DFEnitely different: Genome-wide characterization of differences in mutation fitness effects between populations

Alyssa Lyn Fortier[1], Alec J. Coffman[2], Travis J. Struck[3], Megan N. Irby[3], Jose E. L. Burguete[4], Aaron P. Ragsdale[5], Ryan N. Gutenkunst[3*]

[1]Stanford University, [2]University of Pennsylvania, [3]University of Arizona,

[4]National Autonomous University of Mexico, [5]McGill University

[*]To whom correspondence should be addressed: rgutenk@email.arizona.edu

Abstract

The effect of a mutation on fitness may differ between populations, depending on environmental and genetic context. Experimental studies have shown that such differences exist, but little is known about the broad patterns of such differences or the factors that drive them. To quantify genome-wide patterns of differences in mutation fitness effects, we extended the concept of a distribution of fitness effects (DFE) to a joint DFE between populations. To infer the joint DFE, we fit parametric models that included demographic history to genomic data summarized by the joint allele frequency spectrum. Using simulations, we showed that our approach is statistically powerful and robust to many forms of model misspecification. We then applied our approach to populations of *Drosophila melanogaster*, wild tomatoes, and humans. We found that mutation fitness effects are overall least correlated between populations in tomatoes and most correlated in humans, corresponding to overall genetic differentiation. In *D. melanogaster* and tomatoes, mutations in genes involved in immunity and stress response showed the lowest correlation of fitness effects, consistent with environmental influence. In *D. melanogaster* and humans, deleterious mutations showed a lower correlation of fitness effects than tolerated mutations, hinting at the complexity of the joint DFE. Together, our results show that the joint DFE can be reliably inferred and that it offers extensive insight into the genetics of population divergence.

Introduction

Mutations that alter organismal fitness are the key input into the evolutionary process. Given a new mutation's effect on fitness, population genetics theory can predict a great deal; for example, how likely the mutation is to be lost from or fix in the population. But population genetics theory cannot predict how likely a new mutation is to have a given effect on fitness. It is known that typically the majority of mutations are deleterious (reduce fitness) or nearly neutral (negligible effect on fitness), so only a small minority are adaptive (increase fitness). But these three categories encompass a continuum of fitness effects. This continuum is quantified by the distribution of fitness effects (DFE) among new mutations (Eyre-Walker & Keightley 2007). The DFE is central to many fundamental evolutionary topics, such as the maintenance of genetic variation (Charlesworth 1994) and the evolution of recombination (Barton 1995). It is also key to many practical evolutionary topics, such as the emergence of pathogens (Gandon et al. 2013) and the genetic architecture of complex disease (Eyre-Walker 2010).

Here we focus on the deleterious component of the DFE, which both theory and experiment have contributed to understanding. Fisher's geometrical model (Fisher 1930; Tenaillon 2014), which considers evolution in an abstract trait space, has been used to predict the shape of the DFE (Martin & Lenormand 2006). Although Fisher's model is simple, some of its predictions have found empirical support (Bank et al. 2014; Huber et al. 2017). Experimental approaches to measure the DFE using either random mutagenesis (Elena et al. 1998) or mutation accumulation (Fry et al. 2002) are typically limited to studying a small number of

mutations. But they have shown that the deleterious portion of the DFE has a long tail (Sanjua et al. 2004). Recent approaches based on high-throughput sequencing promise to capture many more mutations, at least in microbes (Bank et al. 2014).

In non-microbes, most of our knowledge regarding the DFE has come from model-based inferences based on contemporary patterns of natural genetic variation. In the typical approach, a model of demographic history (potentially including population size change, divergence, and migration) is first fit to putatively neutral variants, then a model of demographic history plus selection is fit to another class of variants to estimate their DFE (Eyre-Walker et al. 2006; Keightley & Eyre-Walker 2007; Boyko et al. 2008). In general, these analyses summarize genetic data using the allele frequency spectrum (AFS). This approach has been applied to numerous organisms, including humans (Kim et al. 2017), *Arabidopsis thaliana* (Huber et al. 2018), *Drosophila melanogaster* (Keightley & Eyre-Walker 2007), and orangutans (Ma et al. 2013).

To date, inference approaches have not considered the key point that the fitness effect of a mutation may depend on genetic or environmental context. Experimental studies in numerous model organisms, including *D. melanogaster* (Kondrashov & Houle 1994; Wang et al. 2014) and *Daphnia pulex* (Latta et al. 2015), have shown that the fitness effect of a mutation may differ between environments or genetic backgrounds. Population genetic studies have compared DFEs inferred for multiple populations of humans (Boyko et al. 2008), orangutans (Ma et al. 2013), and chimpanzees (Tataru & Bataillon 2019), finding some evidence for differences between populations. These studies, however, have been limited by the implicit assumption that the fitness effects of a given mutation in different populations are independent draws from distinct DFEs. Biologically, the fitness effects of a given mutation in different contexts are correlated, but it is unclear how large that correlation typically is and what factors affect it.

Here, we developed and applied an approach for inferring the joint distribution of fitness effects of new mutations between pairs of populations from the joint allele frequency spectrum. Considering a joint DFE allowed us to quantify the correlation of mutation fitness effects between populations. Using simulations, we showed that our approach requires only modest sample sizes and is robust to many forms of model misspecification. We applied our approach to data from humans, *D. melanogaster*, and wild tomato. We found that the correlation in fitness effects is lowest in tomato and highest in humans. In tomato and *D. melanogaster*, we found biologically interpretable differences in the correlation of the DFE among genes with different functions. In all three cases, we found that mutations with more deleterious effects exhibit lower correlations. More broadly, our results show that the joint DFE offers new insight into the population genetics of these populations.

Results

In the simplest case, the joint DFE is defined between two populations that have recently diverged, one of which may have entered a new environment (Fig. 1A). Any given mutation has selection coefficient s_1 in the ancestral population and s_2 in the recently diverged population. We based our inference on the joint allele frequency spectrum (AFS). For two populations, the joint AFS is a matrix in which each entry i, j corresponds to the number of variants observed at frequency i in population 1 and j in population 2 in a sequenced sample of individuals from the two populations. Different combinations of s_1 and s_2 lead to distinct patterns in the AFS (Fig. 1B). We were interested in inferring the joint probability distribution for (s_1, s_2), which we refer to as the joint distribution of fitness effects (Fig 1C). To calculate the expected AFS for a given demographic model and DFE, we first cached calculations of the expected AFS for a grid of selection coefficient pairs. Assuming independence among sites, the expectation for the full DFE is then an integration over values of s_1, s_2, weighted by the DFE (Fig. S1) (Ragsdale et al. 2016; Kim et al. 2017).

Little is known about the shape of the joint DFE, so we considered multiple parametric models. The best-fitting models for single populations tend to be lognormal or gamma distributions (Boyko et al. 2008), although discrete distributions may sometimes fit better (Kousathanas & Keightley 2013). First, we considered a bivariate lognormal distribution (Fig. 1C), because it has an easily interpretable correlation coefficient. (Gamma distributions are also popular for modeling DFEs, but there are multiple ways of defining a bivariate gamma distribution (Nadarajah & Gupta 2006).) But the bivariate lognormal distribution can be

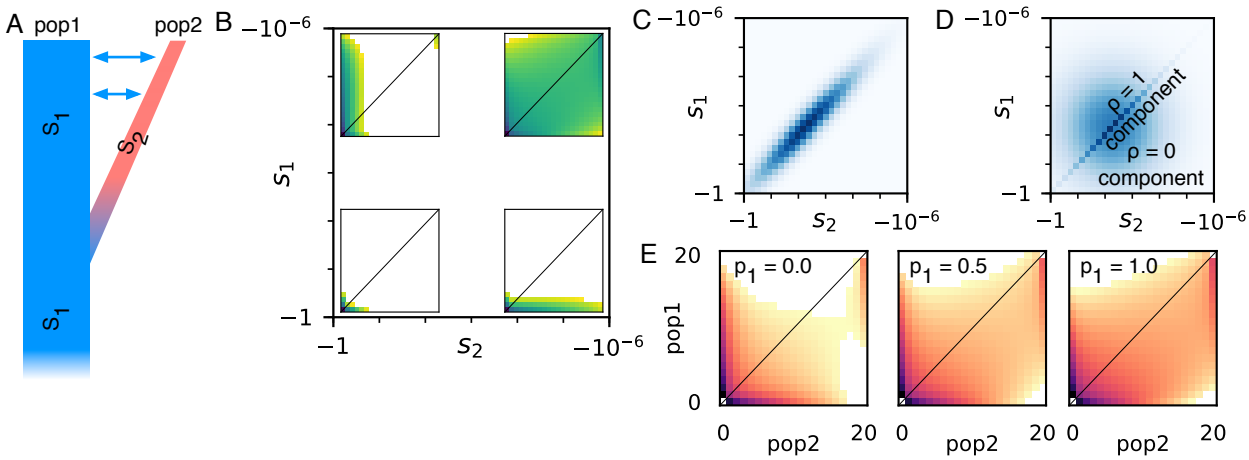

Figure 1: **The joint allele frequency spectrum and joint distribution of fitness effects.** A: Populations that have recently diverged or have gene flow between them will share genetic variants. Some of those variants will have a different effect on fitness in the diverged population (s_2) than in the ancestral population (s_1). B: The joint distribution of fitness effects (DFE) is defined over pairs of selection coefficients (s_1, s_2). Insets show joint allele frequency spectra corresponding to selection coefficients for pairs of variants that are strongly or weakly deleterious in each population. In each frequency spectrum, the number of segregating variants at a given pair of allele frequencies is exponential with the color depth. C: One potential model for the joint DFE is a bivariate lognormal distribution, illustrated here for strong correlation. D: Another potential model is a mixture of components corresponding to equality ($\rho = 1$) and independence ($\rho = 0$) of fitness effects. E: As illustrated by these simulated allele frequency spectra, stronger correlations of mutation fitness effects lead to more shared polymorphism. Here p_1 is the weight of the $\rho = 1$ component in the mixture model.

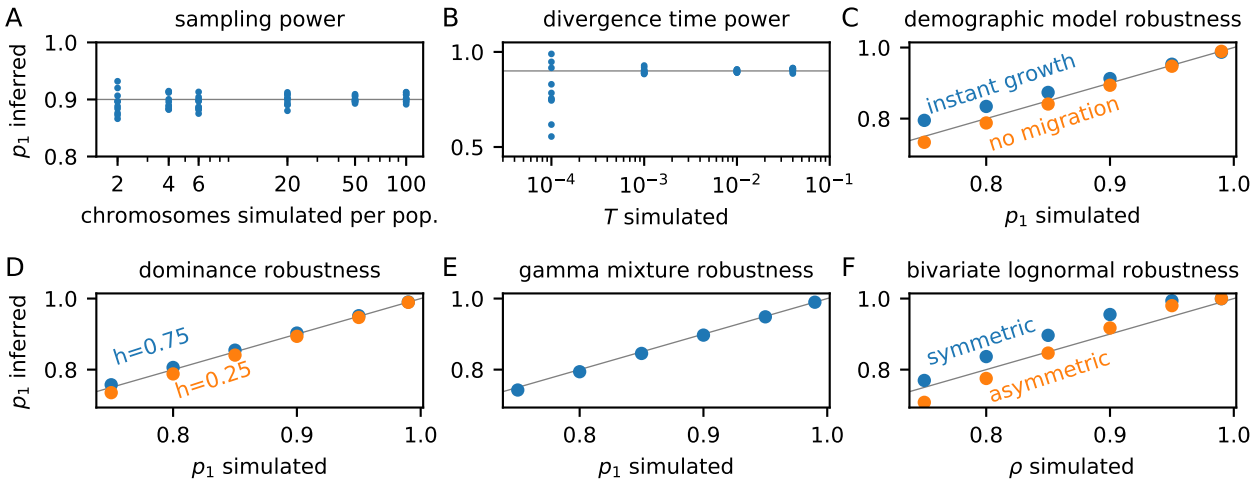

Figure 2: **Power and robustness of joint DFE inference.** Using simulated data, we tested the statistical power of joint DFE inference and its robustness to model misspecification, focusing on inferences of the mixture proportion p_1 for the perfectly correlated component of the joint DFE (Fig. 1D). See Methods for simulation details. A: Precise inference of p_1 was possible even for modest sample sizes. B: Precise inference of p_1 was also possible even when the populations were only recently diverged. Here T is the population-scaled divergence time. C: The bias in inferred values of p_1 was modest when data were simulated with a demographic model including both exponential growth and migration but analyzed assuming instantaneous growth or no migration. D: The bias in p_1 was also modest when data were simulated with either dominant ($h = 0.75$) or recessive ($h = 0.25$) derived alleles, but analyzed assuming additive alleles ($h = 0.5$). E: The bias in p_1 was also modest when data were simulated under a mixture model with gamma-distributed components, but analyzed assuming lognormal components. F: When data were simulated under a bivariate lognormal model with given correlation coefficient ρ, the inferred mixture proportion p_1 was similar to the simulated ρ, for both symmetric and asymmetric bivariate lognormal simulations.

numerically poorly behaved as the correlation coefficient approaches one and the distribution becomes very thin. We thus also considered a mixture model that consisted of a component corresponding to perfect correlation, with weight p_1, and a component corresponding to zero correlation, with weight $(1 - p_1)$ (Fig. 1D). The mixture proportion p_1 can be interpreted similarly to the correlation coefficient in the bivariate normal distribution.

The correlation of the joint DFE profoundly affects the expected AFS (Fig. 1E). Qualitatively, if the correlation is low, there is little shared high-frequency polymorphism. In this case, alleles that are nearly neutral in one population are often deleterious in the other, driving their frequencies lower in that population. If the correlation of the joint DFE is larger, more shared polymorphism is preserved. To calculate the expected AFS for a given joint DFE and model of demographic history, we extended the `fitdadi` framework developed by Kim et al. (2017), which is itself built upon our `dadi` software (Gutenkunst et al. 2009).

Methodological

Before analyzing real data, we carried out a number of tests on simulated data, to evaluate the statistical power for inferring the joint DFE from the AFS and robustness to potential sources of bias.

To evaluate statistical power, we simulated unlinked SNPs under a known demographic model and a lognormal mixture model for the joint DFE (Fig. 1D); then we fit the joint DFE to those data. We found that the variance of the estimated mixture coefficient p_1 grew only slowly as the sample size decreased (Fig. 2A). This suggests that only modest sample sizes are necessary to confidently infer the correlation of the joint DFE, similar to inference for the univariate DFE (Keightley & Eyre-Walker 2010). As expected,

statistical power is lower for population pairs that have diverged for less time, although power to infer p_1 is ample for divergence times much smaller than that, for example, between continental human populations (Fig. 2B).

A key assumption in the DFE inference procedure is the form of the demographic model, so we tested how imperfect modeling of demographic history would bias inference. To do so, we simulated both neutral and selected data under a true demographic history that included divergence, exponential growth in both populations, and asymmetric migration. We then fit the neutral data with models that lacked migration or that modeled instantaneous growth. We then used these misspecified models as the basis for fitting the joint DFE. For both misspecified demographic models, we found that inferences of the mixture coefficient were not strongly biased (Fig. 2C).

Dominance is a potential confounding factor when inferring the joint DFE, since dominance influences allele frequencies differently in populations that have and have not undergone a bottleneck (Balick et al. 2015). Typically, mutation fitness effects in diploids are assumed to be additive, corresponding to a dominance coefficient of $h = 0.5$, although recent work suggests that the average dominance coefficient may vary with the fitness effect of the mutation (Huber et al. 2018). To test the effects of dominance on our inference, we simulated frequency spectra with dominance coefficients $h = 0.25$ and $h = 0.75$ and then optimized joint DFE parameters under the assumption that $h = 0.5$. An incorrect assumption about dominance did bias the inferred mean μ and variance σ of the marginal DFEs (data not shown), but it did not substantially bias the mixture proportion p_1 (Fig 2D).

The parametric form assumed for the joint DFE is another potential source of model misspecification. To test how this might bias inference, we first simulated a true mixture model in which the marginal distributions were gamma, rather than lognormal. In this case, we found that inference of the mixture proportion was not substantially biased (Fig. 2E). We also considered fitting the mixture lognormal model (Fig. 1D) to data simulated under a bivariate lognormal model (Fig. 1C). In this case, we found that the inferred values of p_1 we larger than the simulated correlation coefficient ρ, although they were similar (Fig. 2F). The mixture model assumes symmetric marginal distributions between the two populations, but the bivariate lognormal model is more general and permits asymmetric marginal distributions. When we simulated data under a bivariate model with asymmetric means and variances of the marginal distributions, but fit with a symmetric mixture model, we found only slight bias, similar to the symmetric bivariate case (Fig. 2F).

Together, our tests on simulated data suggest that the joint allele frequency spectrum has good power for inferring the joint distribution of fitness effects. Moreover, inference of the mixture proportion of our joint DFE model is robust to multiple confounding factors, including misspecification of the demographic and selection models.

Biological

We applied our joint DFE inference approach to *D. melanogaster*, wild tomatoes, and humans, using data from the Drosophila Genome Nexus (Lack et al. 2016), Beddows et al. (2017), and the 1000 Genomes Project (The 1000 Genomes Project Consortium 2015). For *D. melanogaster*, we considered the joint DFE between Zambian and French populations, because the Zambian population is representative of the ancestral population (Lack et al. 2015) and France is a distinct environment. For tomatoes, we considered the joint DFE between two closely related species, *Solanum chilense* Dunal and *Solanum peruvianum* L., because they still share substantial polymorphism and have overlapping ranges. For humans, we considered the joint DFE between Yoruba (YRI) and Northern and Western European (CEU) populations, because the Yoruba are a well-studied proxy for the ancestral human population and European populations parallel the history of *D. melanogaster*.

To begin our analysis, we first fit models of demographic history to synonymous variants in each population pair (Fig. 3A,D,&G). (Although linked selection may bias demographic inference from synonymous variants, subsequent inferences of selection are expected to be relatively unbiased (Boyko et al. 2008; Messer & Petrov 2013).) For all three analyses, we fit relatively simple models of divergence with exponential growth and gene flow, although for humans we also found it necessary to include pre-divergence population growth. Broadly, these models fit the data well (Fig. 3B,E,&H, Table S1,S2,&S3).

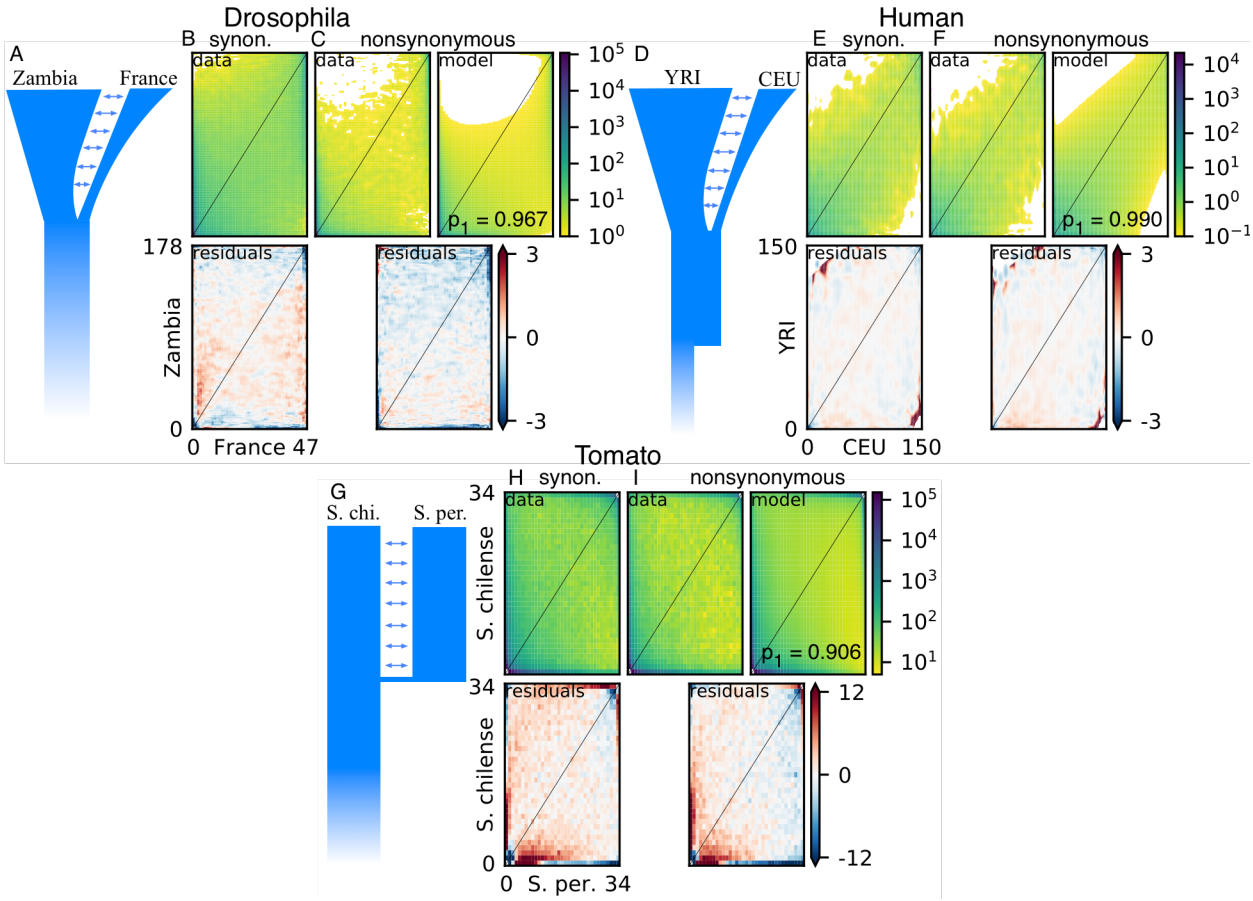

Figure 3: **Model fits to allele frequency spectra.** A) For Drosophila, we fit an isolation-with-migration demographic model to the synonymous data. B) That model fit the data well, as evidenced by the small and mostly uncorrelated residuals. C) The nonsynonymous data showed a substantial reduction in shared polymorphism. Those data were best fit by a joint DFE mixture model (Fig. 1D) with $p_1 = 0.97$, and the resulting model had small residuals. D) For humans, we fit an isolation-with-migration demographic model that included growth before divergence. E) That model also fit the data well. (In these plots, we projected down to a smaller sample size, to smooth variation due to the sparsity of the spectrum.) F) The nonsynonymous human data were well-fit by a joint DFE mixture model with $p_1 = 0.99$. G) For wild tomatoes, we fit an isolation-with-migration model to data from two closely related species. H) That model fit the data moderately well. I) When we fit the nonsynonymous tomato data with a joint DFE mixture model, we inferred $p_1 = 0.91$, with similar fit quality.

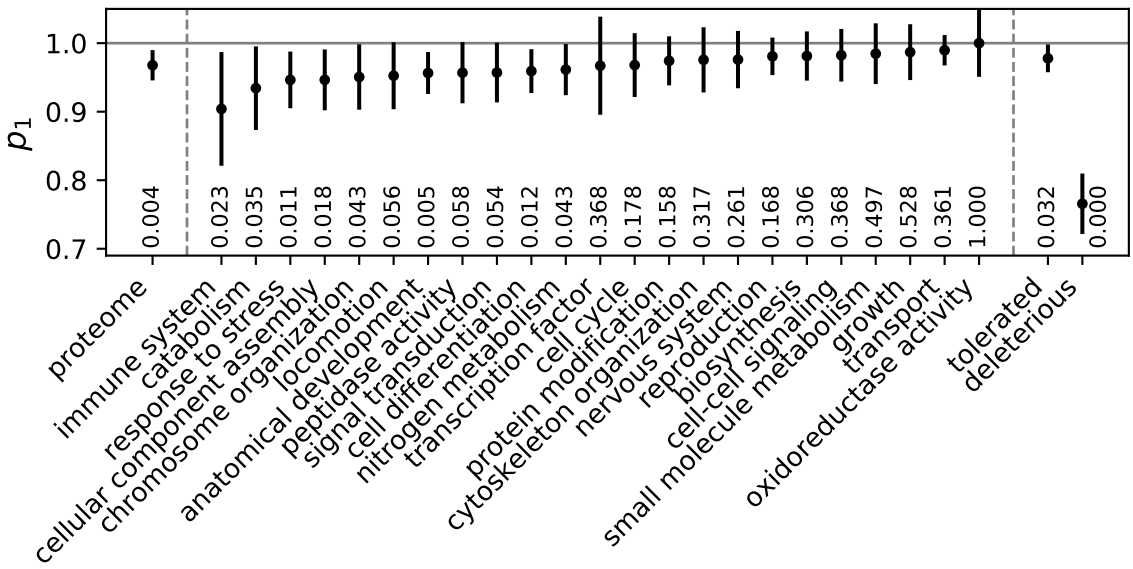

Figure 4: **Joint DFE mixture proportions inferred between Zambian and French *D. melanogaster* populations.** Plotted are maximum likelihood inferences with 95% confidence intervals. Numbers indicate p-values from two-sized z-tests. 'Proteome' refers to all nonsynonymous SNPs. Inferences are also plotted for the largest Gene Ontology slim terms. 'Tolerated' and 'deleterious' refer to SNPs classified by functional impact using SIFT (Vaser et al. 2016).

We next estimated the joint DFE using all nonsynonymous variants in each proteome, using our lognormal mixture model for the joint DFE (Fig. 1D). For *D. melanogaster*, we found that mutation fitness effects between Zambian and French populations were highly correlated, with $p_1 = 0.967 \pm 0.022$ (Table S4). For humans we found an even higher correlation that was statistically indistinguishable from perfect correlation, $p_1 = 0.990 \pm 0.010$ (Table S5). For tomatoes, we found a substantially smaller correlation, $p_1 = 0.906 \pm 0.019$. In all three cases, the resulting models fit the nonsynonymous joint frequency spectrum well, with similar patterns of residuals to the demographic models fit to synonymous data (Fig. 3C,F,&I).

To investigate the biological basis of the joint DFE, we then considered genes of different function based on Gene Ontology (GO) terms (The Gene Ontology Consortium 2000). For *D. melanogaster*, we found a wide range of inferred correlations, with the lowest corresponding to mutations in genes involved in the immune system at $p_1 = 0.904 \pm 0.088$ (Fig. 4, Table S4). Although the p-values for many individual terms were modest, considering all tested GO terms, Fisher's combined probability test yielded a p-value of $p \sim 10^{-6}$, suggesting that our results are highly enriched for deviations below $p_1 = 1$. For tomatoes, we found an even wider range of inferred correlations, with the lowest being genes involved in responding to abiotic stimulus at $p_1 = 0.858 \pm 0.093$ (Fig. 5, Table S6). For humans, we found that all GO terms yielded values of p_1 that were statistically indistinguishable from one (Fig. S2). Among the *D. melanogaster* terms, we found no correlation between the inferred mixture proportion p_1 and other parameters in the joint DFE (Fig. S3), suggesting that the variation we see in p_1 is not driven simply by variation in overall constraint.

In humans, we further explored the biological context of the joint DFE by considering genes that are involved in disease and that interact with pathogens. We found a weak signal that genes involved in Mendelian disease tend to have less correlated selection coefficients than other genes (Fig. 6). We also found that genes that interacted with multiple viruses had highly correlated selection coefficients, but those that were known to interact with only a single virus exhibited a significantly lower correlation (Fig. 6).

We used the variation among our inferences for *D. melanogaster* GO terms to test the robustness of our analyses to various modeling choices. We first fit simpler models of demographic history with instantaneous growth or no migration to the synonymous data (Table S1), then used those models as the basis of joint DFE

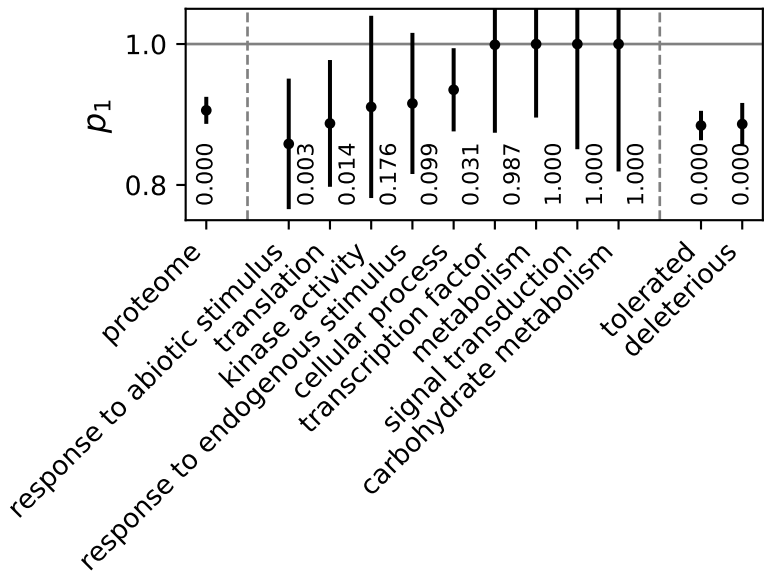

Figure 5: **Joint DFE mixture proportions inferred between wild tomato species.** As in Fig. 4, but for two species of wild tomato, *Solanum chilense* Dunal and *Solanum peruvianum* L. (Beddows et al. 2017).

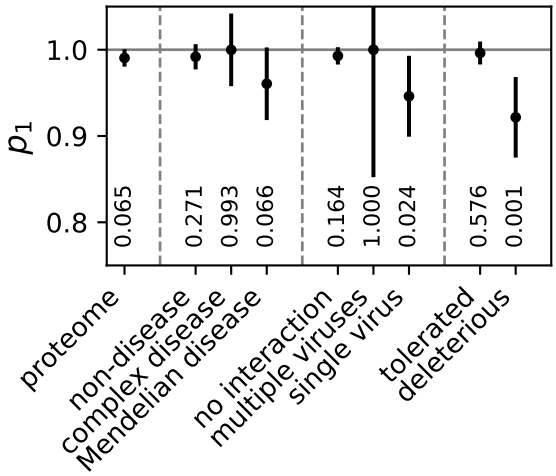

Figure 6: **Joint DFE mixture proportions inferred between Yoruba and Western European human populations.** As in Fig. 4, but genes were also classified by association with disease (Struck et al. 2018) and known physical interactions with viral proteins (Enard & Petrov 2018).

analysis. Although these demographic models fit the data much less well than our main model, the inferred values of p_1 for the GO terms were highly correlated with those from our main model (Fig. S4). Using our main demographic model, we also tested different models for the joint DFE. Our first test was to include a point mass of positive selection in the model, which has been supported by previous analyses of the DFE in *D. melanogaster* (Ragsdale et al. 2016). Including positive selection had very little effect on our inferences, because we inferred the proportion of positively selected mutations to be small (Fig. S5A, Table S7). Our second test was to use a bivariate lognormal model, rather than the mixture model. The inferred values for ρ in the bivariate model were highly correlated with the values for the mixture proportion p_1, although systematically smaller (Fig. S5B). Together, these results suggest that the robustness we observed in simulated data (Fig. 2C&F) holds true for real data.

For simplicity, our model of the joint DFE assumed that the mixture proportion is constant throughout the distribution, but the proportion may depend on how deleterious the mutation is. To test this assumption, rather than adding complexity to the DFE model, we instead segregated our data. To do so, we applied SIFT (Vaser et al. 2016), which uses evolutionary conservation to predict whether a nonsynonymous mutation is likely to be tolerated or deleterious. We then fit DFE models to the SNPs in each class. As expected, we inferred a larger mean fitness effect for the deleterious class than the tolerated class (Tables S4,S5,&S6). Moreover, we found that the mixture proportion p_1 was dramatically smaller for the deleterious class than the tolerated class in humans and *D. melanogaster* (Fig. 4&6), but not in wild tomatoes (Fig. 5).

Discussion

We developed an approach for inferring the joint distribution of fitnesses effects of new mutations between pairs of populations. We tested our approach with simulation studies, finding that it is statistically powerful and robust to many forms of model misspecification (Fig. 2). We then applied it to *D. melanogaster*, wild tomatoes, and humans (Fig. 3). Among these examples, we found the lowest correlation of fitness effects in tomatoes and the highest in humans. In *D. melanogaster* and tomatoes, we found variation in the correlation of the DFE among genes with different functions (Fig. 4&5). In *D. melanogaster* and humans, we found that the correlation of the DFE is lower for mutations that are more deleterious (Fig. 4&6). Lastly, in humans, we found that genes involved in interactions with specific viruses also exhibited lower correlation (Fig. 6). Together, these results illustrate the biological insights that can be gained by considering the joint DFE between populations.

Each of our analyses began by fitting a model of demographic history, although our selection inferences are robust to details of that model (Fig. 2C&S4). Nevertheless, our inferred demographic models agree well with other inferences. For *D. melanogaster*, our inferred population sizes and divergence time for African and European populations agree with those of Arguello et al. (2019), although we used different populations and different models. For humans, our demographic parameters were also similar to previous analyses (Gravel et al. 2011). Lastly, for tomatoes we reproduced the demographic inference results of Beddows et al. (2017).

A key assumption in our analysis is the parametric form for the joint DFE. Recently, Martin & Lenormand (2015) extended Fisher's Geometrical Model to consider the relationship between mutation fitness effects in two different environments, represented by two optima in trait space. Unfortunately, they could not derive an analytic joint DFE for their model. Thus we focused our analysis on simpler generic models for the joint DFE. All our models assume that the correlation or mixture proportion is constant throughout the distribution. But segmenting our data with SIFT suggests that in *D. melanogaster* and humans the correlation of fitness effects may be smaller for more deleterious variants (Fig. 4&6). Moreover, including positive selection was not necessary to explain the data we analyzed (Fig. S5A), but it may be necessary in other cases. Overall, our simple models of the joint DFE fit the data well, but more complex models may be more informative.

In our models for the joint DFE, we assumed that the marginal DFE was identical between the two populations. In other words, we assumed that the mean and the variance of mutation fitness effects did not differ between the two populations. Experiments suggest that stressful environments can alter the marginal DFE (Wang et al. 2014). And previous population genetic studies have found evidence for differences in

marginal DFEs between populations of humans (Boyko et al. 2008; Lopez et al. 2018) and other primates (Ma et al. 2013; Tataru & Bataillon 2019). But those studies found only slight differences, and our simulation study suggests that inferences of the correlation of the DFE are robust to relatively large differences in marginal DFEs (Fig. 2F).

In both humans and *D. melanogaster*, we found the lowest correlation of mutation fitness effects in genes related to immunity. In *D. melanogaster*, genes annotated as functioning in the immune system had the lowest correlation among all GO terms tested (Fig. 4). It is well established that immune system genes exhibit more rapid adaptive evolution than other functional categories in Drosophila (Schlenke & Begun 2003; Sackton et al. 2007), potentially due to host-parasite coevolution. A study on genetic variation in *D. melanogaster* revealed significant enrichment of highly differentiated genes between Malawi and North Carolina populations for several GO terms, including immune function (Langley et al. 2012). Given that North American *D. melanogaster* populations are predominantly of European origin, our results align well with these findings. In humans, we found that genes known to physically interact with only a single virus showed significantly lower correlation of fitness effects than genes known to interact with no or multiple viruses (Fig. 6). This may be because genes that are known to interact with only a single virus are more likely to be affected by differences in pathogen content between environments. Thus both our human and *D. melanogaster* results are consistent with differences in selection due to distinct pathogen environments in Africa and Europe.

Beyond immunity, our analyses in *D. melanogaster*, wild tomatoes, and humans all suggest that the joint DFE differs between genes of different functions (Fig. 4, 5, & 6). In some cases, these gene sets may naturally correspond to different environmental factors, such as "response to stress" in flies and "response to abiotic stimulus" in tomatoes. In other cases, the biological explanation is less clear. Nevertheless, our results suggest new directions for exploring the environmental responsiveness of different phenotypes in these species.

The fitness effect of a mutation may differ between populations due to differences in both environmental and genetic context. The wild tomato species we analyzed overlap in range and are more genetic differentiated than the *D. melanogaster* and human populations we studied. So in this case we speculate that differences in fitness effects are primarily driven by differences in genetic background. Out of the species we studied, humans exhibited the highest correlation of mutation fitness effects. It is unclear whether this is caused by our relatively low genetic differentiation or our ability to control our local environment. Over the long term, assessing the joint DFE between multiple populations of multiple species may reveal the relative importance of environmental and genetic context in determining the effects of mutations on fitness.

Materials and Methods

Calculating expected AFS under the joint DFE

To calculate the expected AFS given a joint DFE, we first cached the expected AFS for the given demographic model for pairs of population-scaled selection coefficients γ_1, γ_2 (Ragsdale et al. 2016). The cached values of γ_1, γ_2 were from 50 points logarithmically spaced within $[-10^{-4}, -2000]$, for a total of 2500 cached spectra (Fig. S1). The expected AFS for a given joint DFE is then an integral over these cached spectra, and we carry out this integration numerically using the trapezoid rule. In general, probability density for the joint DFE will extend outside the range of cached spectra. To account for this density, we integrated outward from the sampled domain to $\gamma = 0$ or $-\infty$ to estimate the excluded weight of the joint DFE. We then weighted the closest cached AFS by the result and added it to the expected AFS. For the edges of the domain, this was done using the SciPy method `quad`, and for the corners it was done using `dblquad` (Oliphant 2007). For the mixture model, the expected $\rho = 1$ component was calculated simply as a one-dimensional integral over a cache of $\gamma_1 = \gamma_2$ spectra. When a point mass of positive selection was included in the mixture model, the corresponding joint density for the $\rho = 0$ component was included and integrated over, corresponding to population-specific adaptation. Our implementation of this approach was based on `fitdadi` (Kim et al. 2017) and has been incorporated into `dadi` (Gutenkunst et al. 2009).

Simulated data tests

Unless otherwise specified, for Fig. 2, the "truth" simulations were performed with an isolation-with-migration (IM) demographic model with parameters as in Table S8, a joint DFE model with marginal $\mu = 3.6$ and $\sigma = 5.1$, and with sample sizes of 218 for population 1 and 198 for population 2. For Fig. 2A and 2B, data were simulated with $p_1 = 0.9$, by Poisson sampling from the expected joint AFS, with the population-scaled mutation rate θ held constant at 13,842.5. For Fig. 2A, the resulting average number of segregating polymorphisms varied with sample size, ranging from 7,373 for sample size 2 chromosomes to 44,897 for sample size 100. For Fig. 2B, the sample size was fixed at 20 chromosomes per population. For the remaining tests (Fig. 2C-F), we were interested in bias rather than variance, so misspecified models were fit directly to the expected frequency spectrum under the true model, without Poisson sampling noise. For Fig. 2C, the best fit model with no migration had $s = 0.937$, $\nu_1 = 3.025$, $\nu_2 = 3.219$, $T = 0.0639$, $m = 0$, and the best-fit model with instantaneous growth and symmetric migration had $\nu_1 = 2.4$, $\nu_2 = 0.92$, $T = 0.23$, $m = 0.42$. For Fig. 2E, the true joint DFE was a mixture model with marginal gamma distributions with $\alpha = 0.4$, $\beta = 1400$. For Fig. 2F, the true joint DFE was a bivariate lognormal specified correlation coefficient ρ, and for the asymmetric case, $\mu_1 = 3.6$, $\sigma_1 = 5.1$, $\mu_2 = 4.5$, $\sigma_2 = 6.8$.

Genomic data

We obtained Zambian and French *D. melanogaster* genomic data from the Drosophila Genome Nexus (Lack et al. 2016). The Zambian sequences were 197 haploids from the DPGP3 (Lack et al. 2015), and the French were from 96 inbred individuals (Lack et al. 2016). In these data, many SNPs were not called in all individuals. We thus projected downward to obtain a consensus AFS with the maximal number of segregating sites. For these data, that was to a sample size of 178 Zambian and 47 French chromosomes. Calling success was uneven across the genome, so our projection excludes many sites from Chr3L and Chr3R (Fig.S6). To annotate SNPs to their corresponding genes and as synonymous or nonsynonymous, we used ANNOVAR (Wang et al. 2010) with default settings and the `dm6` genome build.

We obtained YRI and CEU human genomic data from the 1000 Genomes Phase 3 (The 1000 Genomes Project Consortium 2015), considering only individuals with phase 3 genotypes. We only considered SNPs that were within the exome targeted regions, had as their most severe functional annotation `synonymous_variant` or `missense_variant`, and had a reported ancestral allele. We projected the data down to 216 YRI and 198 CEU chromosome samples for analysis.

We obtained *S. chilense* and *S. peruvianum* RNA sequencing data from Beddows et al. (2017) and followed their scheme for assigning individuals to species. To more easily apply SIFT, we used the NCBI genome remapping service to convert the data from SL2.50 coordinates to SL2.40.

Fitting demographic history models

We used `dadi` to fit models for demographic history to spectra for synonymous mutations (Gutenkunst et al. 2009), including a parameter for ancestral state misidentification (Ragsdale et al. 2016). For the *Drosophila* analysis, we used `dadi` grid points of [300,400,500], and we found that an isolation-with-migration (IM) model fit the data well (Table S1). We also analyzed simpler models assuming instantaneous growth or no migration. For the human analysis, we used `dadi` grid points of [800,1000,1200], and we found that an IM model with growth in the ancestral population fit the data well (Table S3). For the tomato analysis, we used `dadi` grid points of [300,400,500], and fit the same isolation-with-migration model that Beddows et al. (2017) fit to the data, obtaining very similar parameter estimates (our Table S2 vs. their Table S4).

For comparison with other studies, we converted model parameters for *D. melanogaster* and humans from genetic units to physical units. For *Drosophila*, we assumed an overall mutation rate of 2.8×10^{-9} per generation (Keightley et al. 2014), that a fraction $1/(1+2.85)$ of exonic mutations were synonymous (Huber et al. 2017), and 10 generations per year (Arguello et al. 2019). For *Drosophila*, there were 10,643,343 bases at which at which at least 47 individuals were called in France and 178 in Zambia. For humans, we assumed an overall mutation rate of 1.5×10^{-8} (Ségurel et al. 2014), that a fraction $1/(1+2.31)$ of exonic mutations

were synonymous (Huber et al. 2017), and a generation time of 25 years. In both analyses, we estimated the ancestral population size using $\theta_{syn} = 4N_A\mu_{syn}L$.

Fitting joint DFEs

Cached allele frequency spectra were created for the corresponding demographic models, using the same grid points settings. Models of the joint DFE were then fit to nonsynonymous data by maximizing the likelihood of the data, assuming a Poisson Random Field (Sawyer & Hartl 1992). In these fits, the population-scaled mutation rate for nonsynonymous mutations θ_{NS} was held at fixed at a given ratio to the rate for synonymous mutations θ_S in the same subset of genes, as inferred from our demographic history model. For *D. melanogaster* this ratio was 2.85 and for humans it was 2.31 (Huber et al. 2017). For tomato, this ratio was assumed to be 2.5. For the lognormal mixture model, the three parameters of interest are the weight of the $\rho = 1$ component, p_1, and the mean and variance of the marginal distributions, μ and σ. In addition, we included a separate parameter for ancestral state misidentification for each subset of the data tested, because the rate of misidentification depends on the strength of selection acting on the sites of interest. To include positive selection in our *D. melanogaster* analysis, we first estimated an appropriate value for the strength of positive selection of $\gamma_+ = 12.035$ by fitting a model with perfectly correlated selection ($\rho = 1$). We then created a spectra cache that included that γ_+ value (Fig. S1).

We separately analyzed SNPs from genes associated with different Gene Ontology terms. We used terms from the Generic GO subset, which is a set of curated terms that are applicable to a range of species (The Gene Ontology Consortium 2000). We considered the direct children of GO:0008510 "Biological Process", and any gene annotated with a child of a given term was assumed to also be annotated by the parent term. Thus in our analysis a given gene may be present in multiple GO term subsets. We used Ensembl Biomart (Cunningham et al. 2019) to retrieve the annotated GO terms for each gene.

We also separately analyzed SNPs classified by SIFT as deleterious or tolerated (Vaser et al. 2016). To do so, we considered only sites assigned "regular" confidence and ignored "low" confidence sites. To carry out our DFE analysis, we needed to estimate an appropriate population-scaled mutation rate θ_{NS} for deleterious and tolerated mutations. To do so, we randomly inserted mutations into the reference genome of each species, then ran SIFT and scaled the whole genome θ_{NS} by the proportion of simulated mutations that SIFT classified as deleterious or tolerated. For *D. melanogaster*, we used the nucleotide mutation rates of Singh et al. (2007). For humans, we used the trinucleotide mutations rates of Jónsson et al. (2017). For tomato, we used the *Arabidopsis thaliana* nucleotide mutation rates of Ossowski et al. (2010), because mutation rates have not been directly measured in tomato.

For humans, we also divided genes into classes based on their role in disease and interactions with viruses. Following Struck et al. (2018), we classified genes as associated with Mendelian disease, complex disease, or no disease using Online Mendelian Inheritance in Man (OMIM, Amberger et al. 2015) and the European Bioinformatics Institute's genome-wide association studies (GWAS) catalog (MacArthur et al. 2016). We used the data of Enard & Petrov (2018) to annotate 4,534 genes as encoding virus-interacting proteins (VIPs). We defined the set of non-VIP genes as the 17,603 Ensembl genes that were not annotated as encoding VIPs. We identified 1,728 genes as known to interact with 2 or more viruses, leaving 2,806 genes known to interact with only a single virus.

To estimate the uncertainty of our inferences, we used an approach based on the Godambe Information Matrix (Coffman et al. 2016). To generate the requisite bootstrap data sets, we divided the reference genomes into 1 Mb chunks. Because gene content varied among bootstraps, θ_{NS} also needed to vary. To estimate the appropriate θ_{NS} for each bootstrap, we scaled corresponding θ_{NS} from the real data by the ratio of the number of segregating sites in the AFS of the bootstrap versus real data. We found good agreement between the uncertainties estimated by the Godambe approach and those from directly fitting the bootstrap data sets (Fig. S7). Note that this process does not propagate uncertainty in the demographic parameter inference, so our uncertainties are somewhat underestimated.

Acknowledgments

We thank Laura Rose and Thorsten Kloesges for sharing their tomato data and assistance with the analysis. We also thank David Enard for sharing his virus-interacting-proteins data. Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R01GM127348.

References

Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A (2015) OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an Online catalog of human genes and genetic disorders. Nucleic Acids Res 43:D789.

Arguello JR, Laurent S, Clark AG (2019) Demographic history of the human commensal Drosophila melanogaster. Genome Biol Evol 11:844.

Balick DJ, Do R, Cassa CA, Reich D, Sunyaev SR (2015) Dominance of Deleterious Alleles Controls the Response to a Population Bottleneck. PLoS Genet 11:1.

Bank C, Hietpas RT, Wong A, Bolon DN, Jensen JD (2014) A Bayesian MCMC approach to assess the complete distribution of fitness effects of new mutations: Uncovering the potential for adaptive walks in challenging environments. Genetics 196:841.

Barton NH (1995) A general model for the evolution of recombination. Genet Res (Camb) 65:123.

Beddows I, Reddy A, Kloesges T, Rose LE (2017) Population genomics in wild tomatoes-the interplay of divergence and admixture. Genome Biol Evol 9:3023.

Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, Lohmueller KE, Adams MD, Schmidt S, Sninsky JJ, Sunyaev SR, White TJ, Nielsen R, Clark AG, Bustamante CD (2008) Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet 4:e1000083.

Charlesworth B (1994) The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet Res 63:213.

Coffman AJ, Hsieh P, Gravel S, Gutenkunst RN (2016) Computationally efficient composite likelihood statistics for demographic inference. Mol Biol Evol 33:591.

Cunningham F, Achuthan P, Akanni W, Allen J, Amode MR, Armean IM, Bennett R, Bhai J, Billis K, Boddu S, Cummins C, Davidson C, Dodiya KJ, Gil L, Grego T, Haggerty L, Gall A, Garc C, Haskell E, Hourlier T, Izuogu OG, Janacek SH, Juettemann T, Kay M, Laird MR, Lavidas I, Liu Z, Marug C, Loveland JE, Maurel T, Mcmahon AC, Moore B, Morales J, Mudge JM, Nuhn M, Ogeh D, Parker A, Parton A, Patricio M, Abdul AI, Schmitt BM, Schuilenburg H, Sheppard D, Sparrow H, Stapleton E, Szuba M, Taylor K, Threadgold G, Thormann A, Vullo A, Walts B, Winterbottom A, Zadissa A, Chakiachvili M, Frankish A, Hunt SE, Kostadima M, Langridge N, Martin FJ, Muffato M, Perry E, Ruffier M, Staines DM, Trevanion SJ, Aken BL, Yates AD, Zerbino DR, Flicek P (2019) Ensembl 2019. Nucleic Acids Res 47:D745.

Elena SF, Ekunwe L, Hajela N, Oden SA, Lenski RE (1998) Distribution of fitness effects caused by random insertion mutations in Escherichia coli. Genetica 102-103:349.

Enard D, Petrov DA (2018) Evidence that RNA Viruses Drove Adaptive Introgression between Neanderthals and Modern Humans. Cell 175:360.

Eyre-Walker A (2010) Evolution in health and medicine Sackler colloquium: Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. Proc Natl Acad Sci U S A 107 Suppl:1752.

Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. Nat Rev Genet 8:61061.

Eyre-Walker A, Woolfit M, Phelps T (2006) The distribution of fitness effects of new deleterious amino acid mutations in humans. Genetics 173:891.

Fisher RA (1930) *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

Fry JD, Keightley PD, Heinsohn SL, Nuzhdin SV (2002) New estimates of the rates and effects of mildly deleterious mutation in Drosophila melanogaster. Proc Natl Acad Sci U S A 96:574.

Gandon S, Hochberg M, Holt RD, Martin G, Day T (2013) What limits the evolutionary emergence of pathogens? Philos Trans R Soc B 368:20120086.

Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, Yu F, Gibbs RA, The 1000 Genomes Project, Bustamante CD (2011) Demographic history and rare allele sharing among human populations. Proc Natl Acad Sci USA 108:11983.

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:e1000695.

Huber CD, Durvasula A, Hancock AM, Lohmueller KE (2018) Gene expression drives the evolution of dominance. Nat Commun 9:2750.

Huber CD, Kim BY, Marsden CD, Lohmueller KE (2017) Determining the factors driving selective effects of new nonsynonymous mutations. Proc Natl Acad Sci USA 114:4465.

Jónsson H, Sulem P, Kehr B, Kristmundsdottir S, Zink F, Hjartarson E, Hardarson MT, Hjorleifsson KE, Eggertsson HP, Gudjonsson SA, Ward LD, Arnadottir GA, Helgason EA, Helgason H, Gylfason A, Jonasdottir A, Jonasdottir A, Rafnar T, Frigge M, Stacey SN, Th Magnusson O, Thorsteinsdottir U, Masson G, Kong A, Halldorsson BV, Helgason A, Gudbjartsson DF, Stefansson K (2017) Parental influence on human germline de novo mutations in 1,548 trios from Iceland. Nature 549:519.

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

Keightley PD, Eyre-Walker A (2010) What can we learn about the distribution of fitness effects of new mutations from DNA sequence data? Philos Trans R Soc B Biol Sci 365:1187.

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 196:313.

Kim BY, Huber CD, Lohmueller KE (2017) Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. Genetics 206:345.

Kondrashov AS, Houle D (1994) Genotype-environment interactions and the estimation of the genomic mutation rate in Drosohpila melanogaster. Proc R Soc B 258:221.

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

Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA, Langley CH, Pool JE (2015) The Drosophila Genome Nexus: A population genomic resource of 623 Drosophila melanogaster genomes, including 197 from a single ancestral range population. Genetics 199:1229.

Lack JB, Lange JD, Tang AB, Corbett-Detig RB, Pool JE (2016) A Thousand Fly Genomes: An Expanded Drosophila Genome Nexus. Mol Biol Evol 33:3308.

Langley CH, Stevens K, Cardeno C, Lee YCG, Schrider DR, Pool JE, Langley SA, Suarez C, Corbett-Detig RB, Kolaczkowski B, Fang S, Nista PM, Holloway AK, Kern AD, Dewey CN, Song YS, Hahn MW, Begun DJ (2012) Genomic variation in natural populations of Drosophila melanogaster. Genetics 192:533.

Latta LC, Peacock M, Civitello DJ, Dudycha JL, Meik JM, Schaack S (2015) The Phenotypic Effects of Spontaneous Mutations in Different Environments. Am Nat 185:243.

Lopez M, Kousathanas A, Quach H, Harmant C, Mouguiama-Daouda P, Hombert JM, Froment A, Perry GH, Barreiro LB, Verdu P, Patin E, Quintana-Murci L (2018) The demographic history and mutational load of African hunter-gatherers and farmers. Nat Ecol Evol 2:721.

Ma X, Kelley JL, Eilertson K, Musharoff S, Degenhardt JD, Martins AL, Vinar T, Kosiol C, Siepel A, Gutenkunst RN, Bustamante CD (2013) Population genomic analysis reveals a rich speciation and demographic history of orang-utans (Pongo pygmaeus and Pongo abelii). PLoS One 8:e77175.

MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, Pendlington ZM, Welter D, Burdett T, Hindorff L, Flicek P, Cunningham F, Parkinson H (2016) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res 45:D896.

Martin G, Lenormand T (2006) A General Multivariate Extension of Fisher's Geometrical Model and the Distribution of Mutation Fitness Effects Across Species. Evolution 60:893.

Martin G, Lenormand T (2015) The fitness effect of mutations across environments: Fisher's geometrical model with multiple optima. Evolution 69:1433.

Messer PW, Petrov DA (2013) Population genomics of rapid adaptation by soft selective sweeps. Trends Ecol Evol 28:659.

Nadarajah S, Gupta AK (2006) Some bivariate gamma distributions. Appl Math Lett 19:767.

Oliphant TE (2007) Python for Scientific Computing Python. Comput Sci Eng May/June:10.

Ossowski S, Schneeberger K, Lucas-Lledo JI, Warthmann N, Clark RM, Shaw RG, Weigel D, Lynch M (2010) The Rate and Molecular Spectrum of Spontaneous Mutations in Arabidopsis thaliana. Science 327:92.

Ragsdale AP, Coffman AJ, Hsieh P, Struck TJ, Gutenkunst RN (2016) Triallelic population genomics for inferring correlated fitness effects of same site nonsynonymous mutations. Genetics 203:513.

Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG (2007) Dynamic evolution of the innate immune system in Drosophila. Nat Genet 39:1461.

Sanjua R, Myoa A, Elena SF (2004) The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proc Natl Acad Sci USA 101:8396.

Sawyer SA, Hartl DL (1992) Population genetics of polymorphism and divergence. Genetics 132:1161.

Schlenke TA, Begun DJ (2003) Natural selection drives Drosophila immune system evolution. Genetics 164:1471.

Ségurel L, Wyman MJ, Przeworski M (2014) Determinants of Mutation Rate Variation in the Human Germline. Annu Rev Genomics Hum Genet 15:47.

Singh ND, Dumont VLB, Hubisz MJ, Nielsen R, Aquadro CF (2007) Patterns of Mutation and Selection at Synonymous Sites in Drosophila. Mol Biol Evol 24:2687.

Struck TJ, Mannakee BK, Gutenkunst RN (2018) The impact of genome-wide association studies on biomedical research publications. Hum Genomics 12:38.

Tataru P, Bataillon T (2019) polyDFEv2.0: testing for invariance of the distribution of fitness effects within and across species. Bioinformatics .

Tenaillon O (2014) The Utility of Fisher's Geometric Model in Evolutionary Genetics. Annu Rev Ecol Evol Syst 45:179.

The 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. Nature 526:68.

The Gene Ontology Consortium (2000) Gene Ontology: tool for the unification of biology. Nat Genet 25:25.

Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC (2016) SIFT missense predictions for genomes. Nat Protoc 11:1.

Wang AD, Sharp NP, Agrawal AF (2014) Sensitivity of the distribution of mutational fitness effects to environment, genetic background, and adaptedness: A case study with Drosophila. Evolution 68:840.

Wang K, Li M, Hakonarson H (2010) ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38:e164.

Supporting Information

Table S1: Demographic model parameters for Zambian and French *D. melanogaster* populations.

model	θ_S	s	ν_{Zm}	ν_{Fr}	T	$m_{Zm\leftarrow Fr}$	$m_{Fr\leftarrow Zm}$	p_{misid}
IM	55137	0.959	3.490	0.209	0.083	0.799	3.560	0.0498
instant growth, symmetric mig.	51596		2.267	0.164	0.125	1.725		0.0515
no migration	60096	0.975	5.601	0.399	0.032			0.0482

model	N_A	N_{Zm}	N_{Fr}	T
IM	1.8×10^6	6.2×10^6	3.7×10^5	30×10^3 years

Table S2: Demographic model parameters for wild tomato populations.

model	θ_S	ν_{chi}	ν_{per}	T	$m_{chi\leftarrow per}$	$m_{per\leftarrow chi}$	p_{misid}
instant growth, asymmetric mig.	38943	1.317	3.708	1.447	0.2669	0.1672	0.039

Table S3: Demographic model parameters for African YRI and European CEU human populations.

model	θ_S	ν_{pre}	T_{pre}	s	ν_{YRI}	ν_{CEU}	T	$m_{1,2}$	$m_{2,1}$	p_{misid}
IMpre	5537	1.85	0.52	0.96	3.34	4.28	0.099	0.29	0.26	0.015

model	N_A	N_{pre}	T_{pre}	N_{YRI}	N_{CEU}	T
IMpre	8.0×10^3	15×10^3	2.1×10^5 years	27×10^3	34×10^3	40×10^3 years

Table S4: Inferred joint DFE parameters for *D. melanogaster*.

group	SNP count	μ	σ	p_1	p_{misid}	GO definition
proteome	399146	6.258 ± 0.100	3.991 ± 0.116	0.968 ± 0.022	0.027 ± 0.002	
GO:0002376	16863	6.755 ± 0.203	3.822 ± 0.217	0.904 ± 0.083	0.023 ± 0.004	immune system process
GO:0009056	26927	6.829 ± 0.209	3.990 ± 0.227	0.934 ± 0.061	0.028 ± 0.006	catabolic process
GO:0006950	37690	6.937 ± 0.176	4.035 ± 0.187	0.946 ± 0.041	0.023 ± 0.003	response to stress
GO:0022607	37837	6.962 ± 0.157	3.799 ± 0.159	0.946 ± 0.044	0.021 ± 0.003	cellular component assembly
GO:0051276	20478	6.527 ± 0.224	3.979 ± 0.259	0.951 ± 0.048	0.032 ± 0.006	chromosome organization
GO:0040011	29170	6.994 ± 0.176	3.856 ± 0.182	0.952 ± 0.049	0.016 ± 0.003	locomotion
GO:0048856	98813	6.760 ± 0.109	3.776 ± 0.115	0.956 ± 0.031	0.020 ± 0.003	anatomical structure development
GO:0008233	18481	5.761 ± 0.245	3.983 ± 0.346	0.957 ± 0.045	0.038 ± 0.006	peptidase activity
GO:0007165	52642	6.925 ± 0.142	3.878 ± 0.147	0.957 ± 0.044	0.019 ± 0.002	signal transduction
GO:0030154	71178	6.766 ± 0.147	3.745 ± 0.155	0.959 ± 0.032	0.022 ± 0.003	cell differentiation
GO:0034641	77809	6.506 ± 0.118	3.859 ± 0.135	0.961 ± 0.037	0.025 ± 0.003	cellular nitrogen compound metabolic process
GO:0003700	15408	6.181 ± 0.180	3.483 ± 0.211	0.967 ± 0.072	0.028 ± 0.011	DNA binding transcription factor activity
GO:0007049	27114	6.520 ± 0.298	3.785 ± 0.333	0.968 ± 0.047	0.025 ± 0.004	cell cycle
GO:0006464	39082	6.972 ± 0.200	3.974 ± 0.210	0.974 ± 0.036	0.024 ± 0.003	cellular protein modification process
GO:0007010	29806	6.518 ± 0.156	3.703 ± 0.171	0.976 ± 0.048	0.020 ± 0.003	cytoskeleton organization
GO:0050877	21226	6.735 ± 0.267	4.093 ± 0.299	0.976 ± 0.042	0.018 ± 0.004	nervous system process
GO:0000003	51700	6.653 ± 0.199	4.010 ± 0.225	0.981 ± 0.027	0.030 ± 0.005	reproduction
GO:0009058	64870	6.720 ± 0.128	3.900 ± 0.140	0.981 ± 0.036	0.025 ± 0.004	biosynthetic process
GO:0007267	22906	7.032 ± 0.243	3.786 ± 0.241	0.982 ± 0.038	0.022 ± 0.008	cell-cell signaling
GO:0044281	25023	7.268 ± 0.238	4.108 ± 0.239	0.985 ± 0.044	0.018 ± 0.003	small molecule metabolic process
GO:0040007	15579	7.218 ± 0.174	3.711 ± 0.164	0.987 ± 0.041	0.016 ± 0.004	growth
GO:0006810	59789	6.920 ± 0.151	3.873 ± 0.155	0.990 ± 0.022	0.022 ± 0.003	transport
GO:0016491	16379	6.785 ± 0.320	4.152 ± 0.360	1.000 ± 0.049	0.022 ± 0.004	oxidoreductase activity
SIFT tolerated	278620	4.900 ± 0.079	3.943 ± 0.146	0.978 ± 0.020	0.034 ± 0.002	
SIFT deleterious	61960	7.312 ± 0.092	3.052 ± 0.074	0.766 ± 0.044	0.006 ± 0.001	

Table S5: Inferred joint DFE parameters for humans.

group	SNP count	μ	σ	p_1	p_{misid}
proteome	129299	2.041 ± 0.021	4.248 ± 0.210	0.990 ± 0.010	0.009 ± 0.001
SIFT tolerated	76075	-0.887 ± 0.989	4.562 ± 2.233	0.996 ± 0.013	0.013 ± 0.003
SIFT deleterious	37019	3.523 ± 0.027	2.620 ± 0.108	0.922 ± 0.047	0.001 ± 0.002
non-disease	80121	1.735 ± 0.034	4.159 ± 0.207	0.992 ± 0.015	0.009 ± 0.001
complex disease	32401	2.556 ± 0.015	4.502 ± 0.355	1.000 ± 0.042	0.008 ± 0.002
Mendelian disease	16109	2.556 ± 0.014	4.004 ± 0.312	0.961 ± 0.042	0.007 ± 0.002
non-VIP	98652	1.722 ± 0.038	4.515 ± 0.249	0.993 ± 0.010	0.009 ± 0.001
multiple VIP	9333	3.321 ± 0.055	3.271 ± 0.282	1.000 ± 0.148	0.008 ± 0.002
single VIP	17254	2.639 ± 0.012	3.423 ± 0.232	0.946 ± 0.047	0.006 ± 0.002

Table S6: Inferred joint DFE parameters for wild tomatoes.

group	SNP count	μ	σ	p_1	p_{misid}	GO definition
proteome	1151584	3.710 ± 0.393	5.346 ± 0.288	0.906 ± 0.019	0.026 ± 0.001	
GO:0009628	15454	3.373 ± 4.301	4.315 ± 4.672	0.858 ± 0.093	0.027 ± 0.018	response to abiotic stimulus
GO:0016043	54683	3.680 ± 1.160	4.861 ± 1.306	0.887 ± 0.090	0.025 ± 0.014	translation
GO:0016301	14428	5.970 ± 0.645	6.976 ± 0.948	0.911 ± 0.129	0.034 ± 0.009	kinase activity
GO:0009719	13108	4.108 ± 1.760	4.926 ± 1.672	0.916 ± 0.100	0.025 ± 0.021	response to endogenous stimulus
GO:0009987	272501	3.788 ± 1.145	5.002 ± 2.547	0.935 ± 0.059	0.026 ± 0.016	cellular process
GO:0003700	11078	2.809 ± 1.747	4.029 ± 3.766	0.999 ± 0.125	0.027 ± 0.028	DNA binding transcription factor activity
GO:0008152	21097	3.919 ± 6.324	5.349 ± 7.087	1.000 ± 0.104	0.024 ± 0.033	metabolic process
GO:0007165	12783	4.148 ± 2.889	6.553 ± 3.775	1.000 ± 0.149	0.042 ± 0.014	signal transduction
GO:0005975	10282	4.612 ± 0.800	5.398 ± 2.912	1.000 ± 0.181	0.024 ± 0.010	carbohydrate metabolic process
SIFT tolerated	906379	1.966 ± 0.398	5.853 ± 0.444	0.884 ± 0.021	0.030 ± 0.001	
SIFT deleterious	199262	4.556 ± 0.347	2.946 ± 0.152	0.886 ± 0.030	0.010 ± 0.001	

Table S7: Alternative proteome-wide joint DFE models for *D. melanogaster*.

Model	μ	σ	p_1	ρ	p_{pos}	p_{misid}
mixture	6.26	3.99	0.97			0.027
mixture w/ pos. selection	6.26	4.01		0.97	0.0047	0.027
bivariate	6.25	3.95		0.91		0.027

Table S8: Demographic model parameters for simulation studies.

Model	s	ν_1	ν_2	T	$m_{1\leftarrow 2}$	$m_{2\leftarrow 1}$
IM	0.93	2.9	2.83	0.084	0.47	0.35

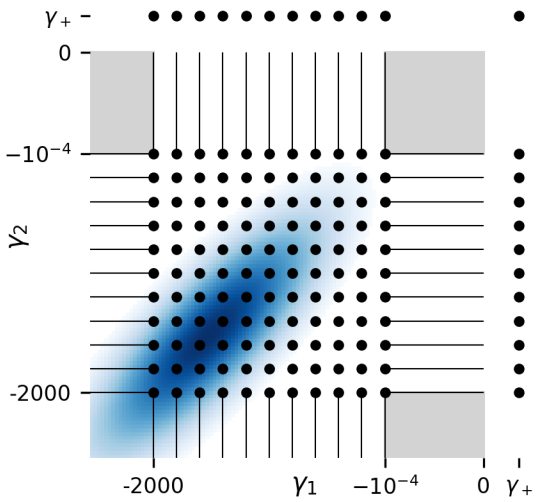

Figure S1: **Illustration of computational approach for calculating expected AFS for a given joint DFE.** Dots represent cached frequency spectra. Horizontal and vertical lines indicate single-variable semi-analytic integrations to estimate DFE density outside the sampled domain, and gray regions indicate corresponding double-variable integrations.

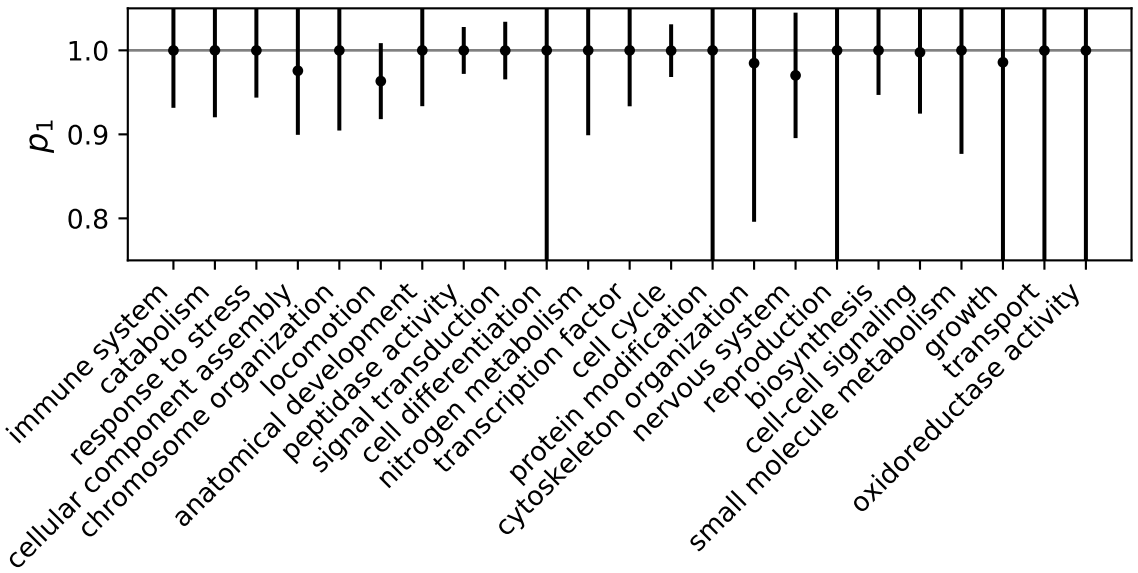

Figure S2: **Joint DFE mixture proportions inferred for human gene functional groups** As in Fig. 6, but for Gene Ontology terms.

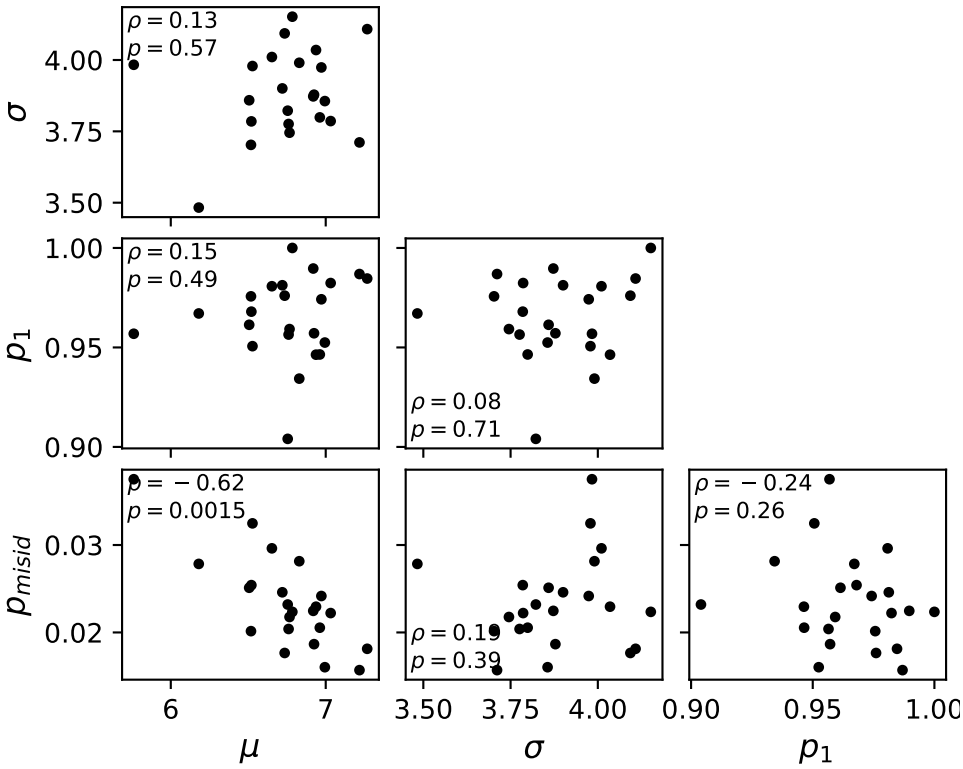

Figure S3: **Parameter correlations among fits to *D. melanogaster* GO slim groups.** The only significant correlation is between μ and p_{misid}, which is driven by the likelihood of secondary mutation on outgroup branch.

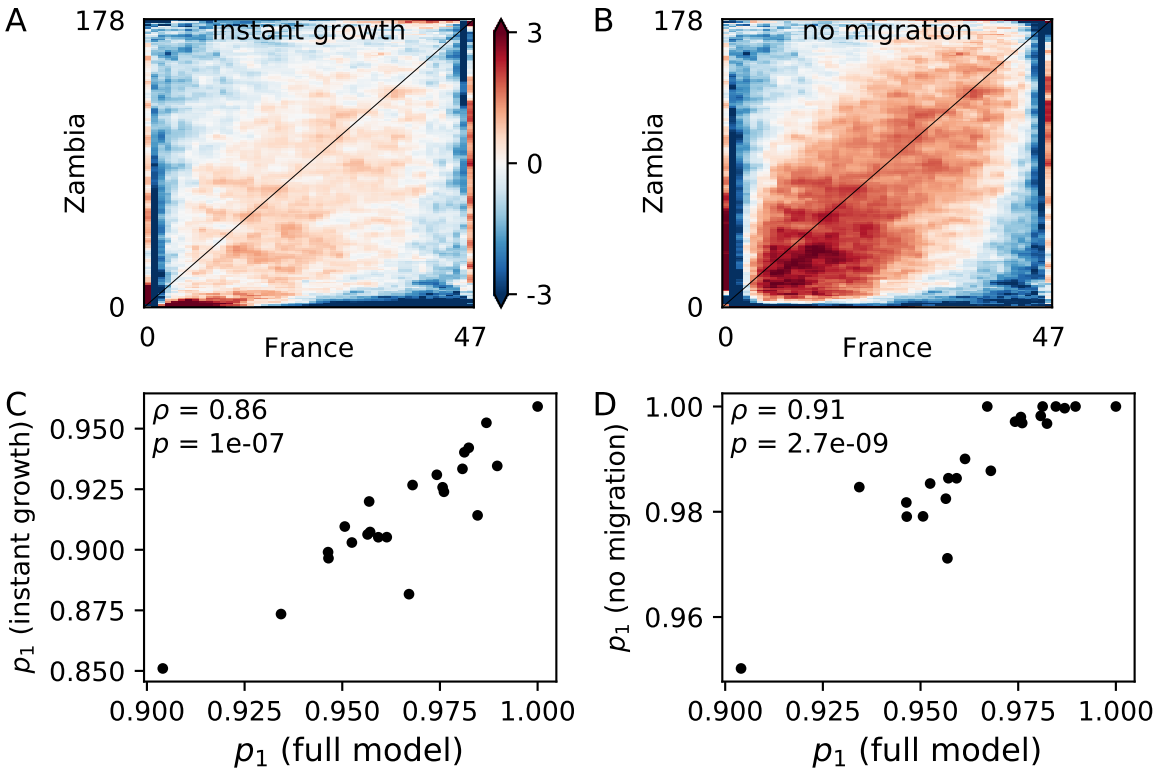

Figure S4: ***D. melanogaster* results assuming simpler demographic models.** A) A model with instantaneous growth fits the data less well than our full model with exponential growth. (Compare residuals with panel in Fig. 3B.) B) A model with no migration fits even worse. C&D) Yet the inferred values of the mixture proportion for GO terms are highly correlated between our full model and these simpler models.

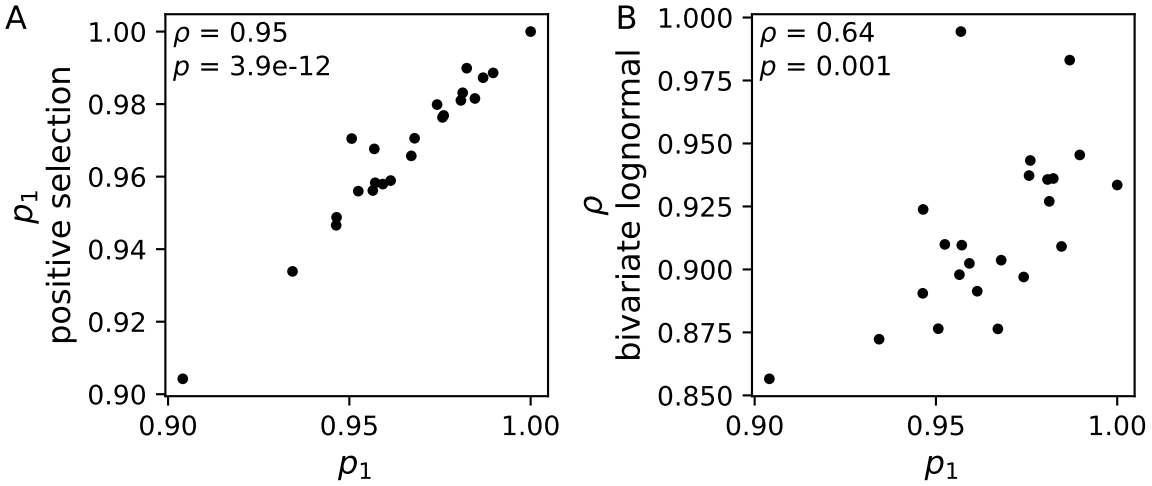

Figure S5: ***D. melanogaster* results assuming different DFE models.** For the *D. melanogaster* Gene Ontology terms, inferences from the lognormal mixture model are compared with A) a model including a point mass of positive selection and B) a bivariate lognormal model.

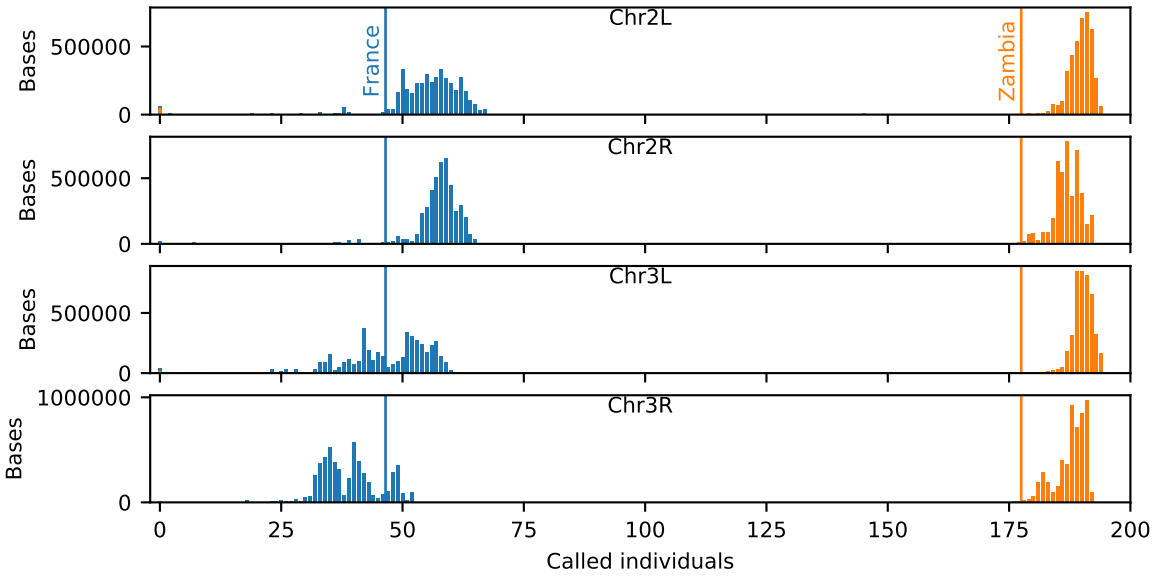

Figure S6: **Calling rate for different chromosome arms in analyzed *D. melanogaster* data.** For each arm, histograms indicate the number of bases at which a given number of individuals were called in each population. The vertical lines indicate the projection sizes for the AFS analysis.

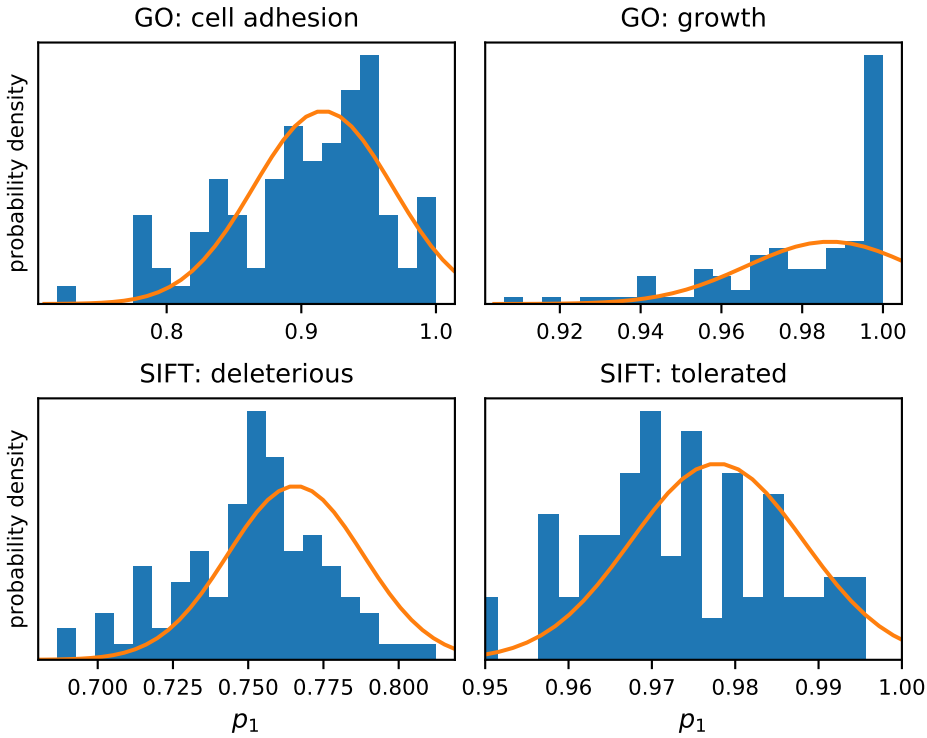

Figure S7: **Comparison between uncertainties estimated with the Godambe Information Matrix and bootstrap fitting.** For the *D. melanogaster* data, each panel shows a different subset of genes/mutations. In each panel, the histogram shows results from conventional bootstrap fitting, while the smooth curve is a normal distribution centered at the maximum likelihood inferred value and standard deviation estimated using the Godambe approach.