Determining the factors driving selective effects of new nonsynonymous mutations

Christian D. Huber[1,], Bernard Kim[1], Clare D. Marsden[1], Kirk E. Lohmueller[1,2,3,*]*

Affiliations:
[1] Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA.
[2] Interdepartmental Program in Bioinformatics, University of California, Los Angeles, CA 90095, USA.
[3] Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA.

1. Text S1-S7
2. Figures S1-S14
3. Tables S1-S5

Text S1

Genomic Data

We used published next generation sequencing data sets to extract the synonymous and nonsynonymous SFS. For humans, we used the sample of 112 individuals from Yoruba in Ibadan, Nigeria (YRI) from the 1000 Genomes Project (1). We downloaded the 1000 Genomes phase 3 dataset from the 1000 Genomes ftp site (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase3/, accessed Sept 2014). Using information in the sample information PED file, related individuals were removed and for each trio or family group only the mother and father were used. The SNPs were also filtered for whether they were within the exome capture array region and in the strict mask part of the human genome, as defined by the 1000 Genomes Project. The genotypes of YRI individuals were extracted and annotated using the SeattleSeq annotation pipeline (http://snp.gs.washington.edu/SeattleSeqAnnotation138/). For *Drosophila melanogaster*, we used the DPGP phase 3 data of a sample of 197 lines originating from Zambia, Africa (2). We accessed whole genome genotype data for the 197 genomes from the Pool lab (http://johnpool.net/genomes.html). These data were provided in non-standard vcf format (vcf sites file, downloaded August 2014), therefore we first converted these to a standard vcf format with the BDGP5.75 genome as the reference using a custom python script. We then merged all the individual vcf files and removed any sites with evidence of identity by descent or admixture using the masking package provided by the Pool lab. Only the 2L, 2R, 3L and 3R chromosome arms were used in our analyses. We then conducted variant annotation using SnpEeff v3.6 using the BDGP5.75 database.

We filtered both datasets for sites with sample size > 99 and down-sampled all sites with larger sample size than 100 to a sample size of 100 using the hypergeometric probability distribution. Further, we selected only sites that were in exons and computed an exon length $L_{exon,i}$ for each gene *i*. The nonsynonymous and synonymous sequence lengths (L_{NS}, L_S) depend on the transition/transversion ratio and CpG mutational bias. We assumed a transition:transversion ratio of 2:1 in *Drosophila* (3, 4) and 3:1 for human exons (5, 6). Further, we assumed a 10x mutational bias at CpG sites in humans, but no such effect in *Drosophila* (7). This leads to multipliers of L_{NS} = 2.85 x L_S in *Drosophila*, and L_{NS} = 2.31 x L_S in humans. We calculated the synonymous and nonsynonymous SFS, and the respective sequence lengths ($L_{NS,i}$, $L_{S,i}$), for each gene *i*. For all further inference, we used the folded SFS to avoid correcting for misidentification of the ancestral state. Ancestral misidentification could lead to unwanted and difficult to control biases (8).

To study the effect of gene expression on the DFE, we used two recent gene expression datasets from humans (9) and *Drosophila* (10) that provide mRNA expression level estimates in 27 and 29 different tissues, respectively. For both datasets, we transformed the 'fragments per kilobase of exon per million fragments' (FPKMs) by computing log(FPKM+1) and quantile normalizing this value over all tissues using 'normalize.quantiles' of the R package 'preprocessCore', resulting in an expression level *S*. We computed τ as a measure of tissue specificity for each gene: $\tau = \sum_{j=1}^{n} 1 - \frac{log(S_j)}{\log(S_{max})(n-1)}$. Here, *n* is the number of tissues, S_j is the expression level in tissue *j* and S_{max} is the largest expression level over all tissues. We used τ to

classify genes as tissue specific ($\tau > 0.6$) or broadly expressed ($\tau < 0.4$). We further classified genes as low ($\bar{S} < 2$), intermediate ($2 < \bar{S} < 3$) and highly expressed ($\bar{S} > 3$), where $\bar{S} = \sum S_j / n$. This classification leads to strongly different gene expression profiles between classes (SI Appendix, Fig. S2), but still enough data in every class to be able to reliably estimate the DFE ($\theta_S > 100$ in humans and $\theta_S > 900$ in *Drosophila*).

To infer the DFE in *Mus musculus castaneus* (mouse) and *Saccharomyces paradoxus* (yeast), we used the synonymous and nonsynonymous SFS data from Gossmann et al. (11). Our estimates of proportions of mutations in different $N_e s$ bins (SI Appendix, Fig. S10) are concordant with what has been reported previously (11–13). We then used mutation rate estimates for yeast (14) and mouse (15), respectively, to estimate N_e and transform the DFE from $N_e s$ to s.

Text S2

Estimating demography and DFE

We infer the parameters of a single size change model using the synonymous site frequency spectrum (SFS) under the Poisson Random Field framework (16). In this framework, the multinomial likelihood quantifies how well the empirical SFS fits to an expected SFS that is derived from specific demographic parameters (17). Assume that Θ_D is a vector of demographic parameters (i.e., time and strength of a population size change), X_i is the count of SNPs with frequency i, P_i is the proportion of SNPs at frequency i, θ is the population mutation rate, and n is the sample size. The distribution of allele frequency q in the population ($g[q|\Theta_D]$) can be computed by numerically solving the diffusion approximation to the Wright-Fisher model, and can also incorporate selection (17–19). We used ∂a∂i (17) to numerically approximate $g(q|\Theta_D)$. Further, the expected number of SNPs at frequency i in a sample of size n is $E[X_i|\Theta_D,\theta] = \theta \int_0^1 \binom{n}{i} q^i (1-q)^{n-1} g(q|\Theta_D) dq$. The relative proportion of SNPs at frequency i can then be calculated as $P_i(\Theta_D) = \frac{E[X_i|\Theta_D,\theta]}{\sum_{j=1}^{n-1} E[X_j|\Theta_D,\theta]}$, and the formula for the multinomial likelihood is $L(\Theta_D) = \prod_{i=1}^{n-1} P_i^{X_i}$. To determine the maximum likelihood estimate of Θ_D ($\widehat{\Theta_D}$) we maximized $L(\Theta_D)$.

Conditional on the estimated demographic parameters, we estimate the DFE for new nonsynonymous mutations using the nonsynonymous SFS. In short, our approach utilizes the fact that more deleterious mutations segregate in lower numbers and at lower frequencies than less deleterious or neutral mutations. Thus, we do not directly quantify the deleteriousness of any specific mutation, but indirectly summarize the fitness effects over many sites by estimating the parameters of a DFE that fits the SFS. We used the Poisson likelihood instead of the multinomial likelihood to estimate the vector of parameters of the DFE (Θ_{DFE}). We found that this strongly improves the precision of the estimate of the scale parameter of the gamma distribution compared to using the multinomial likelihood since the Poisson likelihood uses information from both the absolute number of SNPs as well as the curvature of the SFS (18, 20). Note however that we do not make use of fixed differences to an outgroup. Including information from fixed differences hardly improves inferring the DFE of deleterious mutations (21), which is the main focus of our paper. However, we do not filter out variants from the SFS that might have been segregating already for a long time (e.g. trans-specific polymorphisms). The reason is that the Poisson Random Field approach is designed to use overall patterns of polymorphism segregating within a species to estimate the distribution of selection coefficients of new mutations. Including old polymorphisms in our analysis is important because their age suggests that those mutations are neutral. Filtering them would bias our estimation by reducing the proportion of new mutations that are predicted to be effectively neutral. The likelihood of Θ_{DFE} was thus calculated as $L(\Theta_{DFE}|\Theta_D,\theta) = \prod_{i=1}^{n-1} \frac{E(X_i|\Theta_D,\Theta_{DFE},\theta)^{X_i}}{X_i!} e^{-E(X_i|\Theta_D,\Theta_{DFE},\theta)}$. We set Θ_D here to the maximum likelihood estimates of the demographic parameters $\widehat{\Theta_D}$, and θ to the nonsynonymous population mutation rate $\theta_{NS} = 4N_e\mu L_{NS}$. We estimated θ_{NS} from θ_S by accounting for the difference in synonymous and nonsynonymous sequence length.

The formula of the Poisson likelihood depends on $E(X_i|\Theta_D,\Theta_{DFE},\theta)$, i.e. on the expected SFS given the demography, θ_{NS} and some distribution of $N_e s$ with parameters Θ_{DFE}. However, ∂a∂i only allows computing the expected SFS $E(X_i|\Theta_D,\mathrm{N}_e s,\theta)$ for a single selection coefficient

$N_e s$ (and some arbitrary demography). Thus, we extend ∂a∂i's functionality by computing the expected SFS for a grid of 1000 $N_e s$ values on an exponentially distributed grid between -15000 and -10^{-4}. This set of site frequency spectra is further used to calculate the expected SFS for an arbitrary distribution of $N_e s$ values. This is done by numerically integrating over the respective spectra weighted by the gamma distribution. The numerical integration was done using the 'numpy.trapz' function as implemented in ∂a∂i. Due to numerical instabilities for strongly skewed distributions, we did not integrate all the way towards 0, but computed the weight of $N_e s$ values between -10^{-4} and 0 and added the product of this weight with the neutral SFS to the expected SFS. Mutations with $N_e s < -15000$ are expected not to contribute to the SFS since they are strongly selected against.

Numerical optimization is used to find the parameters of the DFE that maximize the Poisson likelihood. For this optimization step, we use the BFGS algorithm as implemented in the 'optimize.fmin_bfgs' function of scipy. To avoid finding local optima, we repeated every estimation approach (for both the simulations and the real data) from 50 uniformly distributed random starting parameters. Standard errors were based on the Hessian matrix of the log-likelihood function, numerically computed at the maximum likelihood estimates using the 'Hessian.hessian' function of ∂a∂i (17). They were computed as the square root of the diagonal elements of the inverse of the negative Hessian matrix (22). Confidence intervals were approximated as plus/minus two times the standard errors, except where specified otherwise.

Our approach allows us to estimate the parameters of any arbitrary distribution of $N_e s$ values. We implemented the gamma distribution, log-normal distribution, the formula of Piganeau and Eyre-Walker (23), eq. 7, assuming gamma distributed effect sizes, and the formula of Lourenço et al. (24), eq. 15. The formula of Lourenço et al. (24) provides an explicit solution to the DFE for Fisher's geometrical model under fitness equilibrium. It is a function of three parameters: population size, effect size, and the average number of phenotypes affected by a mutation (pleiotropy). The DFE of Lourenço et al. (24) and Piganeau and Eyre-Walker (23) are distributions with some proportion of slightly beneficial mutations. In models with some proportion of beneficial mutations, those mutations are expected to segregate in the population and thus influence both the shape of the SFS and the absolute number of SNPs. We use this expectation to infer the full DFE (beneficial plus deleterious mutations) from the SFS, similar to Tataru et al. (25). To do this, we also integrate over beneficial mutations with $N_e s$ from 0 to 15000.

Note that population genetic methods for estimating the DFE from the SFS can only estimate the composite parameter of selection coefficient s with effective population size N_e, since the effect of selection on the SFS depends on $N_e s$ and not s alone. However, the distribution of s can be derived from the distribution of $N_e s$ by scaling it by $1/N_e$ (e.g. multiplying the scale parameter of a gamma distribution of $N_e s$ by $1/N_e$). Fitting the demographic model to the synonymous SFS provided an estimate of $\theta_S = 4 N_e \mu L_S$ for synonymous sites, where μ is the neutral per base-pair mutation rate and L_S is the synonymous sequence length. Using this formula, we estimated N_e by setting the neutral mutation rate to either 2.5×10^{-8} for humans and 1.5×10^{-9} for *Drosophila* (phylogenetic estimates (26–28)) or to 1.5×10^{-8} for humans and 3×10^{-9} for *Drosophila* (current estimates (27, 29, 30)). Note that when partitioning our data into different gene categories and estimating the DFE for each category separately, we also allow for a

different ancestral N_e and demographic estimates in those categories to control for different levels of background selection in different genomic regions (31, 32).

Text S3

Robustness of DFE inference to possible confounding factors

We inferred the shape (α) and scale (β) parameters of a gamma-distributed DFE conditional on the estimated demographic parameters in each species (SI Appendix, Text S2 and Table S2). We first fit a null model where the shape and scale parameters were constrained to be the same in both species (i.e., $\alpha_H = \alpha_D$ and $\beta_H = \beta_D$), where H denotes human and D denotes *Drosophila*. This corresponds to a model where the DFE of *s* is the same in both species. Importantly, although the DFE is constrained to be the same, we condition on the inferred demographic model for each species when estimating the DFE. As such, this approach appropriately controls for the differences in population size between species. Further, by fitting a size change model, we also aim to control for any bias of the DFE parameter estimates caused by background selection, as suggested by simulation studies (20, 25) and theoretical work (33). We next estimated parameters in a full model where each species was allowed to have its own DFE (i.e., $\alpha_H \neq \alpha_D$ and $\beta_H \neq \beta_D$). Because the constrained model uses a subset of the parameters of the full model, the models are nested and we can compare the fit of the two models to our data using a likelihood ratio test (LRT), where the test statistic (Λ) is asymptotically chi-square distributed with two degrees of freedom. We find that $\Lambda > 920$ ($p < 10^{-16}$), even after employing various data quality filters, or using alternate mutation rate estimates (SI Appendix, Table S2). Similarly large Λ values are found when assuming a log-normal instead of a gamma distributed DFE (SI Appendix, Table S3). Individually, the gamma distribution fits better than the log-normal distribution in both species (SI Appendix, Table S2 and S3). Further, the estimated parameters of the gamma DFE are comparable to those from previous studies (18, 34, 35).

The results based on both gamma and log-normal DFEs suggest that mutations are on average about 80-fold more deleterious in humans than in *Drosophila*. However, due to the long tail of the gamma distribution, the scale parameter is difficult to estimate and potentially sensitive to the actual functional form of the DFE. Therefore, we tested the robustness of our findings by examining a range of alternative functional forms of the DFE. We tested the following additional distributions: 1) gamma + neutral point mass, 2) a DFE based on eq. 7 in Piganeau and Eyre-Walker (23), 3) a DFE based on eq. 8 in Lourenço et al. (24) 4) a DFE based on eq. 15 in Lourenço et al. (24) and 5) a DFE based on eq. 5 in Martin and Lenormand (36). For all tested cases, the average selection coefficient is at least 22 times more deleterious in humans than in *Drosophila* (SI Appendix, Fig. S8). Simulations under a null model suggest only a, at most, 1.5-fold difference at the 99% confidence level due to uncertainty in estimated parameters. In other words, we observe more deleterious mutations in humans than in *Drosophila* for all of the different functional forms of the DFE assumed during the inference. In addition to these distributions, we also tested a nonparametric discretized distribution to infer the properties of the DFE in humans and *Drosophila* (SI Appendix, Fig. S14). The distribution assumes four mutational effect classes, and each class is modeled as a uniform distribution. The classes are placed consecutive to each other, with the following boundaries on *s*: [0, -1e-5), [-1e-5, -1e-4), [-1e-4, -1e-3), [-1e-3, -1]. The probabilities of a mutation falling into each class are the four estimated parameters of the DFE. This and similar nonparametric distributions were shown to correctly approximate the general form of the underlying DFE even if the true DFE is multi-modal (13, 20). We found that humans have four-fold fewer mutations with $|s| < 10^{-5}$, but

almost three-fold more mutations with $|s| > 10^{-3}$ than *Drosophila*, again indicating a larger proportion of strongly deleterious mutations in humans than in *Drosophila* (SI Appendix, Fig. S14).

In conclusion, these results indicate that a model with distinct DFEs in humans and *Drosophila* fits the data significantly better than a model with the same DFE in both species ($p < 10^{-16}$). Further, mutations are estimated to be more deleterious in humans than in *Drosophila*, and this result is highly robust to the assumed functional form of the DFE.

Several factors could lead us to falsely reject the null hypothesis of the same DFE in both species. First, the Poisson Random Field approach for calculating the likelihood assumes that allele frequencies at different SNPs are independent. Violations of this assumption can lead to LRTs being too liberal (37, 38). Second, our inferences do not incorporate uncertainty in the demographic parameter estimates. Uncertainty in the demographic parameters might further broaden the distribution of Λ. Third, we numerically optimize the likelihood, which might result in finding suboptimal solutions to the true maximum likelihood estimate. Finally, the Poisson Random Field approach assumes that there is no interference between selected sites, or between selected and neutral sites. However, background selection, selective sweeps, and interference between selected sites can bias estimates of demographic parameters (39) and could lead to biased estimates of DFE parameters.

To test if these four factors in combination lead to false-rejection of the null hypothesis, we performed forward simulations under the Wright-Fisher model including realistic levels of background selection and linkage disequilibrium. We assumed the estimated demographic models (SI Appendix, Table S1) and the shape and scale estimates of both the full model and the constrained model (SI Appendix, Table S2, All Data). We then performed our inference procedure on each simulated dataset and tabulated the distribution of Λ. Our forward simulations assumed a spatial distribution of selected elements that reflects the empirical distribution of coding and conserved non-coding (CNC) sequence in the genome of humans and *Drosophila*. We further simulated under realistic maps of recombination rate across the two genomes (SI Appendix, Text S4). Mutations in CNC regions are assumed to be selected with gamma distributed selection coefficients for humans (40) and *Drosophila* (41). The simulations resulted in considerable amounts of background selection, with average reduction in neutral diversity of 10% in humans and 12% in *Drosophila*. However, when we estimated the DFE from the simulations of the full model, the estimates were unbiased (Fig. 3A). This suggests that the size change model fit to synonymous polymorphisms successfully controls for the effects of background selection (SI Appendix, Fig. S3, see also (20, 25)). As expected, the null distribution of Λ derived from simulations under the constrained model is broader than the chi-square distribution with two degrees of freedom (Fig. 3C). However, all of the 300 Λ values that we simulated were smaller than 34, suggesting the probability of seeing a Λ value bigger than 920 is substantially less than 0.33% under the null.

Since selective sweeps were suggested to be a major determinant of genetic diversity in *Drosophila* (42), we also examined the effect of recurrent selective sweeps on our inference. As estimated by Keightley et al. (35), we modified the DFE used to simulate the data such that 0.5% of new nonsynonymous mutations were beneficial with $N_e s$=12. Note that although the proportion of beneficial mutations seems small, it is in line with McDonald-Kreitman table based estimates that 50% of amino acid substitutions are positively selected (35). We then inferred the

demography using synonymous sites and then, conditional on the demographic parameter estimates, inferred the DFE for nonsynonymous mutations. Importantly, to mimic the inference done on the empirical data, the DFE that we fit to the new nonsynonymous mutations was a gamma distribution that only included deleterious mutations. We did the inference of the DFE of nonsynonymous mutations in two ways. First, we removed all the beneficial segregating nonsynonymous variants from the simulated data. This scenario examines the indirect effects of positive selection on segregating deleterious mutations (i.e. the effect of linkage of a deleterious mutation to a positively selected one). In line with other studies (25), we found that selective sweeps, similar to background selection, do not significantly bias our DFE estimates when correcting for the effect of demography using the observed SFS at neutral sites (SI Appendix, Fig. S9). Second, we repeated our inference of the DFE of nonsynonymous mutations leaving the segregating beneficial variants in the simulated SFS. Biologically, this scenario allows for some segregating nonsynonymous polymorphisms to be under positive selection. We then fit a gamma DFE that included only deleterious mutations. Here, the presence of beneficial nonsynonymous polymorphisms in the SFS slightly biases estimates due to model misspecification. However, this bias is only small and cannot explain the difference in estimates that we observe between *Drosophila* and humans (SI Appendix, Fig. S9).

Another potential confounder of our inference is strong selection on synonymous sites. A recent study suggested that selection on synonymous sites in *Drosophila* could be strong, and that synonymous diversity is reduced by 22% due to this effect of strong selection (43). This study is based on comparing patterns of genetic diversity between synonymous sites and short introns. Other studies suggested that the positions 8-30 of short introns (<=65bp) can be used as a neutral reference that is free of the influence of weak or strong selection (44, 45). However, differences in the mutation rate between synonymous and intronic sites make direct use of the intronic SFS as neutral standard difficult (45). Therefore, to test the proposed effect of strong selection on synonymous sites, we generated a truly neutral synonymous SFS, assuming that the study of Lawrie et al. (43) is correct and 22% of synonymous diversity is missing due to strong selection. We generated this new neutral synonymous SFS by 1) estimating the shape (i.e. the proportional SFS, or the proportions of SNPs at each frequency bin) of the SFS from data from short introns according to the definition in Parsch et al. (44), and 2) setting the total synonymous SNP count to a factor of 1/(1-0.22) larger than what is observed for synonymous sites in the data. We use this new synonymous SFS as neutral standard for estimating synonymous θ_S, demographic parameters, and effective population size. We then use this new estimate of N_e and θ_S to infer the distribution of *s* on nonsynonymous mutations from the estimates of 2 $N_e s$. We see no qualitative difference in the estimated proportions of mutations in different '*s*' or '$N_e s$' bins compared to the results using plain synonymous sites as neutral reference (SI Appendix, Fig. S7). In particular, the difference in the DFE we observe between humans and *Drosophila* is robust to which neutral standard is used.

Another concern is that the observed differences in the DFE between species could be solely due to analyzing different sets of genes under distinct selective pressures in one species compared to the other species. Such a scenario might generate a difference in the DFE of *s* not due to selection acting differently in the two species, but by selection acting on different gene classes. To address this issue, we repeated our analyses of the DFE considering sets of orthologous genes in both humans and *Drosophila*. To find orthologous genes between humans

and *Drosophila*, we integrate 10 different tools for predicting orthologous gene relationship by using the DIOPT diseases and traits query tool (DIOPT-DIST; http://www.flyrnai.org/diopt-dist). We identify a highly confident set of orthologous genes by requiring the orthologous relationship to be supported by at least 4 different prediction tools, resulting in retaining about half of the genes in both humans (7356/14245) and *Drosophila* (5827/12304). Inferences based on this set of genes were similar to those described in the main text (Figure 2), and revealed a significant difference in the DFE between humans and *Drosophila* (Λ = 8,370, $p < 10^{-16}$; see SI Appendix, Table S2 and S3 and Fig. S5A and S6A).

Another potential concern is that studies of the rate of protein sequence evolution suggest that levels of evolutionary constraint, and therefore the DFE, strongly depend on expression level and tissue specificity of the genes (46–48). To investigate how differences in gene expression could affect our results, we classify genes into sets with different gene expression profiles (SI Appendix, Fig. S2 and Text S1). We use two recent gene expression datasets from humans (9) and *Drosophila* (10) that provide mRNA expression level estimates in 27 and 29 different tissues, respectively. We computed average expression level, and τ as a measure of tissue specificity for each gene. We further classified genes as low, medium, or highly expressed, and tissue specific ($\tau > 0.6$) or broadly expressed ($\tau < 0.4$). We found differences in the nonsynonymous over synonymous polymorphism ratio for genes with different expression profiles in both species, suggesting differences in constraint, and thus the DFE, between genes (SI Appendix, Fig. S1). However, fitting a gamma DFE to each set of genes using our method, we still find that the average selection coefficient E[*s*] is about 70-110 fold more deleterious for humans than for *Drosophila*, regardless of the overall expression level or tissue specificity of the genes (SI Appendix, Fig. S12A).

In summary, a combination of confounding factors of linkage, uncertainty in demographic parameters, background selection, selective sweeps, interference, selection on synonymous sites, poor numerical optimization, and analyzing different sets of genes cannot account for our finding of different DFEs between human and *Drosophila*.

Text S4

Forward simulations

To compute the null distribution of the likelihood ratio test statistic, Λ, we performed forward simulations under the estimated demographic models for humans and *Drosophila*. Selection coefficients for nonsynonymous mutations were drawn from a gamma distribution with shape and scale parameters estimated from the constrained model (i.e., $\alpha_H = \alpha_D$ and $\beta_H = \beta_D$). We assume a spatial distribution of selected elements that reflects the empirical distribution of coding and conserved non-coding (CNC) sequence in the genome. Further, we simulate varying recombination across the genomes that is based on empirical high-resolution recombination maps (49, 50). Mutations in CNC regions are assumed to be selected with gamma distributed selection coefficients taken from Torgerson et al. (40) for humans and Casillas et al. (41) for *Drosophila*. The exon element ranges where taken from GENCODE v14 (51) for humans and BDGP 6.79 FlyBase gene annotation (52) for *Drosophila*. To define CNC ranges in both species, we used predicted conserved elements by phastCons (53), downloaded from the UCSC genome browser. All forward simulations were carried out using the simulation software 'SLiM' (54). For both species, we simulated under a single size change model with the empirically estimated parameters (SI Appendix, Table S1). Since *Drosophila* has a prohibitively large population size for forward simulations, we simulated both species with an ancestral effective population size of 10,000 and scaled mutation rate, recombination rate, selection coefficients and demographic parameters accordingly (55). To assess power, we performed a different set of simulations assuming the gamma DFE parameter estimates from the full model (SI Appendix, Table S2).

Further, to allow quasi genome-wide simulations, we followed a bootstrapping approach by first simulating 1000 x 7 Mb large regions that were selected randomly from the respective genome. We then selected a centered 3 Mb window from the simulated 7 Mb region and discarded the rest of the sequence to remove edge effects, notably the lower strength of background selection at the edges (56). From those 1000 x 3 Mb regions, we resampled regions until we arrived at a full genome data set, i.e. similar numbers of synonymous and nonsynonymous SNPs as seen in the actual data. That way, we simulated data of 300 independent genomes. In both species, the simulations resulted in considerable amounts of background selection, with average reduction in neutral diversity in the 7Mb region of 10% in humans and 12% in *Drosophila*. For each simulated genome data set, we first estimated the demographic model from the synonymous SFS and then the DFE parameters from the nonsynonymous SFS conditional on the estimated demographic parameters.

Text S5

Theoretical models of DFE evolution

The factors that drive differences in the DFE between humans and *Drosophila* are unclear. As discussed in the main text, several theoretical models make predictions regarding the factors that are influencing the evolution of the DFE, but they have not been tested with population genetic data from natural populations. These five categories of models lead to contrasting predictions regarding DFE differences between *Drosophila* and humans. Here we describe the assumptions and the predictions of the five models that we use to discriminate among them.

Functional importance model

The functional importance of a protein is generally thought to be a major determinant of its evolutionary rate (57), although a number of recent studies have challenged this view (48). Our functional importance model posits that the importance of a protein for fitness is the major determinant of fitness effects of mutations in a protein. The more important a protein is for the fitness of an organism, the more deleterious are mutations at functional sites in that protein. Under the assumption that the distribution of functional importance of proteins stays constant over evolutionary time, this model then suggests that the DFE of *s* of protein-changing mutations is the same between different species. Note that individual proteins may shift in their importance over evolutionary time as long as those shifts cancel each other out such that the overall distribution of functional importance stays the same.

Protein stability models

In contrast to the functional importance model, the basic idea behind protein stability models is that much of the selection pressure on coding regions involves maintaining thermodynamic stability of the proteins (46, 47, 58). Fitness of a protein is a concave function of "protein folding stability", the difference in free energy between the folded and the unfolded protein state. The shape of the function is defined by the fraction of proteins that fold at equilibrium (58), the cytotoxic effects of protein misfolding (59), or the effect of enzyme stability on metabolic flux (59). The distribution of $N_e s$ values that is generated by such a one-dimensional fitness-phenotype relationship was shown to be gamma distributed (59) and independent of the effective population size (N_e) when at equilibrium (58). Thus, this model predicts that $N_e s$ is the same across taxa. Since the effective population size (N_e) in *Drosophila* is estimated to be 120-415-fold larger in humans, this model predicts that $|s|$ must be at least 120-fold smaller (i.e. less deleterious) in *Drosophila*. We observe a factor of less than 100 (SI Appendix, Fig. S8). The discrepancy is even more extreme in a comparison of humans with mouse, where complexity might be considered more comparable. Here, effective population size is estimated to be about 40-fold larger than in humans, thus $|s|$ must be 40-fold less deleterious in mouse. We estimate a factor of only 4 (Fig. 4A), which is inconsistent with the prediction of the protein stability model.

Note that a major assumption made by protein stability models is equimutability (60), meaning that the effect size distribution of a mutation on protein stability ($\Delta\Delta G$) is independent of the stability of the wild-type (ΔG). This assumption has been verified experimentally and with simulations (61–63), however, some authors propose a negative correlation between $\Delta\Delta G$ and ΔG (59, 64). Another key assumption is constant population size. Long-term fluctuating population size was shown to have an effect on the DFE from protein stability models when

averaged over time (58). Further studies on the relevance of such deviations from model assumptions for the expected DFE under protein stability models are warranted.

Back-mutation models

Back-mutations restore the ancestral state after a previous mutation has occurred. For example, after an A to G mutation has fixed in a population, a new G to A mutation would be referred to as a back-mutation. Back-mutation models rest on the assumption that if there is a category of slightly deleterious mutations that fix in the population, then there should also be a category of slightly advantageous back-mutations, i.e. mutations back to the ancestral state (65). These models usually make two assumptions (65–68): 1) the back-mutation has the same absolute value of the selection coefficient as the forward-mutation, but with opposite sign, and 2) there is no epistasis, i.e. the selection coefficient of a mutation is independent of the genetic background. Back-mutations models predict that in small populations, the proportion of slightly beneficial mutations is greater than in large populations, because more slightly deleterious mutations can become fixed in small populations, leading to more opportunities for new beneficial back-mutations. However, this mechanism is slow, since it relies on the occurrence of new mutations and their fixation. Thus, this model predicts that long-term effective population size ($N_{e,long\text{-}term}$) is a key factor determining the DFE.

Piganeau and Eyre-Walker (23), eq. 7, derived a formula for the equilibrium DFE as a function of population size (see also eq. 6 in Rice et al. (68)). To test predictions of the back-mutation model, we use the formula of Piganeau and Eyre-Walker within our Poisson Random Field method by assuming gamma distributed effect sizes (i.e., $|s| \sim \Gamma(\alpha, \beta)$). Under the back-mutation model, any difference in the DFE between two populations is the result of the effect of different long-term population sizes on the DFE, leading to a shift in the proportion of deleterious vs. beneficial mutations. The fixation of deleterious mutations in small populations leads to an increase in beneficial back-mutations with same selection coefficient, but opposite sign (i.e. same effect size $|s|$). Importantly, this process does not change the effect size distribution or the average effect size of new mutations (i.e. $|s|$). Thus, we can test this model by testing if $E[|s|]$ is the same in both populations.

Mutational robustness models

Mutational robustness models postulate that more robust organisms have, on average, less deleterious mutations (69). There are several mechanisms that could lead to increased or decreased levels of robustness. Kimura was one of the first to suggest that more complex organisms have a higher level of physiological homeostasis than less complex organisms, which should lead to a larger proportion of neutral mutations and thus more robustness (70). Later theoretical work and computer simulations supported this idea by showing that robustness emerges directly as a property of complex metabolic and regulatory networks, and that more highly connected networks have higher robustness (71–74). In other studies, robustness is not an intrinsic property of the system, but evolves directly under natural selection. For example, two-locus models have been developed where mutations at the first (modifier) locus reduce the deleterious selection coefficient of mutations at the second locus (75). Modifier mutations fix because they reduce the mutational load of the robust lineages. A specific example of such modifier mutations is mutations that increase the expression level of heat shock proteins. Heat shock proteins aid in correct folding and enhance stability of proteins. They allow mutated proteins to retain their correct function and thus reduce the deleterious effect of the mutation

(76). In the context of Fisher's Geometrical Model (see the next model description), evolution of robustness is modeled by allowing modifier mutations to affect the flatness of the fitness function, such that the same mutation has a smaller effect on fitness in a more robust organisms than in less robust ones (77). In such models, smaller populations tend to evolve higher levels of robustness (78), and this tendency is increased by higher phenotypic complexity and more positive epistasis (77).

Irrespective of which factor is driving the evolution of robustness, these models predict greater mutational robustness in humans than in *Drosophila*, because humans are more complex and have a smaller effective population size compared to *Drosophila*. Further, robustness models predict that less pleiotropic mutations are more deleterious, since the smaller effective complexity of such mutations impedes the evolution of robustness.

Fisher's Geometrical Model (FGM)

In the final model, FGM, phenotypes are represented as points in a multidimensional space, and fitness is a decreasing function of the distance to the optimal phenotype (69). The dimensionality of the phenotype space is termed "complexity". Mutations are represented as random vectors that change the current multivariate phenotype to a random new phenotype with a new fitness value. Recent studies have increased the realism of FGM, for example by allowing for mutational and selectional correlations (36, 79), or by relaxing the assumption that every mutation affects every phenotype (restricted pleiotropy; see (24, 36, 80)). Explicit solutions for the shape of the DFE have been derived under the assumption that both the fitness function and the mutational distribution are Gaussian functions (24, 36, 69, 80).

Here, we test three predictions of FGM (see also Fig. 1): 1) Mutations in more complex organisms are on average more deleterious, since mutations are more likely to disrupt something important in a complex organism than in a simple one (81, 82) (see SI Appendix, Text S7 for assumptions that go into this prediction). 2) Smaller populations have a larger proportion of beneficial mutations due to increased fixation of deleterious mutations in smaller populations (drift load; e.g. (24)). 3) Less pleiotropic mutations have a greater variance in fitness effects, i.e. selection coefficients tend to be either close to neutral or very deleterious (36).

The first prediction suggests that humans have more severe deleterious mutations than *Drosophila*, since humans likely have higher complexity (larger number of genes, proteins and protein-protein interactions (83), and cell types (84)). Stronger stabilizing selection and larger effect sizes of mutations in humans than in *Drosophila* could also contribute to more deleterious mutations under FGM, but seem less parsimonious as an explanation (36).

The second prediction suggests that because of their smaller effective population size, humans contain more slightly beneficial mutations than *Drosophila*. Note that, given a certain distance to the optimum, a higher complexity decreases the range of directions in phenotype space that lead toward the phenotypic optimum, decreasing the probability that a random mutation increases fitness. Thus, the higher human complexity might counteract the effect of small population size. However, increasing complexity also has a second, opposite effect: it increases the distance to the optimum because of an increase in drift load, and this in turn increases the proportion of beneficial mutations again (24). In the formulation of Lourenço et al. (24), these two effects of complexity on the proportion of beneficial mutations cancel out and, at equilibrium, there is no effect of complexity on the proportion of beneficial mutations. We tested the prediction of a larger proportion of beneficial mutations in humans by fitting a DFE that is

based on FGM as derived by Lourenço et al. (24). Similar to the back-mutation model of Piganeau and Eyre-Walker (23), in this model a smaller long-term effective population size parameter ($N_{e,long\text{-}term}$) leads to fixation of slightly deleterious mutations due to less effective selection. This increases the distance of the population to the optimum in FGM and thus leads to an increase of the proportion of slightly beneficial compensatory mutations at equilibrium. Note that the long-term effective population size, and therefore effectiveness of selection, could be reduced by recurrent selective sweeps as well as background selection. This would, however, not affect the prediction under FGM that species with strongly different effective population sizes, as estimated from neutral diversity, reside at different distances to the optimum and thus have different proportions of beneficial mutations.

The last prediction of FGM is related to the pleiotropy of mutations, i.e. the effective number of fitness-related phenotypes that are affected by a mutation (24, 80). Pleiotropy changes the "total effect" of a mutation on the phenotypes, which is the length of the random mutational vector in the phenotype space of FGM. The total effect can be shown to follow a generalized chi-distribution, where the degrees of freedom are equal to the pleiotropy (24). Thus, most mutations with large pleiotropy have intermediate effects on phenotype and therefore a smaller variation in s than mutations with little pleiotropy. In fact, the coefficient of variation of selection coefficients CV(s) is inversely related to the pleiotropy of mutations (36). Since pleiotropy is difficult to measure directly in higher organisms, we use tissue specificity of gene expression as a proxy for pleiotropy. Mutations in genes that are expressed in more tissues would be assumed to have more pleiotropic effects than mutations in genes expressed in fewer tissues. Pleiotropy can be estimated by fitting the DFE formula of Lourenço et al. (24), eq. 15, to the data using the Poisson Random Fields approach (see SI Appendix, Text S2). Equation 15 of Lourenço et al. is derived from FGM, assuming a Gaussian fitness function. It contains an explicit pleiotropy parameter (m) that is defined as the number of selected traits affected by a mutation and that is smaller than the total number of selected traits (complexity) n. It considers a haploid population of effective size N_e that is under selection-mutation-drift equilibrium. A non-equilibrium version of the DFE under this model was derived as well (eq. 8 of Lourenço et al. (24)) that can be used to test for deviations from equilibrium conditions.

Text S6

Model choice procedure

Here we present in short the logic of the model choice procedure that we applied to discriminate between the five evolutionary models of Fig. 1 based on our data. The five models make distinct predictions on how the DFE differs between species with different population size and/or complexity. Due to the higher quality and sample size of data from humans and *Drosophila*, we first use those two species to discriminate among the models. However, we then add estimates of the average selection coefficient from additional data from mice and yeast to support our model choice in a larger phylogenetic context. Note, however, that more subtle differences in the shape of the DFE or in the amount of beneficial mutations between species cannot be tested with those datasets (i.e. fitting different functional forms of the DFE result in the same likelihood, see SI Appendix, Table S4).

The first step of our model choice procedure is to show conclusively that the null assumption of the same distribution of s in both species is violated. We develop a likelihood ratio test and use extensive forward simulations to derive the null distribution of the test statistic and show robustness to a multitude of potential confounding factors (see Online Methods and SI Appendix, Text S3 and S4). The same approach can be extended to test the distribution of $N_e s$. These two tests allow evaluation of the functional importance model and the protein stability model, respectively. Similarly, we test for differences in E[|s|], the average absolute selection coefficient. The back-mutation model is inconsistent with large differences in E[|s|] between species because the model predicts that the distribution of |s| is the same between species. Thus, this pattern can be used to test the back-mutation model. Using different functional forms of the DFE to estimate E[|s|], we show that any established difference is not sensitive to the specific functional form of the DFE (SI Appendix, Fig. S8A).

After establishing differences in the distribution of s and |s|, the direction of the difference in E[s] between species discriminates between robustness models and FGM: robustness models predict a less deleterious DFE in the more complex organism (less negative E[s]), whereas FGM predicts a more deleterious DFE in the more complex organism (more negative E[s]). Using different functional forms of the DFE to estimate E[s], we show that any established difference is not sensitive to the specific functional form of the DFE (SI Appendix, Fig. S8B). Using data from mice and yeast we show that the trend observed in humans and *Drosophila* can be replicated in different datasets in a wider phylogenetic context (Fig. 4A). Finally, we use more subtle predictions of the influence of population size and pleiotropy on the shape of the DFE to further validate FGM. In particular, these include predictions about how population size affects the proportion of beneficial mutations, and how pleiotropy affects the variation in s (see SI Appendix, Text S5).

Text S7

The influence of complexity and pleiotropy on the average selection coefficient in FGM

Here we discuss how complexity and pleiotropy can influence the DFE under FGM. First, we define a few key variables. Let n be the total complexity of the organism, or the number of independent and evolvable phenotypes an organism exposes to natural selection. A somewhat related quantity, n_e, is the effective number of traits exposed to selection after accounting for mutational and selective correlations (36). Lastly, mutational pleiotropy, m, refers to the number of traits affected by a particular mutation.

Under the classical version of FGM, mutations are assumed to affect all phenotypic axes similarly and have no preferred direction (universal pleiotropy). In this model, $n=n_e=m$. More specifically, a mutation changes the multivariate phenotype, and the vector that connects the pre-mutant phenotype with the new phenotype is the mutational vector. For a random mutation, this vector has a random direction and length r. Further, fitness follows the Gaussian function, $w = \exp(-z^2)$, where z is the distance from the optimum. It can be shown that under this model, the mean selection coefficient of a random mutation is $E[s|r] = -1/2\ r^2$ (85). Thus, the average selection coefficient $E[s]$ is predicted to be negative, and becomes more negative with increasing r. Increasing complexity means that more and more dimensions are added to the phenotype space. As long as the average effect of the mutation on each individual phenotype does not decrease with increasing complexity, this added dimensionality inevitably leads to an increase of the average length of the mutational vector. Since a longer mutational vector length (i.e. larger r) leads to more negative $E[s]$, this suggests that the DFE becomes more deleterious in more complex organisms.

However, recent high-throughput methods and quantitative genomics approaches seem to reject the concept of universal pleiotropy (reviewed in Wagner and Zhang (86)). They suggest that most mutations only affect a small number of phenotypes (restricted or partial pleiotropy). Furthermore, estimates of the effective number of traits n_e arrive at unrealistically small values (36). Lourenço et al. (24) suggest that the low estimates of n_e are more likely reflecting mutational pleiotropy m than the total complexity of the organism (n). Again, the low estimates of mutational pleiotropy suggest that most mutations effectively change only a small number of phenotypes (24).

Given that universal pleiotropy is unlikely, we would like to know how $E[s]$ depends on the total complexity of the organism under models that relax this assumption. However, different formulations of FGM come to different conclusions about the dependency of $E[s]$ on n, n_e, and m. In the model of Martin and Lenormand (36) (eq. 1) and Chevin et al (80) (eq. 3c), for a given strength of selectional and mutational correlations (ρ_S and ρ_M), E[s] is negatively related to n. Thus, their prediction is matching the prediction of the classical formulation of FGM.

In Lourenço et al. (24), $E[s]$ does not depend on n (see eq. A3 in Lourenço et al. (24)). Instead, here, the DFE depends on the pleiotropy parameter m and the scale parameter σ. However, the model of Martin and Lenormand and the model of Lourenço et al. can be shown to be equivalent (24). The scale parameter of Lourenço, σ, corresponds to the effective scale parameter λ_e ($\sigma = \mathrm{sqrt}(\lambda_e/2)$) of Martin and Lenormand (36) and Chevin et al. (80). The effective complexity, n_e, of Martin and Lenormand corresponds to the pleiotropy parameter of Lourenço et al. ($m = n_e$). In the Lourenço model, the average selection coefficient $E[s]$ is a product of the

scale and pleiotropy parameters ($E[s] = - m \sigma^2 = - 1/2 n_e \lambda_e$). However, whereas the scale parameter (λ_e) increases with total complexity *n* under the model of Martin and Lenormand (see Appendix 2 in Martin and Lenormand (36)), the scale parameter in Lourenço et al. (σ) is fixed and not related to *n* in their study (eq. 2 in Lourenço et al. (24)). Thus, the behavior of $E[s]$ with *n* depends on how the scale parameter changes with total complexity *n*, similar to the result derived in the classical model.

We suggest that it is biologically and experimentally justified to assume that the scale parameter σ (or λ_e) positively depends on complexity (*n)*. Biologically, it is realistic to assume that when total complexity *n* is increased, mutations also affect these new phenotypes. If more phenotypes are affected on average per mutation, this increases the length of the mutational vector, even if this happens in a strongly correlated manner and thus keeping pleiotropy in the sense of 'effective dimensionality' n_e low. Furthermore, experimental data seem to support the assumption of increasing σ with increasing complexity, since mutations that affect more phenotypes also show a larger effect on each individual phenotype (87, 88). Finally, our results also suggest that the scaling parameter σ increases with increasing complexity of the organism, whereas the pleiotropy parameter *m* does not show such a trend (SI Appendix, Fig. S11). In agreement with experimental data, our results also support a model where σ (or λ_e) is not constant, but increases with increasing complexity. Thus, the increase in deleteriousness with increasing complexity of the organism shown in our work and in experimental data from Martin and Lenormand (36) is in line with the prediction of FGM.

Additional References

1. 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491(7422):56–65.

2. Lack JB, et al. (2015) The *Drosophila* genome nexus: a population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. *Genetics* 199(4):1229–1241.

3. Keightley PD, et al. (2009) Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome Res* 19(7):1195–1201.

4. Poh Y-P, Ting C-T, Fu H-W, Langley CH, Begun DJ (2012) Population genomic analysis of base composition evolution in *Drosophila melanogaster*. *Genome Biol Evol* 4(12):1245–1255.

5. Bainbridge MN, et al. (2011) Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. *Genome Biol* 12(7):R68.

6. Wang J, Raskin L, Samuels DC, Shyr Y, Guo Y (2015) Genome measures used for quality control are dependent on gene function and ancestry. *Bioinformatics* 31(3):318–323.

7. Arndt PF, Hwa T (2005) Identification and measurement of neighbor-dependent nucleotide substitution processes. *Bioinformatics* 21(10):2322–2328.

8. Hernandez RD, Williamson SH, Bustamante CD (2007) Context dependence, ancestral misidentification, and spurious signatures of natural selection. *Mol Biol Evol* 24(8):1792–1800.

9. Fagerberg L, et al. (2014) Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics* 13(2):397–406.

10. Li JJ, Huang H, Bickel PJ, Brenner SE (2014) Comparison of *D. melanogaster* and *C. elegans* developmental stages, tissues, and cells by modENCODE RNA-seq data. *Genome Res* 24(7):1086–1101.

11. Gossmann TI, Keightley PD, Eyre-Walker A (2012) The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. *Genome Biol Evol* 4(5):658–667.

12. Koufopanou V, Lomas S, Tsai IJ, Burt A (2015) Estimating the fitness effects of new mutations in the wild yeast *Saccharomyces paradoxus*. *Genome Biol Evol* 7(7):1887–1895.

13. Kousathanas A, Keightley PD (2013) A comparison of models to infer the distribution of fitness effects of new mutations. *Genetics* 193(4):1197–1208.

14. Fay JC, Benavides JA (2005) Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLOS Genet* 1(1):e5.

15. Uchimura A, et al. (2015) Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. *Genome Res* 25(8):1125–1134.

16. Sawyer SA, Hartl DL (1992) Population genetics of polymorphism and divergence. *Genetics* 132(4):1161–1176.

17. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* 5(10):e1000695.

18. Boyko AR, et al. (2008) Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet* 4(5):e1000083.

19. Williamson SH, et al. (2005) Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proc Natl Acad Sci USA* 102(22):7882–7887.

20. Kim BY, Huber CD, Lohmueller KE (2016) Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. bioRxiv 071431 doi:10.1101/071431.

21. Lawrie DS, Petrov DA (2014) Comparative population genomics: power and principles for the inference of functionality. *Trends Genet* 30(4):133–139.

22. Pawitan Y (2013) *In All Likelihood: Statistical Modelling And Inference Using Likelihood* (Oxford University Press, USA, Oxford).

23. Piganeau G, Eyre-Walker A (2003) Estimating the distribution of fitness effects from DNA sequence data: implications for the molecular clock. *Proc Natl Acad Sci USA* 100(18):10335–10340.

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

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

26. Li H, Stephan W (2006) Inferring the demographic history and rate of adaptive substitution in *Drosophila*. *PLoS Genet* 2(10):e166.

27. Ségurel L, Wyman MJ, Przeworski M (2014) Determinants of mutation rate variation in the human germline. *Annu Rev Genomics Hum Genet* 15:47–70.

28. Sharp PM, Li W-H (1989) On the rate of DNA sequence evolution in *Drosophila*. *J Mol Evol* 28(5):398–402.

29. Keightley PD, Ness RW, Halligan DL, Haddrill PR (2014) Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. *Genetics* 196(1):313–320.

30. Keightley PD, et al. (2015) Estimation of the spontaneous mutation rate in *Heliconius melpomene*. *Mol Biol Evol* 32(1):239–243.

31. McVicker G, Gordon D, Davis C, Green P (2009) Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet* 5(5):e1000471.

32. Comeron JM (2014) Background selection as baseline for nucleotide variation across the *Drosophila* Genome. *PLoS Genet* 10(6):e1004434.

33. Nicolaisen LE, Desai MM (2013) Distortions in genealogies due to purifying selection and recombination. *Genetics* 195(1):221–230.

34. 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(4):2251–2261.

35. Keightley PD, Campos JL, Booker TR, Charlesworth B (2016) Inferring the frequency spectrum of derived variants to quantify adaptive molecular evolution in protein-coding genes of *Drosophila melanogaster*. *Genetics* DOI: 10.1534/genetics.116.188102.

36. Martin G, Lenormand T (2006) A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evol Int J Org Evol* 60(5):893–907.

37. Bustamante CD, Wakeley J, Sawyer S, Hartl DL (2001) Directional selection and the site-frequency spectrum. *Genetics* 159(4):1779–1788.

38. Zhu L, Bustamante CD (2005) A composite-likelihood approach for detecting directional selection from DNA sequence data. *Genetics* 170(3):1411–1421.

39. Ewing GB, Jensen JD (2015) The consequences of not accounting for background selection in demographic inference. *Mol Ecol 25(1)*:135–141.

40. Torgerson DG, et al. (2009) Evolutionary processes acting on candidate cis-regulatory regions in humans inferred from patterns of polymorphism and divergence. *PLoS Genet* 5(8):e1000592.

41. Casillas S, Barbadilla A, Bergman CM (2007) Purifying selection maintains highly conserved noncoding sequences in *Drosophila*. *Mol Biol Evol* 24(10):2222–2234.

42. Sella G, Petrov DA, Przeworski M, Andolfatto P (2009) Pervasive natural selection in the *Drosophila* genome? *PLoS Genet* 5(6):e1000495.

43. Lawrie DS, Messer PW, Hershberg R, Petrov DA (2013) Strong purifying selection at synonymous sites in *D. melanogaster*. *PLoS Genet* 9(5):e1003527.

44. Parsch J, Novozhilov S, Saminadin-Peter SS, Wong KM, Andolfatto P (2010) On the utility of short intron sequences as a reference for the detection of positive and negative selection in *Drosophila*. *Mol Biol Evol* 27(6):1226–1234.

45. Clemente F, Vogl C (2012) Unconstrained evolution in short introns? - an analysis of genome-wide polymorphism and divergence data from *Drosophila*. *J Evol Biol* 25(10):1975–1990.

46. Drummond DA, Wilke CO (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134(2):341–352.

47. Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH (2005) Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci USA* 102(40):14338–14343.

48. Zhang J, Yang J-R (2015) Determinants of the rate of protein sequence evolution. *Nat Rev Genet* 16(7):409–420.

49. Comeron JM, Ratnappan R, Bailin S (2012) The many landscapes of recombination in *Drosophila melanogaster*. *PLoS Genet* 8(10):e1002905.

50. Kong A, et al. (2010) Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* 467(7319):1099–1103.

51. Harrow J, et al. (2012) GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 22(9):1760–1774.

52. dos Santos G, et al. (2015) FlyBase: introduction of the *Drosophila melanogaster* Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Res* 43(D1):D690–D697.

53. Siepel A, et al. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 15(8):1034–1050.

54. Messer PW (2013) SLiM: simulating evolution with selection and linkage. *Genetics* 194(4):1037–1039.

55. Aberer AJ, Stamatakis A (2013) Rapid forward-in-time simulation at the chromosome and genome level. *BMC Bioinformatics* 14(1):216.

56. Nordborg M, Charlesworth B, Charlesworth D (1996) The effect of recombination on background selection. *Genet Res* 67(02):159–174.

57. Kimura M, Ohta T (1974) On some principles governing molecular evolution. *Proc Natl Acad Sci USA* 71(7):2848–2852.

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

59. Serohijos AWR, Shakhnovich EI (2014) Contribution of selection for protein folding stability in shaping the patterns of polymorphisms in coding region. *Mol Biol Evol* 31(1):165–176.

60. Cherry JL (1998) Should we expect substitution rate to depend on population size? *Genetics* 150(2):911–919.

61. Bloom JD, Raval A, Wilke CO (2007) Thermodynamics of neutral protein evolution. *Genetics* 175(1):255–266.

62. Goldstein RA (2011) The evolution and evolutionary consequences of marginal thermostability in proteins. *Proteins* 79(5):1396–1407.

63. Tokuriki N, Stricher F, Schymkowitz J, Serrano L, Tawfik DS (2007) The stability effects of protein mutations appear to be universally distributed. *J Mol Biol* 369(5):1318–1332.

64. Serohijos AWR, Rimas Z, Shakhnovich EI (2012) Protein biophysics explains why highly abundant proteins evolve slowly. *Cell Rep* 2(2):249–256.

65. Charlesworth J, Eyre-Walker A (2007) The other side of the nearly neutral theory, evidence of slightly advantageous back-mutations. *Proc Natl Acad Sci USA* 104(43):16992–16997.

66. Bulmer M (1991) The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129(3):897–907.

67. Mustonen V, Lässig M (2009) From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. *Trends Genet* 25(3):111–119.

68. Rice DP, Good BH, Desai MM (2015) The evolutionarily stable distribution of fitness effects. *Genetics* 200(1):321–329.

69. Tenaillon O (2014) The utility of Fisher's geometric model in evolutionary genetics. *Annu Rev Ecol Evol Syst* 45(1):179–201.

70. Kimura M (1979) Model of effectively neutral mutations in which selective constraint is incorporated. *Proc Natl Acad Sci USA* 76(7):3440–3444.

71. Azevedo RBR, Lohaus R, Srinivasan S, Dang KK, Burch CL (2006) Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. *Nature* 440(7080):87–90.

72. Kacser H, Burns JA (1981) The molecular basis of dominance. *Genetics* 97(3-4):639–666.

73. van Nimwegen E, Crutchfield JP, Huynen M (1999) Neutral evolution of mutational robustness. *Proc Natl Acad Sci USA* 96(17):9716–9720.

74. Siegal ML, Bergman A (2002) Waddington's canalization revisited: developmental stability and evolution. *Proc Natl Acad Sci USA* 99(16):10528–10532.

75. Gardner A, Kalinka AT (2006) Recombination and the evolution of mutational robustness. *J Theor Biol* 241(4):707–715.

76. Siegal ML, Leu J-Y (2014) On the nature and evolutionary impact of phenotypic robustness mechanisms. *Annu Rev Ecol Evol Syst* 45(1):495–517.

77. Gros P-A, Tenaillon O (2009) Selection for chaperone-like mediated genetic robustness at low mutation rate: impact of drift, epistasis and complexity. *Genetics* 182(2):555–564.

78. Krakauer DC, Plotkin JB (2002) Redundancy, antiredundancy, and the robustness of genomes. *Proc Natl Acad Sci USA* 99(3):1405–1409.

79. Waxman D (2006) Fisher's geometrical model of evolutionary adaptation--beyond spherical geometry. *J Theor Biol* 241(4):887–895.

80. Chevin L-M, Martin G, Lenormand T (2010) Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution. *Evol Int J Org Evol* 64(11):3213–3231.

81. Orr HA (1998) The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52(4):935–949.

82. Fisher RA (1930) *The genetical theory of natural selection* (Clarendon Press, Oxford, England).

83. Stumpf MPH, et al. (2008) Estimating the size of the human interactome. *Proc Natl Acad Sci USA* 105(19):6959–6964.

84. Valentine JW, Collins AG, Meyer CP (1994) Morphological complexity increase in metazoans. *Paleobiology* 20(2):131–142.

85. Waxman D, Peck JR (1998) Pleiotropy and the preservation of perfection. *Science* 279(5354):1210–1213.

86. Wagner GP, Zhang J (2011) The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nat Rev Genet* 12(3):204–213.

87. Wagner GP, et al. (2008) Pleiotropic scaling of gene effects and the "cost of complexity." *Nature* 452(7186):470–472.

88. Wang Z, Liao B-Y, Zhang J (2010) Genomic patterns of pleiotropy and the evolution of complexity. *Proc Natl Acad Sci USA* 107(42):18034–18039.

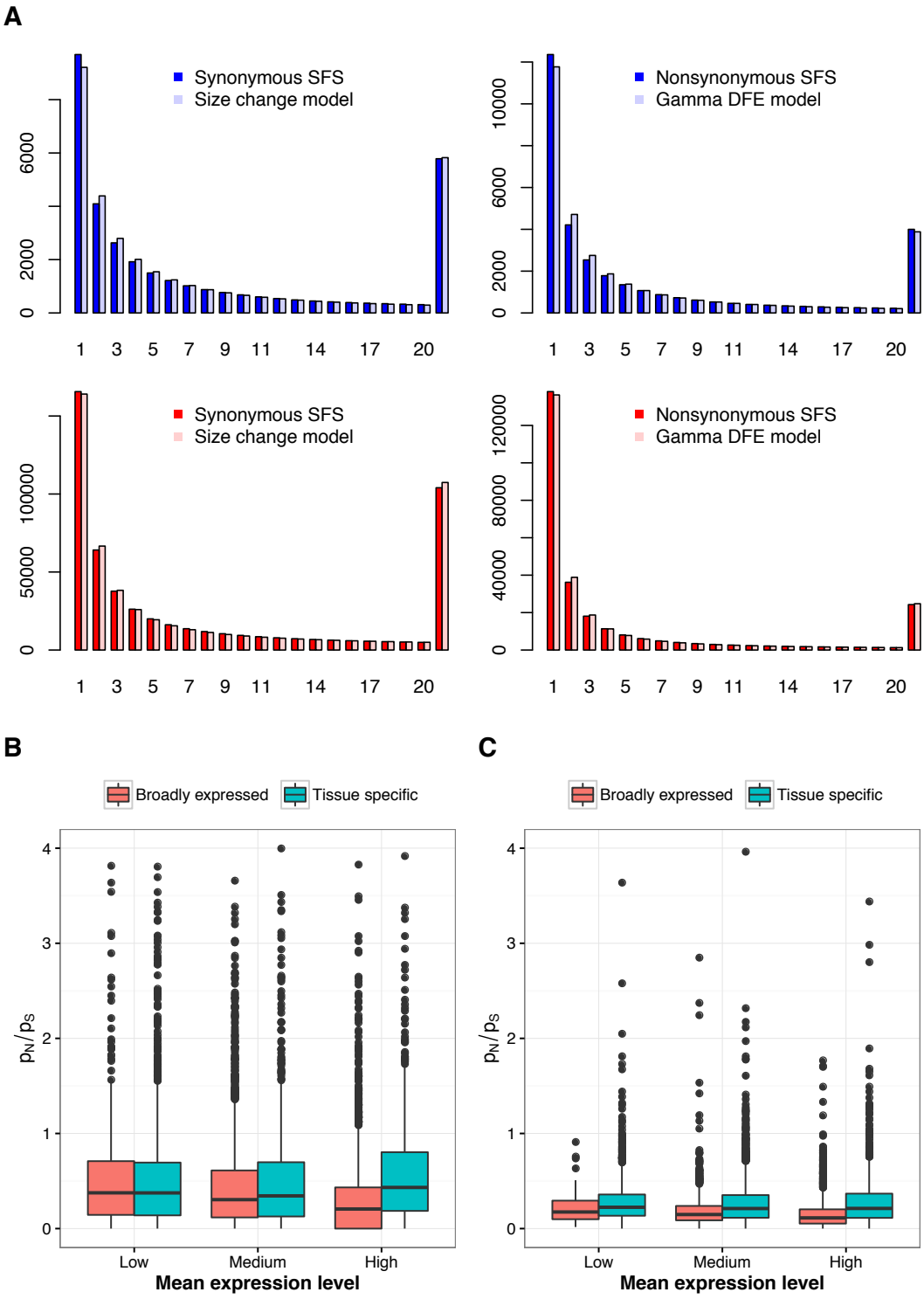

Fig S1. Different diversity patterns between humans and *Drosophila*. (A) The folded synonymous and nonsynonymous SFS for humans (blue) and *Drosophila* (red). The expected SFS under the MLEs of the model parameters are shown in light colors. The x-axis is binned according to the minor allele frequency. Sites with minor allele frequency of 21-50 are combined into the last bin. (B,C) Boxplot of the distribution of nonsynonymous to synonymous polymorphism ratio (p_N/p_S) per gene, for humans (B) and *Drosophila* (C). Results are shown for three different overall expression levels and two levels of tissue specificity (see main text). Broadly expressed genes have $\tau < 0.4$, tissue specific genes have $\tau > 0.6$.

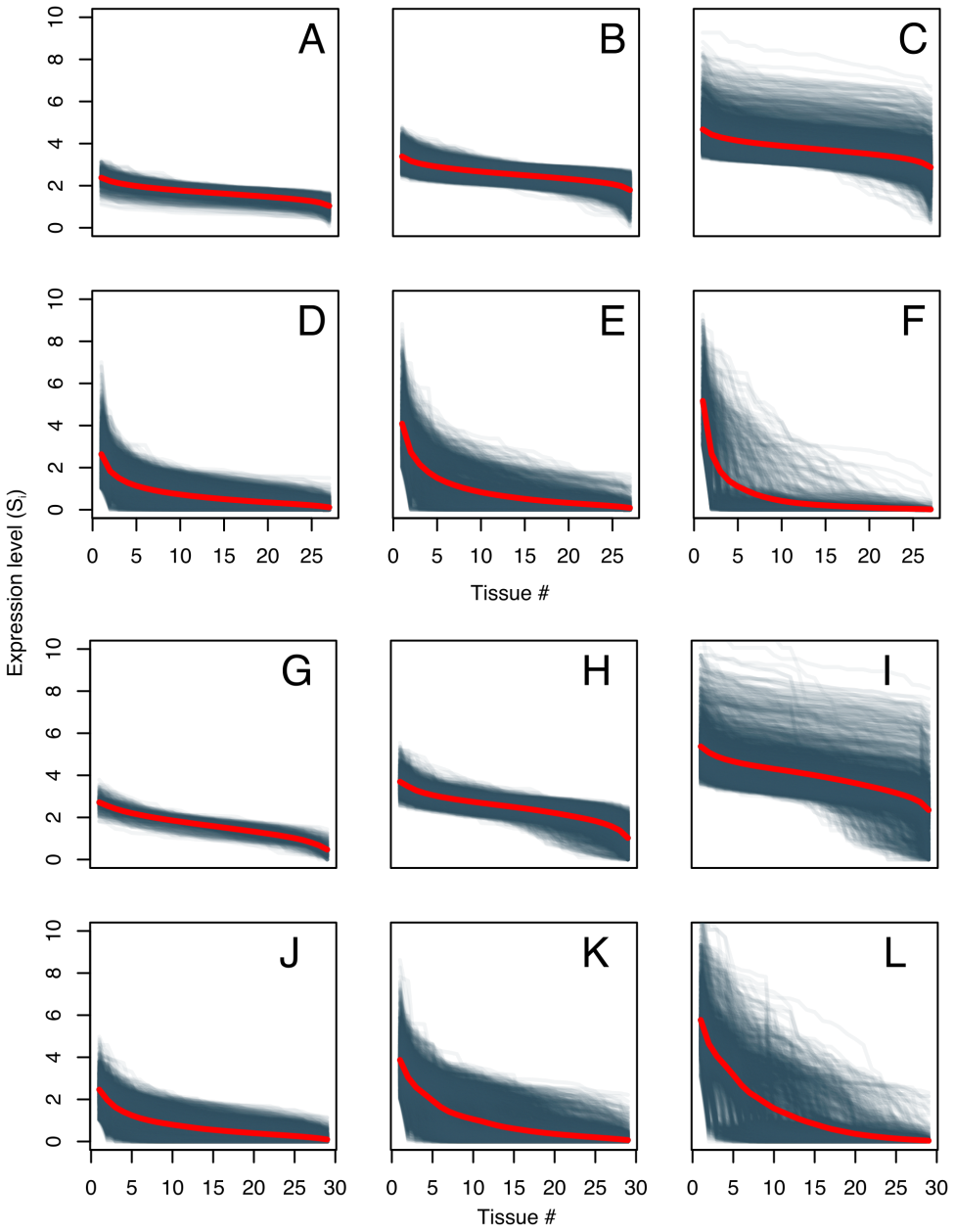

Fig. S2. Expression profiles for human and *Drosophila* genes. (A-F) Expression profiles for human genes. (G-L) Expression profiles for *Drosophila* genes. Each grey line represents a gene. For each gene, the tissue is ordered according to the expression level, i.e. expression level is plotted in decreasing order, beginning with the tissue with the largest expression level. Genes are classified into broadly expressed genes (A, B, C, G, H, I) and tissue-specific genes (D, E, F, J, K, L), and into low (A, D, G, J), intermediate (B, E, H, K), and highly (C, F, I, L) expressed genes (see SI Appendix, Text S1 for definitions). The red line represents the average across genes.

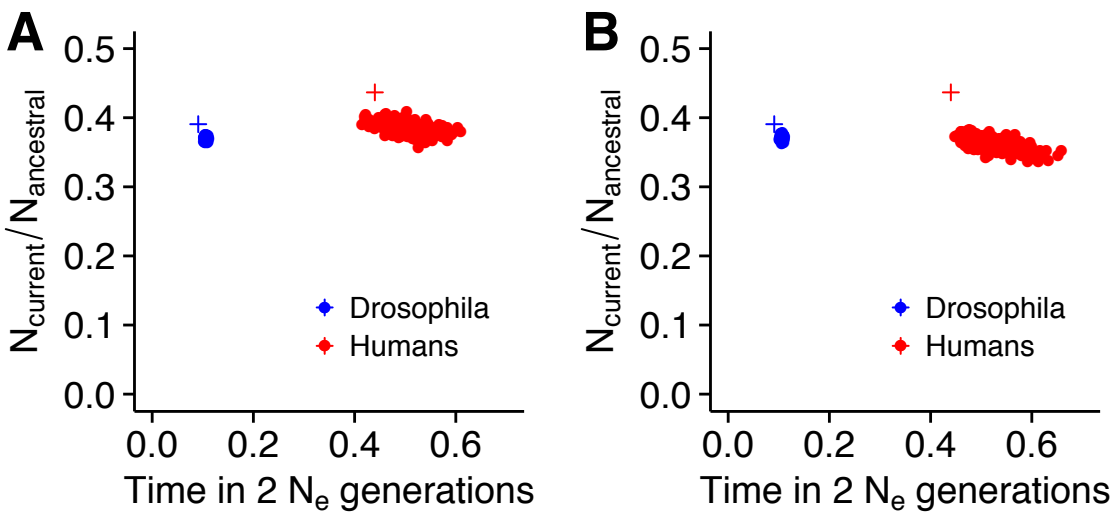

Fig. S3. Demographic parameter estimation for 300 simulated data sets. True parameter values are shown as crosses, estimated parameters as points. (A) Simulations under the full model. (B) Simulations under the constrained model (see main text). Note that the demographic parameter estimates are biased due to background selection and linked selection. However, DFE parameter estimates are unbiased (see Figs. 3A,B).

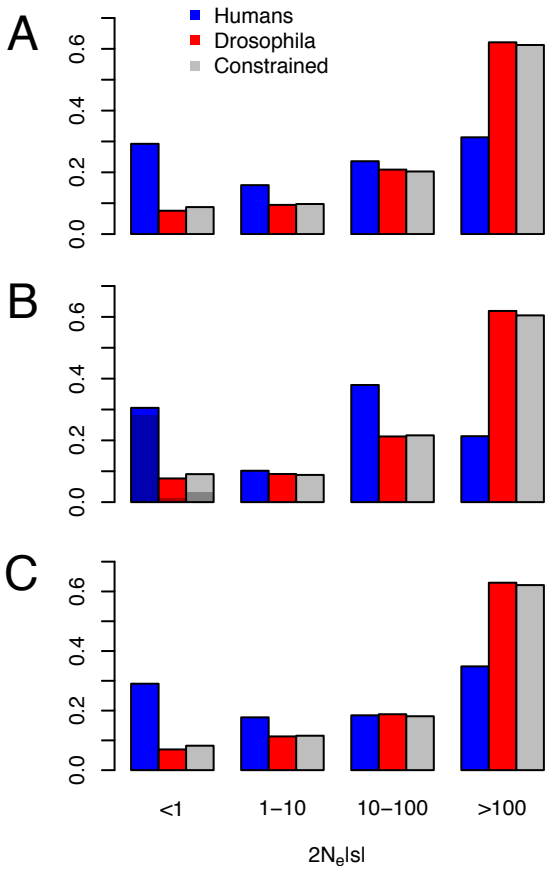

Fig. S4. The proportion of new mutations for various ranges of $2N_e|s|$. Proportions are computed from the estimated (A) gamma distribution, (B) mixture of gamma distribution with neutral point mass, and (C) log-normal distribution. The grey bars indicate the proportions under the null hypothesis of the same distribution of $2N_e|s|$ in both species (constrained model). Darker colors in (B) reflect the estimated proportions of neutral mutations.

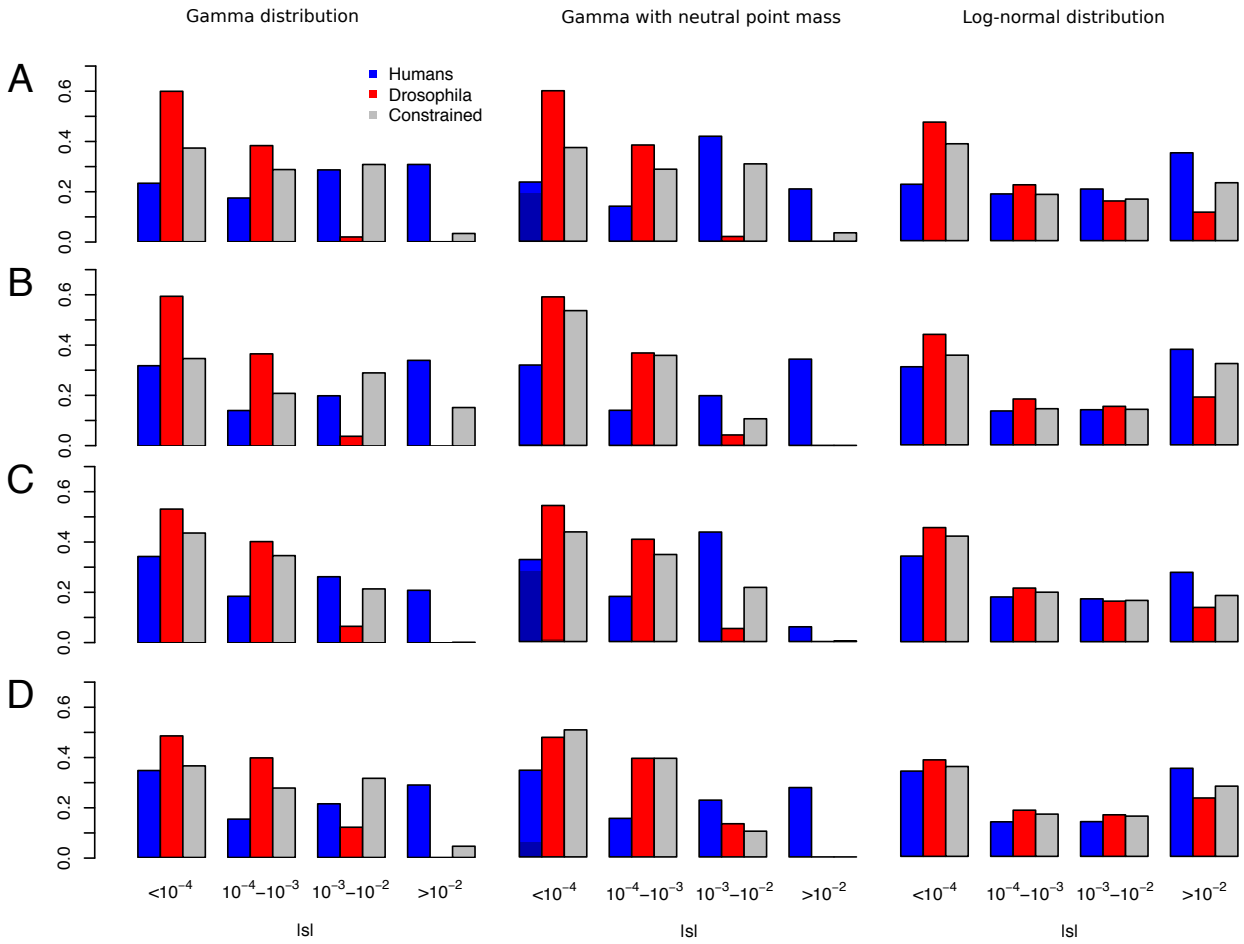

Fig. S5. The proportion of new mutations for various ranges of $|s|$. (A) Using data filtered for genes that have orthologs in both species. (B) Using data after filtering out singletons. (C) Assuming the recent mutation rate estimates (SI Appendix, Text S2). (D) Using the recent mutation rate estimates and filtering out singletons. Proportions are computed from the estimated gamma distribution (left column), mixture of gamma distribution with neutral point mass (middle column), and log-normal distribution (right column). The grey bars indicate the proportions under the null hypothesis of the same distribution of $|s|$ in both species (constrained model). Darker colors in the middle column reflect the estimated proportions of neutral mutations.

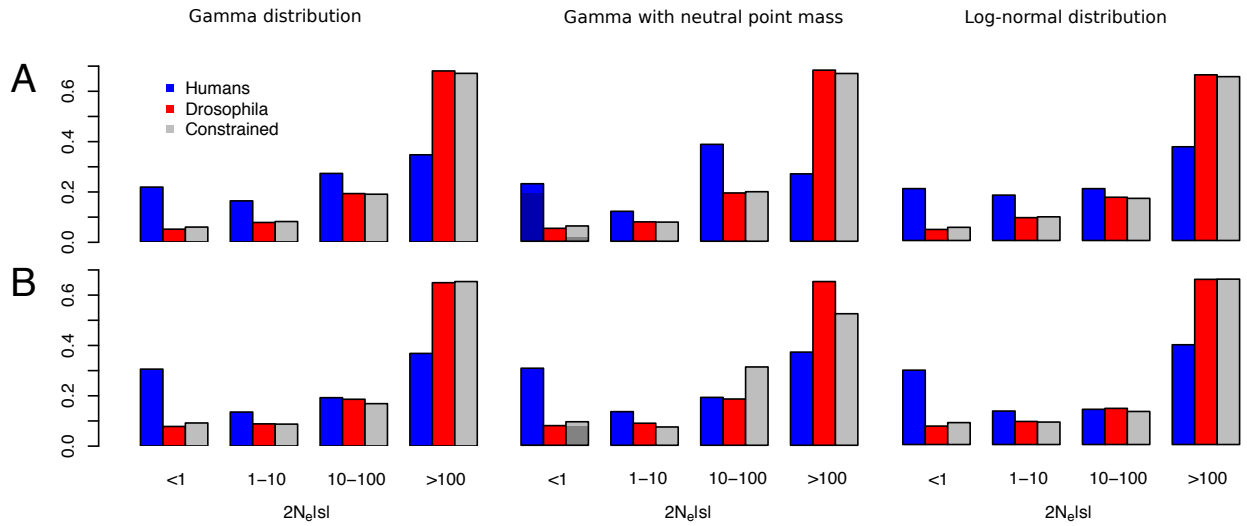

Fig. S6. The proportion of new mutations for various ranges of $2N_e|s|$. (A) Using data filtered for genes that have orthologs in both species. (B) Using data after filtering out singletons. Proportions are computed from the estimated gamma distribution (left column), mixture of gamma distribution with neutral point mass (middle column), and log-normal distribution (right column). The grey bars indicate the proportions under the null hypothesis of the same distribution of $2N_e|s|$ in both species (constrained model). Darker colors in the middle column reflect the estimated proportions of neutral mutations.

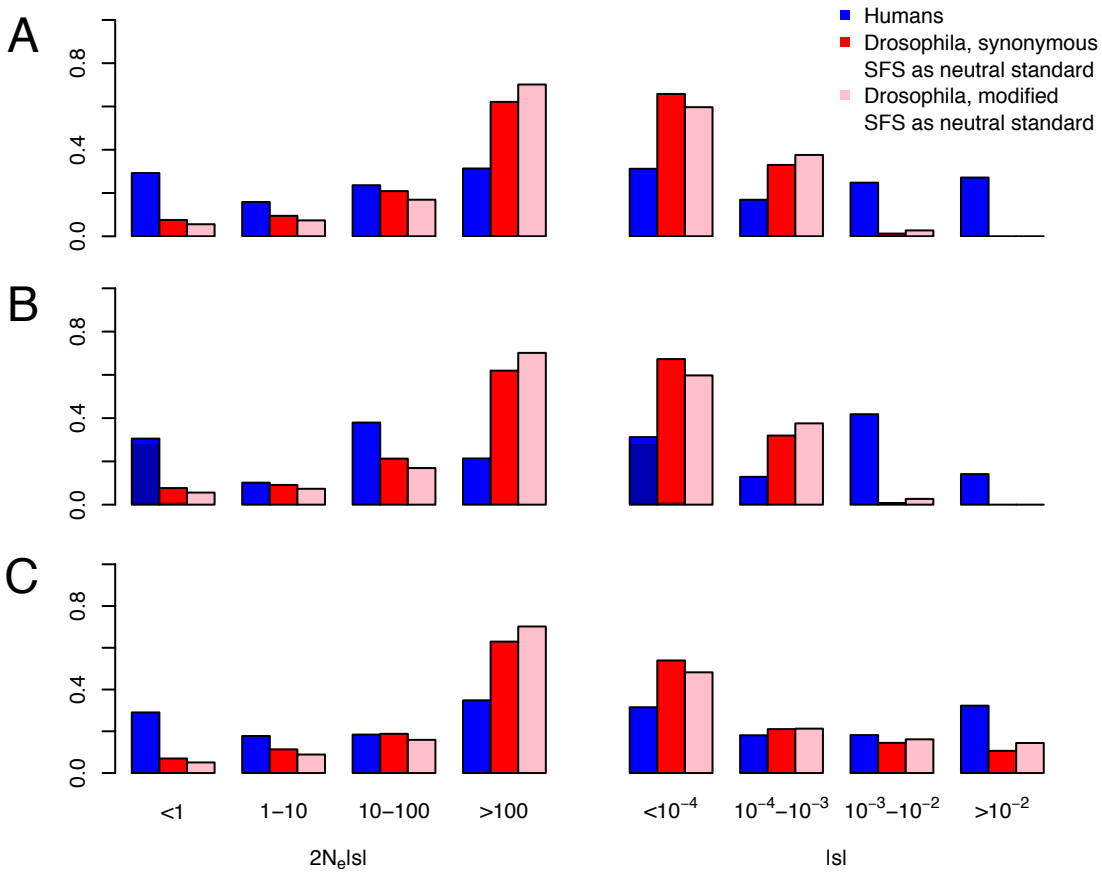

Fig. S7. Examining the possible effect of strong selection on synonymous mutations in *Drosophila* on estimates of the proportion of new nonsynonymous mutations for various ranges of $2N_e|s|$ and s. We generated a modified SFS that accounts for strong selection on synonymous sites. The modified SFS has 1/(1-0.22) times more SNPs than the observed synonymous SFS, and the same shape as the SFS from short introns (see SI Appendix, Text S3). Thus, it represents the truly neutral synonymous SFS when assuming that synonymous diversity is 22% smaller due to strong selection, and mutations in short introns are neutral (43). Proportions of the DFE for nonsynonymous mutations are computed from the estimated (A) gamma distribution, (B) mixture of gamma distribution with neutral point mass, and (C) log-normal distribution. Darker colors in (B) reflect the estimated proportions of neutral mutations. Note that the estimated DFEs for nonsynonymous mutations change only slightly when using the modified SFS as a neutral standard than when using the plain synonymous SFS as a neutral standard.

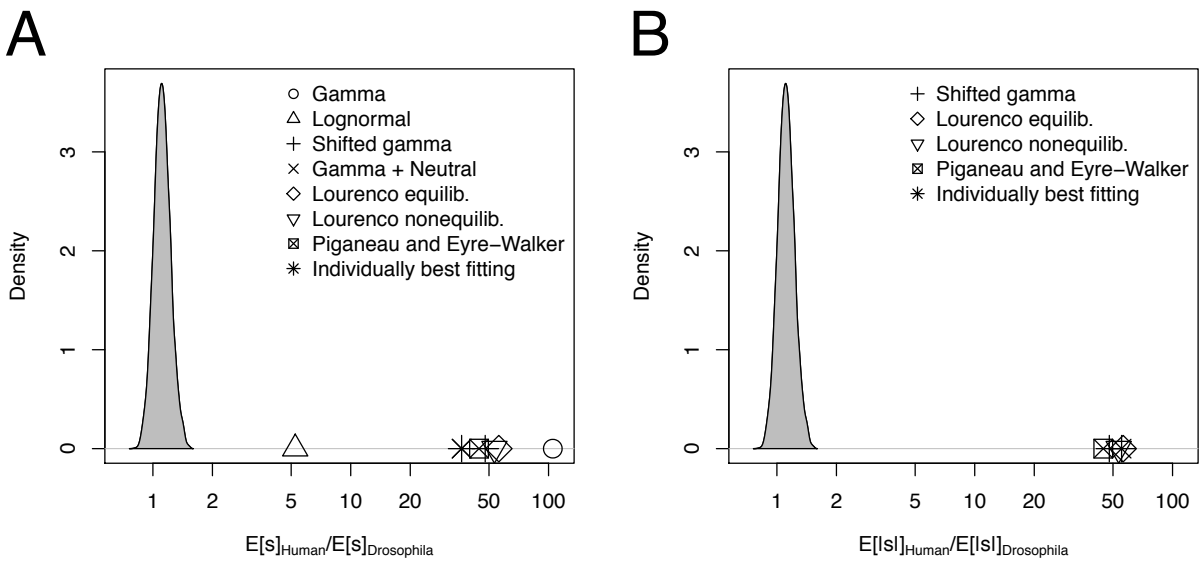

Fig. S8. Robustness of the difference in expected selection coefficient between humans and *Drosophila* to the functional form of the DFE. (A) Estimated average selection coefficient in humans over *Drosophila* ($E[s]_{Human}/E[s]_{Drosophila}$), assuming different functional forms of the DFE. E[*s*] is consistently estimated to be more deleterious in humans than in *Drosophila*. The individually best fitting DFE refers to the Gamma + Neutral distribution in both species, and suggests E[*s*] is 36-fold more deleterious (i.e. negative) in humans than *Drosophila*. (B) Back-mutation models assumes that the distribution of the absolute value of *s* (effect size |s|) is the same between species. Therefore, back-mutation models predict that although E[*s*] might be different due to different proportions of beneficial mutations, the average effect size (E[|*s*|]) should be the same between species. Thus, in (B) we show estimated average effect sizes in humans over *Drosophila* ($E[|s|]_{Human}/E[|s|]_{Drosophila}$), assuming different functional forms of the DFE. Note that all examined DFEs in (B) contain beneficial mutations because this is a central feature of the back-mutation model. The average effect size E[|*s*|] is consistently estimated to be larger in humans than in *Drosophila*, in contradiction to the back-mutation model. The individually best fitting DFE refers to the equilibrium Lourenço et al. distribution in humans and the shifted gamma distribution in *Drosophila*, and suggests E[|*s*|] is 55-fold larger in humans. In both (A) and (B), the null distribution in grey was calculated from forward simulations assuming the same gamma DFE in both species (see SI Appendix, Text S4). These simulations suggest it is unlikely to see $E[s]_{Human}/E[s]_{Drosophila}$ values > 2 assuming the same DFE between both species. Further, $E[s]_{Human}/E[s]_{Drosophila}$ (or $E[|s|]_{Human}/E[|s|]_{Drosophila}$) from the empirical data is consistently > 5 regardless of the functional form of the DFE assumed.

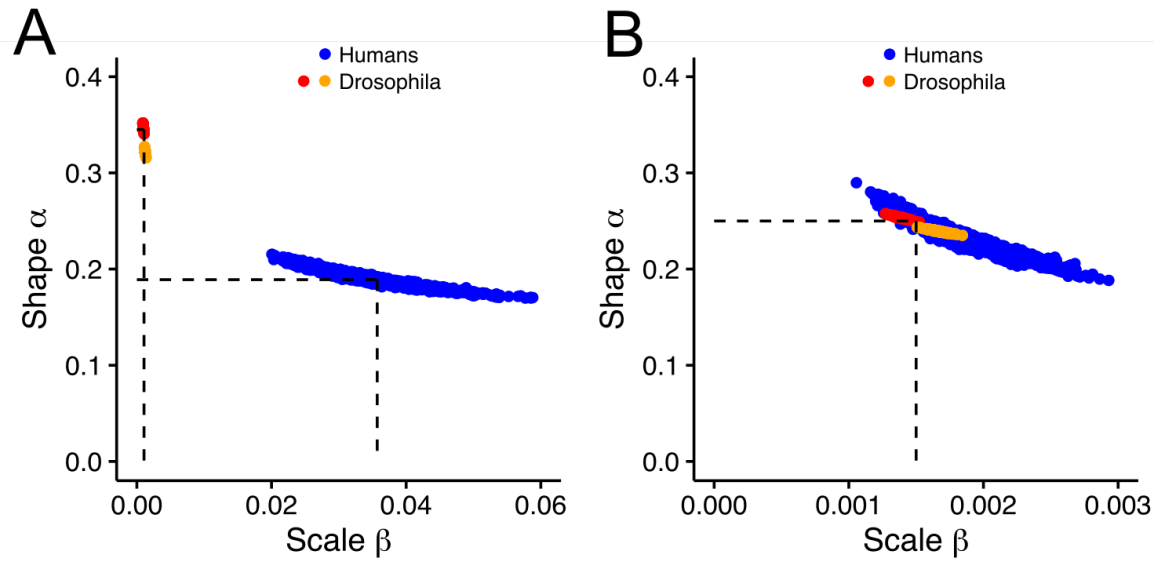

Fig. S9. The effect of positive selection in *Drosophila* on the inference of the deleterious DFE. We assume that there is both positive and negative selection in *Drosophila,* but only negative selection in humans. For simulations with positive selection, 0.5% of new nonsynonymous mutations are positively selected with $N_e s$=12. We estimated shape and scale parameters of a gamma DFE that only includes negative selection from 300 simulations of *Drosophila* (red, orange) data. Human results (blue) are the same as in Fig. 3 and are included only for comparative purposes. (A) Estimates from simulations under the alternative hypothesis (H1), i.e. assuming maximum likelihood gamma parameters in both species (dashed lines). Results show that indirect effects of positive selection (selective sweeps) do not bias our estimates in *Drosophila* (red), and that indirect plus direct effects (i.e. here positively selected nonsynonymous variants are included in the nonsynonymous SFS) of positive selection only slightly bias the estimates to lower shape parameters (orange). (B) Estimates from simulations under the null hypothesis (H0), i.e. assuming a single set of parameters of the deleterious gamma DFE in both species (dashed lines). Results show that, under H0, the indirect effects of positive selection do not bias our estimates in *Drosophila* (red), and that indirect plus direct effects of positive selection only slightly bias the estimates to lower shape and higher scale parameters (orange).

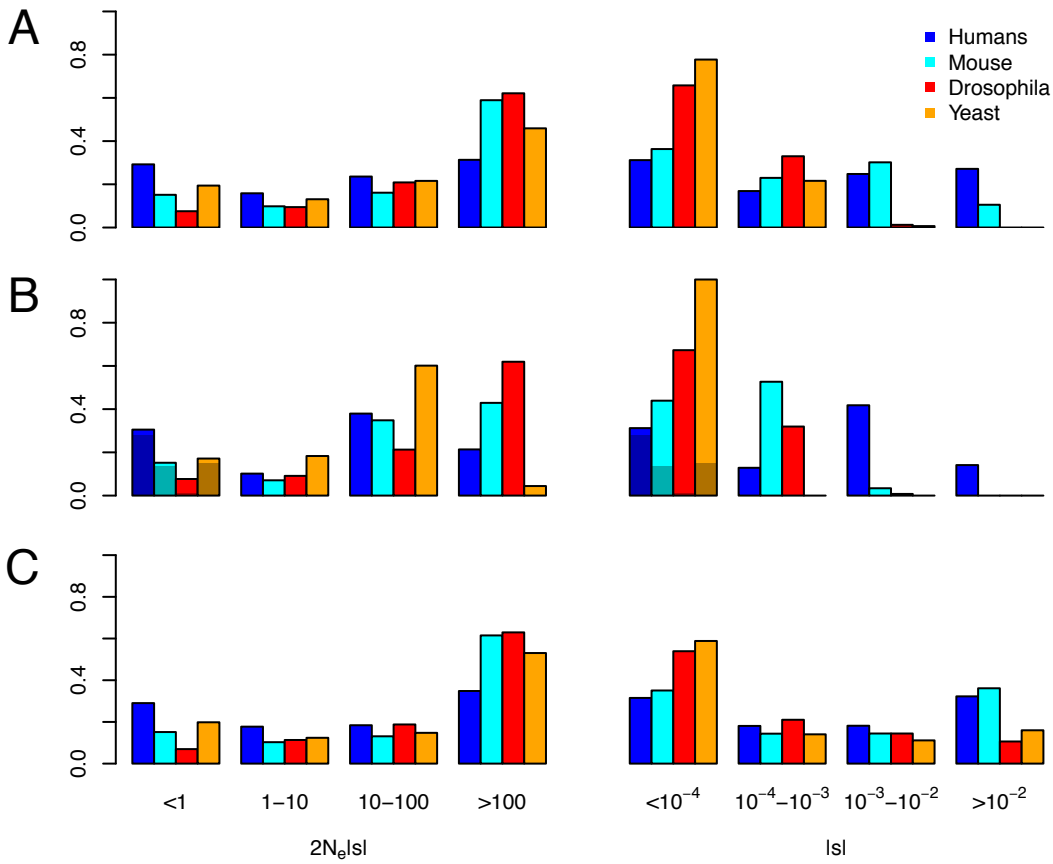

Fig. S10. The proportion of new mutations for various ranges of $2N_e|s|$ and s for humans, *Mus musculus castaneus* (mouse), *Drosophila melanogaster,* and *Saccharomyces paradoxus* (yeast). Proportions are computed from the estimated (A) gamma distribution, (B) mixture of gamma distribution with neutral point mass, and (C) log-normal distribution. Darker colors in (B) reflect the estimated proportions of neutral mutations.

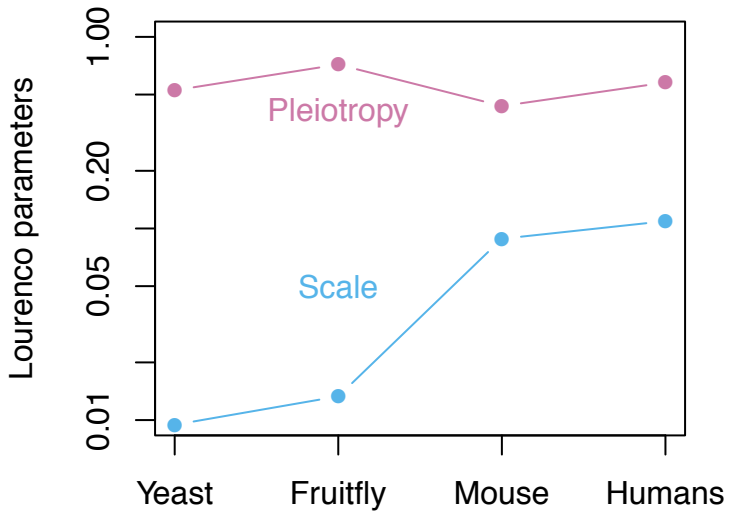

Fig. S11. Parameter estimates of the Lourenço et al. DFE in four species with increasing complexity: *Saccharomyces paradoxus* (yeast), *Drosophila melanogaster* (fruitfly), *Mus musculus castaneus* (mouse), and *Homo sapiens* (humans). 'Scale' refers to the scale parameter σ in Lourenço et al. (24), and 'Pleiotropy' to the pleiotropy parameter *m*. Note that the scale parameter increases with increasing complexity, but the pleiotropy parameter does not.

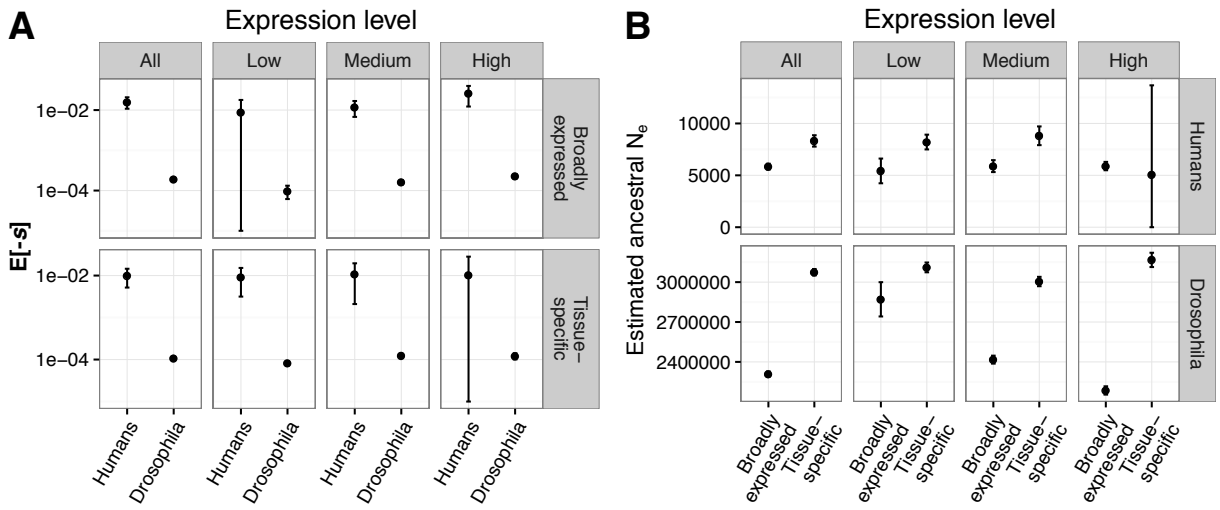

Fig. S12. The effect of gene expression on estimated E[-*s*] under a gamma DFE, and on the estimated effective population size. (A) Average selection coefficient E[*s*] is 70-110 fold less deleterious in Drosophila than in humans, independent of expression level or tissue specificity. (B) Estimated ancestral population size versus expression profiles. The ancestral population size was calculated from estimates of the synonymous population mutation rate θ_S for each category of genes, by fitting separate demographic models to the respective synonymous SFS, but assuming the same neutral per site mutation rate for each category (see SI Appendix, Text S2). Differences in the effective population size between expression categories can be the result of varying levels of linked selection (e.g. background selection, selective sweeps) on synonymous diversity.

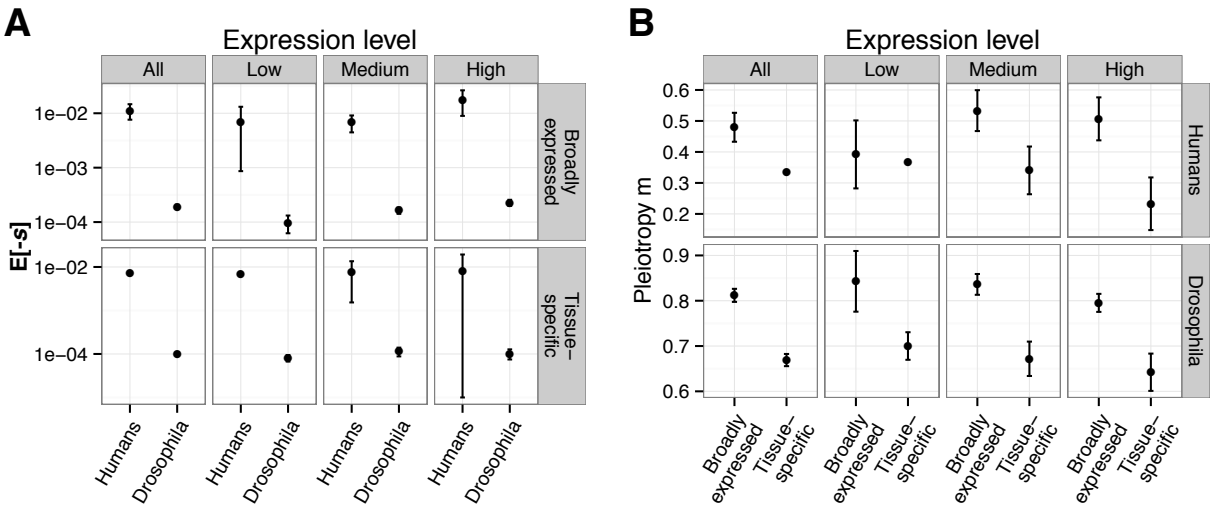

Fig. S13. The effect of gene expression on parameter estimates of the Lourenço et al. DFE. (A) Estimated Lourenço et al. DFE parameters for genes with different gene expression profiles. Average selection coefficient (E[-*s*]) is 70-110 fold less deleterious in *Drosophila* than in humans, independent of expression level or tissue specificity. (B) The pleiotropy parameter of the Lourenço et al. DFE depends on the breadth of gene expression. Tissue-specific genes have a smaller pleiotropy parameter *m* than broadly expressed genes, suggesting less pleiotropy in tissue-specific genes.

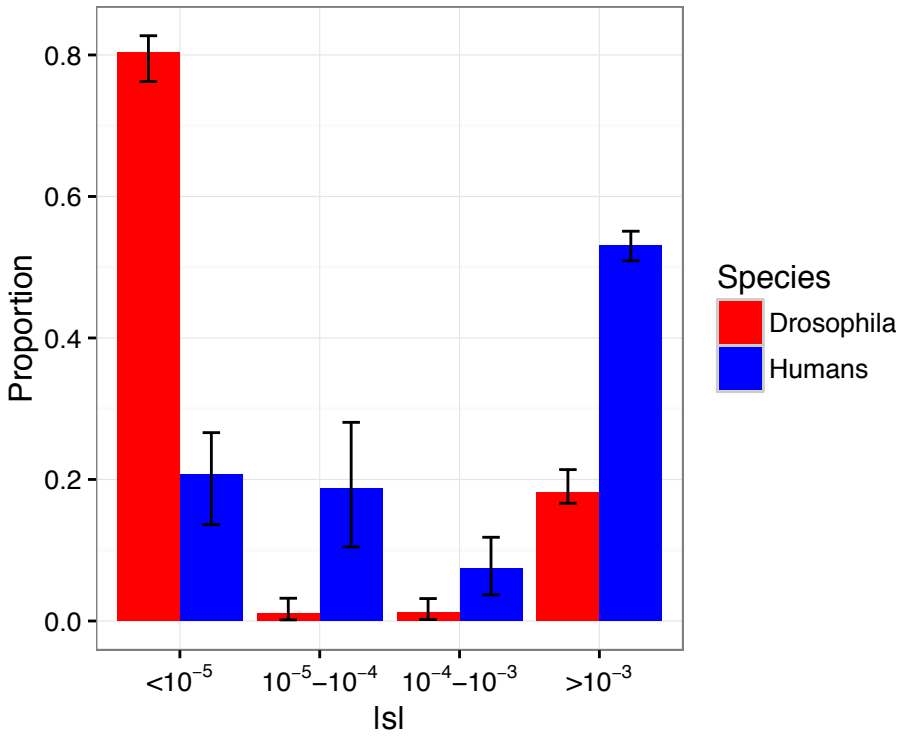

Fig. S14. Estimating the proportions of mutations in different bins of $|s|$ assuming a non-parametric discretized distribution as defined in Kim et al. (20). The DFE assumes a uniform probability mass within each bin. Estimates using this DFE were shown to correctly approximate the general form of the underlying DFE even if the true DFE is multi-modal (20). Errors bars denote 95% confidence intervals obtained from simulations where each entry of the nonsynonymous SFS is drawn from a Poisson distribution with the mean being that expected under the demographic and selection model.

Table S1. Demographic parameter estimates.

Filter	Synonymous sequence length	Species	θ_S	Time in units of $2N_{e,ancestral}$	$N_{e,current}/N_{e,ancestral}$	$N_{e,ancestral}$ [a]	$N_{e,Drosophila}/N_{e,Humans}$ [a]	$N_{e,ancestral}$ [b]	$N_{e,Drosophila}/N_{e,Humans}$ [b]	Log-likelihood
All	5.86E+06	Humans	4,144 (93.1)	0.426 (0.0410)	2.34 (0.0441)	7.07E+03	394	1.18E+04	118	-230.9
	4.74E+06	Drosophila	79,253 (196)	0.0919 (0.00156)	2.73 (0.0187)	2.79E+06		1.39E+06		-471.8
No singletons	5.86E+06	Humans	3,864 (111)	0.631 (0.0689)	2.23 (0.0550)	6.59E+03	416	1.10E+04	125	-188.1
	4.74E+06	Drosophila	77,883 (225)	0.122 (0.00347)	2.11 (0.0292)	2.74E+06		1.37E+06		-318.4
Only common genes	3.37E+06	Humans	2,200 (44.2)	0.420 (0.0304)	2.50 (0.0532)	6.53E+03	414	1.09E+04	124	-212.5
	2.95E+06	Drosophila	47,852 (153)	0.0958 (0.00203)	2.74 (0.0222)	2.70E+06		1.35E+06		-379.4

[a] Assuming phylogenetic mutation rate estimates
[b] Assuming estimates of current mutation rate

Table S2. Testing the null hypothesis of the same gamma DFE in both humans and *Drosophila*. Various types of data filtering are considered, as well as a different set of mutation rate estimates that are based on estimates of the current mutation rate (see SI Appendix, Text S2). The likelihood ratio test statistic $\Lambda = -2*\log(L_{Constrained,max}/L_{Full,max})$ tests the null hypothesis of no difference in shape and scale parameters between humans and *Drosophila*.

GAMMA DFE

	Hypothesis	Species	Shape (α)	Scale (β)	E[|*s*|]	Log - likelihood	Λ	p-value H1/H0
All Data	Full model (H1)	Humans	0.19	7.41E-02	1.40E-02	-245		
		Drosophila	0.35	3.77E-04	1.33E-04	-389		
		Total				-634		
	Constrained model (H0): $DFE(s)_{Humans}=DFE(s)_{Drosophila}$		0.24	4.28E-03	1.02E-03	-6640	12012	<1E-16
	Constrained model (H0): $DFE(N_es)_{Humans}=DFE(N_es)_{Drosophila}$		0.33	2.52E+03		-11502	21734	<1E-16
Only orthologous genes	Full model (H1)	Humans	0.24	5.88E-02	1.43E-02	-203		
		Drosophila	0.40	4.07E-04	1.64E-04	-285		
		Total				-488		
	Constrained model (H0): $DFE(s)_{Humans}=DFE(s)_{Drosophila}$		0.26	6.45E-03	1.68E-03	-4673	8370	<1E-16
	Constrained model (H0): $DFE(N_es)_{Humans}=DFE(N_es)_{Drosophila}$		0.38	2.54E+03		-6236	11497	<1E-16
No singletons	Full model (H1)	Humans	0.16	2.07E-01	3.29E-02	-181		
		Drosophila	0.33	6.01E-04	1.98E-04	-252		
		Total				-433		
	Constrained model (H0): $DFE(s)_{Humans}=DFE(s)_{Drosophila}$		0.21	2.49E-02	5.16E-03	-1715	2564	<1E-16
	Constrained model (H0): $DFE(N_es)_{Humans}=DFE(N_es)_{Drosophila}$		0.29	5.48E+03		-9204	17542	<1E-16
Using recent mutation rate estimates	Full model (H1)	Humans	0.19	4.45E-02	8.39E-03	-245		
		Drosophila	0.35	7.54E-04	2.66E-04	-389		
		Total				-634		
	Constrained model (H0): $DFE(s)_{Humans}=DFE(s)_{Drosophila}$		0.29	2.56E-03	7.31E-04	-3389	5510	<1E-16
	Constrained model (H0): $DFE(N_es)_{Humans}=DFE(N_es)_{Drosophila}$		0.32	2.54E+03		-11502	21734	<1E-16
No singletons & recent mutation rate estimates	Full model (H1)	Humans	0.16	1.25E-01	1.99E-02	-181		
		Drosophila	0.33	1.20E-03	3.96E-04	-252		
		Total				-433		
	Constrained model (H0): $DFE(s)_{Humans}=DFE(s)_{Drosophila}$		0.26	7.64E-03	1.95E-03	-1012	1159	<1E-16
	Constrained model (H0): $DFE(N_es)_{Humans}=DFE(N_es)_{Drosophila}$		0.29	5.51E+03		-9205	17544	<1E-16

Table S3. Testing the null hypothesis of the same log-normal DFE in both humans and *Drosophila*. Various types of data filtering are considered, as well as a different set of mutation rate estimates that are based on estimates of the current mutation rate (see SI Appendix, Text S2). The likelihood ratio test statistic $\Lambda = -2*\log(L_{Constrained,max}/L_{Full,max})$ tests the null hypothesis of no difference in the two parameters of the log-normal distribution (mean and SD) between humans and *Drosophila*.

LOGNORMAL DFE

	Hypothesis	Species	Mean	SD	Median (|s|)	Log - likelihood	Λ	p-value H1/H0
All Data	Full model (H1)	Humans	-6.86	4.89	1.05E-03	-278		
		Drosophila	-9.61	4.01	6.72E-05	-649		
		Total				-927		
	Constrained model (H0): $DFE(s)_{Humans}=DFE(s)_{Drosophila}$		-8.74	4.88	1.59E-04	-3510	5166	<1E-16
	Constrained model (H0): $DFE(N_es)_{Humans}=DFE(N_es)_{Drosophila}$		5.92	4.26		-11844	21834	<1E-16
Only orthologous genes	Full model (H1)	Humans	-6.10	4.02	2.23E-03	-223		
		Drosophila	-9.10	3.79	1.11E-04	-426		
		Total				-649		
	Constrained model (H0): $DFE(s)_{Humans}=DFE(s)_{Drosophila}$		-7.99	4.77	3.38E-04	-2669	4041	<1E-16
	Constrained model (H0): $DFE(N_es)_{Humans}=DFE(N_es)_{Drosophila}$		6.42	4.00		-6423	11547	<1E-16
No singletons	Full model (H1)	Humans	-6.30	5.94	1.84E-03	-181		
		Drosophila	-8.67	4.77	1.71E-04	-261		
		Total				-442		
	Constrained model (H0): $DFE(s)_{Humans}=DFE(s)_{Drosophila}$		-7.19	6.07	7.55E-04	-950	1015	<1E-16
	Constrained model (H0): $DFE(N_es)_{Humans}=DFE(N_es)_{Drosophila}$		7.10	5.30		-9224	17564	<1E-16
Using recent mutation rate estimates	Full model (H1)	Humans	-7.37	4.58	6.32E-04	-278		
		Drosophila	-8.91	4.01	1.34E-04	-649		
		Total				-927		
	Constrained model (H0): $DFE(s)_{Humans}=DFE(s)_{Drosophila}$		-8.49	4.43	2.06E-04	-1809	1765	<1E-16
	Constrained model (H0): $DFE(N_es)_{Humans}=DFE(N_es)_{Drosophila}$		5.92	4.25		-11844	21834	<1E-16
No singletons & recent mutation rate estimates	Full model (H1)	Humans	-6.81	5.94	1.10E-03	-181		
		Drosophila	-7.98	4.77	3.42E-04	-261		
		Total				-442		
	Constrained model (H0): $DFE(s)_{Humans}=DFE(s)_{Drosophila}$		-7.49	5.20	5.58E-04	-532	181	<1E-16
	Constrained model (H0): $DFE(N_es)_{Humans}=DFE(N_es)_{Drosophila}$		7.12	5.32		-9225	17566	<1E-16

Table S4. Maximum likelihood parameter estimates and log-likelihoods for alternative DFE functions. The last column shows the difference in the Akaike Information Criterion (AIC) between the relevant DFE model and the gamma DFE.

Species	DFE	Parameter 1	Parameter 2	Parameter 3	Parameter 4	Log - likelihood	AIC_{Model}	AIC_{Model} - $AIC_{GammaDFE}$
Humans	Gamma	shape=0.19	scale=0.074			-245	494	0
	Lognormal	mean=-6.86	SD=4.89			-278	560	65
	Gamma+Neutral	shape=0.73	scale=0.0082	$p_{neutral}$=0.281		-219	443	-51
	Martin and Lenormand (2006), eq. 5 (shifted gamma)	shape=0.28	scale=0.022	shift=0.66		-232	469	-25
	Piganeau and Eyre-Walker (2003), eq. 7	shape=0.29	scale=0.021	$N_{e,long-term}$=2642		-224	455	-39
	Lourenço et al. (2011), eq.8	z=0.042	n=22.2	m=0.52	sigma=0.12	-225	459	-36
	Lourenço et al. (2011), eq.15	m=0.58	sigma=0.109	$N_{e,long-term}$=2476		-222	451	-43
D. melanogaster	Gamma	shape=0.35	scale=3.8E-04			-389	782	0
	Lognormal	mean=-9.6	SD=4.0			-649	1302	519
	Gamma+Neutral	shape=0.38	scale=0.00032	$p_{neutral}$=0.011		-384	775	-7.5
	Martin and Lenormand (2006), eq. 5 (shifted gamma)	shape=0.36	scale=0.00036	shift=0.027		-388	782	-0.7
	Piganeau and Eyre-Walker (2003), eq. 7	shape=0.35	scale=0.00038	$N_{e,long-term}$= 5.2E+19		-389	784	2.2
	Lourenço et al. (2011), eq.8	z=0.00035	n=14632	m=0.70	sigma=0.014	-389	787	4.6
	Lourenço et al. (2011), eq.15	m=0.71	sigma=0.014	N_e=8.4e7		-390	785	3.0
S. paradoxus	Gamma	shape=0.22	scale=3.7E-04			-11.6	27.3	0
	Lognormal	mean=-10.5	SD=6.0			-11.6	27.3	0
	Gamma+Neutral	shape=1.1	scale=5.5E-6	$p_{neutral}$=0.15		-11.6	29.2	1.9
	Martin and Lenormand (2006), eq. 5 (shifted gamma)	shape=0.24	scale=0.00025	shift=0.086		-11.6	29.2	1.9
	Piganeau and Eyre-Walker (2003), eq. 7	shape=0.56	scale=1.00	$N_{e,long-term}$=43000		-85.3	177	149
	Lourenço et al. (2011), eq.8	z=0.00068	n=5.8	m=76	sigma=0.00014	-13.2	34.5	7.2
	Lourenço et al. (2011), eq.15	m=0.53	sigma=0.0095	N_e=7.3E6		-11.6	29.2	1.9
M. musculus castaneus	Gamma	shape=0.22	scale=0.016			-19.0	42.0	0
	Lognormal	mean=-6.8	SD=6.2			-19.0	42.0	0
	Gamma+Neutral	shape=0.79	scale=0.00036	$p_{neutral}$=0.14		-19.0	43.9	1.9
	Martin and Lenormand (2006), eq. 5 (shifted gamma)	shape=1.2	scale=0.00010	shift=8.0		-18.9	43.9	1.9
	Piganeau and Eyre-Walker (2003), eq. 7	shape=0.31	scale=0.0022	$N_{e,long-term}$= 194000		-19.0	44.0	2.0
	Lourenço et al. (2011), eq.8	z=0.0040	n=4.3	m=1.3	sigma=0.012	-18.9	45.9	3.9
	Lourenço et al. (2011), eq.15	m=0.44	sigma=0.088	N_e=1.98E7		-19.0	44.0	2.0

Table S5. Likelihood ratio (LR) test statistics for all pairwise species comparisons. The LR test statistic tests the null hypothesis of the same DFE in both species, either on scale of s or $N_e s$. It assumes that the true DFE is gamma distributed. A star indicates rejection of the null hypothesis at a 1% significance level, based on the simulation-derived null distribution shown in Fig. 3C.

Null hypothesis	Species pair	Human	Mouse	Drosophila	Yeast
$DFE(s)_{Species\ 'A'} = DFE(s)_{Species\ 'B'}$	Human	0			
	Mouse	5.9	0		
	Drosophila	12011*	11	0	
	Yeast	64*	17	103*	0
$DFE(N_e s)_{Species\ 'A'} = DFE(N_e s)_{Species\ 'B'}$	Human	0			
	Mouse	64*	0		
	Drosophila	21734*	20	0	
	Yeast	48*	3.1	98*	0