path analyses on LHTs, β, s_{mean}, *Ne*, domestication, and genome size. We included all wild species and their domesticated counterparts in this analysis. Although we could also use subpopulations for nondomesticated species as above, there is no *a priori* prediction for the direction of demographic changes, so classification should be based on observed π_4, inducing circularity in the analysis.

The path analysis showed that the DFE was a major contributor to the variance in π_0/π_4 in both animals and plants (fig. 5) but it should be noted that both are estimated with the same data sets. Mating system and longevity had a direct effect on π_0/π_4 in plants and so had longevity in animals. Propagule size had an indirect effect on the DFE through correlation with longevity. For the determinants of genetic diversity, domestication and mutation rate had major effects on π_4, as expected, and the latter was influenced by LHTs in plants. In animals LHTs had an influence on π_4 and μ but contrary to plants variation in μ did not have any significant direct effect on genetic diversity. It is interesting to see that in animals s_{mean} was affected by genetic diversity and longevity but no such effect could be detected in plants. Genome size was correlated to π_4 or, more precisely, affected by N_e (see supplementary fig. S9, Supplementary Material online).

Discussion

All theories of molecular evolution include at least two major components: the effective population size, N_e and the distribution of fitness effects of mutations (DFE). Both are complex and abstract quantities and both are ultimately a reflection of the biological properties of organisms and of their evolutionary histories. The impact of N_e on genetic diversity is now quite well established and accepted. Interest in the effect of the DFE on genetic diversity is more recent and its impact less well understood. Here, using genomewide polymorphisms data from 62 animal and plant species, we clarified the relationship between genetic diversity and N_e, DFE and life history traits, unveiling differences between plants, and lending further support to the nearly neutral theory. Below, we discuss the limitations and implications of these main results.

Comparative Genomics from Existing Data Sets

In contrast to Romiguier et al. (2014) and Galtier (2016) that generated entirely novel RNASeq data sets for their comparative genomics analyses, but like Corbett-Detig et al. (2015), the present study is based on already existing data sets. This certainly entails a higher heterogeneity in the data than in de

novo data sets, with variation in sample size, geographical distribution of the sample and quality of the genomic data. We thus carefully checked data sets'quality and filtered out low quality ones (see supplementary SI text, Supplementary Material online). In general, our estimates were in line with previous ones (Romiguier et al. 2014), and we did not find significant effects of sample size, number of genes, read depth, genotype quality, and SNP calling software on the estimated parameters. Therefore, it seems that existing resequencing data sets can be used for comparative studies, provided that the data quality is preliminary checked.

Genetic Diversity in Plants and Animals: The Importance of Life History Traits

In both animals and plants, longevity had a major impact on genetic diversity: organisms with long lifespan tend to have lower genetic diversity. However, the way longevity affects genetic diversity may vary across species. In animals, the effect of longevity was associated to propagule size (Romiguier et al. 2014), which itself reflects a trade-off between quantity and quality of offspring: *r*-strategists produce a large amount of offspring of lower quality whereas *K*-strategists produce few of high quality. Using analytical derivations, Romiguier et al. (2014) proposed that *r*-strategists have higher long-term N_e than *K*-strategists. One hypothesis is that population size recovers more rapidly after environmental perturbations in *r*- than in *K*-strategists. Straightforwardly, the model also predicts that short-lived species recover more rapidly than long-lived ones (generation-time effect). This simple model could thus explain the strong effect of longevity in both plants and animals. Another nonexclusive hypothesis is that, *conditioned on nonextinction*, *r*-strategists should maintain on average larger population sizes. For instance, *r*-strategists could be more sensitive to Allee effects and can only avoid extinction by being maintained at high density, thus promoting high diversity (Roques et al. 2012). In plants, it is unclear what the equivalent of propagule size would be. Seed mass might come to mind, but, at least in our data set, is not correlated to nucleotide diversity. A proposed equivalent to the *r* and *K*-strategists in plants is Grime's Ruderal, Competitive and Stress-tolerant strategies (*RCS*) (Grime 1977). Perenniality is associated with the *C* strategy but is not sufficient to characterize it (Munoz et al. 2016). So, an interpretative scheme similar to the *r-K* gradient in animal could also work in plants but a better characterization of plant ecological strategies would be necessary which wasn't available for all the species included in the present study.

Mating system also had a strong effect on nucleotide diversity in plants (fig. 1*D*). In animals, we could not compare outcrossers and self-fertilizing species as only *C. briggsae* belonged to the latter, but there is no reason to believe that the results would have been different. Selfing is strongly associated with the *r*-strategy (Munoz et al. 2016) and most selfers are annual. According to the rationale outlined above, selfing species should exhibit high diversity level. However, selfing is associated with low diversity. The strong decrease in diversity in self-fertilizing species is likely resulting partly from linked selection (Glémin and Galtier 2012). A recent study has shown that linked selection affects genetic diversity (Corbett-Detig et al. 2015) but a reanalysis suggested that self-fertilizing species are the only ones where linked selection could play a major role in reducing nucleotide diversity to the extent required to explain the narrow range of diversities observed across species (Coop 2016). In addition, selfing species do not suffer from Allee effect and can thus experience strong and recurrent bottlenecks without jeopardizing population survival, which can also contribute to the low diversity observed in selfers.

π_0/π_4 Is Stable within Species but Variable between Them

We assessed how the efficacy of purifying selection varies among species using π_0/π_4. The usefulness of π_0/π_4 as a measure of selection efficacy within species has been questioned because it is sensitive to nonequilibrium conditions (Gordo and Dionisio 2005; Brandvain and Wright 2016; Gravel 2016): for instance, after a bottleneck, π_0 will reach its equilibrium value much faster than π_4. However, as our results suggest, this is not a major issue for large-scale comparative studies as demographic effects seem to play a lesser role in comparisons among species than in comparison between populations within species. The limited differences in π_0/π_4 between wild species and their domesticated counterparts or among populations within species where there are good evidence that some of the populations did experience severe bottlenecks and other major departures from equilibrium suggest that nonequilibrium conditions are unlikely to obscure the deeper differences in selection efficacy among species resulting from variation in other factors, such as LHTs and DFE. As a matter of fact, it feels fair to say that π_0/π_4 appears as a surprisingly stable quantity for a given species or even for a group of closely related species.

The Nearly Neutral Theory Globally Holds

As expected from the nearly neutral theory, π_0/π_4 decreased linearly as neutral genetic diversity increased on a log–log scale. This is true when considering plants and animals together once polymorphism has been recalibrated with mutation rate. When plants and animals are considered separately the slope is stronger for animals than for plants. This difference is primarily due to insects in animals and trees in plants. Indeed, tree species occupy a singular place among plants, in the π_0/π_4 over π_4 plot with high π_0/π_4 and intermediate π_4 (fig. 3). Harrang et al. (2013) observed a linear relationship between π_N and π_S (their fig. 3) but also noted that oysters departed from other animals: like the trees in our study, oysters tend to have high π_N values relative to π_S. However, they did not depart from other species when π_N was plotted against protein heterozygosity, *H*. Harrang et al. (2013) pointed out that marine invertebrates were biologically similar to trees: they are abundant, relatively long-lived and highly fecund. They then proposed that their observation could be explained by a very large variance in reproductive success, a few individuals contributing most of the next generation offspring. This demographic model had already been evoked to explain the tiny effective population sizes generally observed

in many marine organisms known to have extremely large population sizes (sweepstake's reproductive success). Harrang et al. (2013) speculated that in such populations, deleterious mutations could remain polymorphic for longer periods as they move rapidly towards high frequency if they are lucky enough to be carried by one of the few reproducing individuals. Although such a model could have some merit in some tree species (e.g., *Populus tremula*), it is unlikely to be relevant in others that have much more limited population sizes (e.g., *Populus euphratica* or *P. pruinosa*). In our case, however, the peculiar position of trees seems to have been primarily caused by their comparatively high mutation rates: once nucleotide diversity estimates were rescaled by mutation rate, trees lost their singularity, and animals and plants are now on the same general regression line: hence it seems that higher mutation rates in trees would be sufficient to explain the apparent discrepancy between them and other plant species. We used estimates of the mutation rates taken from the literature and those are of different nature and quality depending on the species. We note that, even in species such as humans where both pedigree-based and phylogeny-based estimates of the mutation rate (per bp and per year) are available there is still a large uncertainty on what value should be used in studies like ours. In humans, phylogeny based estimates are twice as large as estimates based on pedigrees (10^{-9} and 0.5×10^{-9}, respectively) whereas, interestingly, an opposite trend is observed in other organisms where both types of estimates are available (Moorjani et al. 2016). As argued by Moorjani et al. (2016) many factors influence mutation rate estimates and many steps are involved in the conversion of mutation rate estimates from pedigree studies into yearly substitution rates. So variation in mutation rates can lead to some of the variation that is still observed around the regression of π_0/π_4 on π_4 after rescaling by the mutation rate.

Both N_e and DFE Matter for the Efficacy of Purifying Selection

Another major factor affecting π_0/π_4 is the DFE: in both plants and animals there is a strong effect of the DFE on π_0/π_4: species such as Drosophila with a DFE with high values of β yielded lower estimates of π_0/π_4. That the DFE should be related to the relationship between π_0/π_4 and π_4 should not come as a surprise as the DFE itself captures the effect of quantities likely to have an impact on the level and nature of genetic diversity: the fitness landscape, and hence the genomic architecture of the species (complexity, pleiotropy), the process of environmental changes and N_e (Lourenço et al. 2013). Indeed, both Lourenço et al. (2013) and Huber et al. (2016) showed that Fisher Geometric Model provides a good framework to understand differences between species in DFE. Our results clearly point to variation in DFE among species. However, they do not allow to determine the underlying causes of these variations. Relating the properties of the DFE through a fitness landscape model to species biology and ecology is still a highly challenging but desirable goal.

There are some caveats to keep in mind when interpreting estimates of the shape and mean of the DFE. First, they are simply hard to estimate and one generally assumes a certain distribution, here a gamma distribution, when estimating them. In some cases other distributions may perform as well (Kousathanas and Keightley 2013) and it has been shown that quantitative conclusions on the evolutionary process, for instance on the form of the relationship between π_0/π_4 and π_4, depend on the assumed distribution (Welch et al. 2008; Tachida 1996). Second, in the present study, we only estimated the DFE for deleterious mutations, and not the DFE for both deleterious and positive mutations. Ignoring positive mutations biases the size of the effectively neutral mutations since positively selected mutations will now be added to the category of effectively neutral mutations and species with different N_e will be affected differently. Species with small N_e should be far from their fitness optimum and positive mutations will be proportionally more frequent than in species with large N_e. Conversely, species with large N_e are expected to be close to their fitness optimum and mutations are therefore more likely to be deleterious (e.g., fig. 7 in Lourenço et al. 2013). So, the relatively low β value observed in, for instance, *Populus* could suggest that *Populus* does not purge slightly deleterious mutations as efficiently as Drosophila but it could as well indicate that poplar species are further away from their fitness optima than species of flies. Although both hypotheses are likely to be partly true, only estimates of the full DFE will be able to tell apart these two possibilities and in this respect the recent development of statistical methods to do just so from polymorphism data is encouraging (Tataru et al. 2016). Bearing these caveats in mind, we first note that our estimates are in line with those reported in the literature for the species where estimates are available (e.g., Kousathanas and Keightley 2013). The range of β values across species is rather limited and apart from *A. thaliana*, *D. melanogaster*, and cucumber, there is rather limited variation among populations of the same species in estimates of β. Hence, since populations within species have different histories, it seems that the DFE is not too sensitive to changes in selection or demography occurring on that time scale. Similarly, and in line with theoretical work the "negative part" of the DFE does not appear to be too strongly affected by effective population sizes (Lourenço et al. 2013). A multi species survey, hence, reaches a somewhat different conclusion than the one reached by comparing human populations that led authors to emphasized the sensitivity of π_0/π_4 to the details of population histories (see Brandvain and Wright 2016 and references therein). Seen at a larger evolutionary scale, π_0/π_4 appears as a rather stable characterization of species. This does not preclude variation among different classes of genes as suggested by work on the transcriptomic hourglass (Gossmann et al. 2016), a research avenue that remains to be explored.

Conclusion

Genetic diversity and efficacy of purifying selection vary extensively across plant and animal species. Here we showed that a great deal of this variation can be assigned to the variation in life history traits among organisms, in particular longevity. LHT will affect genetic diversity and efficacy of

purifying selection both directly and indirectly, through their impact on effective population size. Variation in mutation rate per generation, which is often assumed to be rather uniform across species, will also need to be considered. Estimates of mutation rates remain scarce, and mostly confined to a handful of model species, but there has been a renewed interest in those and in methods to estimate them (Moorjani et al. 2016). Understanding the mechanisms by which LHT, effective population size and the DFE shape genetic diversity and affect the efficacy of purifying selection will require a general theoretical model linking these different components. At that stage it seems that the Fisher Geometric model provides the best framework. The last decades have witnessed a renaissance of the "adaptationist" program with a large number of studies attempting to estimate the proportion of adaptive mutations in an array of organisms. Our study, as well as Galtier (2016), suggests that more attention should be devoted to deleterious mutations. Recently, Galtier (2016) used the same data as Romiguier et al. (2014) to show that the proportion of adaptive amino-acid changes is positively correlated to effective population size. However, he also showed that this relationship reflects the fact that purging of deleterious mutations is more efficient in large populations than in small ones, rather than, as has often been implicitly assumed, a faster rate of accumulations of adaptive mutations. Hence independently of what the aim is and even if one remains primarily interested in adaptive processes those cannot be properly evaluated if no attention is paid to the load of deleterious genes that all individuals carry.

Materials and Methods

Population Genomic Data Sets

We downloaded single nucleotide polymorphism (SNP) data from 62 species whose genome has been recently resequenced at high coverage with the Illumina HiSeq platform. These included 34 animal species composed of 17 mammals, 12 insects, three fishes, one bird, and one nematode, and 28 plant species composed of 14 herbaceous and 14 woody/shrubby plants. Ten plant species were selfers and 18 were outcrossers. We also sampled domesticated counterparts of nine wild species including maize, sorghum, wheat, rice, cucumber, soybean, watermelon, cassava, and dog. Additionally, we extended our calculations to the population level in species where those were available, namely humans, *D. melanogaster*, *A. thaliana*, *P. trichocarpa*, as well as a few other species. In total 138 species/populations were investigated in this study. For regression analyses of π_0/π_4 and π_4, we only included the 62 wild species whereas for regression analyses we also took the mean value for each family to control the phylogeny. To study the effect of domestication and demographics, we used all data including cultivated species and the population data. Additionally, for comparisons of DFE parameters we only kept species/populations where more than two chromosomes sampled. π_0/π_4 ratios of *Pinus* species were directly taken from Eckert et al. (2013) and only used to show the positions of gymnosperm species in fig. 3.

Species Traits

In total, eight species traits were collected from different data sets for all plant species in this study, including mating system, genome size, longevity, dispersal mode, pollination mode, mutation rate per base per generation, seed mass, and canopy height (see supplementary table S1 in SI text, Supplementary Material online). We also collected five traits for all animal species, including mating system, genome size, longevity, mutation rate, and propagule size (see supplementary table S2 in SI text, Supplementary Material online).

Calculation of π_0/π_4 Ratio

The nucleotide diversity in nonsynonymous and synonymous positions in protein-coding sequences has been used a measurement of the proportion of new deleterious mutations in an organism and as a measure of the efficacy of selection (e.g., Galtier 2016; Eyre-Walker et al. 2006). Defining nonsynonymous and synonymous mutations is fraught with difficulties and to reduce uncertainty, we used the diversity of 0-fold and 4-fold positions as nonsynonymous and synonymous mutations, respectively. To compute genomic diversity, we sampled up to 20 chromosomes or as many as we could. In species where a larger number was available we randomly sampled 20 chromosomes. In cases where the species was represented by more than one population we first pooled equal number of randomly sampled chromosomes from each population and increased the maximum number of chromosomes to 50.

Pairwise nucleotide diversity (π) was calculated for 0-fold and 4-fold positions in protein-coding sequences that start with "ATG" and end with one of the stop codons. To avoid bias from mis-assembly/-annotation, we only calculated π in sequences showing high BLASTP similarity (e value $\leq 1e-10$, bit-score ≥ 100) against the plant/animal protein database of UniProtKB/Swiss-Prot, which has been manually annotated and reviewed. We chose the longest protein-coding sequence as the representative model for the gene. We only kept polymorphic sites with two variants and filtered out sites with ambiguous degeneracy, missing allele, deletion/insertion, as well as heterozygous sites in inbreeding/selfing species, or in haploid genomes. We then corrected the total length of 0-fold and 4-fold positions with the same filtering ratio for the polymorphic sites of that gene. Genes with filtering ratio above 50% were discarded. The final estimate of π_0/π_4 ratio for the species was based on averaged values of π_0 and π_4 across all genes.

Estimate of DFE

The π_0/π_4 ratio is a synthetic statistics summarizing the efficacy of purifying selection. However, its interpretation has been questioned as it can be affected by demography and it is sensitive to nonequilibrium conditions (Gordo and Dionisio 2005; Brandvain and Wright 2016; Gravel 2016). To assess the robustness of results based on π_0/π_4 and to better characterize the efficacy of purifying selection we also estimated the *scaled* DFE. In particular, we estimated the proportion of mutations for which $S < 1$ is an alternative measure of the efficacy of selection that is, in theory at least, devoid of

spurious demographic effects. In addition, characterizing the DFE also allows to test whether it is roughly constant across species and whether differences in the efficacy of purifying selection are only driven by differences in N_e.

For each data set, we applied the method of Eyre-Walker et al. (2006) to the folded 0-fold and 4-fold site frequency spectra (SFS). This method assumes a gamma distribution for $\Phi(S)$ and takes demography (or sampling or any departure from equilibrium) into account by introducing nuisance parameters. Given the wide diversity of species, differences in sampling and demographic history, it appears more flexible than methods based on a single-change in population size (Keightley and Eyre-Walker 2007). Moreover, recent tests have showed that this approach is robust (Tataru et al. 2016). We reimplemented the method in a Mathematica script (see supplementary S2 file, Supplementary Material online).

Statistical Analyses

All statistical analyses were performed in R, with linear regression methods for continuous variables and ANOVA/ANCOVA for discrete variables. To investigate how π_0/π_4 could be affected by all factors studied in this paper, we performed path analyses separately on plant and animal species using the R package "lavaan" (Rosseel 2012). We tested direct effects by creating regression paths between π_0/π_4 and π_4, DFE (β and s_{mean}), LHT (annual/perennial, selfer/outcrosser for plants; longevity and propagule size for animals), and domestication (for plants) on π_0/π_4. Indirect effects were tested through interactions between π_0/π_4 and all possible combinations of factors (see supplementary SI text, Supplementary Material online). We further replaced π_4 and μ with N_e to evaluate its effects using the same method.

Supplementary Material

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

Acknowledgments

This study was supported by grants from FORMAS, the Swedish Research Council and the Swedish Council for Strategical Research (SSF) to Martin Lascoux. S.G. is supported jointly by the French CNRS and the Marie Curie IEF Grant SELFADAPT 623486. We thank Michele Morgante (University of Udine), Tao Ma and Jianquan Liu (Sichuan University) and Jarkko Salojärvi (University of Helsinki) for early access to genomic data of *Populus nigra* and *Vitis vinifera* (M.M.), *Populus euphratica* and *Populus pruinosa* (T.M. and J.L.) and *Betula pendula* (J.S.). We also thank Jing Wang and Pär Ingvarsson (Umeå University) for help with calculations on *Populus tremula* and *Populus tremuloides* that were still unpublished at the time. Finally, we thank Thomas Bataillon, Nicolas Galtier and Antoine Kremer for helpful comments on earlier versions of the manuscript.

References

Akashi H, Osada N, Ohta T. 2012. Weak selection and protein evolution. *Genetics* 192:15–31.
Brandvain Y, Wright SI. 2016. The limits of natural selection in a nonequilibrium world. *Trends Genet.* 32:201–210.
Charlesworth B. 2009. Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. *Nat Rev Genet.* 10:195–205.
Coop G. 2016. Does linked selection explain the narrow range of genetic diversity across species? *BioArXiv*. doi: https://doi.org/10.1101/042598.
Corbett-Detig RB, Hartl DL, Sackton TB. 2015. Natural selection constrains neutral diversity across a wide range of species. *PLoS Biol.* 13:e1002112.
Do R, Balick D, Li H, Adzhubei I, Sunyaev S, Reich D. 2015. No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. *Nat Genet.* 47:126–131.
Eckert AJ, Bower AD, Jermstad KD, Wegrzyn JL, Knaus BJ, Syring JV, Neale DB. 2013. Multilocus analyses reveal little evidence for lineage-wide adaptive evolution within major clades of soft pines (pinus subgenus strobus). *Mol Ecol.* 22:5635–5650.
Ellegren H, Galtier N. 2016. Determinants of genetic diversity. *Nat Rev Genet.* 17:422–433.
Elyashiv E, Bullaughey K, Sattath S, Rinott Y, Przeworski M, Sella G. 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.
Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet.* 8:610–618.
Eyre-Walker A, Woolfit M, Phelps T. 2006. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* 173:891–900.
Felsenstein J. 1985. Phylogenies and the comparative method. *Am Nat.* 125:1–15.
Figuet E, Nabholz B, Bonneau M, Mas Carrio E, Nadachowska-Brzyska K, Ellegren H, Galtier N. 2016. Life history traits, protein evolution, and the nearly neutral theory in amniotes. *Mol Biol Evol.* 33:1517–1527.
Galtier N. 2016. Adaptive protein evolution in animals and the effective population size hypothesis. *PLoS Genet.* 12:e1005774.
Gillespie JH. (2004). Population genetics. A concise guide. Baltimore & London: Johns Hopkins University Press.
Glémin S, Galtier N. 2012. Genome evolution in outcrossing versus selfing versus asexual species. *Methods Mol Biol.* 855:311–335.
Glémin S, Bazin E, Charlesworth D. 2006. Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Philos Trans R Soc Lond B Biol Sci.* 273:3011–3019.
Gordo I, Dionisio F. 2005. Nonequilibrium model for estimating parameters of deleterious mutations. *Phys Rev E Stat Nonlin Soft Matter Phys.* 71:031907.
Gossmann TI, Saleh D, Schmid MW, Spence MA, Schmid KJ. 2016. Transcriptomes of plant gametophytes have a higher proportion of rapidly evolving and young genes than sporophytes. *Mol Biol Evol.* 33:1669–1678.
Gravel S. 2016. When is selection effective? *Genetics* 203:451–462.
Grime JP. 1977. Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. *Am Nat.* 111:1169–1194.
Hamrick JL, Godt M. 1996. Effects of life history traits on genetic diversity in plant species. *Philos Trans R Soc Lond B Biol Sci.* 351:1291–1298.
Harrang E, Lapègue S, Morga B, Bierne N. 2013. A high load of non-neutral amino-acid polymorphisms explains high protein diversity despite moderate effective population size in a marine bivalve with sweepstakes reproduction. *G3-Genes Genom Genet.* 3:333–341.
Henn BM, Botigué LR, Bustamante CD, Clark AG, Gravel S. 2015. Estimating the mutation load in human genomes. *Nat Rev Genet.* 16:333–343.
Huber CD, Kim B, Marsden CD, Lohmueller KD. 2016. Determining the factors driving selective effects of new nonsynonymous mutations. *BioRxiv*. doi: https://doi.org/10.1101/071209.

Keightley PD, Eyre-Walker A. 2007. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics* 177:2251–2261.

Kimura M. 1979. Model of effectively neutral mutations in which selective constraint is incorporated. *Proc Natl Acad Sci USA*. 76:3440–3444.

Kimura M. (1983). The neutral theory of molecular evolution. Cambridge: Cambridge University Press.

Kousathanas A, Keightley PD. 2013. A comparison of models to infer the distribution of fitness effects of new mutations. *Genetics* 193:1197–1208.

Lanfear R, Kokko H, Eyre-Walker A. 2014. Population size and the rate of evolution. *Trends Ecol Evol*. 29:33–41.

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

Leffler EM, Bullaughey K, Matute DR, Meyer WK, Ségurel L, Venkat A, Andolfatto P, Przeworski M. 2012. Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol*. 10:e1001388–e1001388.

Lohmueller KE. 2014. The distribution of deleterious genetic variation in human populations. *Curr Opin Genet Dev*. 29C:139–146.

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

Lourenço JM, Glémin S, Chiari Y, Galtier N. 2012. The determinants of the molecular substitution process in turtles. *J Evol Biol*. 26:38–50.

Lourenço JM, Glémin S, Galtier N. 2013. The rate of molecular adaptation in a changing environment. *Mol Biol Evol*. 30:1292–1301.

Lynch M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc Natl Acad Sci USA*. 104:8597–8604.

Martin G, Lenormand T. 2006. A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution* 60:893–907.

Moorjani P, Gao Z, Przeworski M. 2016. Human germline mutation and the erratic evolutionary clock. *PLoS Biol*. 14:e2000744.

Munoz F, Violle C, Cheptou PO. 2016. Csr ecological strategies and plant mating systems: outcrossing increases with competitiveness but stress-tolerance is related to mixed mating. *Oikos* 125:1296–1303.

Popadin K, Polishchuk LV, Mamirova L, Knorre D, Gunbin K. 2007. Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals. *Proc Natl Acad Sci USA*. 104:13390–13395.

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

Roques L, Garnier J, Hamel F, Klein EK. 2012. Allee effect promotes diversity in traveling waves of colonization. *Proc Natl Acad Sci USA*. 109:8828–8833.

Rosseel Y. 2012. lavaan: An r package for structural equation modeling. *J Stat Softw*. 48:1–36.

Tachida H. 1996. Effects of the shape of distribution of mutant effect in nearly neutral mutation models. *J Genet*. 75:33–48.

Tataru P, Mollion M, Glémin S, Bataillon T. 2016. Inference of distribution of fitness effects and proportion of adaptive substitutions from polymorphism data. *BioarXiv*. doi: 10.1101/062216.

Tenaillon O. 2014. The utility of Fisher's geometric model in evolutionary genetics. *Annu Rev Ecol Evol Syst*. 45:179–201.

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

Whitney KD, Baack EJ, Hamrick JL, Godt MJW, Barringer BC, Bennett MD, Eckert CG, Goodwillie C, Kalisz S, Leitch IJ, Ross-Ibarra J. 2010. A role for nonadaptive processes in plant genome size evolution? *Evolution* 64:2097–2109.

Whitney KD, Boussau B, Baack EJ, Garland T. 2011. Drift and genome complexity revisited. *PLoS Genet*. 7:e1002092.