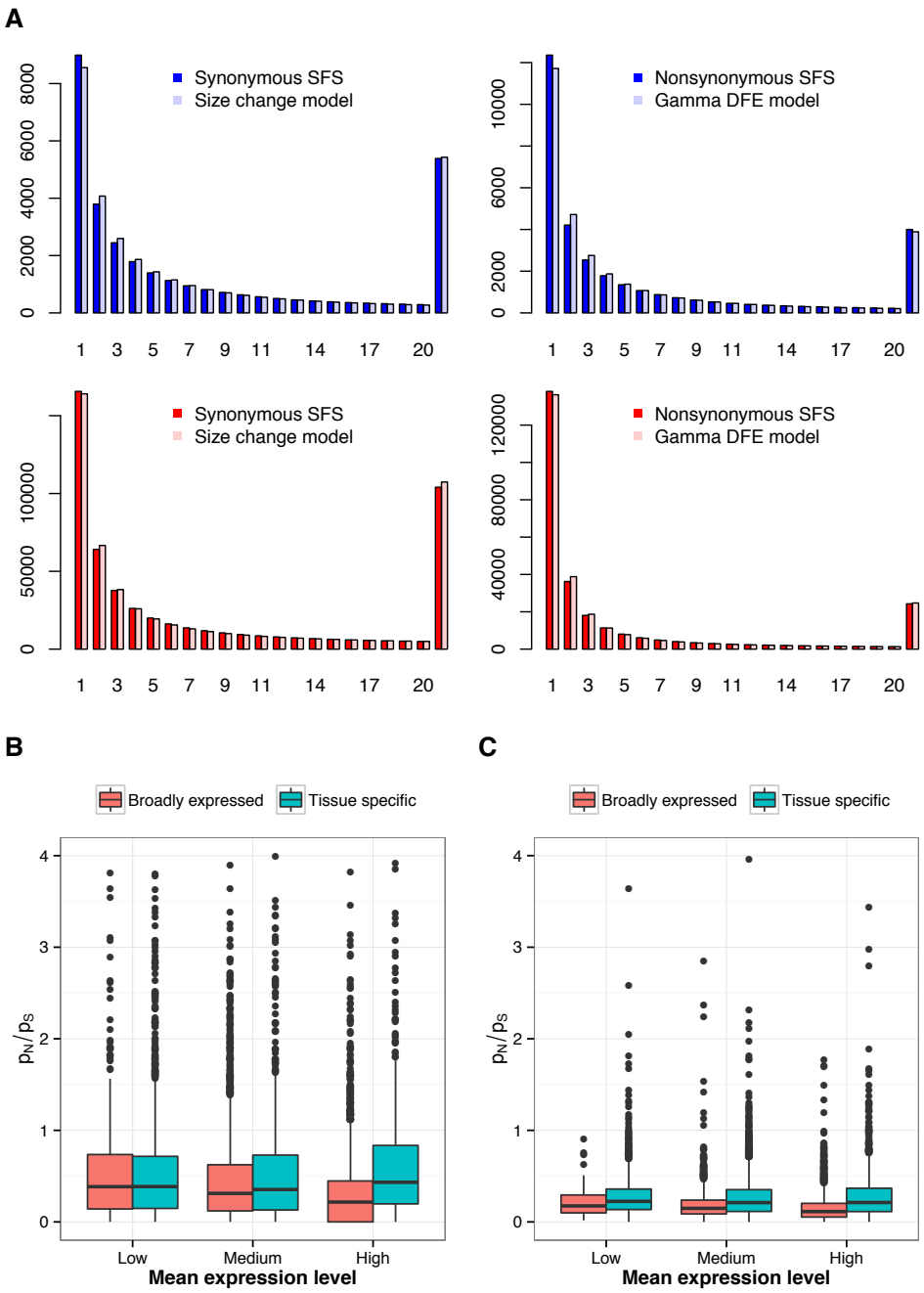
Determining the factors driving selective effects of new nonsynonymous mutations

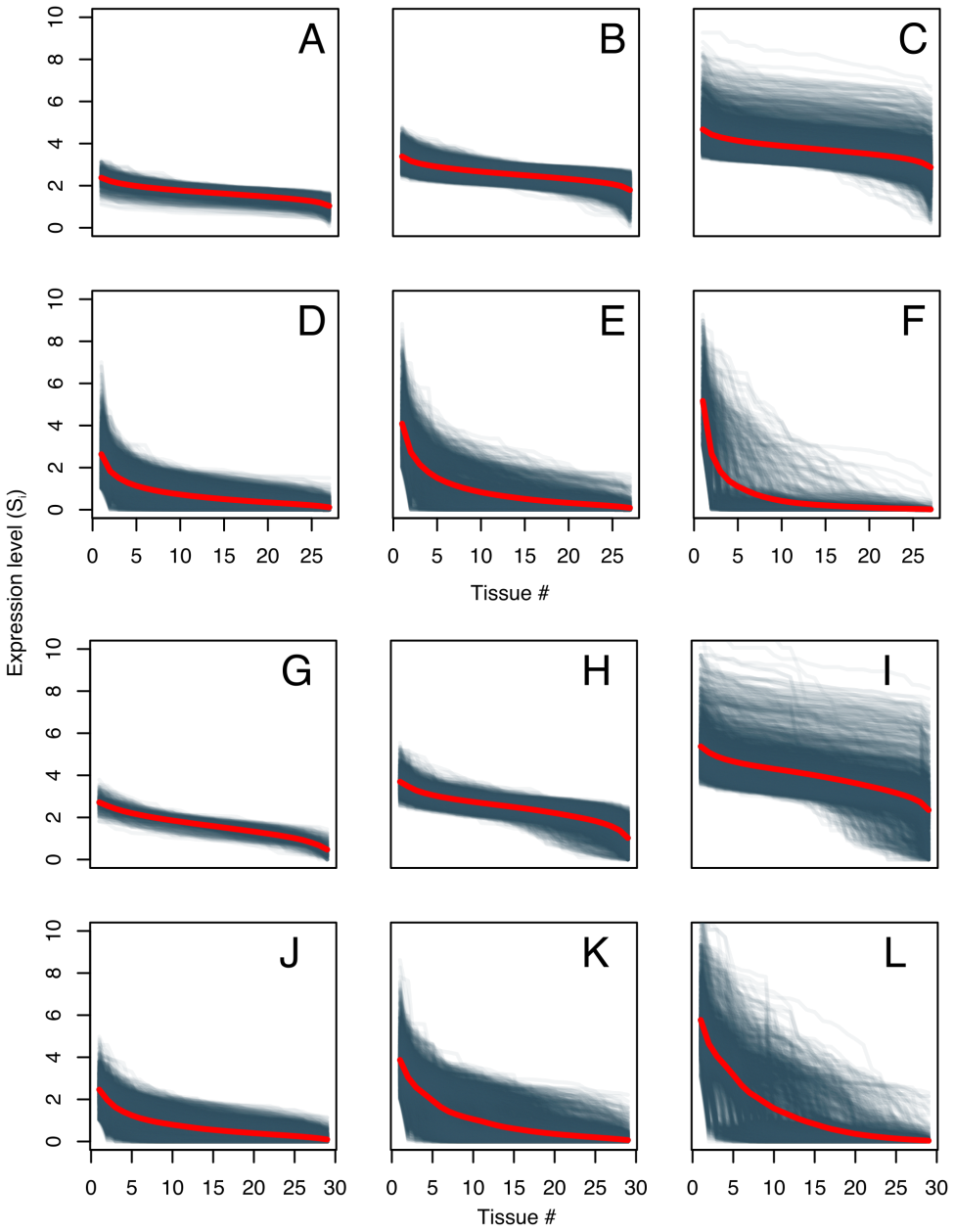
Christian D. Huber[1,], Bernard Kim[1], Clare D. Marsden[1], Kirk E. Lohmueller[1,2,3,*]*

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.

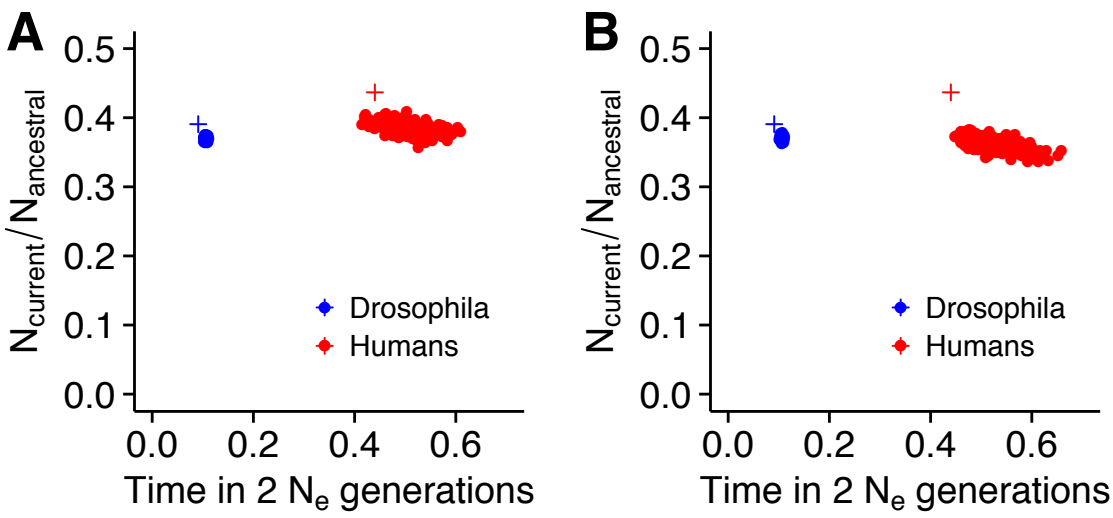
1. Supplementary Figures 1-14
2. Supplementary Tables 1-5
3. Supplementary Notes 1-4



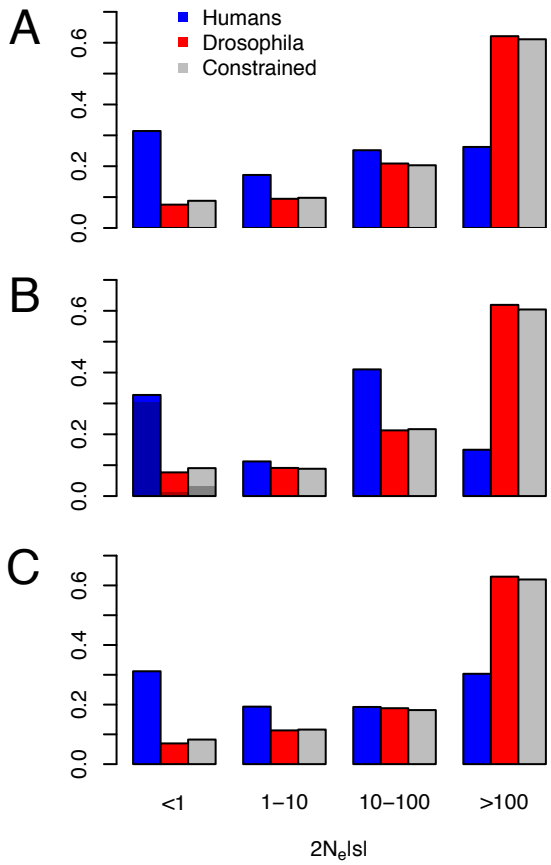
Supplementary Fig 1. Different diversity patterns between humans and Drosophila. (A) The folded synonymous and nonsynonymous SFS for humans (blue) and Drosophila (red). The expected SFS under the MLEs of the model parameters are shown in light colors. The x-axis is binned according to the minor allele 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 $\tau > 0.6$.



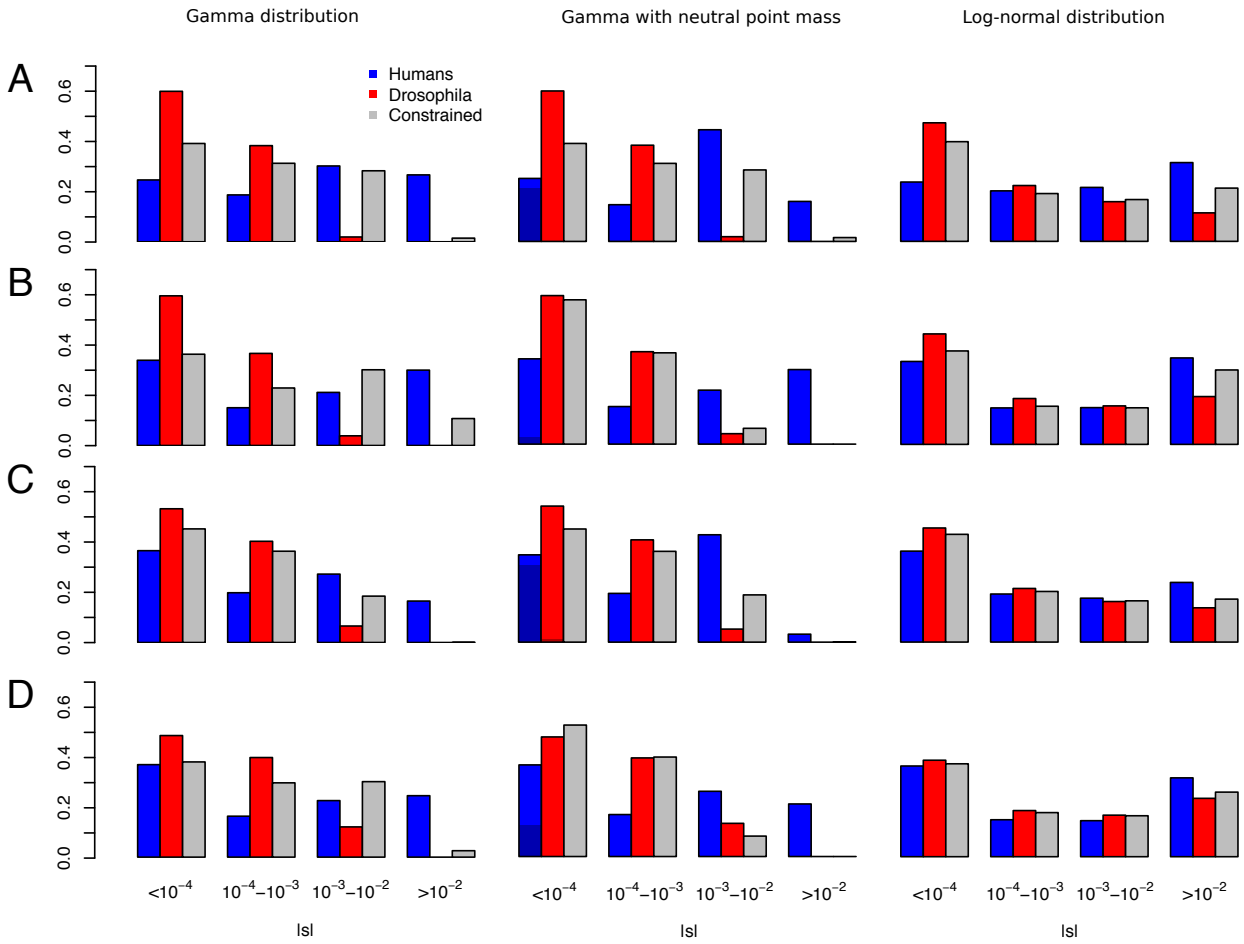
Supplementary Fig 2. Expression profiles for human and Drosophila genes. (A-F) Expression profiles for human genes. (G-L) Expression profiles for Drosophila genes. Each grey line represents a gene. For each gene, the tissue is ordered according to the expression level, i.e. expression level is plotted 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 Online Methods for definitions). The red line represents the average across genes.



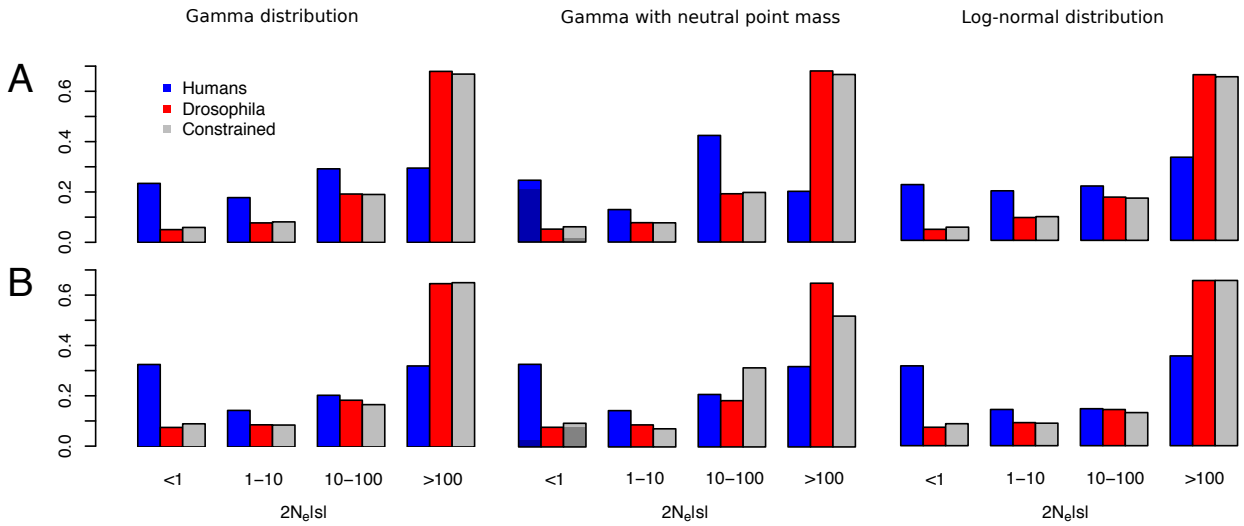
Supplementary Fig 3. Demographic parameter estimation for 300 simulated data sets. True parameter values are shown as crosses, estimated parameters as points. (A) Simulations under the full model. (B) Simulations under the constrained model (see main text). Note that the demographic parameter estimates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



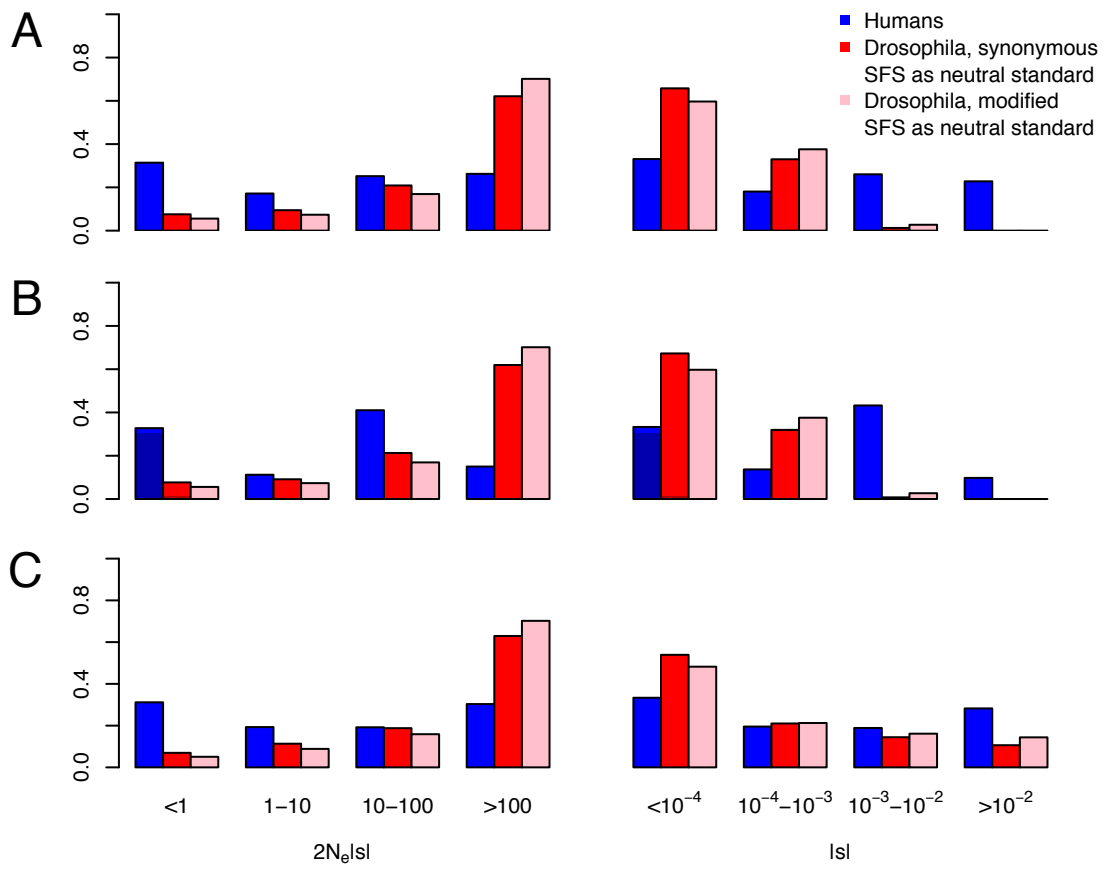
Supplementary Fig 4. The proportion of new mutations for various ranges of $2N_e|s|$. Proportions are computed from the estimated (A) gamma distribution, (B) mixture of gamma distribution with neutral point mass, and (C) log-normal distribution. The grey bars indicate the proportions under the null hypothesis of the same distribution of $2N_e|s|$ in both species (constrained model). Darker colors in (B) reflect the estimated proportions of neutral mutations.



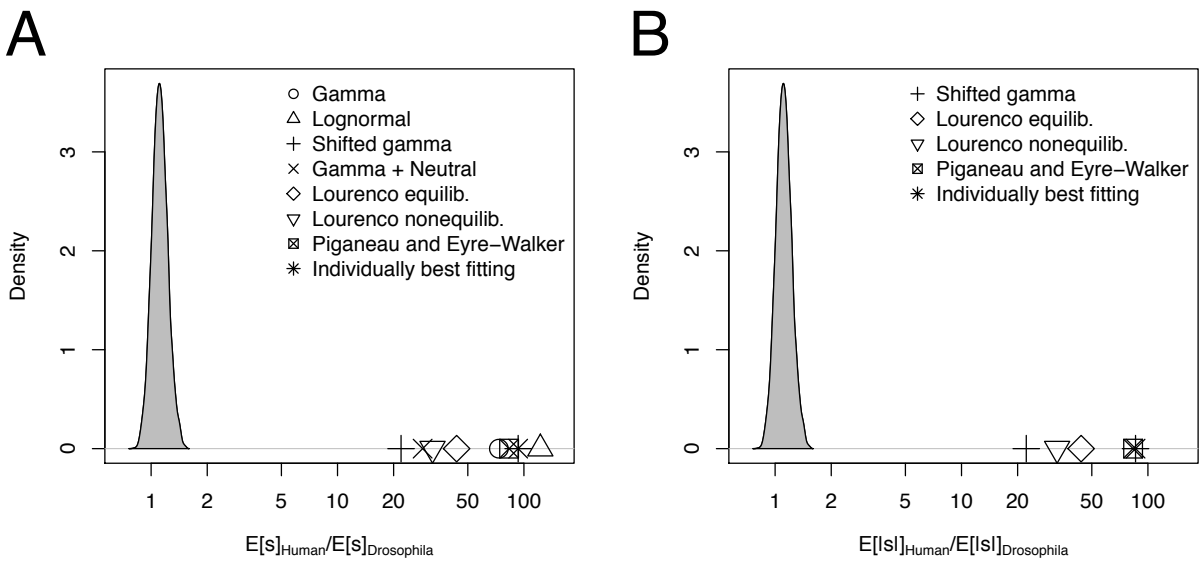
Supplementary Fig 5. The proportion of new mutations for various ranges of $|s|$. (A) Using data filtered for genes that have orthologs in both species. (B) Using data after filtering out singletons. (C) Assuming the recent mutation rate estimates (see main text). (D) Using the recent mutation rate estimates 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.



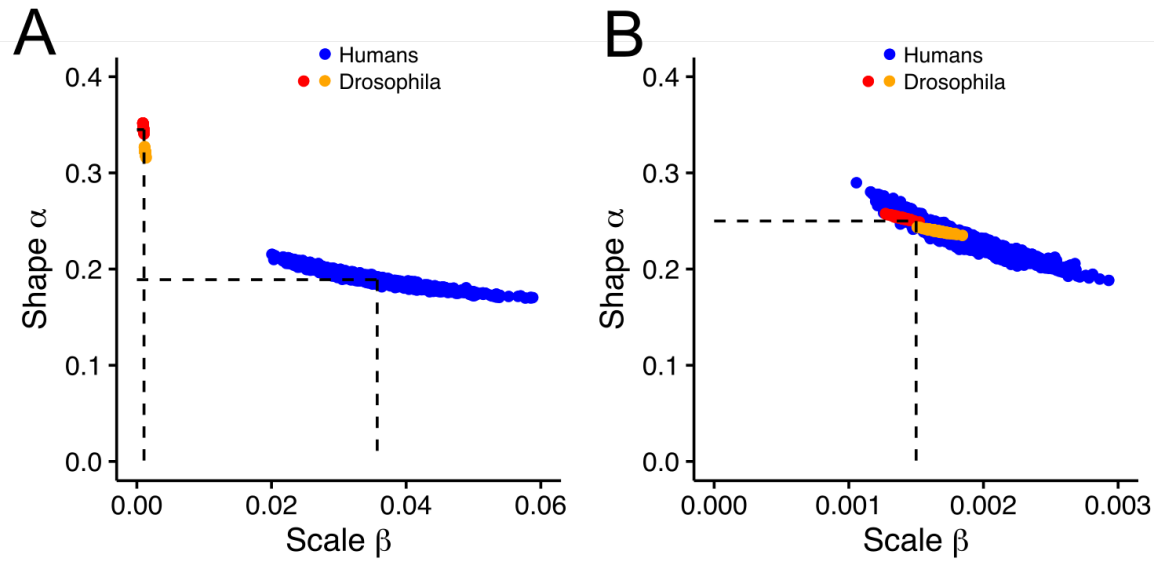
Supplementary Fig 6. The proportion of new mutations for various ranges of $2N_e|s|$. (A) Using data filtered for genes that have orthologs in both species. (B) Using data after filtering out singletons. Proportions are computed from the estimated gamma distribution (left column), mixture of gamma distribution 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.



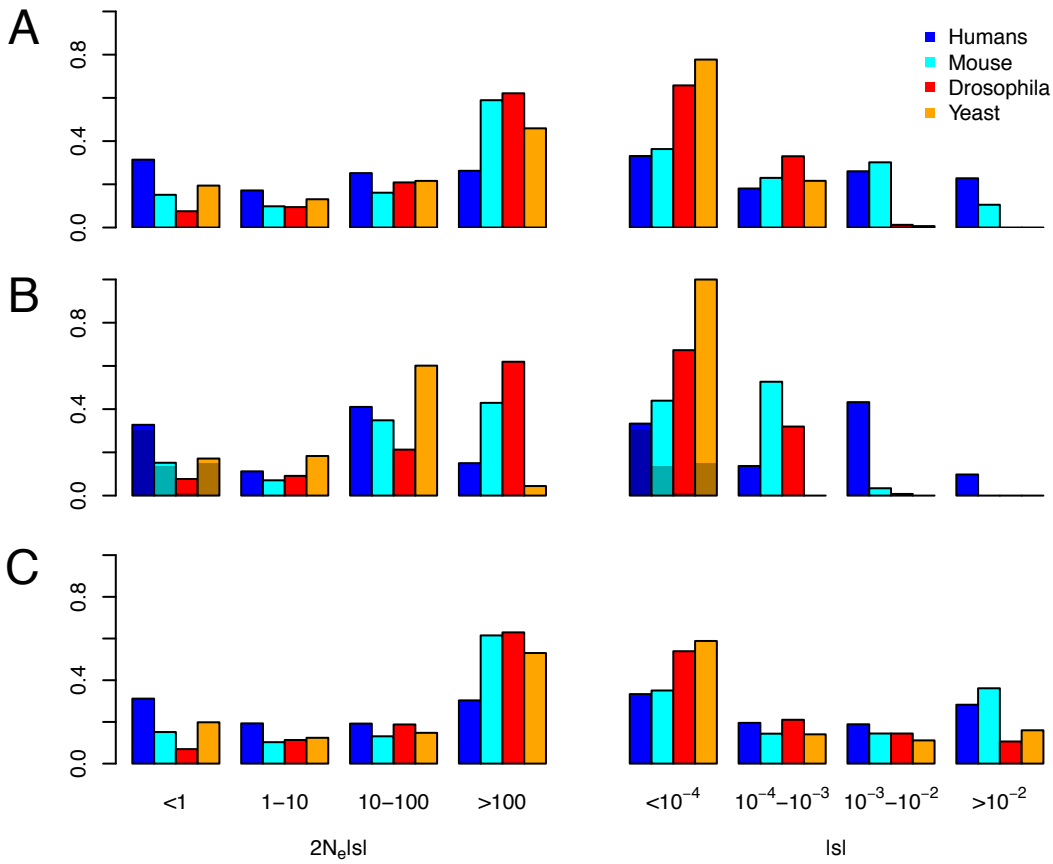
Supplementary Fig 7. Examining the possible effect of strong selection on synonymous mutations in *Drosophila* on estimates of the proportion of new nonsynonymous mutations for various ranges of $2N_e|s|$ and s. We generated a modified SFS that accounts for strong selection on synonymous sites. The modified 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 Online Methods). 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[1]. 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.



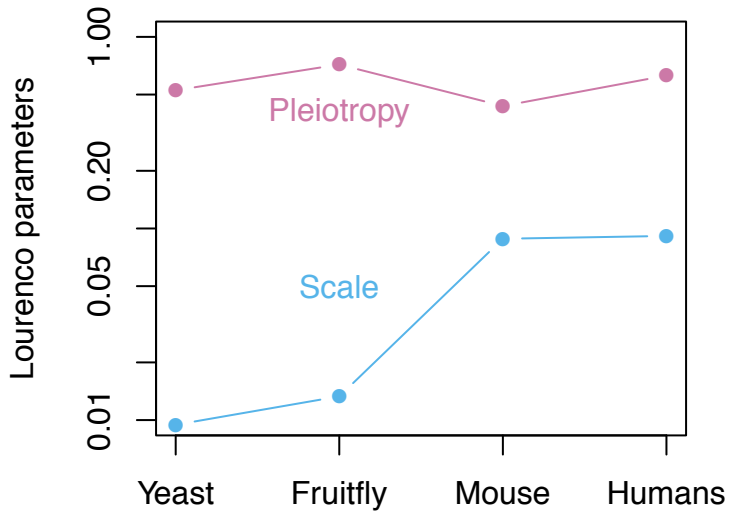
Supplementary Fig 8. Robustness of the difference in expected selection coefficient between humans and *Drosophila* to the functional form of the DFE. (A) Estimated average selection coefficient in humans over *Drosophila* ($E[s]_{Human}/E[s]_{Drosophila}$), assuming different functional forms of the DFE. E[*s*] is consistently estimated to be more negative in humans than in *Drosophila*. The individually best fitting DFE refers to the Piganeau and Eyre-Walker distribution in humans and the Gamma + Neutral distribution in *Drosophila*, and suggests E[*s*] is 93-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 Piganeau and Eyre-Walker distribution in humans and the shifted gamma distribution in *Drosophila*, and suggests E[|*s*|] is 86-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 Online Methods). 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 >20 regardless of the function form of the DFE assumed.



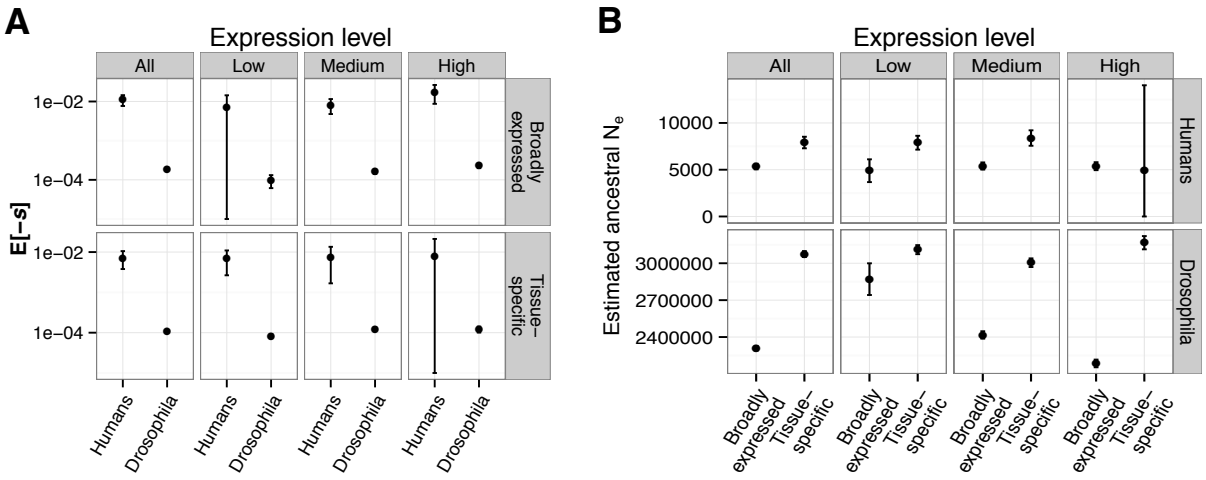
Supplementary Fig 9. The effect of positive selection in *Drosophila* on the inference of the deleterious DFE. We assume that there is both positive and negative selection in *Drosophila,* but only negative selection in humans. For simulations with positive selection, 0.5% of new nonsynonymous mutations 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).



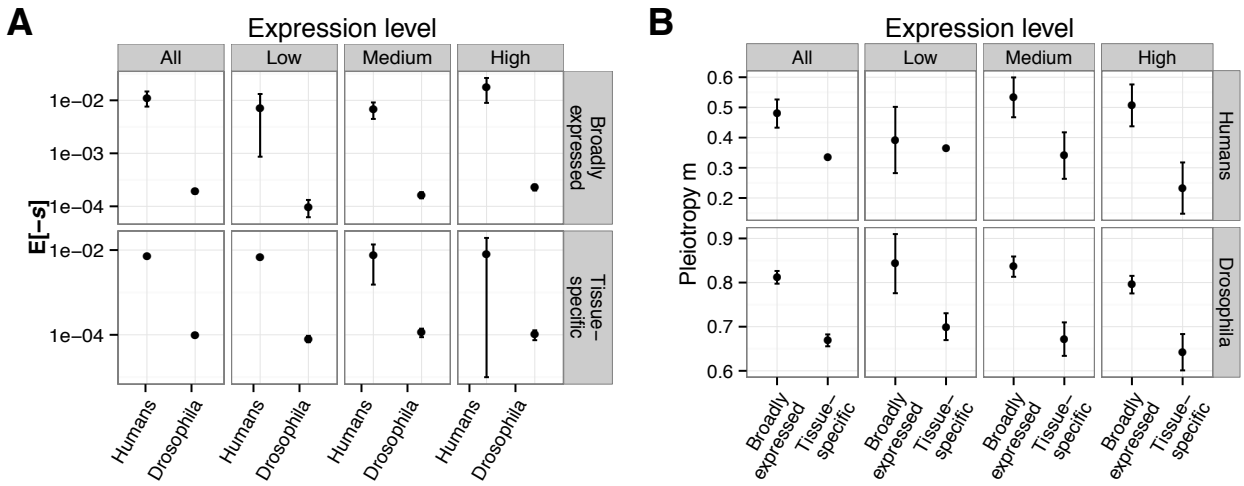
Supplementary Fig 10. The proportion of new mutations for various ranges of $2N_e|s|$ and s for humans, *Mus musculus castaneus* (mouse), *Drosophila melanogaster,* and *Saccharomyces paradoxus* (Yeast). Proportions are computed from the estimated (A) gamma distribution, (B) mixture of gamma distribution with neutral point mass, and (C) log-normal distribution. Darker colors in (B) reflect the estimated proportions of neutral mutations.



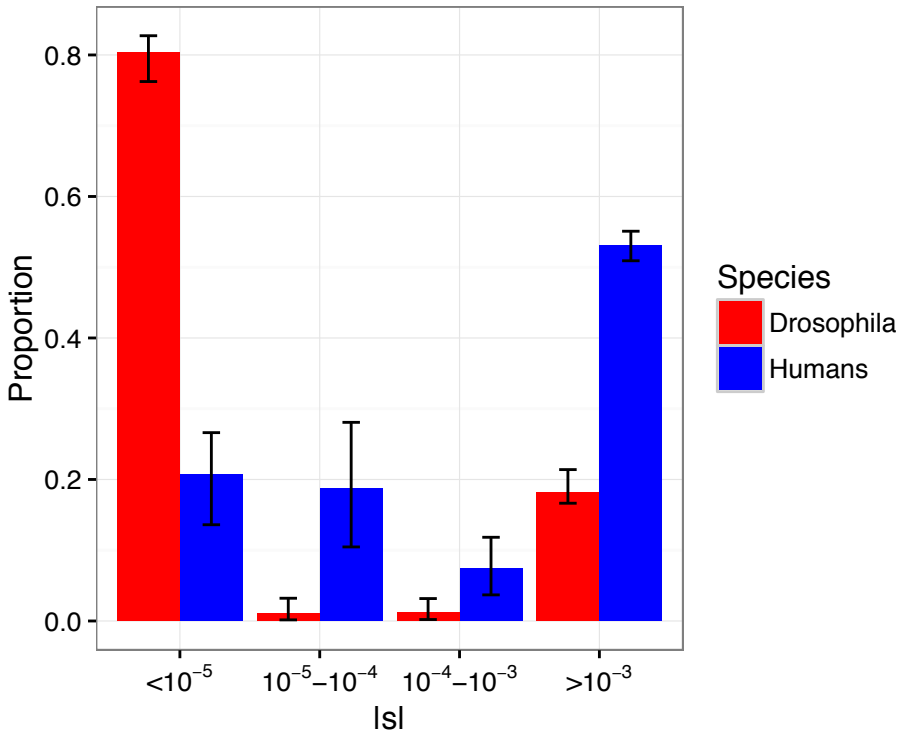
Supplementary Fig 11. Parameter estimates of the Lourenço et al. DFE in four species with increasing complexity: *Saccharomyces paradoxus* (yeast), *Drosophila melanogaster* (fruitfly), *Mus musculus castaneus* (mouse), and *Homo sapiens* (humans). 'Scale' refers to the scale parameter σ in Lourenço et al.[2], and 'Pleiotropy' to the pleiotropy parameter *m*. Note that the scale parameter increases with increasing complexity, but the pleiotropy parameter does not.



Supplementary Fig 12. 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 50-80 fold smaller 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 mutation rate for each category (see Online Methods). 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.



Supplementary Fig 13. 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 50-80 fold smaller 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.



Supplementary Fig 14. Estimating the proportions of mutations in different bins of |*s*| assuming a non-parametric discretized distribution as defined in Kim et al.[3]. 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[3]. 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.

Supplementary Table 1. 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	3,869 (63.6)	0.419 (0.0269)	2.33 (0.0393)	6.60E+03	422	1.10E+04	127	-224.3
	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	3,625 (102)	0.610 (0.0666)	2.21 (0.0544)	6.18E+03	443	1.03E+04	133	-185.7
	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	2,045 (43.2)	0.418 (0.0321)	2.49 (0.0553)	6.08E+03	445	1.01E+04	133	-208.1
	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

Supplementary Table 2. 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 main text). 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	5.13E-02	4.88E-03	-248		
		Drosophila	0.35	3.77E-04	6.65E-05	-389		
		Total				-637		
	Constrained model (H0): $DFE(s)_{Humans}=DFE(s)_{Drosophila}$		0.25	2.87E-03	7.25E-04	-5737	10199	<1E-16
	Constrained model (H0): $DFE(N_es)_{Humans}=DFE(N_es)_{Drosophila}$		0.33	2504	407	-12919	24563	<1E-16
Only orthologous genes	Full model (H1)	Humans	0.25	4.36E-02	5.38E-03	-205		
		Drosophila	0.40	4.07E-04	8.19E-05	-285		
		Total				-490		
	Constrained model (H0): $DFE(s)_{Humans}=DFE(s)_{Drosophila}$		0.27	4.48E-03	1.22E-03	-4174	7369	<1E-16
	Constrained model (H0): $DFE(N_es)_{Humans}=DFE(N_es)_{Drosophila}$		0.38	2521	473	-6945	12911	<1E-16
No singletons	Full model (H1)	Humans	0.16	1.38E-01	1.10E-02	-181		
		Drosophila	0.33	6.01E-04	9.91E-05	-252		
		Total				-433		
	Constrained model (H0): $DFE(s)_{Humans}=DFE(s)_{Drosophila}$		0.22	1.63E-02	3.52E-03	-1546	2227	<1E-16
	Constrained model (H0): $DFE(N_es)_{Humans}=DFE(N_es)_{Drosophila}$		0.29	5514	797	-10156	19447	<1E-16
Using recent mutation rate estimates	Full model (H1)	Humans	0.19	3.08E-02	2.93E-03	-248		
		Drosophila	0.35	7.54E-04	1.33E-04	-389		
		Total				-637		
	Constrained model (H0): $DFE(s)_{Humans}=DFE(s)_{Drosophila}$		0.30	2.01E-03	5.97E-04	-2767	4260	<1E-16
	Constrained model (H0): $DFE(N_es)_{Humans}=DFE(N_es)_{Drosophila}$		0.33	2504	407	-12919	24563	<1E-16
No singletons & recent mutation rate estimates	Full model (H1)	Humans	0.16	8.29E-02	6.62E-03	-181		
		Drosophila	0.33	1.20E-03	1.98E-04	-252		
		Total				-433		
	Constrained model (H0): $DFE(s)_{Humans}=DFE(s)_{Drosophila}$		0.26	5.70E-03	1.51E-03	-893	921	<1E-16
	Constrained model (H0): $DFE(N_es)_{Humans}=DFE(N_es)_{Drosophila}$		0.29	5514	797	-10156	19447	<1E-16

Supplementary Table 3. 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 Online Methods). 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	-7.24	4.58	3.58E-04	-282		
		Drosophila	-9.61	4.01	3.36E-05	-649		
		Total				-930		
	Constrained model (H0): $DFE(s)_{Humans}=DFE(s)_{Drosophila}$		-8.91	4.71	6.73E-05	-2841	3822	<1E-16
	Constrained model (H0): $DFE(N_es)_{Humans}=DFE(N_es)_{Drosophila}$		5.91	4.25	183	-13259	24658	<1E-16
Only ortholog genes	Full model (H1)	Humans	-6.43	4.02	8.05E-04	-226		
		Drosophila	-9.10	3.79	5.57E-05	-426		
		Total				-652		
	Constrained model (H0): $DFE(s)_{Humans}=DFE(s)_{Drosophila}$		-8.17	4.61	1.41E-04	-2262	3220	<1E-16
	Constrained model (H0): $DFE(N_es)_{Humans}=DFE(N_es)_{Drosophila}$		6.40	3.99	302	-7131	12959	<1E-16
No singletons	Full model (H1)	Humans	-6.79	5.94	5.62E-04	-181		
		Drosophila	-8.67	4.77	8.54E-05	-261		
		Total				-442		
	Constrained model (H0): $DFE(s)_{Humans}=DFE(s)_{Drosophila}$		-7.52	5.78	2.70E-04	-787	690	<1E-16
	Constrained model (H0): $DFE(N_es)_{Humans}=DFE(N_es)_{Drosophila}$		7.09	5.30	602	-10174	19466	<1E-16
Using recent mutation rate estimates	Full model (H1)	Humans	-7.75	4.58	2.15E-04	-282		
		Drosophila	-8.91	4.01	6.72E-05	-649		
		Total				-930		
	Constrained model (H0): $DFE(s)_{Humans}=DFE(s)_{Drosophila}$		-8.60	4.32	9.17E-05	-1448	1035	<1E-16
	Constrained model (H0): $DFE(N_es)_{Humans}=DFE(N_es)_{Drosophila}$		5.91	4.25	183	-2841	3822	<1E-16
No singletons & recent mutation rate estimates	Full model (H1)	Humans	-7.30	5.94	3.37E-04	-181		
		Drosophila	-7.98	4.77	1.71E-04	-261		
		Total				-442		
	Constrained model (H0): $DFE(s)_{Humans}=DFE(s)_{Drosophila}$		-7.72	5.00	2.21E-04	-472	61	4.82947 E-14
	Constrained model (H0): $DFE(N_es)_{Humans}=DFE(N_es)_{Drosophila}$		7.09	5.30	602	-10174	19466	<1E-16

Supplementary Table 4. 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=0.051			-248	500	0
	Lognormal	mean=-7.24	SD=4.58			-282	567	67
	Gamma+Neutral	shape=0.79	scale=0.0062	$p_{neutral}$=0.304		-223	452	-48
	Martin and Lenormand (2006), eq. 5 (shifted gamma)	shape=0.31	scale=0.014	shift=0.81		-238	481	-19
	Piganeau and Eyre-Walker (2003), eq. 7	shape=0.74	scale=0.015	$N_{e,long-term}$=4760		-220	445	-55
	Lourenço et al. (2011), eq.8	z=0.0092	n=0.74	m=0.63	sigma=0.083	-228	463	-37
	Lourenço et al. (2011), eq.15	m=0.63	sigma=0.091	$N_{e,long-term}$=2100		-225	457	-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

Supplementary Table 5. 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	0.11	0		
	Drosophila	10199*	11	0	
	Yeast	58*	17	103*	0
$DFE(N_e s)_{Species\ 'A'} = DFE(N_e s)_{Species\ 'B'}$	Human	0			
	Mouse	58*	0		
	Drosophila	24563*	20	0	
	Yeast	43*	3.1	98*	0

Supplementary Note 1

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 (Supplementary Table 2; see Online Methods). 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 Messer and Petrov[4] and by theoretical work[5]. 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 (Supplementary Table 2). Similarly large Λ values are found when assuming a log-normal instead of a gamma distributed DFE (Supplementary Table 3). Individually, the gamma distribution fits better than the log-normal distribution in both species (Supplementary Table 2, Supplementary Table 3).

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[6], 3) a DFE based on eq. 8 in Lourenco et al.[2] 4) a DFE based on eq. 15 in Lourenco et al.[2] and 5) a DFE based on eq. 5 in Martin and Lenormand[7]. For all tested cases, the average selection coefficient is at least 22 times more deleterious in humans than in *Drosophila* (Supplementary Fig. 8). Simulations under a null model suggest only an 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* (Supplementary Fig. 14). 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[3,8]. 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* (Supplementary Fig. 14). In conclusion, these results indicate that a model with distinct gamma 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[9,10]. 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[11] 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 (Supplementary Table 1) and the shape and scale estimates of both the full model and the constrained model (Supplementary Table 2, 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 (see Online Methods). Mutations in CNC regions are assumed to be selected with gamma distributed selection coefficients for humans[12] and *Drosophila*[13]. 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 (Supplementary Fig. 3, see also[10]). 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*[14], we also examined the effect of recurrent selective sweeps on our inference. As estimated by Keightley et al.[15], 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 MacDonald-Kreitman table based estimates that 50% of amino acid substitutions are positively selected[15]. We then inferred the demography using synonymous sites and then, conditional on the demographic parameter estimates,

inferred the DFE on 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[16], 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 (Supplementary Fig. 9). 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 (Supplementary Fig. 9).

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[1]. 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[17,18]. However, differences in the mutation rate between synonymous and intronic sites make direct use of the intronic SFS as neutral standard difficult[18]. 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.[1] is correct and 22% of synonymous diversity is missing due to strong selection. We generate 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.[17], 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 (Supplementary Fig. 7). In particular, the difference in the DFE we observe between humans and *Drosophila* is robust to which neutral standard is used.

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

Another concern is that the observed differences in the DFE between species could be solely due to genes that are specific to one species that might be under more or less constraint than genes that are common to both species. Further, 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[19–21]. Thus, we investigated how these factors affect our results.

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 above, and revealed a significant difference in the DFE between humans and *Drosophila* ($\Lambda = 7{,}369$, $p < 10^{-16}$; see Supplementary Table 2, Supplementary Table 3, Supplementary Fig. 5A, Supplementary Fig. 6A).

To investigate how differences in gene expression could affect our results, we classify genes into sets with different gene expression profiles (Supplementary Fig. 2; see Online Methods). We use two recent gene expression datasets from humans[22] and *Drosophila*[23] 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 (Supplementary Fig. 1). 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 50-80 fold more negative for humans than for *Drosophila*, regardless of the overall expression level or tissue specificity of the genes (Supplementary Fig. 12A). In summary, different sets of genes have very different DFEs. However, this effect cannot explain our findings that selection coefficients are more deleterious in humans than in *Drosophila*.

Supplementary Note 2

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 four categories of models lead to contrasting predictions regarding DFE differences between *Drosophila* and humans. Here we describe the assumptions and the predictions of the four models that we use to discriminate among them.

Protein stability models

The basic idea behind protein stability models is that much of the selection pressure on coding regions involves maintaining thermodynamic stability of the proteins[19,20,24]. 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[24], the cytotoxic effects of protein misfolding[25], or the effect of enzyme stability on metabolic flux[25]. The distribution of $N_e s$ values that is generated by such a one-dimensional fitness-phenotype relationship was shown to be gamma distributed[25] and independent of the effective population size (N_e) when at equilibrium[24]. 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 130-420 times larger in humans, this model predicts that $|s|$ must be at least 130 times smaller in *Drosophila*. We observe a factor of less than 100 (Supplementary Fig. 8). 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 43 times larger than in humans, thus $|s|$ must be 43 times smaller in mouse. We estimate a factor of only 2.8 (Fig. 4A), inconsistent with the protein stability prediction.

Note that a major assumption made by protein stability models is equimutability[26], 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[27–29], however, some authors propose a negative correlation between $\Delta\Delta G$ and ΔG[25,30]. 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[24]. 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[31]. These models usually make two assumptions[31–34]: 1) the back-mutation has the same absolute value of 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[6], eq. 7, derived a formula for the equilibrium DFE as a function of population size (see also eq. 6 in Rice et al.[34]). To test predictions of the back-mutation model, we use the formula of Piganeau and Eyre-Walker with our Poisson Random Field method by assuming gamma distributed effect sizes (i.e., $|s| \sim \Gamma(\alpha, \beta)$). Importantly, because the back-mutation model does not make any prediction about how this effect size distribution (the distribution of $|s|$) differs between species, we assume that under the null model it is the same between species, and all the differences in the DFE we see between species are due to different proportions of beneficial vs. deleterious mutations. Thus, we can test the back-mutation model by testing for differences in the effect size distribution between species, for example by testing for differences in $E[|s|]$.

Mutational robustness models

Mutational robustness models postulate that more robust organisms have, on average, less deleterious mutations[35]. 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[36]. 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[37–40]. 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[41]. 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[42]. 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[43]. In such models, smaller populations tend to evolve higher levels of robustness[44], and this tendency is increased by higher phenotypic complexity and more positive epistasis[43].

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[35]. 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 relaxing the assumption that every mutation affects every phenotype (restricted pleiotropy[29–31]), or accounting for correlations in the fitness function[7,46]. 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[35,45,2,7].

Here, we test three predictions of FGM (see also Fig. 1): 1) More complex organisms have more deleterious mutations, since mutations are more likely to disrupt something important in a complex organism than in a simple one[47] (see Supplementary Note 3 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[30]). 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[7].

The first prediction suggests that humans have more deleterious mutations than *Drosophila*, since humans likely have higher complexity (larger number of genes, proteins and protein-protein interactions[48], and cell types[49]). 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[7].

The second prediction suggests that because of their smaller effective population size, humans contain more slightly beneficial mutations than *Drosophila*. We tested this prediction by fitting a distribution that is based on FGM as derived by Lourenço et al.[2]. Similar to the back-mutation model of Piganeau and Eyre-Walker[6], 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 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[32,33]. 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[2]. 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[7]. 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.[2], eq. 15, to the data using the Poisson Random Fields approach (see Online Methods). 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.[2]) that can be used to test for deviations from equilibrium conditions.

Supplementary Note 3

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 Supplementary Table 4).

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 factors (see Online Methods and Supplementary Note 1). 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-folding 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 (Supplementary Fig. 8A).

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 (Supplementary Fig. 8B). 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 Supplementary Note 2).

Supplementary Note 4

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[7]. Lastly, mutational pleiotropy, *m*, refers to the number of traits affected by a particular mutation.

Under a simple model of universal pleiotropy, mutations are assumed to affect all phenotypic axes similarly and have no preferred direction. In this model, $n=n_e=m$. Additionally, the average selection coefficient E[*s*] is predicted to be negative, and becomes more negative with increasing phenotypic complexity *n* and thus increasing pleiotropy *m*[35]. Hence, one would predict 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[50]). 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[7]. Lourenco et al.[2] 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[2].

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[7] (eq. 1) and Chevin et al[45] (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. Note that E[*s*] is independent of the distance to the optimum[7], therefore this model does not rely on any assumptions about the degree of maladaptation.

Importantly, however, the dependence of n_e on *n* is more complex. Specifically, $\mathrm{n_e} = \frac{\mathrm{n}}{1+\mathrm{n}(\rho_S^2+\rho_M^2)} \xrightarrow{\mathrm{n}\to\infty} \frac{1}{\rho_S^2+\rho_M^2}$, where ρ_S and ρ_M measure the strength of selective and mutational correlations averaged over all traits (eq. 8 in Martin and Lenormand[7]). Thus, Martin and Lenormand find that as *n* increases indefinitely, n_e reaches a limiting value that depends only on phenotypic correlations, but is independent of complexity *n*. It is possible then, under the assumption that complexity *n* is generally large in most species, that organisms with increasing complexity (increasing *n*) may not show an increasing n_e, as supported by their empirical studies.

In Lourenco et al[2], E[*s*] does not depend on *n* (see eq. A3 in Lourenco et al.[2]). Instead, here, the DFE depends on the pleiotropy parameter *m* and the scale parameter σ. However, there are some similarities between the Martin and Lenormand and the Lourenco models. Both models lead to similar predictions regarding the shape of the DFE (a gamma distribution at the optimum) and thus can be shown to have equivalent parameters. The scale parameter of

Lourenco, σ, corresponds to the effective scale parameter λ_e ($\sigma = \mathrm{sqrt}(\lambda_e/2)$) of Martin and Lenormand[7] and Chevin et al.[45]. The effective complexity, n_e, of Martin and Lenormand corresponds to the pleiotropy parameter of Lourenco et al. ($m = n_e$). In the Lourenco 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[7]), the scale parameter in Lourenco et al. (σ) is fixed and does not depend on other parameters of the model (eq. 2 in Lourenco et al.[2]).

The independence of E[*s*] from *n* under the model of Lourenco et al. is therefore a result of the assumption that the scaling parameter σ is considered fixed and does not depend on *n* in their formulation. Thus, by design, rather than biology, their model does not include such a relationship. However, biologically, it seems unrealistic that when total complexity is increased (i.e. there is an evolutionary increase in the total number of selected phenotypes, or in other words, *n* increases), mutations would not 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 contradict the assumption of constant σ and suggest that mutations that affect more phenotypes also show a larger effect on each individual phenotype[51,52]. Thus, constancy of the scaling parameter σ as defined by Lourenco et al. (the standard deviation in effect size per trait) is not supported by experimental data.

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 (Supplementary Fig. 11). 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[7] is in line with the prediction of FGM.

References

1. Lawrie, D. S., Messer, P. W., Hershberg, R. & Petrov, D. A. Strong purifying selection at synonymous sites in *D. melanogaster*. *PLoS Genet* **9,** e1003527 (2013).
2. Lourenço, J., Galtier, N. & Glémin, S. Complexity, pleiotropy, and the fitness effect of mutations. *Evol. Int. J. Org. Evol.* **65,** 1559–1571 (2011).
3. Kim, B. Y., Huber, C. D. & Lohmueller, K. E. Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. (2016).
4. Messer, P. W. & Petrov, D. A. Frequent adaptation and the McDonald-Kreitman test. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 8615–8620 (2013).
5. Nicolaisen, L. E. & Desai, M. M. Distortions in genealogies due to purifying selection and recombination. *Genetics* **195,** 221–230 (2013).
6. Piganeau, G. & Eyre-Walker, A. Estimating the distribution of fitness effects from DNA sequence data: implications for the molecular clock. Proc. Natl. Acad. Sci. U. S. A. 100, 10335–10340 (2003).
7. Martin, G. & Lenormand, T. A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evol. Int. J. Org. Evol.* **60,** 893–907 (2006).
8. Kousathanas, A. & Keightley, P. D. A comparison of models to infer the distribution of fitness effects of new mutations. *Genetics* **193,** 1197–1208 (2013).
9. Bustamante, C. D., Wakeley, J., Sawyer, S. & Hartl, D. L. Directional selection and the site-frequency spectrum. *Genetics* **159,** 1779–1788 (2001).
10. Zhu, L. & Bustamante, C. D. A composite-likelihood approach for detecting directional selection from DNA sequence data. *Genetics* **170,** 1411–1421 (2005).
11. Ewing, G. B. & Jensen, J. D. The consequences of not accounting for background selection in demographic inference. *Mol Ecol.* **25,** 135–141 (2016).
12. Torgerson, D. G. *et al.* Evolutionary processes acting on candidate cis-regulatory regions in humans inferred from patterns of polymorphism and divergence. *PLoS Genet* **5,** e1000592 (2009).
13. Casillas, S., Barbadilla, A. & Bergman, C. M. Purifying selection maintains highly conserved noncoding sequences in Drosophila. *Mol. Biol. Evol.* **24,** 2222–2234 (2007).
14. Sella, G., Petrov, D. A., Przeworski, M. & Andolfatto, P. Pervasive natural selection in the *Drosophila* genome? *PLoS Genet.* **5,** (2009).
15. Keightley, P. D., Campos, J. L., Booker, T. R. & Charlesworth, B. Inferring the frequency spectrum of derived variants to quantify adaptive molecular evolution in protein-coding genes of *Drosophila melanogaster*. *Genetics* (2016). doi:10.1534/genetics.116.188102
16. Tataru, P., Mollion, M., Glemin, S. & Bataillon, T. Inference of distribution of fitness effects and proportion of adaptive substitutions from polymorphism data. *bioRxiv* 062216 (2016). doi:10.1101/062216
17. Parsch, J., Novozhilov, S., Saminadin-Peter, S. S., Wong, K. M. & Andolfatto, P. On the utility of short intron sequences as a reference for the detection of positive and negative selection in *Drosophila*. *Mol. Biol. Evol.* **27,** 1226–1234 (2010).
18. Clemente, F. & Vogl, C. Unconstrained evolution in short introns? - an analysis of genome-wide polymorphism and divergence data from *Drosophila*. *J. Evol. Biol.* **25,** 1975–1990 (2012).
19. Drummond, D. A. & Wilke, C. O. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **134,** 341–352 (2008).
20. Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O. & Arnold, F. H. Why highly expressed proteins evolve slowly. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 14338–14343 (2005).

21. Zhang, J. & Yang, J.-R. Determinants of the rate of protein sequence evolution. *Nat. Rev. Genet.* **16,** 409–420 (2015).
22. Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* **13,** 397–406 (2014).
23. Li, J. J., Huang, H., Bickel, P. J. & Brenner, S. E. Comparison of *D. melanogaster* and *C. elegans* developmental stages, tissues, and cells by modENCODE RNA-seq data. *Genome Res.* **24,** 1086–1101 (2014).
24. Goldstein, R. A. Population size dependence of fitness effect distribution and substitution rate probed by biophysical model of protein thermostability. *Genome Biol. Evol.* **5,** 1584–1593 (2013).
25. Serohijos, A. W. R. & Shakhnovich, E. I. Contribution of selection for protein folding stability in shaping the patterns of polymorphisms in coding region. *Mol Biol Evol* **31,** 165–176 (2014).
26. Cherry, J. L. Should we expect substitution rate to depend on population size? *Genetics* **150,** 911–919 (1998).
27. Bloom, J. D., Raval, A. & Wilke, C. O. Thermodynamics of neutral protein evolution. *Genetics* **175,** 255–266 (2007).
28. Goldstein, R. A. The evolution and evolutionary consequences of marginal thermostability in proteins. *Proteins Struct. Funct. Bioinforma.* **79,** 1396–1407 (2011).
29. Tokuriki, N., Stricher, F., Schymkowitz, J., Serrano, L. & Tawfik, D. S. The stability effects of protein mutations appear to be universally distributed. *J. Mol. Biol.* **369,** 1318–1332 (2007).
30. Serohijos, A. W. R., Rimas, Z. & Shakhnovich, E. I. Protein biophysics explains why highly abundant proteins evolve slowly. *Cell Rep.* **2,** 249–256 (2012).
31. Charlesworth, J. & Eyre-Walker, A. The other side of the nearly neutral theory, evidence of slightly advantageous back-mutations. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 16992–16997 (2007).
32. Bulmer, M. The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129,** 897–907 (1991).
33. Mustonen, V. & Lässig, M. From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. *Trends Genet.* **25,** 111–119 (2009).
34. Rice, D. P., Good, B. H. & Desai, M. M. The evolutionarily stable distribution of fitness effects. *Genetics* **200,** 321–329 (2015).
35. Tenaillon, O. The utility of Fisher's geometric model in evolutionary genetics. *Annu. Rev. Ecol. Evol. Syst.* **45,** 179–201 (2014).
36. Kimura, M. Model of effectively neutral mutations in which selective constraint is incorporated. *Proc. Natl. Acad. Sci. U. S. A.* **76,** 3440–3444 (1979).
37. Azevedo, R. B. R., Lohaus, R., Srinivasan, S., Dang, K. K. & Burch, C. L. Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. *Nature* **440,** 87–90 (2006).
38. Kacser, H. & Burns, J. A. The molecular basis of dominance. *Genetics* **97,** 639–666 (1981).
39. van Nimwegen, E., Crutchfield, J. P. & Huynen, M. Neutral evolution of mutational robustness. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 9716–9720 (1999).
40. Siegal, M. L. & Bergman, A. Waddington's canalization revisited: developmental stability and evolution. *Proc. Natl. Acad. Sci.* **99,** 10528–10532 (2002).
41. Gardner, A. & Kalinka, A. T. Recombination and the evolution of mutational robustness. *J. Theor. Biol.* **241,** 707–715 (2006).
42. Siegal, M. L. & Leu, J.-Y. On the nature and evolutionary impact of phenotypic robustness mechanisms. *Annu. Rev. Ecol. Evol. Syst.* **45,** 495–517 (2014).
43. Gros, P.-A. & Tenaillon, O. Selection for chaperone-like mediated genetic robustness at low mutation rate: impact of drift, epistasis and complexity. *Genetics* **182,** 555–564 (2009).

44. Krakauer, D. C. & Plotkin, J. B. Redundancy, antiredundancy, and the robustness of genomes. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 1405–1409 (2002).
45. Chevin, L.-M., Martin, G. & Lenormand, T. Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution. *Evol. Int. J. Org. Evol.* **64,** 3213–3231 (2010).
46. Waxman, D. Fisher's geometrical model of evolutionary adaptation--beyond spherical geometry. *J. Theor. Biol.* **241,** 887–895 (2006).
47. Orr, H. A. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* **52,** 935–949 (1998).
48. Stumpf, M. P. H. *et al.* Estimating the size of the human interactome. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 6959–6964 (2008).
49. Valentine, J. W., Collins, A. G. & Meyer, C. P. Morphological complexity increase in metazoans. *Paleobiology* **20,** 131–142 (1994).
50. Wagner, G. P. & Zhang, J. The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nat. Rev. Genet.* **12,** 204–213 (2011).
51. Wagner, G. P. et al. Pleiotropic scaling of gene effects and the 'cost of complexity'. *Nature* **452,** 470–472 (2008).
52. Wang, Z., Liao, B.-Y. & Zhang, J. Genomic patterns of pleiotropy and the evolution of complexity. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 18034–18039 (2010).