bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Inferring genome-wide correlations of mutation fitness effects between populations

Xin Huang[1], Alyssa Lyn Fortier[2], Alec J. Coffman[3], Travis J. Struck[1], Megan N. Irby[1], Jennifer E. James[1], José E. León-Burguete[4], Aaron P. Ragsdale[5], Ryan N. Gutenkunst[1,*]

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

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

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

Abstract

The effect of a mutation on fitness may differ between populations depending on environmental and genetic context, but little is known about the factors that underlie such differences. To quantify genome-wide correlations in mutation fitness effects, we developed a novel concept called a joint distribution of fitness effects (DFE) between populations. We then proposed a new statistic w to measure the DFE correlation between populations. Using simulation, we showed that inferring the DFE correlation from the joint allele frequency spectrum is statistically precise and robust. Using population genomic data, we inferred DFE correlations of populations in humans, *Drosophila melanogaster*, and wild tomatoes. In these specices, we found that the overall correlation of the joint DFE was inversely related to genetic differentiation. In humans and *D. melanogaster*, deleterious mutations had a lower DFE correlation than tolerated mutations, indicating a complex joint DFE. Altogether, the DFE correlation can be reliably inferred, and it offers extensive insight into the genetics of population divergence.

Introduction

New mutations that alter fitness are the key input into the evolutionary process. Typically, the majority of new mutations are deleterious or nearly neutral, and only a small minority are adaptive. These three categories constitute a continuum of fitness effects—the distribution of fitness effects (DFE) of new mutations (Eyre-Walker and Keightley 2007). The DFE is central to many theoretical evolutionary topics, such as the maintenance of genetic variation (Charlesworth 1994) and the evolution of recombination (Barton 1995), in addition to being key to applied evolutionary topics, such as the emergence of pathogens (Gandon et al. 2013) and the genetic architecture of complex disease (Durvasula and Lohmueller 2019).

The DFE can be quantified by either experimental approaches or statistical inference. Experimental approaches measure the DFE using random mutagenesis (Elena et al. 1998) or mutation accumulation (Fry et al. 2002); however, these approaches are limited to studying a small number of mutations. Most of our knowledge regarding the DFE has come from statistical inferences based on contemporary patterns of natural genetic variation. In these inferences, genetic data are typically summarized by the allele frequency spectrum (AFS). In some methods, a demographic model is inferred from the AFS of putatively neutral variants, and the DFE is estimated from the AFS of variants under selection, conditional on the best fit demographic model (Eyre-Walker et al. 2006; Keightley and Eyre-Walker 2007; Boyko et al. 2008; Kim et al. 2017). In other methods, the background pattern of variation is accounted for by the inclusion of nuisance parameters when fitting a DFE model to the AFS of variants under selection (Eyre-Walker et al. 2006; Tataru et al. 2017; Barton and Zeng 2018). In an alternative approach, a recent study applied approximate Bayesian computation to simultaneously infer the DFE and a demographic model (Johri et al. 2020). Moreover, a linear regression method can be used to infer the DFE from nucleotide diversity (James et al. 2017). These approaches has been applied to numerous organisms, including plants (Chen et al. 2017; Huber et al. 2018;

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Chen et al. 2020), *Drosophila melanogaster* (Keightley and Eyre-Walker 2007; Huber et al. 2017; Castellano et al. 2017; Barton and Zeng 2018; Johri et al. 2020), and primates (Boyko et al. 2008; Huber et al. 2017; Kim et al. 2017; Ma et al. 2013; Castellano et al. 2019).

Using these inference methods, several studies have found evidence for differences in DFEs among different populations (Boyko et al. 2008; Ma et al. 2013; Kim et al. 2017; Castellano et al. 2019; Tataru and Bataillon 2019). 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 this correlation is and what factors affect it. Moreover, these studies only compared DFEs from the AFS of single populations and therefore cannot investigate differences in fitness effects in new environments after population divergence.

Here, we developed a novel concept called the joint DFE of new mutations, which can be inferred from the joint AFS of pairs of populations. We then defined the correlation of mutation fitness effects between populations using the joint DFE. With simulation, we showed that inferring the joint DFE and correlation requires only modest sample sizes and is robust to many forms of model misspecification. We then applied our approach to data from humans, *D. melanogaster*, and wild tomatoes. We found that the correlation of mutation fitness effects between populations is lowest in wild tomatoes and highest in humans. In *D. melanogaster* and wild tomatoes, we found differences in the correlation among genes with different functions. We also found that mutations with more deleterious effects exhibit lower correlations. Together, our results show that the joint DFE and correlation of mutation fitness effects offer new insight into the population genetics of these species.

Results

Definition

To define the joint DFE, we considered two populations that have recently diverged, one of which may have entered a new environment (Fig. 1A). We also considered that a mutation has selection coefficient s_1 in the ancestral population and s_2 in the recently diverged population. 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 joint AFS (Fig. 1B). We refer to the joint probability distribution for (s_1, s_2) as the joint DFE (Fig. 1C), and we refer to the marginal probability distributions for s_1 or s_2 as the marginal DFEs for population 1 or population 2, respectively. The observed AFS from a pair of populations results from integrating spectra for different values of s_1 and s_2 over the joint DFE.

Little is known about the shape of the joint DFE, so we considered multiple parametric models. The best fit DFEs for single populations tend to be lognormal or gamma distributions (Boyko et al. 2008), although discrete distributions may sometimes fit better (Kousathanas and Keightley 2013; Johri et al. 2020) We first considered a bivariate lognormal distribution (Fig. 1C), because it has an easily interpretable correlation coefficient. However, the bivariate lognormal distribution can be numerically poorly behaved as the correlation coefficient approaches one and the distribution becomes very thin. We also considered another popular probability distribution for modeling DFEs, the gamma distribution, but there are multiple ways of defining a bivariate gamma distribution (Nadarajah and Gupta 2006). We thus focused on a mixture model that consisted of a component corresponding to perfect correlation with weight w, and a component corresponding to zero correlation with weight $(1 - w)$ (Fig. 1D). To limit the complexity of the model, we assumed that the marginal DFEs were identical for both populations. In this case, the correlation of the overall distribution is equal to the mixture proportion w. We thus interpret and discuss w as a DFE correlation coefficient.

The DFE correlation 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

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

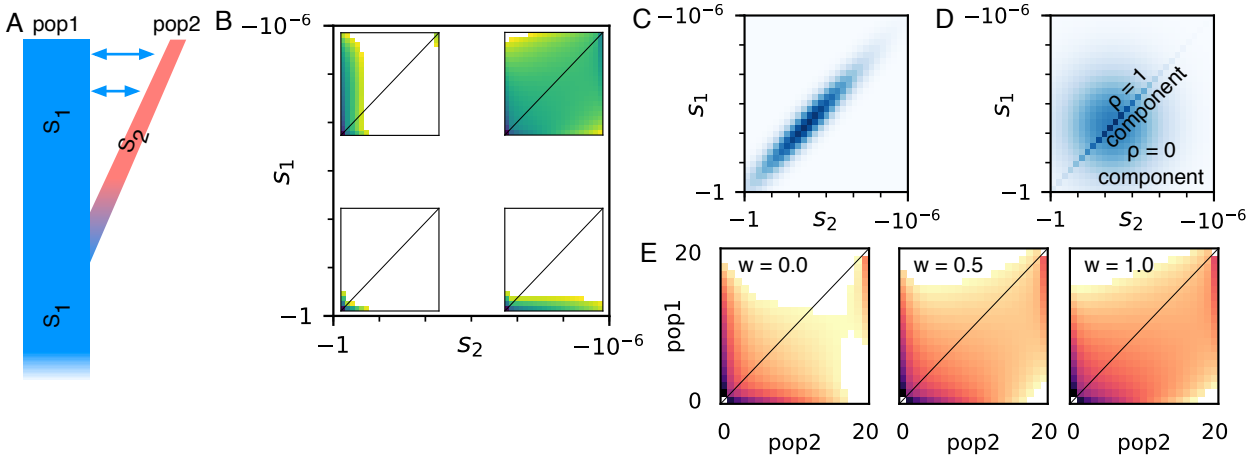

Figure 1: **The joint allele frequency spectrum (AFS) and joint distribution of fitness effects (DFE). A**: We considered populations that have recently diverged with gene flow between them. Some genetic variants will have a different effect on fitness in the diverged population (s_2) than in the ancestral population (s_1). **B**: The joint DFE is defined over pairs of selection coefficients (s_1, s_2). Insets show the joint AFS for pairs of variants that are strongly or weakly deleterious in each population. In each 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**: We focus on a model in which the joint DFE 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 w is the weight of the $\rho = 1$ component in the mixture model.

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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). We based our approach on the `fitdadi` framework developed by Kim et al. (2017), and our approach is integrated into our `dadi` software (Gutenkunst et al. 2009). More detail can be found in the Methods section.

Simulation

To evaluate the precision of our approach, we first stochastically simulated unlinked single nucleotide polymorphisms (SNPs) under a known demographic model (Table S1 & Fig. S2) and a symmetric lognormal mixture model for the joint DFE (Fig. 1; Eq. 4 & 6). We then inferred the three joint DFE parameters: the mean μ and standard deviation σ of the marginal lognormal distributions and the DFE correlation w. The demographic and joint DFE parameters for these simulations were similar to those we later inferred for human populations under a demographic model of divergence, growth, and migration. When we fit the joint DFE to these simulated data, we found that the variance of the inferred parameters grew only slowly as the sample size decreased (Fig. S3A). This suggests that only modest sample sizes are necessary to confidently infer the joint DFE, similar to how only modest sample sizes are necessary to infer the mean and variance of the univariate DFE (Keightley and Eyre-Walker 2010).

Because our inference approach focuses on shared variation, we expected precision to depend on the divergence time between the populations. To test this, we simulated data sets with sample size similar to our real Drosophila data and varied the divergence time in the demographic model. We found that the variance of the inferred μ and σ parameters was always small (Fig. S3B & C), but the variance of the inferred DFE correlation w depended on the divergence time (Fig. S3B & C). That variance was large for small divergence times ($T = 10^{-4}$). This is expected, because in this case selection has had little time act differently in the two populations. That variance was also large if the divergence time was large and there was no migration between the populations (Fig. S3C). This is also expected, because in this scenario there is little shared variation between populations. However, the variance of the inferred DFE correlation w was small when the divergence time was between 10^{-3} and 10^{0} (Fig. S3B & C).

Having found good precision for our inference, we then turned to testing the robustness of our inference to model misspecification. Since these tests focused on biases in the average inference, we did not stochastically sample data for these analyses, but rather used the expected AFS under each scenario as the data.

The demographic model is a key assumption of our joint DFE inference procedure. To test how imperfect modeling of demographic history would bias our inference, we simulated both neutral and selected data under a demographic model that included divergence, exponential growth in both populations, and asymmetric migration between populations (Fig. S2B). We then fit models that either lacked migration or that modeled instantaneous growth and symmetric migration to the neutral data (Fig. S2C). We then used these misspecified models to infer the DFE correlation w from the selected data. For both misspecified demographic models, although the inferred μ and σ were biased, we found that the inferred w was not strongly biased, particularly for large correlations (Fig. 2A).

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 $h = 0.5$. To test the effects of dominance on our inference, we simulated nonsynonymous frequency spectra with dominance coefficients of $h = 0.25$ and $h = 0.75$ and then optimized joint DFE parameters under the assumption that $h = 0.5$. We found that an incorrect assumption about dominance did not substantially bias the inferred w, although it did bias the inferred μ and σ (Fig 2B).

The probability distribution assumed for the joint DFE is another potential confounding factor. To test how this might bias inference, we first simulated a true mixture model in which the marginal distributions were gamma (Eq. 7), rather than lognormal (Eq. 6). In this case, we found that inferred w was not substantially biased (Fig. 2C). We also considered fitting the lognormal mixture model (Fig. 1D) to data simulated under a bivariate lognormal model (Fig. 1C & Eq. 8). In this case, we found that the inferred mixture component w was larger than the simulated bivariate lognormal correlation coefficient ρ, although they were

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

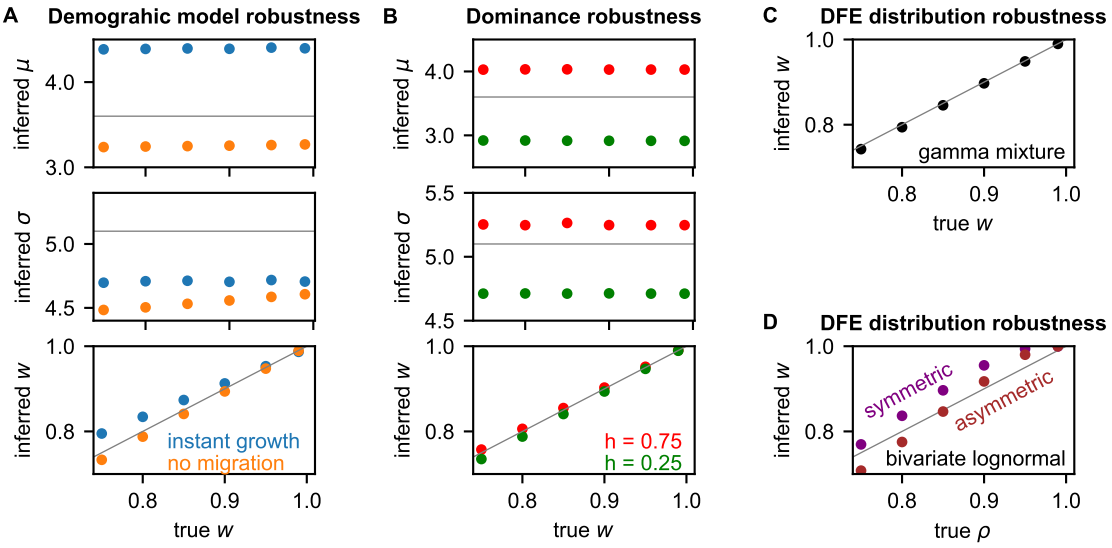

Figure 2: **Robustness of joint DFE inference to model misspecification.** Simulated neutral and selected data were generated under a demographic model with exponential growth and migration (Table S1), and lognormal mixture DFE models were fit to the data. The DFE parameters are: μ, the mean log population-scaled selection coefficient; σ, the standard deviation of those log coefficients; and w, the correlation of the DFE. The gray lines indicate true values, and the data plotted in these figures can be found in Table S4–S6. **A**: In this case, simpler demographic models with instantaneous growth or symmetric migration were fit to the neutral data. The resulting misspecified model was then used when inferring the DFE. This misspecification strongly biased μ and σ, but not w. **B**: In this case, selected data were simulated assuming dominant or recessive mutations, but the DFE was inferred assuming no dominance ($h = 0.5$). Again, μ and σ are strongly biased, but w is not. **C**: In this case, selected data were simulated using a mixture of gamma distributions. When these data were fit using our mixture of lognormal distributions, w was not biased. **D**: In this case, selected data were simulated using bivariate lognormal models, with either symmetric or asymmetric marginal distributions. When these data were fit using our symmetric mixture of lognormal distributions, w was only slightly biased.

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

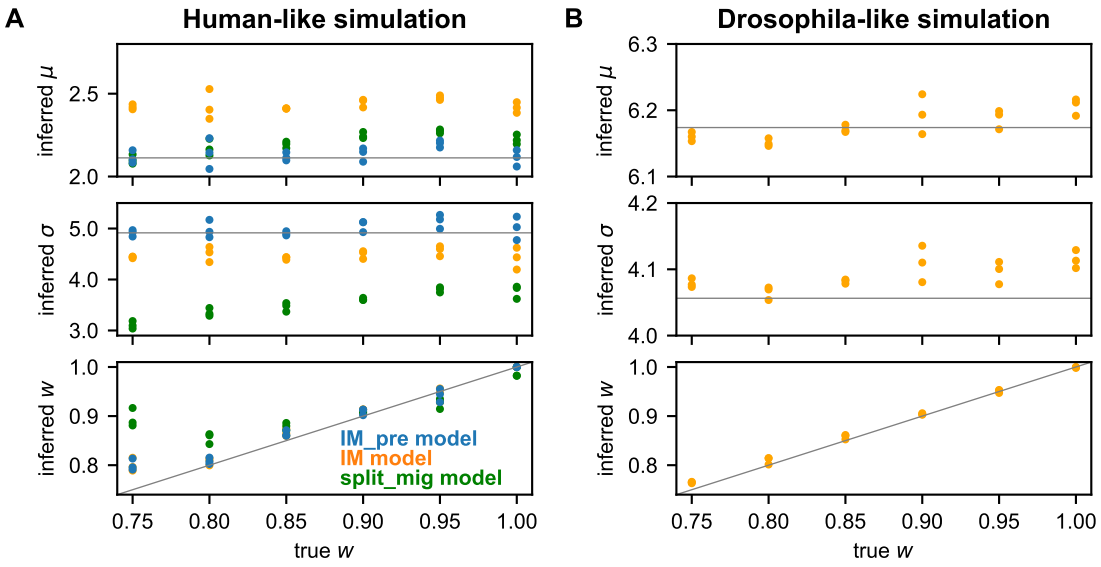

Figure 3: **Robustness of joint DFE inference to background selection.** Simulated genome-scale data were generated with background selection and different DFE correlations. **A**: Data were simulated using the best fit demographic model for humans in Fig. S5A with $\mu = 2.113$ and $\sigma = 4.915$. Beside fitting the true model, simpler demographic models (Fig. S2) were also fitted to test the robustness of model misspecification in the presence of background selection. **B**: Data were simulated using the best fit demographic model for *D. melanogaster* in Fig. S5B with $\mu = 6.174$ and $\sigma = 4.056$. Points indicate inferences from distinct data sets and colors indicate different demographic models using in the inference. Gray lines indicate true values. The data plotted in these figures can be found in Table S7.

similar (Fig. 2D). 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. 2D).

Finally, background selection may also bias our joint DFE inference. To examine the effects of background selection on our inference, we simulated data with linkage using SLiM 3 (Haller and Messer 2019). We simulated genome-scale data for both human- and Drosophila-like scenarios using the best fit demographic models we inferred for our real data (Fig. S5A & B). For each data set, we fit a demographic model to the simulated synonymous mutations then used that demographic model to infer the joint DFE from the simulated nonsynonyous mutations. For human-like simulations, we also carried out the analysis using simpler demographic models in the inference. As expected, we found that background selection (BGS) biased our demographic model inferences. For example, if we used the same human demographic model in the inference and simulation, the inferred divergence time increased as the DFE correlation w decreased (Table S7). As w decreased, the strength of BGS increased (Table S8). However, we found that the joint DFE correlation w could be robustly inferred in the presence of background selection (Fig. 3). The inferred μ and σ were biased if the demographic model was misspecified (Fig. 3A). But the inferred w was somewhat overestimated only if w was less than 0.8 with misspecified demographic models (Fig. 3A). In our Drosophila-like simulations, no bias in inferred w was observed (Fig. 3B). Because the strength of background selection in our simulations was much stronger than estimated in empirical studies (Fig. S4), these simulations suggest that our analysis of the real data is robust to background selection.

Together, our tests on simulated data suggest that inferring the DFE correlation w from the joint AFS can be done with high precision and is robust to multiple confounding factors, including misspecification of

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

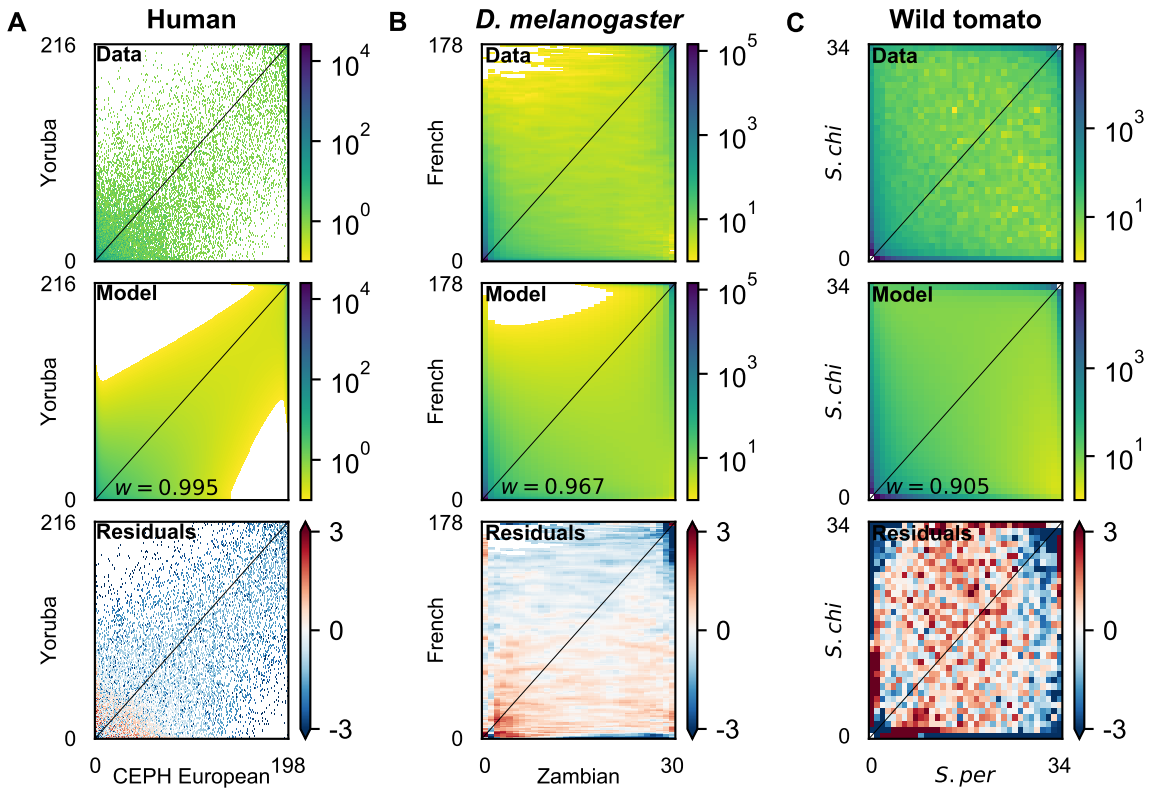

Figure 4: **Model fits to joint allele frequency spectra (AFS) using nonsynonymous data. A**: Joint AFS for the human nonsynonymous data, the best fit model with DFE correlation $w = 0.995$, and the residuals between model and data. **B**: Joint AFS for the *D. melanogaster* nonsynonymous data and the best fit model with DFE correlation $w = 0.967$. **C**: Joint AFS for the wild tomato nonsynonymous data and the best fit model with DFE correlation $w = 0.905$.

the demographic model and DFE distribution as well as the presence of background selection.

Application

We applied our joint DFE inference approach to humans, *D. melanogaster*, and wild tomatoes. For humans, we considered the joint DFE between Yoruba in Ibadan (YRI) and Utah residents (CEPH) with Northern and Western European ancestry (CEU) populations, because the Yoruba are a well-studied proxy for the ancestral human population and European populations parallel the history of French *D. melanogaster*. 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 wild tomatoes, we considered the joint DFE between two closely related species, *Solanum chilense* and *Solanum peruvianum*, because they still share substantial polymorphism and have overlapping ranges.

We first fit demographic models to synonymous variants in each population pair. For all the three species, 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. S5 & S6).

We next estimated the joint DFE using all nonsynonymous variants in the whole exome data from

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

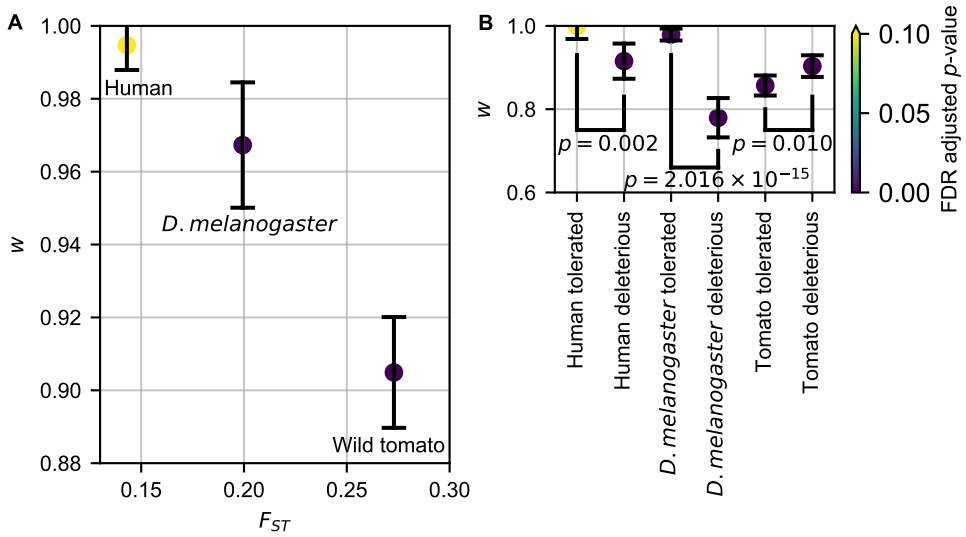

Figure 5: **Exome-wide DFE correlations.** **A**: Plotted are maximum likelihood inferences of the DFE correlation w with 95% confidence intervals versus genetic divergence F_{ST} of the considered population pair. **B**: Plotted are maximum likelihood inferences of the DFE correlation w with 95% confidence intervals for nonsynonymous SNPS with different predicted effects from SIFT. Colors indicate FDR adjusted p-values from two-tailed z-tests as to whether the confidence interval overlaps $w = 1$. F_{ST} was estimated using whole-exome synonymous mutations.

each species with our lognormal mixture model (Fig. 1D). In all the 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. 4 & S6). For humans, we found the highest DFE correlation $w = 0.995 \pm 0.007$ in our study (Fig. 5 & Table S9), which was statistically indistinguishable from perfect correlation $w = 1$. For *D. melanogaster*, we found that mutation fitness effects between Zambian and French populations were highly correlated, with $w = 0.967 \pm 0.017$ (Fig. 5 & Table S10). For wild tomatoes, we found the lowest DFE correlation, $w = 0.905 \pm 0.015$ (Fig. 5 & Table S11). We also inferred DFE correlations using mutations from non-CpG regions in humans and *D. melanogaster*, because the mutation rate may differ between non-CpG and CpG regions. The resulting estimates of w (Fig. S10 & Table S9 & S10) were statistically indistinguishable from those using the whole exome data. Among these three population pairs, the inferred DFE correlation was negatively related to genetic divergence, as measured by F_{ST} (Fig. 5A).

For simplicity, we assumed that the DFE correlation w is constant throughout the distribution, but the correlation 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 by applying SIFT scores to predict whether a nonsynonymous mutation is likely to be tolerated or deleterious based on evolutionary conservation (Vaser et al. 2016). We then fit DFE models to the SNPs in each class. As expected, we inferred a more negative mean fitness effect for the deleterious class than the tolerated class (Fig. S10 & Table S9–S11). Moreover, we found that the DFE correlation w was dramatically smaller for the deleterious class than the value from the tolerated class in humans and *D. melanogaster*, but not in wild tomatoes (Fig. 5B). To test whether this effect extended beyond individual mutations to whole genes, we also separated our data by the dN/dS ratio in humans and *D. melanogaster*. We found no significant difference in DFE correlations among genes with different dN/dS ratios (Fig. S10). However, we did observe that the average strength of purifying selection increases as the dN/dS ratio decreases (Fig. S10).

To investigate the biological basis of the joint DFE, we considered genes of different function based on Gene Ontology (GO) terms (The Gene Ontology Consortium 2000). For *D. melanogaster*, we found a wide

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

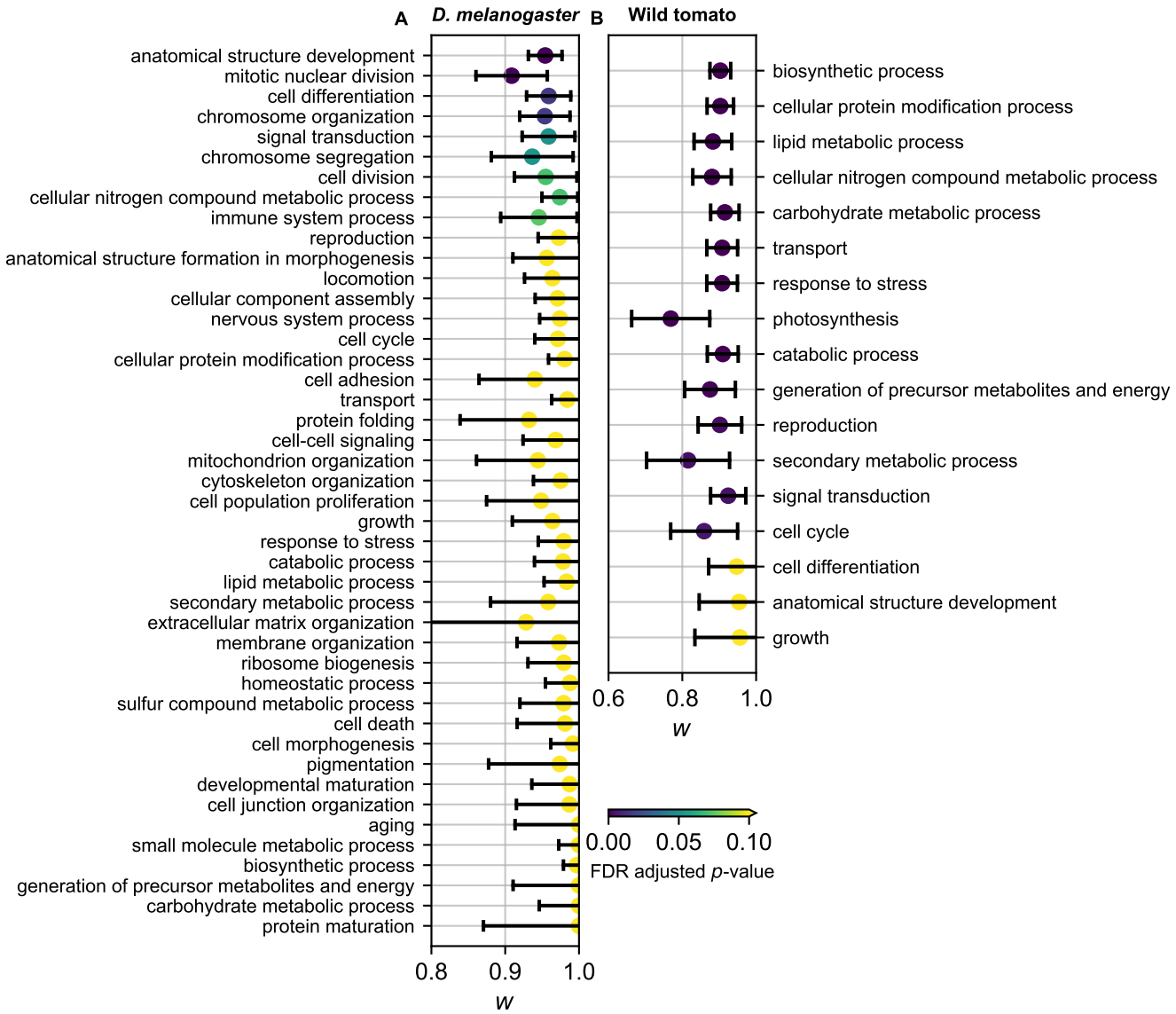

Figure 6: **DFE correlation for different GO terms in *D. melanogaster* and wild tomatoes.** Plotted are maximum likelihood inferences with 95% confidence intervals. Colors indicate FDR-adjusted p-values from two-tailed z-tests as to whether the confidence interval overlaps $w = 1$. The data plotted in these figures can be found in Table S10 & S11. **A**: Inferred DFE correlation in *D. melanogaster*. **B**: Inferred DFE correlation in wild tomatoes.

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

range of inferred DFE correlations, with the lowest maximum likelihood estimate corresponding to mutations in genes involved in the mitotic nuclear division at $w = 0.901 \pm 0.048$ (Fig. 6 & Table S10). For wild tomatoes, we found an even wider range of inferred DFE correlations, with the lowest maximum likelihood estimate being genes involved in photosynthesis at $w = 0.769 \pm 0.106$ (Fig. 6 & Table S11). For humans, we found that all GO terms yielded values of w that were statistically indistinguishable from one (Table S9 & Fig. S7). Among the *D. melanogaster* GO terms, we found no correlation between the inferred w and the mean and standard deviation of the marginal DFEs (Fig. S11), suggesting that the variation we see in w 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 viral pathogens. We found no statistically significant differences in DFE correlations among these gene groups, although we did find that the DFE for genes involved in disease or that interact with viruses was shifted toward more negative selection (Table S9 & Fig. S10).

To test the robustness of our analyses in the real data to various modeling choices, we used the variation among our inferences among *D. melanogaster* GO terms. We fit simpler models of demographic history with instantaneous growth in the two diverged populations with and without symmetric migration to the synonymous data and used those models as the basis of joint DFE analysis. Although these demographic models fit the data much less well than our main model (Fig. S6 & S12), the inferred values of w for the GO terms were highly correlated with those from our main model (Fig. S13A & B). We also tested our approach using a DFE model with a bivariate lognormal model instead of a lognormal mixture model. The inferred values for ρ in the bivariate model were highly correlated with the values for the inferred w (Fig. S13C). Together, these results suggest that the robustness we observed in simulated data (Fig. 2) holds true for real data.

Discussion

In this study, we introduced the concept of a joint distribution of fitness effects between pairs of populations, and we developed and applied an approach for inferring it. We tested our approach with simulation studies and found that inferring the DFE correlation between populations does not require excessive data and is robust to many forms of model misspecification (Fig. S3, 2, & 3). We then applied our approach to humans, *D. melanogaster*, and wild tomatoes. Among these species, we found the lowest exome-wide DFE correlation in wild tomatoes and the highest in humans (Fig. 5A). In humans and *D. melanogaster*, we found that the DFE correlation is lower for deleterious mutations than tolerated mutations (Fig. 5B). And in *D. melanogaster* and tomatoes, we found that the DFE correlation varied with gene function (Fig. 6). These results illustrate the biological insights that can be gained by considering the joint DFE between populations.

The first step of our analyses is fitting a demographic model, although our correlation inferences are robust to details of that model (Fig. 2A & S12). Nevertheless, our inferred demographic models (Fig. S5) agree well with other inferences. For humans, our demographic parameters were similar to those of Gravel et al. (2011). 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 wild tomatoes, we obtained a demographic model close to the result of Beddows et al. (2017).

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 genetically differentiated than the *D. melanogaster* or human populations we studied. In this case, we speculate that differences in fitness effects are primarily driven by differences in genetic background, although *Solanum chilense* does exhibit adaptations for more arid habitats (Moyle 2008). Among the species we studied, humans exhibited the highest correlation of mutation fitness effects, which was statistically indistinguishable from perfect correlation $w = 1$, suggesting little difference in mutation fitness effects between YRI and CEU populations. Huang et al. (2021) also estimated the genome-wide differences of selection coefficients between Africans and Europeans were almost 0 with a different approach (He et al. 2015). It is unclear whether this is caused by our relatively low genetic differentiation or our ability to control our local environment. Experi-

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

ments suggest that stressful environments can alter DFEs between populations (Wang et al. 2014). Previous population genetic studies also have found evidence for differences in marginal DFEs between populations of humans (Boyko et al. 2008; Lopez et al. 2018) and also between populations of other primates (Ma et al. 2013; Castellano et al. 2019; Tataru and Bataillon 2019). Although we assumed that the mean and the variance of mutation fitness effects did not differ between the two populations in our models for the joint DFE, those previous studies found only slight differences and our simulation study suggests that inferences of the DFE correlation are robust to relatively large differences in marginal DFEs (Fig. 2D). Recently, Martin and 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, so we could not apply it here. Overall, our simple models of the joint DFE fit the data well, but more complex models may be more informative. 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 mutation fitness effects.

We focused on the deleterious component of the DFE in this study, and positive selection or local adaptation may affect joint DFE inference. However, Castellano et al. (2019) found that including beneficial mutations or not did not affect the DFE model for the deleterious components in humans. Moreover, Zhen et al. (2021) estimated the proportion of new beneficial mutations to be around 1.5% in humans and close to 0 in *D. melanogaster*. Therefore, we do not expect beneficial mutations to significantly affect our inference in humans and *D. melanogaster*. Further studies that include local adaptation when inferring the joint DFE may improve our analysis of populations with low DFE correlations, such as wild tomatoes.

Finally, the concept of a joint DFE could be widely applicable. For example, we recently inferred a joint DFE between mutations at the same protein site, using triallelic variants (Ragsdale et al. 2016). Remarkably, we found that biochemical experiments in a variety of organisms yielded a similar correlation of pairwise fitness effects to the value we inferred from *D. melanogaster* population genetic data. Other potential applications of a joint DFE include modeling ancient DNA data to infer DFE correlations across time and modeling linkage to infer DFE correlations across genomic positions. We thus anticipate that extending the concept of the DFE from one population to two or more will significantly advance our understanding of population evolution and have broad impact in population genetics.

Methods

Inferring joint DFE from joint allele frequency spectrum (AFS)

If we sample n_1 chromosomes from population 1 and n_2 chromosomes from population 2, then the joint AFS for these two populations can be written as

$$\mathbf{X} = \{X_{i,j}, 0 \leq i \leq n_1, 0 \leq j \leq n_2, 0 < i + j < n_1 + n_2\}. \tag{1}$$

Here, $X_{i,j}$ denotes the number of mutations in the sample that have i copies of derived alleles among the n_1 chromosomes from population 1 and j copies of derived alleles among the n_2 chromosomes from population 2. We denote the joint spectra for neutral and selected mutations as $\mathbf{N} = \{N_{i,j}\}$ and $\mathbf{S} = \{S_{i,j}\}$, respectively.

Let $\mathbf{F}(\gamma_1, \gamma_2|\Theta_{\text{demo}}) = \{F_{i,j}(\gamma_1, \gamma_2|\Theta_{\text{demo}})\}$ be the expected joint AFS for demographic parameters Θ_{demo}, population-scaled selection coefficients γ_1 in the ancestral and first contemporary population and γ_2 in the second contemporary population, and population-scaled mutation rate $\theta = 1$. Then the expected neutral joint AFS is

$$E\left(N_{i,j}|\Theta_{\text{demo}}\right) = \theta_{\text{neu}} F_{i,j}(\gamma_1 = 0, \gamma_2 = 0|\Theta_{\text{demo}}), \tag{2}$$

where θ_{neu} is the population-scaled neutral mutation rate (Gutenkunst et al. 2009). The expected selected joint AFS is

$$E\left(S_{i,j}|\Theta_{\text{demo}}, \Theta_{\text{DFE}}\right) = \theta_{\text{sel}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_{i,j}(\gamma_1, \gamma_2|\Theta_{\text{demo}})\, G(\gamma_1, \gamma_2|\Theta_{\text{DFE}})\, \mathrm{d}\gamma_1\, \mathrm{d}\gamma_2. \tag{3}$$

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Here θ_{sel} is the population-scaled mutation rate for selected mutations, and $G(\gamma_1, \gamma_2|\Theta_{\text{DFE}})$ is the joint DFE.

In most of our analyses, we modeled the joint DFE as a mixture of two components, $G_{1\text{d}}$ and $G_{2\text{d}}$, where $G_{1\text{d}}$ is a DFE with equal selection coefficients in the two populations, and $G_{2\text{d}}$ is a DFE with statistically independent selection coefficients and marginal distributions $G_{1\text{d}}$. Letting w be the mixture proportion of $G_{1\text{d}}$, we have

$$G_{\text{mix}} = w\, G_{1\text{d}} + (1-w)\, G_{2\text{d}},\ 0 \leq w \leq 1. \tag{4}$$

And considering only deleterious mutations we have

$$\begin{aligned} E\left(S_{i,j}|\Theta_{\text{demo}}, \Theta_{\text{DFE}}\right) = w\, \theta_{\text{sel}} \int_{-\infty}^{0} F_{i,j}(\gamma, \gamma|\Theta_{\text{demo}}) G_{1\text{d}}(\gamma|\Theta_{\text{DFE}})\, \mathrm{d}\gamma \\ + (1-w)\, \theta_{\text{sel}} \int_{-\infty}^{0}\int_{-\infty}^{0} F_{i,j}(\gamma_1, \gamma_2|\Theta_{\text{demo}}) G_{2\text{d}}(\gamma_1, \gamma_2|\Theta_{\text{DFE}})\, \mathrm{d}\gamma_1\, \mathrm{d}\gamma_2. \end{aligned} \tag{5}$$

We typically worked with lognormal distributions, so

$$\begin{aligned} G_{1\text{d}}(\gamma) &= \frac{1}{\gamma\sigma\sqrt{2\pi}}\ \exp\left(-\frac{\left(\ln(-\gamma)-\mu\right)^2}{2\sigma^2}\right), \\ G_{2\text{d}}(\gamma_1, \gamma_2) &= \frac{1}{\gamma_1\gamma_2\sigma^2 2\pi}\ \exp\left(-\frac{\left(\ln(-\gamma_1)-\mu\right)^2 + \left(\ln(-\gamma_2)-\mu\right)^2}{2\sigma^2}\right). \end{aligned} \tag{6}$$

Here μ and σ are the mean and standard deviation of the logs of the population-scaled selection coefficients, respectively.

To test the robustness of our approach, we also considered other models for the joint DFE. When using a mixture of gamma distributions,

$$\begin{aligned} G_{1\text{d}}(\gamma) &= \frac{1}{\beta^{\alpha}\Gamma(\alpha)}(-\gamma)^{\alpha-1}\exp(\gamma/\beta) \\ G_{2\text{d}}(\gamma_1, \gamma_2) &= \frac{1}{\beta^{2\alpha}\Gamma(\alpha)^2}(\gamma_1\gamma_2)^{\alpha-1}\exp((\gamma_1+\gamma_2)/\beta). \end{aligned} \tag{7}$$

Here α is the shape parameter and β is the scale parameter. When using a bivariate lognormal distribution, which is potentially asymmetric,

$$\begin{aligned} G(\gamma_1, \gamma_2) =& \frac{1}{2\pi\sigma_1\sigma_2\gamma_1\gamma_2\sqrt{1-\rho^2}} \\ &\times \exp\left(-\frac{1}{2}\left(\frac{\left(\ln(-\gamma_1)-\mu_1\right)^2}{\sigma_1^2} + \frac{\left(\ln(-\gamma_2)-\mu_2\right)^2}{\sigma_2^2} - \frac{2\rho\left(\ln(-\gamma_1)-\mu_1\right)\left(\ln(-\gamma_2)-\mu_2\right)}{\sigma_1\sigma_2}\right)\right). \end{aligned} \tag{8}$$

Here ρ is the correlation coefficient.

Calculating the expected selected joint AFS (Eq. 3 & 5) is computationally expensive, because spectra $\mathbf{F}(\gamma_1, \gamma_2|\Theta_{\text{demo}})$ must be calculated for many pairs of selection coefficients. Simultaneously inferring the demographic Θ_{demo} and the DFE Θ_{DFE} parameters is thus infeasible. We thus first inferred the demographic parameters using the putative neutral data and then held those parameters constant while inferring the DFE parameters.

We inferred the demographic parameters $\hat{\Theta}_{\text{demo}}$ by maximizing the composite likelihood of the neutral joint AFS, including θ_{neu} as a free parameter (Gutenkunst et al. 2009). To then infer the DFE parameters Θ_{DFE}, we modeled the selected joint AFS as a Poisson Random Field (Sawyer and Hartl 1992) and maximized the composite likelihood

$$\mathcal{L}(\mathbf{S}|\hat{\Theta}_{\text{demo}}, \Theta_{\text{DFE}}) = \prod_{i,j} \frac{\exp\left[-E(S_{i,j}|\hat{\Theta}_{\text{demo}}, \Theta_{\text{DFE}})\right] E(S_{i,j}|\hat{\Theta}_{\text{demo}}, \Theta_{\text{DFE}})^{S_{i,j}}}{S_{i,j}!}. \tag{9}$$

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Here, $\hat{\Theta}_{\text{demo}}$ represents the demographic parameters inferred from the neutral data. And in this step we fixed θ_{sel} to a multiple of θ_{neu} determined by the expected ratio of new selected to new neutral mutations, based on base-specific mutation rates and genome composition.

Numerically, to calculate the expected selected joint AFS, we first cached expected spectra $\mathbf{F}(\gamma_1, \gamma_2|\hat{\Theta}_{\text{DFE}})$ for a range of selection coefficient pairs. 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). We then evaluated Eq. 3 using the trapezoid rule over these cached points. For the mixture model (Eq. 5), the $G_{1\text{d}}$ component was calculated as a one-dimensional integral over a cache of $\gamma_1 = \gamma_2$ spectra. Probability density for the joint DFE may 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 joint AFS $\mathbf{F}$ by the result and added it to the expected joint AFS. For the edges of the domain, this was done using the SciPy method `quad`, and for the corners it was done using `dblquad` (Virtanen et al. 2020).

Simulated data

For our precision tests (Fig. S3), we used `dadi` to simulate data sets without linkage. Unless otherwise specified, for Fig. S3 and Fig. 2, the "truth" simulations were performed with an isolation-with-migration (IM) demographic model (Fig. S2B) with parameters as in Table S1, a joint lognormal mixture DFE model with marginal mean $\mu = 3.6$ and standard deviation $\sigma = 5.1$, and with sample sizes of 216 for population 1 and 198 for population 2. For Fig. S3, data were simulated with $w = 0.9$ and the nonsynonymous population-scaled mutation rate $\theta_{NS} = 13842.5$ by Poisson sampling from the expected joint AFS. For Fig. S3A the resulting average number of segregating polymorphisms varied with sample size, ranging from 6,953 for sampling two chromosomes to 45,691 for sampling 100 chromosomes. For Fig. S3B & C, the sample size was fixed at 20 chromosomes per population.

For our robustness tests (Fig. 2), 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. 2A, 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. 2C, the true joint DFE was a mixture model with marginal gamma distributions with $\alpha = 0.4$, $\beta = 1400$. For Fig. 2D, the true joint DFE was a symmetric bivariate lognormal distribution with $\mu = 3.6$ and $\sigma = 5.1$, and for the asymmetric case in Fig. 2D, $\mu_1 = 3.6$, $\sigma_1 = 5.1$, $\mu_2 = 4.5$, $\sigma_2 = 6.8$. We then simulated data with different correlation coefficients ρ to examine the relationship between ρ and the DFE correlation w.

To examine the effects of background selection, we used SLiM 3 (Haller and Messer 2019) to simulate data with linkage. We replicated our simulation and inference three times for each w with different demographic models in the human simulations and an IM model in the *D. melanogaster* simulations (Fig. S2). For humans, we simulated the exome in chromosome 21 using the demographic parameters in Fig. S5A, the joint DFE parameters μ and σ from the whole human exome in Table S9 with $w = 0.75, 0.8, 0.85, 0.9, 0.95, 1$, and sample sizes of 216 for population 1 and 198 for population 2. We assumed the mutation rate was 1.5×10^{-8} per nucleotide per generation (Ségurel et al. 2014). We further assumed the ratio of the nonsynonymous to synonymous mutations in humans was 2.31 (Huber et al. 2017). In our simulation, we used the human exome based on the reference genome hg19 from UCSC Genome Browser and the deCODE human genetic map (Kong et al. 2010). For each w, we first simulated human chromosome 21 twenty times, then obtained 20 synonymous frequency spectra and 20 nonsynonymous frequency spectra from these sequences. We combined these 20 synonymous frequency spectra into a single one and inferred the demographic models. We then combined the 20 nonsynoymous frequency spectra into one spectrum and inferred the joint DFEs. We inferred the joint DFEs using both the true (IM_pre model) and wrong (IM model with asymmetric migration & split_mig model without migration) demographic models (Fig. S2). For *D. melanogaster*, we simulated small sequences instead of a whole chromosome, because the large population size of *D. melanogaster* made our simulation extremely slow. We used the demographic parameters for the IM model in Fig. S5B, the joint DFE parameters μ and σ from the whole *D. melanogaster* exome in Table S10 with $w = 0.75, 0.8, 0.85, 0.9, 0.95, 1$,

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

and sample sizes of 178 for population 1 and 30 for population 2. For each w, we simulated 20,000 small sequences with a length of 1000 bp, then obtained 20,000 synonymous frequency spectra and 20,000 nonsynonymous frequency spectra. We combined these 20,000 synonymous frequency spectra into a single one and inferred the demographic models. We then combined the 20,000 nonsynonymous frequency spectra into one spectrum and inferred the joint DFEs. This was equivalent to a total sequence size of 20 Mb. We assumed that the mutation rate was 2.8×10^{-9} per nucleotide per generation and that the recombination rate was 5×10^{-9} per nucleotide per generation (Keightley et al. 2014). We also assumed the ratio of the nonsynonymous to synonymous mutations in *D. melanogaster* was 2.85 (Huber et al. 2017). To accelerate our simulation, we used a factor of 1000 to re-scale the population size, mutation rate, and recombination rate (Hoggart et al. 2007). To quantify the strength of BGS in our simulations, we simulated data under neutral models and compared the expected number of pairwise differences between two chromosomes in the non-neutral scenarios with the neutral ones (Hudson and Kaplan 1995). The strength of BGS (Fig. S4) in the simulated data for both humans and *D. melanogaster* was stronger than the estimated strength from the empirical studies (Charlesworth 2013).

Genomic data

In all analyses, we only considered biallelic SNPs from automosomes. For humans, we obtained 108 and 99 unrelated individuals (216 and 198 haplotypes) from YRI and CEU populations in the 1000 Genomes Project Phase 3 genotype data (The 1000 Genomes Project Consortium 2015). We removed those regions that were not in the 1000 Genomes Project phase 3 strict mask file. We only considered biallelic exonic SNPs that were annotated as `synonymous_variant` or `missense_variant` by the 1000 Genomes Project. We further excluded SNPs without reported ancestral alleles. We also used the CpG table from the UCSC Genome Browser to distinguish SNPs in CpG regions.

For *D. melanogaster*, 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 and the French were 87 inbred individuals. We removed those SNPs in the IBD and/or admixture masks. In these data, many SNPs were not called in all individuals. We thus projected downward to obtain a consensus AFS with maximal genome coverage. For these data, that was to a sample size of 178 Zambian and 30 French haplotypes (Fig.S14). We used *D. simulans* as the outgroup and downloaded the alignment between the reference genome for *D. simulans* (drosim1) and the reference genome for *D. melanogaster* (dm3) from UCSC Genome Browser to determine the ancestral allele of each SNP. We then used GATK (version: 4.1.4.1) (McKenna et al. 2010) to liftover the genomic coordinates from dm3 to dm6 with the liftover chain file from the UCSC Genome Browser. To annotate SNPs to their corresponding genes and as synonymous or nonsynonymous mutations, we used ANNOVAR (version: 20191024) (Wang et al. 2010) with default settings and the `dm6` genome build. We downloaded the CpG table from the UCSC Genome Browser to distinguish SNPs in CpG regions.

For wild tomatoes, we obtained *Solanum chilense* and *Solanum peruvianum* DNA sequencing data from Beddows et al. (2017) and followed their scheme for assigning individuals to species. We only analyzed 17 *Solanum chilense* and 17 *Solanum peruvianum* individuals sequenced by Beddows et al. (2017) because of their high quality. We used an *Solanum lycopersicoides* individual sequenced by Beddows et al. (2017) to determine the ancestral allele of each SNP. We further removed variants with heterozygous genotype in this *Solanum lycopersicoides* individual. 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 models to genomic data

We used `dadi` to fit models for demography to spectra for synonymous mutations (Gutenkunst et al. 2009), including a parameter for ancestral state misidentification (Ragsdale et al. 2016). For the human analysis, we used `dadi` with grid points of [226,236,246], and we found that an isolation-with-migration model with an instantaneous growth in the ancestral population (IM_pre) fit the data well (Fig. S5A). For the *D. melanogaster* analysis, we used `dadi` with grid points of [188,198,208], and we found that an IM model

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

fit the data well (Fig. S5B). For the wild tomato analysis, we used `dadi` with grid points of [44,54,64] and fit a split-migration model with asymmetric migration (Fig. S5C), as Beddows et al. (2017) did.

Fitting joint DFEs to genomic data

Cached allele frequency spectra were created for the corresponding demographic models. For humans and *D. melanogaster*, we used the same grid points settings as the grid points used when inferring demographic models. For wild tomatoes, we used `dadi` with grid points of $[300, 400, 500]$ to generate caches with selection. Models of the joint DFE were then fit to nonsynonymous data by maximizing the likelihood of the data, assuming a Poisson Random Field (Sawyer and Hartl 1992). In these fits, the population-scaled mutation rate for nonsynonymous mutations θ_{NS} was held 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 wild tomatoes, this ratio was assumed to be 2.5, which was between the ratios in humans and *D. melanogaster*. For the lognormal mixture model, the three parameters of interest are the DFE correlation w as well as the mean μ and standard deviation σ of the marginal distributions. 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 mitigate the effect of background selection, we separately inferred demographic parameters for each subset of the data (Table S9–S11) with the best fit demographic model inferred from the whole exome data (Fig. S5).

We separately analyzed SNPs from genes associated with different Gene Ontology (GO) terms. We downloaded the Generic GO subset from http://geneontology.org/docs/download-ontology/ on August 12, 2020. This 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, a given gene may be present in multiple GO terms in our analysis. We used Ensembl Biomart (Cunningham et al. 2019) to retrieve the annotated GO terms for each gene. For humans, we downloaded the GO annotation from https://grch37.ensembl.org/biomart/martview/ with Ensembl Genes 101 database and Human genes (GRCh37.p13) on August 19, 2020. For *D. melanogaster*, we downloaded the GO annotation from https://www.ensembl.org/biomart/martview/ with Ensembl Genes 101 database and *D. melanogaster* genes (BDGP6.28) on September 10, 2020. For tomatoes, we downloaded the GO annotation from https://jul2018-plants.ensembl.org/biomart/martview/ with Ensembl Plants Genes 40 database and Solanum lycopersicum genes (SL2.50) on September 26, 2020. To ensure convergence in our inference, we removed those GO terms with less than 2,000 either synonymous or nonsynonymous mutations (Table S9–S10).

We also separately analyzed SNPs classified by SIFT as deleterious (SIFT score ≤ 0.05) or tolerated (SIFT score > 0.05) (Vaser et al. 2016). We downloaded SIFT predictions from https://sift.bii.a-star.edu.sg/sift4g/ on October 2, 2020. We used the SIFT prediction data with GRCH37.74 for humans, with BDGP6.83 for *D. melanogaster*, and with SL2.40.26 for tomatoes. To carry out our DFE analysis, we needed to estimate an appropriate population-scaled nonsynonymous mutation rate θ_{NS} for deleterious and tolerated mutations. To do so, we estimated the proportions of deleterious and tolerated mutations in the downloaded SIFT prediction datasets. This is because all the possible mutations and their SIFT scores were predicted in the downloaded datasets. We then obtained the population-scaled mutation rates for deleterious and tolerated mutations by multiplying θ_{NS} from the whole exome data with the proportions of deleterious and tolerated mutations respectively.

We further considered differences between regions of the genome that experience different levels of evolutionary conservation, as estimated from the ratio of nonsynonymous to synonymous divergence dN/dS. For humans, we separated SNPs into categories based on the estimated dN/dS values of the gene in which they are found from a previous study (Gayà-Vidal and Albà 2014). For *D. melanogaster*, we separated SNPs based on the dN/dS estimate of the surrounding 10 kb genomic region from PopFly (Hervas et al. 2017).

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

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Bioinformatics Institute's genome-wide association studies (GWAS) catalog (MacArthur et al. 2017). We used the data of Enard and 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), which is computationally more efficient than conventional bootstrap parameter optimization. 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. S15). Note that this process does not propagate uncertainty in the demographic parameter inference, so our uncertainties are somewhat underestimated.

To estimate p-values for inferred DFE correlation w, we used the two-tailed z-test by assuming $w = 1$ under the null hypothesis and using the standard deviation estimated from the Godambe approach. To compare inferred DFE correlations between tolerated and deleterious mutations, we used two-tailed z-tests to calculate p-values by assuming no difference under the null hypothesis and using the standard deviations estimated from the Godambe approach. For multiple testing correction, we estimated false discovery rate (FDR) adjusted p-values by the Benjamini–Hochberg procedure (Benjamini and Hochberg 1995). These multiple hypothesis tests are from different types of data, including whole-exome data, whole-exome data without CpG regions, different GO terms, genes with different dN/dS values, genes with different SIFT scores, genes associated with no/simple/complex diseases, and genes associated with no/single/multiple VIPs (Table S9–S11).

Data availability

The software tools used in this study are dadi (`https://bitbucket.org/gutenkunstlab/dadi`) and SLiM 3 (`https://github.com/MesserLab/SLiM/`). The supplementary tables can be found at `https://doi.org/10.6084/m9.figshare.13936985`.

Acknowledgments

We thank Laura Rose and Thorsten Kloesges for sharing their tomato data and assistance with the analysis. We thank David Enard for sharing his virus-interacting-proteins data. We thank Christian Huber and Kirk Lohmuller for discussing SLiM simulation with background selection. 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.

Author contribution

R.N.G designed the study. X.H, A.L.F, A.J.C, T.J.S., M.N.I, J.E.J, J.E.L, A.P.R, and R.N.G designed the analysis method and simulated and analyzed the data. X.H., A.L.F., T.J.S., J.E.J, and R.N.G wrote the manuscript.

Competing interests

The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F., and Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an Online catalog of human genes and genetic disorders. *Nucleic Acids Research*, 43(D1):D789–D798.

Arguello, J. R., Laurent, S., and Clark, A. G. (2019). Demographic history of the human commensal Drosophila melanogaster. *Genome Biology and Evolution*, 11(3):844–854.

Balick, D. J., Do, R., Cassa, C. A., Reich, D., and Sunyaev, S. R. (2015). Dominance of Deleterious Alleles Controls the Response to a Population Bottleneck. *PLoS Genetics*, 11(8):1–23.

Barton, H. J. and Zeng, K. (2018). New methods for inferring the distribution of fitness effects for INDELs and SNPs. *Molecular Biology and Evolution*, 35(6):1536–1546.

Barton, N. H. (1995). A general model for the evolution of recombination. *Genetics Research*, 65:123–144.

Beddows, I., Reddy, A., Kloesges, T., and Rose, L. E. (2017). Population genomics in wild tomatoes-the interplay of divergence and admixture. *Genome Biology and Evolution*, 9(11):3023–3028.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B*, 57(1):289–300.

Boyko, A. R., Williamson, S. H., Indap, A. R., Degenhardt, J. D., Hernandez, R. D., Lohmueller, K. E., Adams, M. D., Schmidt, S., Sninsky, J. J., Sunyaev, S. R., White, T. J., Nielsen, R., Clark, A. G., and Bustamante, C. D. (2008). Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet*, 4(5):e1000083.

Castellano, D., James, J., and Eyre-Walker, A. (2017). Nearly neutral evolution across the Drosophila melanogaster genome. *Mol Biol Evol*, 35(11):2685–2694.

Castellano, D., MacIà, M. C., Tataru, P., Bataillon, T., and Munch, K. (2019). Comparison of the full distribution of fitness effects of new amino acid mutations across great apes. *Genetics*, 213(3):953–966.

Charlesworth, B. (1994). The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genetical Research*, 63(3):213–227.

Charlesworth, B. (2013). Background selection 20 years on.

Chen, J., Glémin, S., and Lascoux, M. (2017). Genetic diversity and the efficacy of purifying selection across plant and animal species. *Molecular Biology and Evolution*, 34(6):1417–1428.

Chen, J., Glémin, S., and Lascoux, M. (2020). From drift to draft: How much do beneficial mutations actually contribute to predictions of Ohta's slightly deleterious model of molecular evolution? *Genetics*, 214(4):1005–1018.

Coffman, A. J., Hsieh, P., Gravel, S., and Gutenkunst, R. N. (2016). Computationally efficient composite likelihood statistics for demographic inference. *Molecular Biology and Evolution*, 33(2):591–593.

Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M. R., Armean, I. M., Bennett, R., Bhai, J., Billis, K., Boddu, S., Cummins, C., Davidson, C., Dodiya, K. J., Gil, L., Grego, T., Haggerty, L., Gall, A., Garc, C., Haskell, E., Hourlier, T., Izuogu, O. G., Janacek, S. H., Juettemann, T., Kay, M., Laird, M. R., Lavidas, I., Liu, Z., Marug, C., Loveland, J. E., Maurel, T., Mcmahon, A. C., Moore, B., Morales, J., Mudge, J. M., Nuhn, M., Ogeh, D., Parker, A., Parton, A., Patricio, M., Abdul, A. I., Schmitt, B. M., 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, S. E., Kostadima, M., Langridge, N., Martin, F. J., Muffato, M., Perry, E., Ruffier, M., Staines, D. M., Trevanion, S. J., Aken, B. L., Yates, A. D., Zerbino, D. R., and Flicek, P. (2019). Ensembl 2019. *Nucleic Acids Research*, 47:D745–D751.

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Durvasula, A. and Lohmueller, K. E. (2019). Negative selection on complex traits limits genetic risk prediction accuracy between populations. *bioRxiv*, page https://doi.org/10.1101/721936.

Elena, S. F., Ekunwe, L., Hajela, N., Oden, S. A., and Lenski, R. E. (1998). Distribution of fitness effects caused by random insertion mutations in Escherichia coli. *Genetica*, 102-103(1-6):349–58.

Enard, D. and Petrov, D. A. (2018). Evidence that RNA Viruses Drove Adaptive Introgression between Neanderthals and Modern Humans. *Cell*, 175(2):360–371.

Eyre-Walker, A. and Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, 8(8):61061–8.

Eyre-Walker, A., Woolfit, M., and Phelps, T. (2006). The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics*, 173(2):891–900.

Fry, J. D., Keightley, P. D., Heinsohn, S. L., and Nuzhdin, S. V. (2002). New estimates of the rates and effects of mildly deleterious mutation in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*, 96(2):574–579.

Gandon, S., Hochberg, M., Holt, R. D., Martin, G., and Day, T. (2013). What limits the evolutionary emergence of pathogens? *Philosophical Transactions of the Royal Society B*, 368:20120086.

Gayà-Vidal, M. and Albà, M. (2014). Uncovering adaptive evolution in the human lineage. *BMC Genomics*, 15:599–undefined.

Gravel, S., Henn, B. M., Gutenkunst, R. N., Indap, A. R., Marth, G. T., Clark, A. G., Yu, F., Gibbs, R. A., The 1000 Genomes Project, and Bustamante, C. D. (2011). Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences USA*, 108:11983–11988.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., and Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, 5(10):e1000695.

Haller, B. C. and Messer, P. W. (2019). SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Molecular Biology and Evolution*, 36(3):632–637.

He, Y., Wang, M., Huang, X., Li, R., Xu, H., Xu, S., and Jin, L. (2015). A probabilistic method for testing and estimating selection differences between populations. *Genome Research*, 25(12):1903–1909.

Hervas, S., Sanz, E., Casillas, S., Pool, J. E., and Barbadilla, A. (2017). PopFly: the Drosophila population genomics browser. *Bioinformatics*, 33(17):2779–2780.

Hoggart, C. J., Chadeau-Hyam, M., Clark, T. G., Lampariello, R., Whittaker, J. C., De Iorio, M., and Balding, D. J. (2007). Sequence-Level Population Simulations Over Large Genomic Regions. *Genetics*, 177(3):1725–1731.

Huang, X., Wang, S., Jin, L., and He, Y. (2021). Dissecting dynamics and differences of selective pressures in the evolution of human pigmentation. *Biology Open*, 10(2):bio.056523.

Huber, C. D., Durvasula, A., Hancock, A. M., and Lohmueller, K. E. (2018). Gene expression drives the evolution of dominance. *Nature Communications*, 9(1):2750.

Huber, C. D., Kim, B. Y., Marsden, C. D., and Lohmueller, K. E. (2017). Determining the factors driving selective effects of new nonsynonymous mutations. *Proceedings of the National Academy of Sciences USA*, 114(17):4465–4470.

Hudson, R. R. and Kaplan, N. L. (1995). Deleterious background selection with recombination. *Genetics*, 141(4):1605–1617.

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

James, J., Castellano, D., and Eyre-Walker, A. (2017). DNA sequence diversity and the efficiency of natural selection in animal mitochondrial DNA. *Heredity*, 118(1):88–95.

Johri, P., Charlesworth, B., and Jensen, J. D. (2020). Toward an evolutionarily appropriate null model : Jointly inferring demography and purifying selection. *Genetics*, 215(1):173–192.

Keightley, P. D. and Eyre-Walker, A. (2007). Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics*, 177(4):2251–61.

Keightley, P. D. and Eyre-Walker, A. (2010). What can we learn about the distribution of fitness effects of new mutations from DNA sequence data? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1544):1187–1193.

Keightley, P. D., Ness, R. W., Halligan, D. L., and Haddrill, P. R. (2014). Estimation of the spontaneous mutation rate per nucleotide site in a Drosophila melanogaster full-sib family. *Genetics*, 196(1):313–320.

Kim, B. Y., Huber, C. D., and Lohmueller, K. E. (2017). Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. *Genetics*, 206(1):345–361.

Kong, A., Thorleifsson, G., Gudbjartsson, D. F., Masson, G., Sigurdsson, A., Jonasdottir, A., Walters, G. B., Jonasdottir, A., Gylfason, A., Kristinsson, K. T., Gudjonsson, S. A., Frigge, M. L., Helgason, A., Thorsteinsdottir, U., and Stefansson, K. (2010). Fine-scale recombination rate differences between sexes, populations and individuals. *Nature*, 467(7319):1099–1103.

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

Lack, J. B., Cardeno, C. M., Crepeau, M. W., Taylor, W., Corbett-Detig, R. B., Stevens, K. A., Langley, C. H., and Pool, J. E. (2015). The Drosophila Genome Nexus: A population genomic resource of 623 Drosophila melanogaster genomes, including 197 from a single ancestral range population. *Genetics*, 199(4):1229–1241.

Lack, J. B., Lange, J. D., Tang, A. B., Corbett-Detig, R. B., and Pool, J. E. (2016). A Thousand Fly Genomes: An Expanded Drosophila Genome Nexus. *Mol Biol Evol*, 33(12):3308–3313.

Lopez, M., Kousathanas, A., Quach, H., Harmant, C., Mouguiama-Daouda, P., Hombert, J. M., Froment, A., Perry, G. H., Barreiro, L. B., Verdu, P., Patin, E., and Quintana-Murci, L. (2018). The demographic history and mutational load of African hunter-gatherers and farmers. *Nature Ecology and Evolution*, 2(4):721–730.

Ma, X., Kelley, J. L., Eilertson, K., Musharoff, S., Degenhardt, J. D., Martins, A. L., Vinar, T., Kosiol, C., Siepel, A., Gutenkunst, R. N., and Bustamante, C. D. (2013). Population genomic analysis reveals a rich speciation and demographic history of orang-utans (Pongo pygmaeus and Pongo abelii). *PLoS ONE*, 8(10):e77175.

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., Pendlington, Z. M., Welter, D., Burdett, T., Hindorff, L., Flicek, P., Cunningham, F., and Parkinson, H. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic acids research*, 45(D1):D896–D901.

Martin, G. and Lenormand, T. (2015). The fitness effect of mutations across environments: Fisher's geometrical model with multiple optima. *Evolution*, 69(6):1433–1447.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9):1297–1303.

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Moyle, L. C. (2008). Ecological and evolutionary genomics in the wild tomatoes (Solanum Sect. Lycopersicon). *Evolution*, 62(12):2995–3013.

Nadarajah, S. and Gupta, A. K. (2006). Some bivariate gamma distributions. *Applied Mathematics Letters*, 19(8):767–774.

Ragsdale, A. P., Coffman, A. J., Hsieh, P., Struck, T. J., and Gutenkunst, R. N. (2016). Triallelic population genomics for inferring correlated fitness effects of same site nonsynonymous mutations. *Genetics*, 203(1):513–523.

Sawyer, S. A. and Hartl, D. L. (1992). Population genetics of polymorphism and divergence. *Genetics*, 132(4):1161–76.

Ségurel, L., Wyman, M. J., and Przeworski, M. (2014). Determinants of Mutation Rate Variation in the Human Germline. *Annual Review of Genomics and Human Genetics*, 15(1):47–70.

Struck, T. J., Mannakee, B. K., and Gutenkunst, R. N. (2018). The impact of genome-wide association studies on biomedical research publications. *Human Genomics*, 12(1):38.

Tataru, P. and Bataillon, T. (2019). polyDFEv2.0: testing for invariance of the distribution of fitness effects within and across species. *Bioinformatics*, 35(16):2868–2869.

Tataru, P., Mollion, M., Glémin, S., and Bataillon, T. (2017). Inference of distribution of fitness effects and proportion of adaptive substitutions from polymorphism data. *Genetics*, 207(3):1103–1119.

The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature*, 526(7571):68–74.

The Gene Ontology Consortium (2000). Gene Ontology: tool for the unification of biology. *Nat Genet*, 25(1):25–29.

Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M., and Ng, P. C. (2016). SIFT missense predictions for genomes. *Nature Protocols*, 11(1):1–9.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, , Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., Vijaykumar, A., Bardelli, A. P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C. N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D. A., Hagen, D. R., Pasechnik, D. V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G. A., Ingold, G. L., Allen, G. E., Lee, G. R., Audren, H., Probst, I., Dietrich, J. P., Silterra, J., Webber, J. T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J. L., de Miranda Cardoso, J. V., Reimer, J., Harrington, J., Rodríguez, J. L. C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N. J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P. A., Lee, P., McGibbon, R. T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T. J., Robitaille, T. P., Spura, T., Jones, T. R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y. O., and Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272.

Wang, A. D., Sharp, N. P., and Agrawal, A. F. (2014). Sensitivity of the distribution of mutational fitness effects to environment, genetic background, and adaptedness: A case study with Drosophila. *Evolution*, 68(3):840–853.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, 38(16):e164.

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Zhen, Y., Huber, C. D., Davies, R. W., and Lohmueller, K. E. (2021). Greater strength of selection and higher proportion of beneficial amino acid changing mutations in humans compared with mice and Drosophila melanogaster. *Genome Research*, 31(1):110–120.

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Supporting Information

Table S1: Model parameters for unlinked simulation

Model	s	ν_1	ν_2	T	$m_{1\leftarrow 2}$	$m_{2\leftarrow 1}$	θ_{NS}
IM	0.93	2.9	2.83	0.084	0.47	0.35	13,842.5

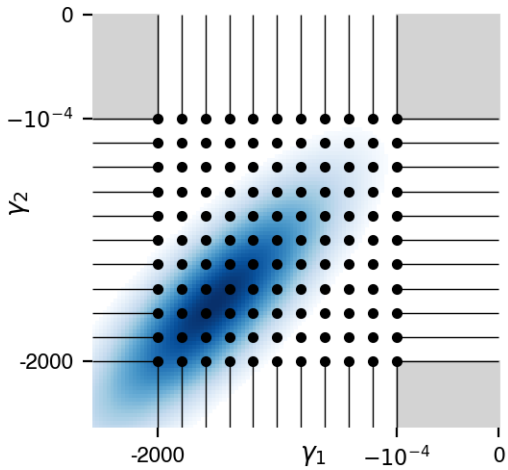

Figure S1: **Illustration of computational approach for calculating expected joint AFS for a given joint DFE.** Dots represent cached frequency spectra. Horizontal and vertical lines indicate single-variable semi-analytic integrations to estimate DFE density outside the sampled domain, and gray regions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

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

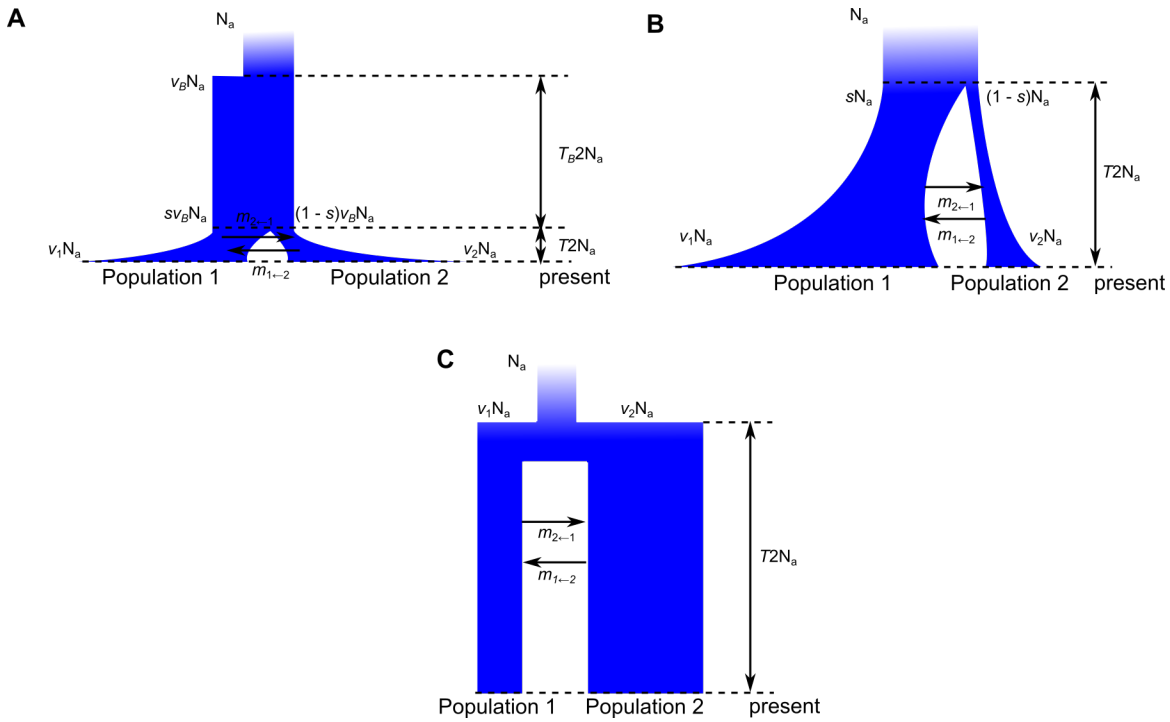

Figure S2: **Demographic models used in this study.** In these models, the population sizes were scaled by the ancestral population size N_a and the times were in units of $2N_a$ generations. The migration rates between Population 1 and 2 can be symmetric ($m_{1\leftarrow 2} = m_{2\leftarrow 1}$) or asymmetric ($m_{1\leftarrow 2} \neq m_{2\leftarrow 1}$). If there was no migration, then $m_{1\leftarrow 2} = m_{2\leftarrow 1} = 0$. **A**: IM_pre model. In this model, the ancestral population experienced an instantaneous population size change at time $T + T_B$ before present. After the change, its population size became $\nu_B N_a$ and remained constant until time T before present. Population 1 and 2 diverged at time T before present and then grew exponentially. For Population 1, its initial population size was $s\nu_B N_a$ and its final population size was $\nu_1 N_a$ at present. For Population 2, its initial population size was $(1-s)\nu_B N_a$ and its final population size was $\nu_2 N_a$ at present. **B**: IM model. This model corresponds to the IM_pre model with no ancestral growth event. **C**: split_mig model. In this model, Population 1 and 2 diverged at time T before present. Their population sizes remained constant after the divergence with $\nu_1 N_a$ for Population 1 and $\nu_2 N_a$ for Population 2.

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

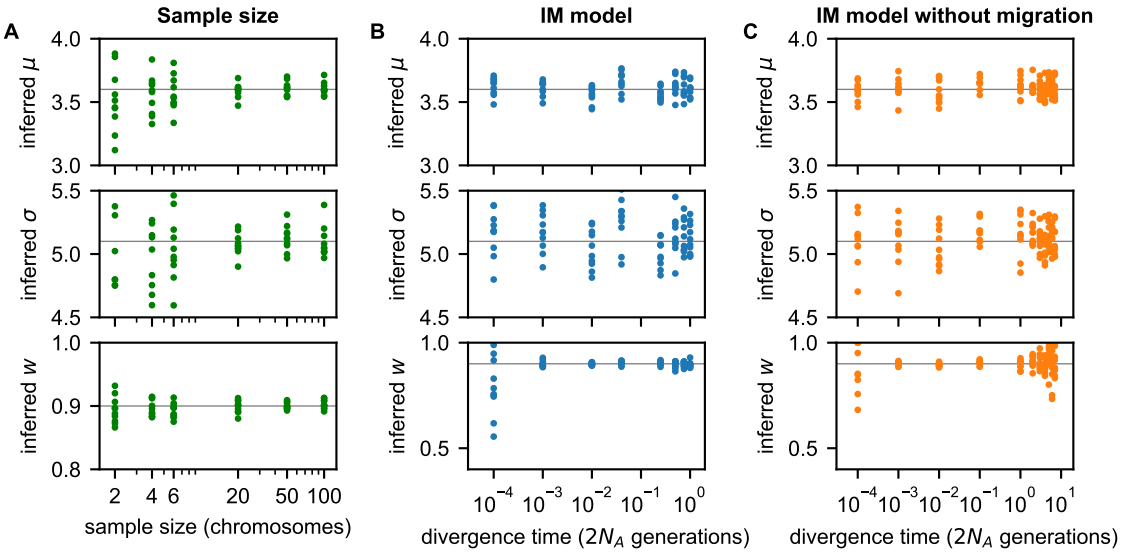

Figure S3: **Precision of joint DFE inference versus sample sizes and divergence time.** Simulated data were generated without linkage and with different sample sizes or divergence times, using the demographic model parameters in Table S1 with logornal $\mu = 3.6$ and $\sigma = 5.1$ and DFE correlation $w = 0.9$. In each panel, points represent inferences from individual data sets and the gray line indicates the true value. **A**: Inferred DFE parameters for different sample sizes. For small sample sizes, w is still inferred precisely, even though μ and σ are highly uncertain. **B**: Inferred DFE parameters for different divergence times, in a model with migration. Precision of w inference is low for small divergence times. **C**: Inferred DFE parameters for different divergence times, in a model without migration. In this case, precision of w inference is low for both small and large divergence times.

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

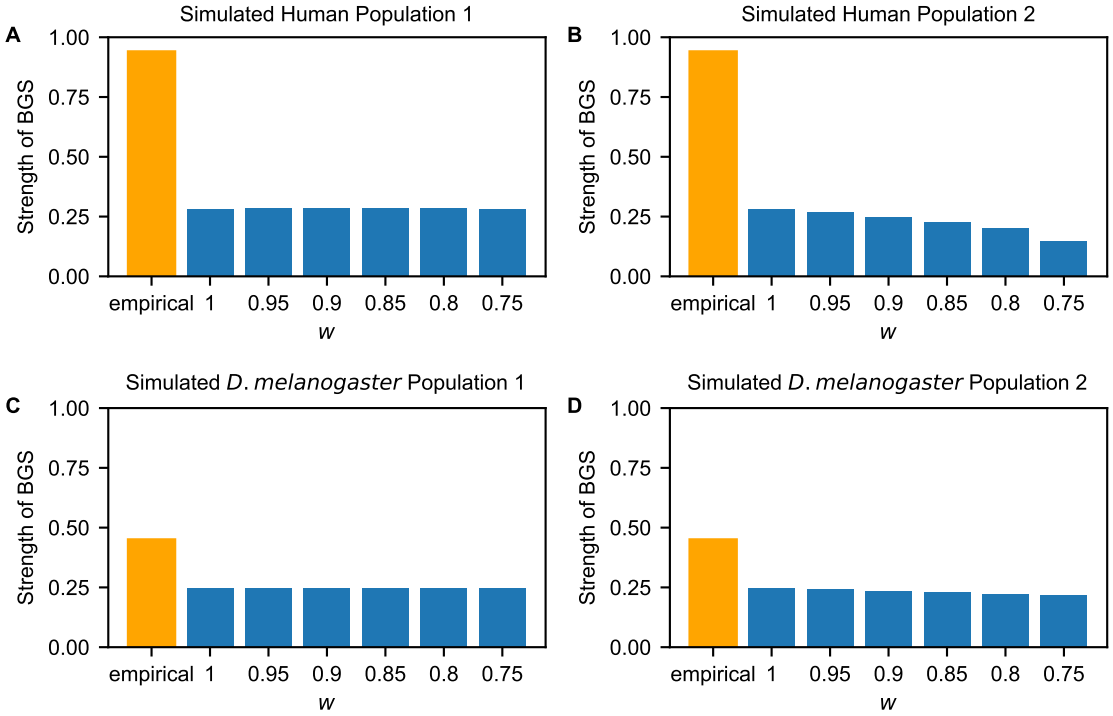

Figure S4: **Strength of background selection (BGS) between simulation and empirical studies.** The strengths of BGS in empirical studies are from Charlesworth (2013). The strengths in simulation are estimated from the observed number of pairwise differences between two chromosomes in the non-neutral scenarios versus the neutral ones. The data plotted in these figures can be found in Table S8.

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

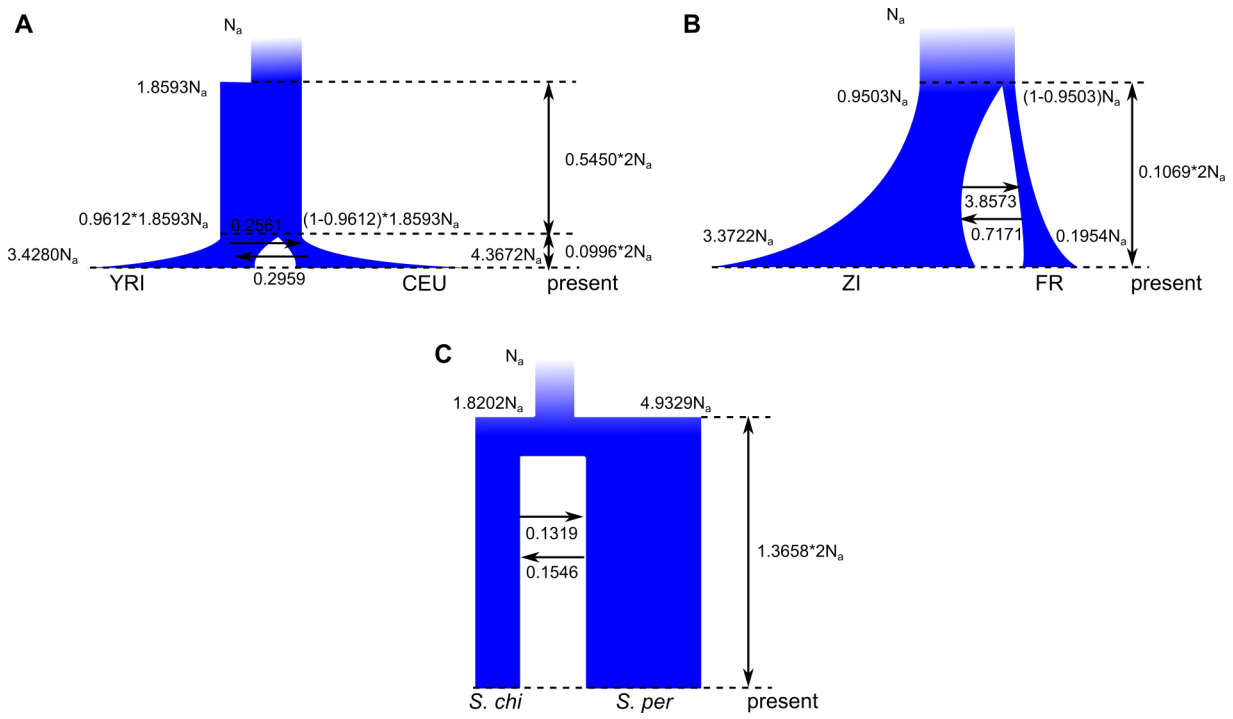

Figure S5: **Best fit demographic models. A**: The best fit demographic model for humans. **B**: The best fit demographic model for *D. melanogaster*. **C**: The best fit demographic model for wild tomatoes.

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

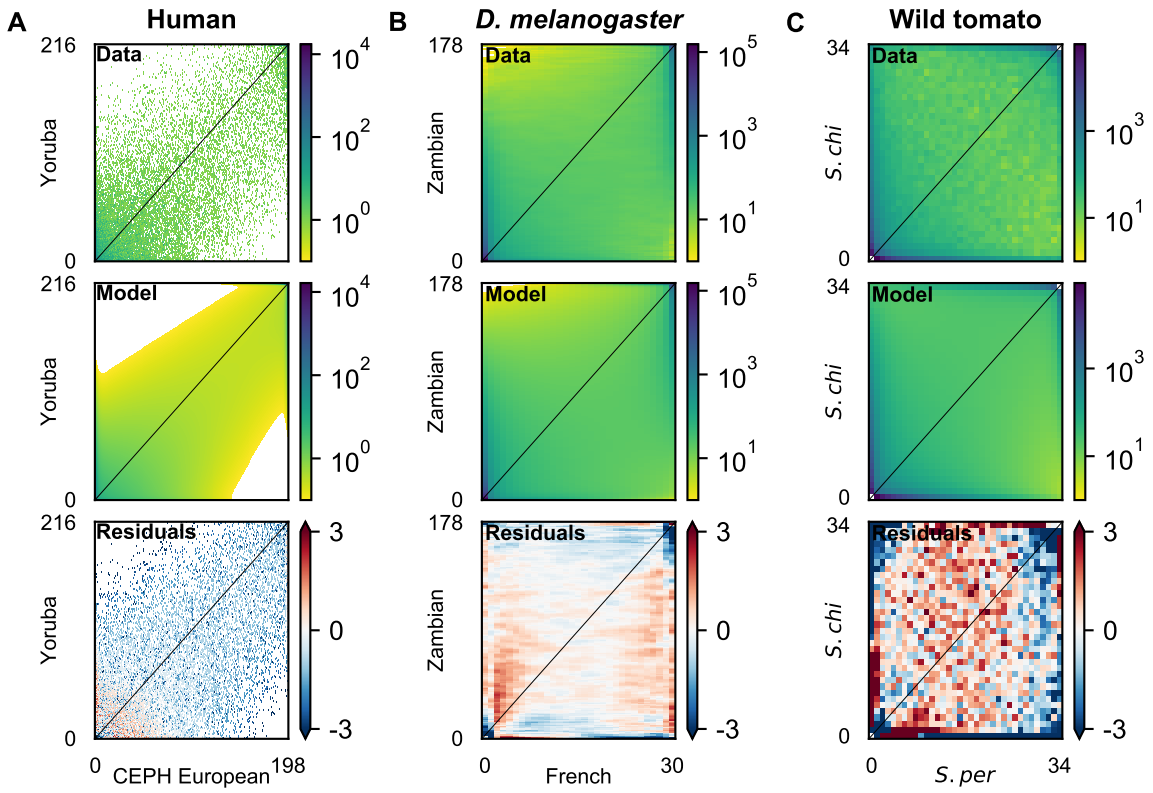

Figure S6: **Model fits to joint allele frequency spectra (AFS) using synonymous data.** **A**: Joint AFS for the human synonymous data and the best fit model (Fig. S5A). **B**: Joint AFS for the *D. melanogaster* synonymous data and the best fit model (Fig. S5B). **C**: Joint AFS for the wild tomato synonymous data and the best fit model (Fig. S5C).

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

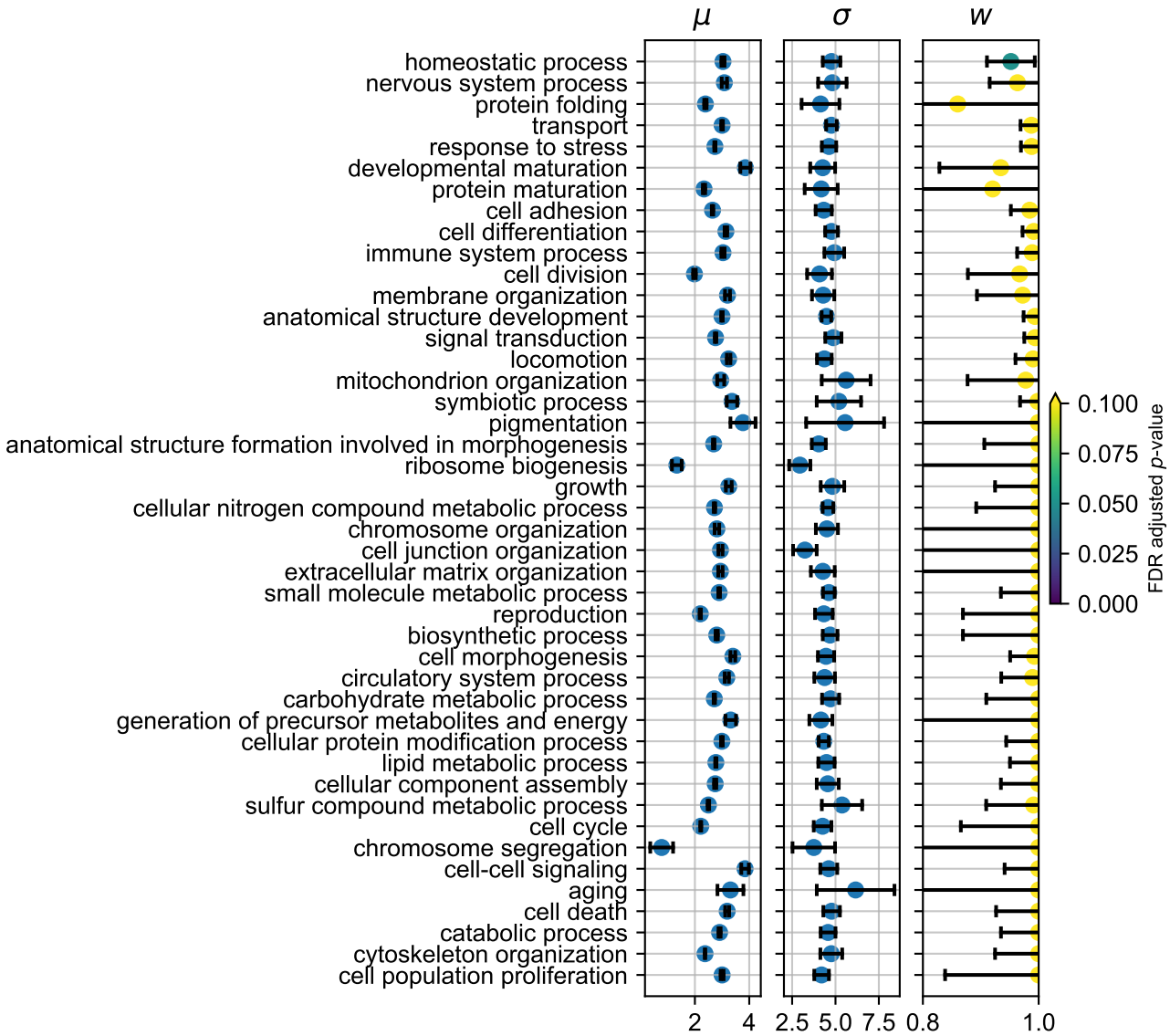

Figure S7: **Joint DFE inference for different GO terms in humans.** Plotted are maximum likelihood inferences with 95% confidence intervals. For inferred w, colors indicate FDR-adjusted p-values from two-tailed z-tests as to whether the confidence interval overlaps $w = 1$. The data plotted in these figures can be found in Table S9.

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

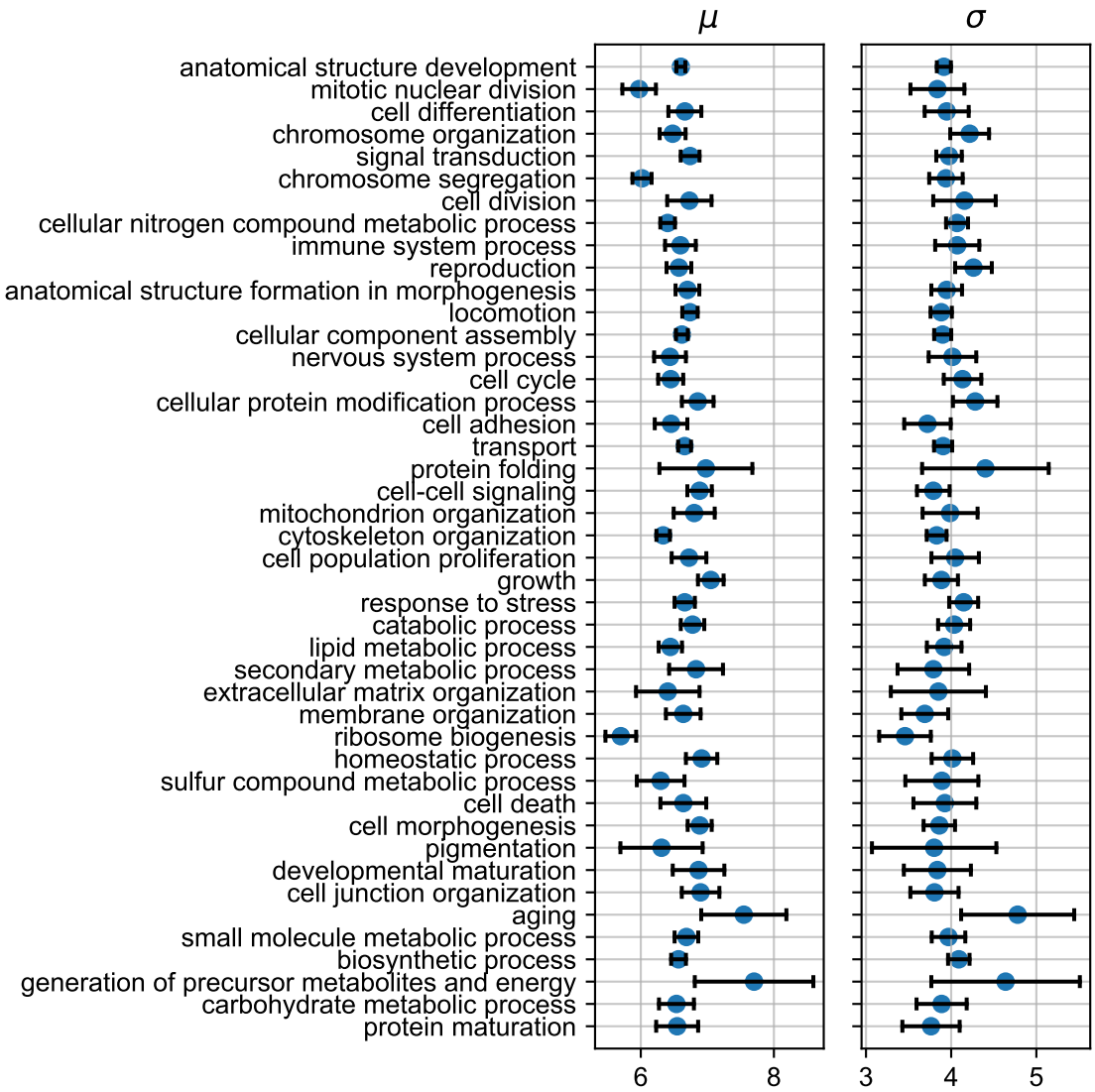

Figure S8: **Inferred μ and σ from the joint DFE inference for different GO terms in *D. melanogaster*.** Plotted are maximum likelihood inferences with 95% confidence intervals. The data plotted in these figures can be found in Table S10.

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

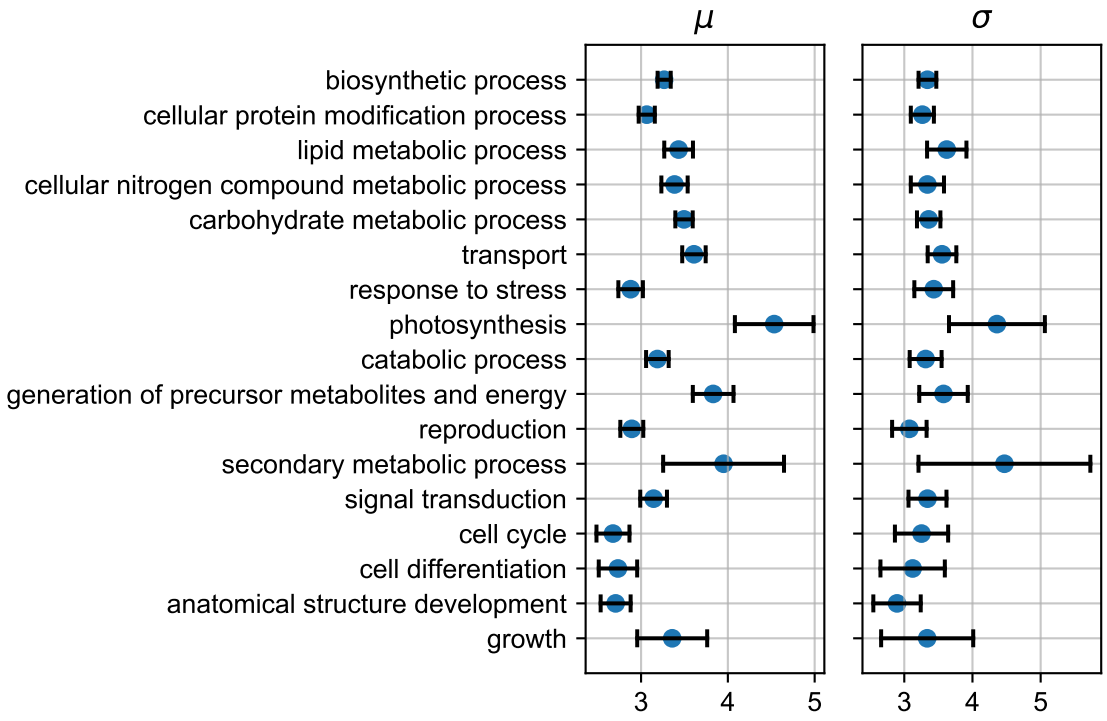

Figure S9: **Inferred μ and σ from the joint DFE inference for different GO terms in wild tomatoes.** Plotted are maximum likelihood inferences with 95% confidence intervals. The data plotted in these figures can be found in Table S11.

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

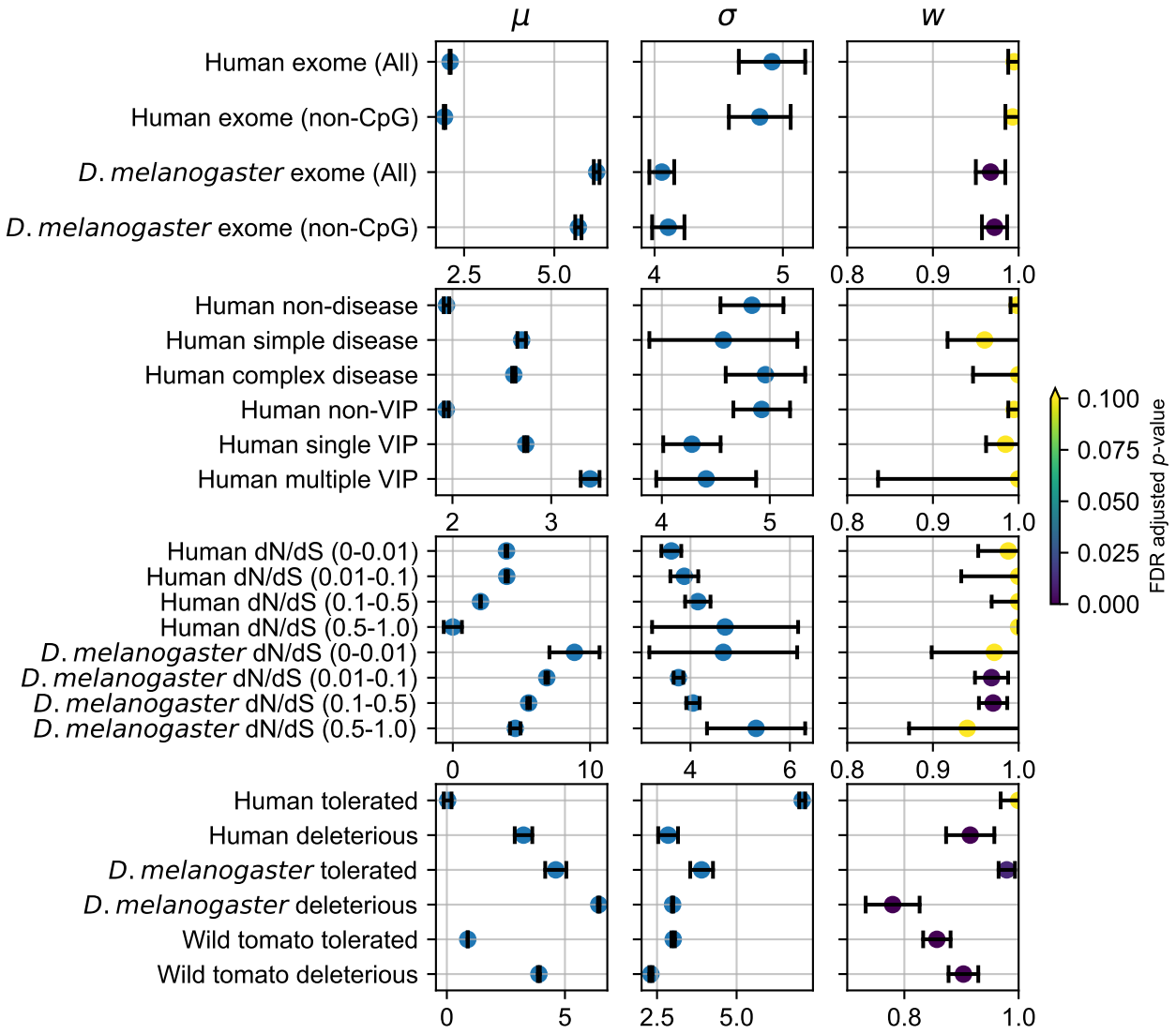

Figure S10: **Joint DFE inference for different types of data in humans and *D. melanogaster*.** Plotted are maximum likelihood inferences with 95% confidence intervals. For inferred w, colors indicate FDR-adjusted p-values from two-tailed z-tests as to whether the confidence interval overlaps $w = 1$. The data plotted in these figures can be found in Table S9, S10 and S11.

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

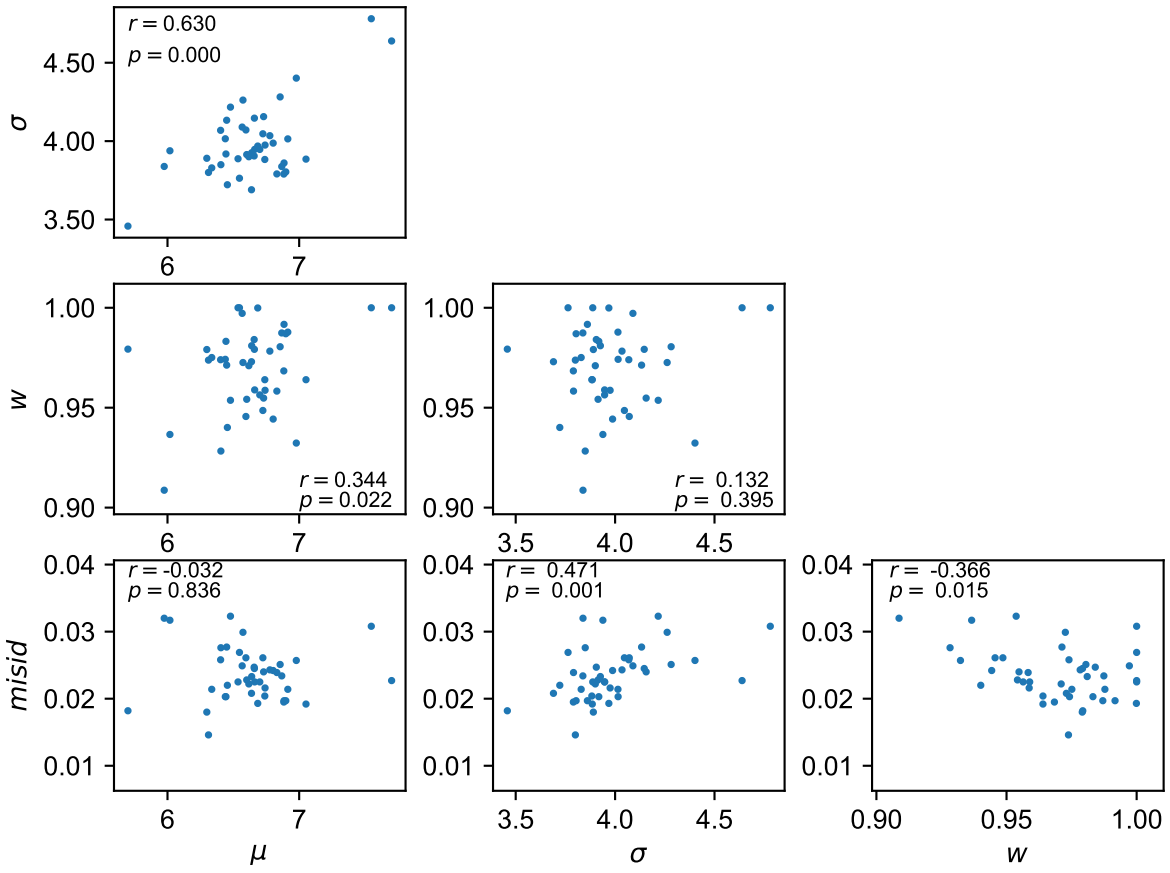

Figure S11: **Relationships among among fitted parameters to *D. melanogaster* GO terms.** Insets indicate Pearson correlations. The data plotted in these figures can be found in Table S10.

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

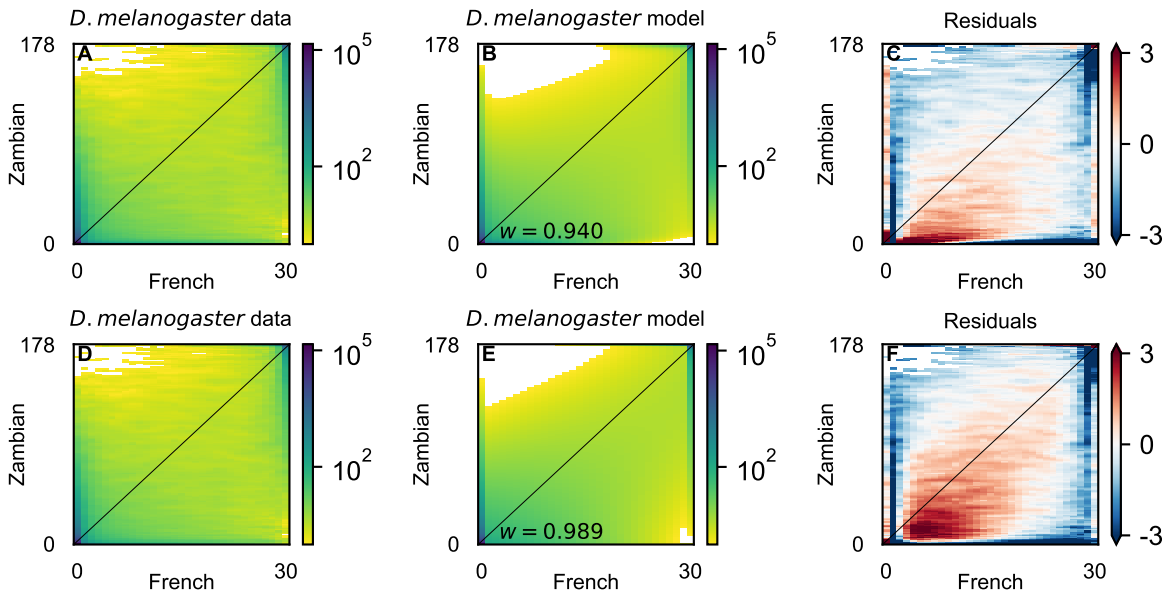

Figure S12: ***D. melanogaster* results assuming simpler demographic models. A, B & C**: A split_mig model (Fig. S2C) with symmetric migration fits the data less well than our full model with exponential growth and asymmetric migration. (Compare residuals with Fig. S6F.) **D, E & F**: A split_mig model without migration fits even worse.

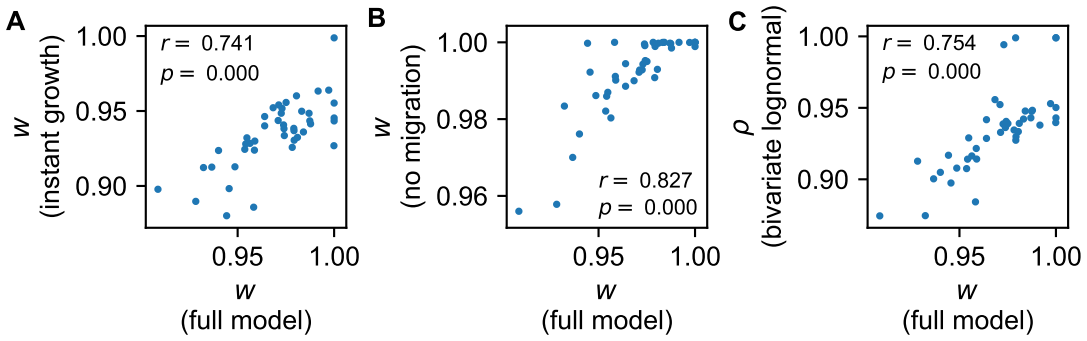

Figure S13: ***D. melanogaster* results assuming different models.** For the *D. melanogaster* Gene Ontology terms, inferences from the full demographic model and lognormal mixture DFE model are compared with **A**: A split_mig model with symmetric migration, **B**: A split_mig model withnot migration and **C**: A DFE model with a bivariate lognormal distribution. The data plotted in these figures can be found in Table S12.

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

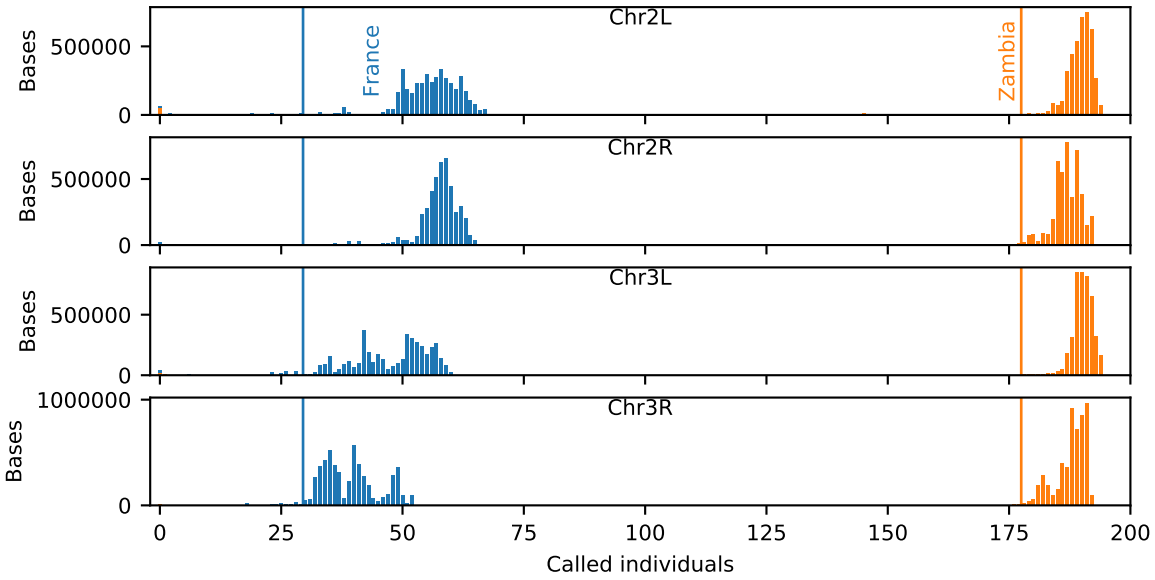

Figure S14: **Calling rate for different chromosome arms in analyzed *D. melanogaster* data.** For each arm, histograms indicate the number of bases at which a given number of individuals were called in each population. The vertical lines indicate the projection sizes for the AFS analysis.

bioRxiv preprint doi: https://doi.org/10.1101/703918; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

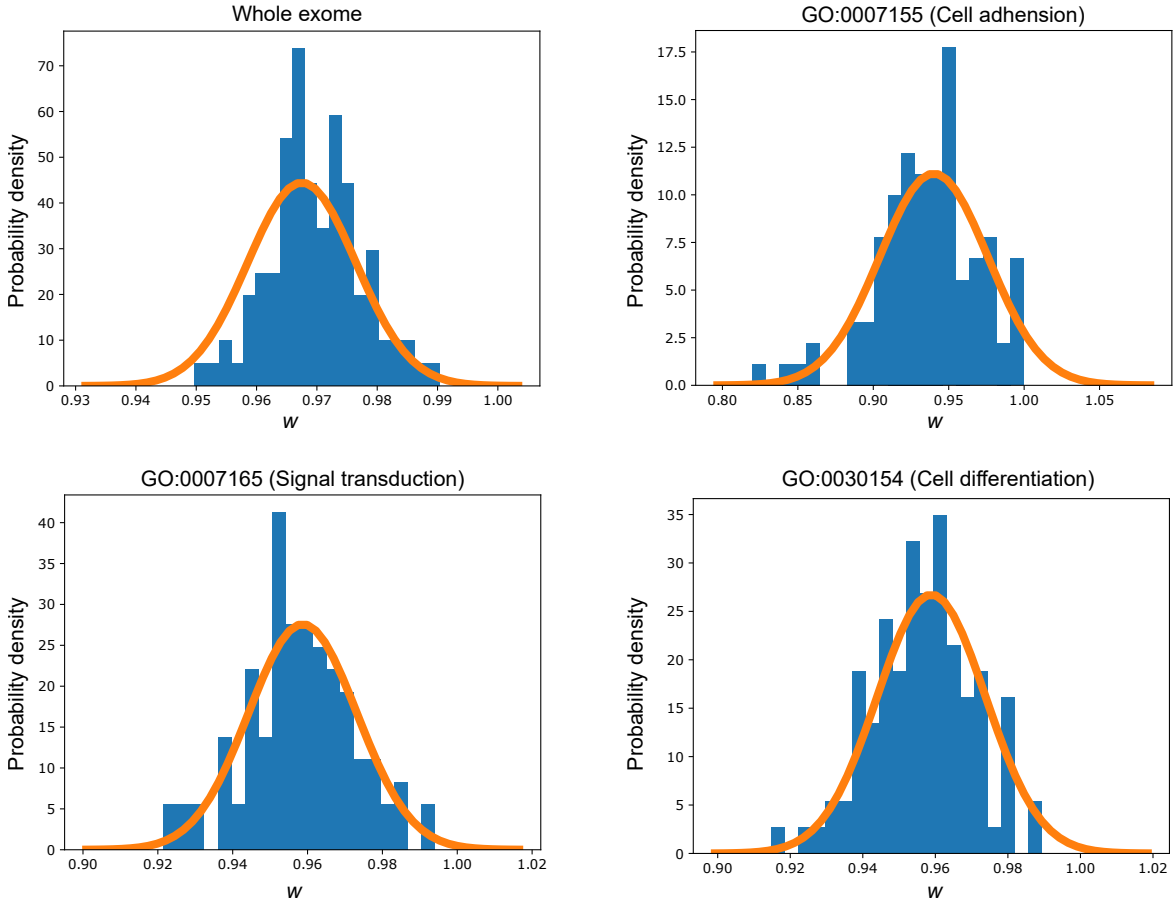

Figure S15: **Comparison between uncertainties estimated with the Godambe Information Matrix and bootstrap fitting.** For the *D. melanogaster* data, each panel shows a different subset of genes/mutations. In each panel, the histogram shows results from conventional bootstrap fitting, while the smooth curve is a normal distribution centered at the maximum likelihood inferred value and standard deviation estimated using the Godambe approach.