

Inference of the Distribution of Selection Coefficients for New Nonsynonymous Mutations Using Large Samples

Bernard Y. Kim,^{*} Christian D. Huber,^{*} and Kirk E. Lohmueller^{*,†,‡,1}

^{*}Department of Ecology and Evolutionary Biology, [†]Interdepartmental Program in Bioinformatics, and [‡]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California 90095

ABSTRACT The distribution of fitness effects (DFE) has considerable importance in population genetics. To date, estimates of the DFE come from studies using a small number of individuals. Thus, estimates of the proportion of moderately to strongly deleterious new mutations may be unreliable because such variants are unlikely to be segregating in the data. Additionally, the true functional form of the DFE is unknown, and estimates of the DFE differ significantly between studies. Here we present a flexible and computationally tractable method, called Fit $\partial a\partial i$, to estimate the DFE of new mutations using the site frequency spectrum from a large number of individuals. We apply our approach to the frequency spectrum of 1300 Europeans from the Exome Sequencing Project ESP6400 data set, 1298 Danes from the LuCamp data set, and 432 Europeans from the 1000 Genomes Project to estimate the DFE of deleterious nonsynonymous mutations. We infer significantly fewer (0.38–0.84 fold) strongly deleterious mutations with selection coefficient $|s| > 0.01$ and more (1.24–1.43 fold) weakly deleterious mutations with selection coefficient $|s| < 0.001$ compared to previous estimates. Furthermore, a DFE that is a mixture distribution of a point mass at neutrality plus a gamma distribution fits better than a gamma distribution in two of the three data sets. Our results suggest that nearly neutral forces play a larger role in human evolution than previously thought.

KEYWORDS deleterious mutations; diffusion theory; population genetics; site frequency spectrum

A fundamental concept in population genetics is the distribution of fitness effects (DFE) of new mutations. The DFE refers to the proportion of new mutations that have particular effects on fitness. The DFE is a crucial quantity in evolutionary genetics because it determines how selection affects genetic variation (Eyre-Walker and Keightley 2007), the conditions under which recombination could evolve (Keightley and Otto 2006), and the spectrum of mutations potentially involved in genetic diseases (Eyre-Walker 2010). Further, an accurate DFE is required for robust inference of the amount of adaptive evolution across taxa (Boyko *et al.*

2008; Gossmann *et al.* 2012; Castellano *et al.* 2016; Galtier 2016); a topic of widespread interest. Because this distribution is so important, considerable effort has been put forth toward estimating it in several species.

In organisms that allow experimental manipulation, the DFE can be directly estimated. Here, the DFE is either derived from direct measurements of fitness from a collection of single-step mutants, or indirectly inferred from observed changes in population fitness in mutation accumulation (MA) experiments (Eyre-Walker and Keightley 2007; Bataillon and Bailey 2014). The first approach, in combination with high-throughput methods, has been successfully applied to examine the full spectrum of (even weak) selection coefficients of mutations in small mutational target regions in a number of viral, bacterial, and yeast systems (Fowler *et al.* 2010; Hietpas *et al.* 2011; Boucher *et al.* 2014). They frequently report a gamma or unimodal, similarly shaped distribution of fitness effects (Bataillon and Bailey 2014), or a bimodal distribution with a second cluster of highly deleterious mutations (Acevedo *et al.* 2014; Bank *et al.* 2014;

Copyright © 2017 by the Genetics Society of America

doi: <https://doi.org/10.1534/genetics.116.197145>

Manuscript received October 23, 2016; accepted for publication February 14, 2017; published Early Online February 27, 2017.

Available freely online through the author-supported open access option.

Supplemental material is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.197145/-/DC1.

¹Corresponding author: Department of Ecology and Evolutionary Biology, University of California, Los Angeles, 621 Charles E. Young Dr. South, Los Angeles, CA 90095-1606. E-mail: klohmueller@ucla.edu

Boucher *et al.* 2014). The second approach infers the DFE from fitness trajectories of a collection of populations over time in MA experiments, without directly identifying the mutations involved. Assuming that the true DFE is gamma distributed, they estimate the parameters of a gamma DFE that best fit to the observed changes in the mean and variance of fitness among replicate populations (Halligan and Keightley 2009). These studies point to a shape of the DFE that is less leptokurtic than an exponential distribution, with the mode different from zero. This could indicate that the true underlying DFE is more complex than the gamma distribution (Halligan and Keightley 2009), or reflect a bias of MA-based methods toward mutations with large fitness effects (Eyre-Walker and Keightley 2007). In summary, experimental estimates of the DFE suggest that a substantial proportion of new mutations are strongly deleterious. However, due to the inherent limitations of these methods, inference of the exact functional form of the genome-wide DFE is challenging.

A second category of methods to infer the DFE involves examining patterns of neutral and putatively deleterious genetic variation in natural populations, and finding the model of demographic history and purifying selection that can match the observed level of variation. This framework has been applied to many species including humans (Eyre-Walker *et al.* 2006; Keightley and Eyre-Walker 2007; Boyko *et al.* 2008; Li *et al.* 2010), *Drosophila* (Keightley and Eyre-Walker 2007; Kousathanas and Keightley 2013), yeast (Koufopanou *et al.* 2015), orangutans (Ma *et al.* 2013), gorillas (McManus *et al.* 2015), and mice (Halligan *et al.* 2013). Many of these studies suggest that the DFE has a strongly leptokurtic distribution, conflicting with the MA-based estimates. Consistent with the bimodal DFE found by many site-directed mutagenesis studies (Bataillon and Bailey 2014; Boucher *et al.* 2014), they find a large proportion of nearly neutral mutations, as well as many strongly deleterious mutations. For example, previous studies in humans (Eyre-Walker *et al.* 2006; Boyko *et al.* 2008) have estimated the parameters of a gamma distribution for the DFE of new nonsynonymous mutations. These studies have found ~57–61% of new nonsynonymous mutations to be moderately to strongly deleterious ($|s| \geq 10^{-3}$), about 15–16% to be weakly deleterious ($10^{-4} \leq |s| < 10^{-3}$), and the remainder (24–28%) to be nearly neutral (Figure 1).

The estimates of the DFE from genetic variation data for humans by Eyre-Walker *et al.* (2006) and Boyko *et al.* (2008) have been widely used in human population genetic studies. For example, these DFEs were used to estimate differences in the genetic load across human populations (Henn *et al.* 2016), to model the ancient introgression of Neanderthal alleles into humans (Harris and Nielsen 2016), as a model for the frequency spectrum of deleterious polymorphisms in simulating data for disease studies (Uricchio *et al.* 2016), to evaluate the contribution of background selection to diversity on the Y chromosome in humans (Wilson Sayres *et al.* 2014), and to estimate the strength of selection acting on disease genes (Moon and Akey 2016). While the Boyko *et al.* (2008)

study has had considerable impact in the field, it is important to appreciate that the estimates of the DFE were made using a sample of a small number of individuals. As such, most of the variation segregating in those samples is likely to be neutral or nearly neutral. Inferences about the proportion of moderately and strongly deleterious mutations largely come from assuming the gamma distribution approximates the DFE of new mutations well, and then tabulating the proportion in those categories predicted by the gamma distribution. In other words, the second mode of strongly deleterious and lethal mutations observed by experimental studies is unlikely to be directly observed in polymorphism data sets, and these proportions are extrapolated from the long tail of the DFE.

This extrapolation of the proportion of strongly deleterious mutations may not be accurate. A more recent study using exome sequencing data from 200 Danish individuals (Li *et al.* 2010) estimated a DFE that differs considerably from that inferred in Boyko *et al.* (2008) or from the experimental estimates in lower organisms. Specifically, Li *et al.* (2010) found a mixture distribution consisting of a neutral point mass and gamma distribution fit best to their data (Figure 1). Additionally, they estimated that only 1% of new mutations have $|s| > 10^{-4}$ (compared to 57% in Boyko *et al.* 2008), and 78% of new mutations fall in the $10^{-4} \leq |s| < 10^{-3}$ range (compared to 15% in Boyko *et al.* 2008). Li *et al.* (2010) attributed this difference in the DFEs to their study considering a larger sample of individuals. As such, they surmised that they sampled more weakly deleterious variants, allowing more accurate inferences. However, this explanation has not been tested using simulations or larger data sets. Thus, the proportions of moderately vs. strongly deleterious nonsynonymous mutations in humans, as well as the functional form of the DFE, remain elusive.

Due to large-scale genome and exome sequencing projects, samples of hundreds to thousands of individuals are available (Tennessen *et al.* 2012; Fu *et al.* 2013; Lohmueller *et al.* 2013; The 1000 Genomes Project Consortium 2015). These large data sets should yield more reliable inferences of the DFE because moderately deleterious polymorphisms should be segregating, albeit at low frequency, in these samples (Supplemental Material, Figure S1 in File S1). As such, it should be possible to determine the functional form of the DFE and directly estimate the proportion of moderately and strongly deleterious mutations.

However, a major roadblock to using these new data sets for inference of the DFE is a lack of suitable software for inference from large samples. Generally, methods to infer the DFE summarize the allele frequency information of two classes of sites, one assumed to be neutral and the other selected by the site frequency spectrum (SFS). Then, they find the DFE that, under the inferred model of demography fit to the SFS from neutral sites, fits the observed SFS from selected sites. The method of Keightley and Eyre-Walker (2007), DFE-alpha, models demography using a Wright–Fisher transition matrix. It can only consider demographic models with one or

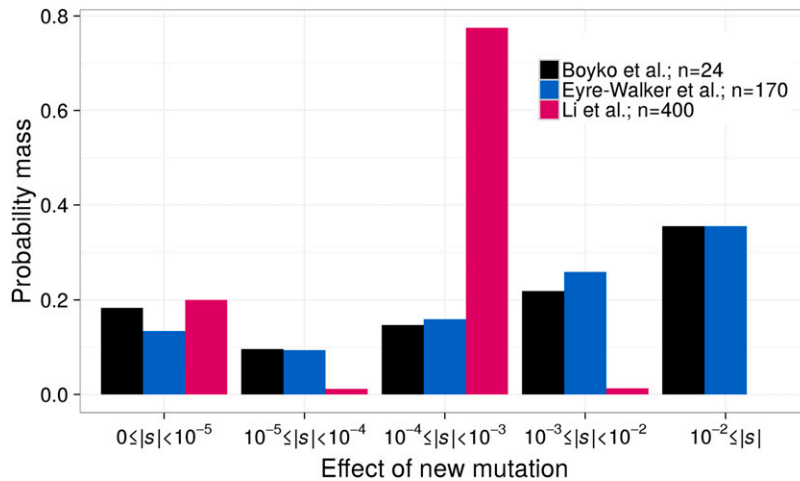


Figure 1 Previously inferred DFEs differ across studies. We rescaled the DFE in terms of the population size assumed or inferred in each study. A population size of 10,000 diploids is used to rescale the distribution of $2Ns$ to s for Eyre-Walker *et al.* (2006). For Boyko *et al.* (2008) and Li *et al.* (2010), we rescale the DFE from $2Ns$ to s using population sizes of 25,636 and 52,097 diploids, respectively (see *Materials and Methods*).

two size changes due to computational complexity, and it can be slow for the two-size-change model. This is particularly limiting in large samples of human genetic variation since a single-size-change demographic model is insufficient for capturing the excess of rare variation in human populations (Keightley and Eyre-Walker 2007; Kousathanas and Keightley 2013). Another class of methods to infer the DFE uses the Poisson random field (PRF) approach (Sawyer and Hartl 1992; Hartl *et al.* 1994; Williamson *et al.* 2005; Eyre-Walker *et al.* 2006; Boyko *et al.* 2008; Li *et al.* 2010). This approach has been implemented in the program PRFREQ (Boyko *et al.* 2008), but that implementation becomes numerically unstable when applied to samples larger than a few hundred individuals. The program *daði* (Gutenkunst *et al.* 2009) uses a similar framework, but implementing a DFE is slow due to the way that the DFE is repeatedly integrated (Figure S2 in File S1). Thus, there is a need for a new software tool to infer the DFE from large samples.

In this study, we first extend the program *daði* to analyze arbitrary DFEs in a computationally efficient manner. We implement these features in a module for *daði*, which we call *Fitdaði*. We then use this approach to estimate the DFE of deleterious, nonsynonymous mutations from multiple large human data sets. We consider several different functional forms for the DFE. We find that across the multiple data sets, a mixture distribution where a proportion of mutations are neutral and the remainder are gamma distributed fits best. Analysis of multiple data sets suggests there are fewer strongly deleterious mutations where $|s| > 10^{-2}$ (0.38–0.84 fold) than previously estimated in Boyko *et al.* (2008) (35%), regardless of the functional form of the DFE. Further, our results are not consistent with a model where 99% of new mutations have a selection coefficient weaker than 10^{-3} , as suggested by Li *et al.* (2010). Because we anticipate that our estimates of the DFE will be useful in subsequent simulation studies, we provide SFS_CODE (Hernandez 2008) and SLiM (Messer 2013) commands to simulate data where mutations have fitness effects from these DFEs.

Materials and Methods

Fitdaði: software to infer the DFE

Here we present our new software, *Fitdaði*, to infer distributions of selection coefficients of new mutations under the PRF model using the SFS. *Fitdaði* is a module that extends the functionality of the Python package, *daði* (Gutenkunst *et al.* 2009). Specifically, *daði* uses diffusion theory to compute the expected SFS for a set of demographic parameters and selection coefficients. *Fitdaði* offers a substantial computational improvement over the existing implementation of *daði* for models involving more than a single selection coefficient. To do this, *Fitdaði* computes SFSs for a range of selection coefficients and saves each SFS into an array. Then, subsequent integrations of the DFE are done using the array of precomputed SFSs. This process results in a substantial improvement in computational speed compared to the existing implementation of *daði*, which recomputes the SFS for each selection coefficient in each step of the optimization process (Figure S2 in File S1). *Fitdaði* also allows parallel computation of the SFS by using multiple cores or a cluster. Importantly, *Fitdaði* leverages the modular nature of *daði* to allow the user to define any arbitrary demographic model and DFE, including DFEs that are complex mixture distributions. Lastly, we incorporated an optimization routine that allows for constrained optimization of complex mixture distributions. Below we describe our inference procedure in greater detail, starting with inference of demography, followed by the details of *Fitdaði*. We then discuss a simulation study to assess its performance, both under the assumptions of the PRF model as well as when some are violated, and then the data sets that we use to infer the DFE in humans.

Inference using the SFS

We inferred demography and selection from segregating sites in a maximum likelihood framework (Williamson *et al.* 2005; Boyko *et al.* 2008; Gutenkunst *et al.* 2009). Because both demography and selection affect patterns of deleterious mutations, our inference of the DFE begins (as done in Williamson *et al.* 2005 and Boyko *et al.* 2008) by first

estimating a demographic model for putatively neutral, synonymous sites. Then, conditional on the demographic parameter estimates, we infer the parameters for the DFE of nonsynonymous mutations.

To do this, we summarized synonymous and nonsynonymous sites with the SFS. The SFS can be described as a vector, $\mathbf{X} = [X_1, X_2, \dots, X_{n-1}]$, in which each entry X_i describes the number of SNPs at frequency i in a data set of size n chromosomes. In the PRF framework, each entry in the SFS is assumed to be comprised of independent sites (Sawyer and Hartl 1992; Hartl *et al.* 1994).

Additionally, we folded the SFS to avoid difficulties with misidentification of the ancestral state (Hernandez *et al.* 2007). This form of the SFS counts the number of SNPs of minor allele frequencies (MAFs) 1 to $n/2$ without taking the ancestral state into account. The folded SFS has been shown to perform well at inferring the DFE of deleterious mutations (Keightley and Eyre-Walker 2007; Boyko *et al.* 2008; Tataru *et al.* 2016).

Inference of demography

A demographic model, the parameters of which are denoted as Θ_D , was fit to the SFS of synonymous sites with $\partial a\partial i$ (Gutenkunst *et al.* 2009). Here, $\partial a\partial i$ uses a diffusion approximation to compute the distribution of allele frequencies given some demographic model, $f(x; \Theta_D)$. Then, the expected number of SNPs at frequency i in a sample of size n chromosomes can be written as:

$$E[X_i] = \frac{\theta}{2} \int_0^1 x^i (1-x)^{n-i} f(x; \Theta_D) dx. \quad (1)$$

The multinomial likelihood, computed with the folded SFS, is maximized to estimate the demographic parameters:

$$\mathcal{L}(\mathbf{X}^* | \Theta_D) = \prod_i \left(\frac{E[X_i^* | \Theta_D]}{\sum_i E[X_i^* | \Theta_D]} \right)^{X_i^*}, \quad (2)$$

where X_i^* denotes the observed count of SNPs at frequency i in the folded SFS. The multinomial likelihood uses the proportions of SNPs at a particular frequency in the sample rather than the counts from the model. Therefore, it does not require an *a priori* assumption about the mutation rate or ancestral population size. The mutation rate of synonymous sites, denoted θ_S , was then computed as the scaling factor difference between the optimized SFS and the data.

When fitting models incorporating periods of rapid exponential growth with $\partial a\partial i$, we set the program parameter *dadi.Inference.set_timescale_factor* = 10^{-6} , in contrast to the default setting of 10^{-3} . In $\partial a\partial i$, periods of exponential growth are approximated with a series of instantaneous size changes and, if the time steps are not small enough, parameters related to exponential growth will not be inferred correctly. This causes the demographic model to inaccurately predict the expected numbers of rare variants, biasing downstream inference of selection.

Inference of selection

To infer the DFE, we developed the Fit $\partial a\partial i$ module, which uses $\partial a\partial i$ and some of the methodological improvements of Ragsdale *et al.* (2016). First, we condition on the demographic model that was fit to synonymous sites using the procedure described above. Given that demography, $\partial a\partial i$ is used to compute a distribution of allele frequencies $f(x; \Theta_D, \gamma)$, where Θ_D is a vector containing the demographic parameters inferred from synonymous sites and γ is a single population-scaled selection coefficient. Specifically, $\gamma = 2N_A s$, where N_A is the ancestral population size, but it is rescaled in each time period of the demographic model by the fold size change relative to the ancestral population. A DFE, denoted $g(\gamma)$, can be incorporated by generating $f(x; \Theta_D, \gamma)$ for a range of γ , then weighting the contribution of each of these frequency spectra by $g(\gamma)$ (Boyko *et al.* 2008):

$$E[X_i] = \frac{\theta}{2} \int_{-\infty}^{\infty} \int_0^1 x^i (1-x)^{n-i} f(x; \Theta_D, \gamma) g(\gamma | \Theta_{DFE}) dx d\gamma. \quad (3)$$

In the standard implementation of $\partial a\partial i$, this process is time consuming because the SFS must be computed repeatedly during each step of optimization. In other words, $f(x; \Theta_D, \gamma)$ is computed each time a given value of γ is evaluated in a DFE. This process can be especially slow for large ranges of γ and for large sample sizes. Therefore, similar to Ragsdale *et al.* (2016), we initially computed the SFS for a range of selection coefficients, and then cached these results to avoid recomputing the frequency spectra (Figure S2 in File S1). In addition, we computed many frequency spectra in parallel to save time, added compatibility for user-defined DFEs, modified the optimization routines available in $\partial a\partial i$ to work with cached spectra, and added the option to use constrained optimization for the inference of complex mixture distributions. These extensions are part of the Fit $\partial a\partial i$ module.

To infer selection, we fixed the demographic parameters to the maximum-likelihood estimates (MLEs) inferred from synonymous sites, $\hat{\Theta}_D$. Then, we fit a DFE, the parameters of which are denoted as Θ_{DFE} , to the folded SFS of nonsynonymous sites by maximizing the Poisson likelihood:

$$\mathcal{L}(\mathbf{X}^* | \hat{\Theta}_D, \Theta_{DFE}) = \prod_i \exp \left(- E[X_i^* | \hat{\Theta}_D, \Theta_{DFE}] \right) \times \frac{E[X_i^* | \hat{\Theta}_D, \Theta_{DFE}]}{X_i^*!}. \quad (4)$$

Unlike the multinomial likelihood, the Poisson likelihood requires an *a priori* assumption about the mutation rate for nonsynonymous sites, θ_{NS} . To obtain this, we multiplied our estimate of θ_S by an assumed ratio of the nonsynonymous to synonymous mutation rate, θ_{NS}/θ_S , to obtain θ_{NS} . Specifically, we assumed the ratio to be $\theta_{NS}/\theta_S = 2.31$ (Huber *et al.* 2016), but also estimated the DFE using $\theta_{NS}/\theta_S = 2.5$ to provide a fair comparison to Boyko *et al.* (2008).

Each DFE is defined as an integrable function over a log-spaced range of 600 selection coefficients over intervals between $|s| = [10^{-8}, 0.5]$. We considered any portion of the DFE smaller than $|s| = 10^{-8}$ to be effectively neutral ($|s| = 0$), and any variants of $|s| > 0.5$ to have negligible probability of being found in polymorphism data (i.e., not found in the data). Note that here we only consider the deleterious DFE but this function can easily be extended to incorporate positive selection (Huber *et al.* 2016).

Fit*dadi* includes many of the standard DFEs (Boyko *et al.* 2008; Kousathanas and Keightley 2013), such as a gamma distribution and several mixture distributions. Specifically, we investigated mixture distributions where a proportion of mutations are neutral; with the rest following a gamma distribution as well as a mixture distribution where a fraction is neutral, a fraction is lethal, and the remainder follows an exponential distribution of fitness effects. Lastly, Fit*dadi* includes arbitrary mixture distributions with a fixed number of fitness classes, or bins, where each bin can have its own range of selection coefficients (called the “discrete DFE”). Fit*dadi* infers the proportion of new mutations in each fitness class. For mixture distributions incorporating a point mass at neutrality or lethality, we define the DFE so it can be treated as a single integrable function. We add the area of the point mass to a part of the distribution that is assumed to be neutral or lethal. For example, to add a point mass of neutral mutations to the “neutral+gamma DFE,” we add the probability mass of neutral mutations, p_{neu} , to the $|s| = [0, 10^{-8}]$ portion of the distribution. Then, we integrate the gamma DFE between $|s| = [0, 10^{-8}]$ and sum it with p_{neu} to obtain the total mass of neutral mutations. Additionally, we used the SLSQP algorithm (Kraft 1988) as implemented in SciPy 0.17.1 to perform constrained optimization for mixture distributions incorporating more than two components. Throughout, we will use α and β to denote the shape and scale parameters of the gamma distribution, respectively, and λ to denote the rate parameter of the exponential distribution. The DFEs we report will be scaled by the ancestral population size. To estimate confidence intervals (C.I.’s) for our data, we Poisson resampled the nonsynonymous SFS and refit the DFE to the resampled data (Boyko *et al.* 2008). We note these C.I.’s are likely too narrow because they assume independence between all sites and do not account for the uncertainty in the demographic inference.

Simulations

To assess the performance of Fit*dadi*, we performed forward-in-time simulations under different models of selection and demography. Simulations of independent sites were done using the program PReFerSIM (Ortega-Del Vecchyo *et al.* 2016), which simulates unlinked SNPs under the PRF model. We simulated synonymous sites separately with a population-scaled mutation rate of $\theta_S = 4000$ to approximately match the amount of synonymous genetic diversity in our data sets. We simulated nonsynonymous sites at a ratio of 2.5 nonsynonymous

to 1 synonymous site, in other words using $L_{\text{NS}}/L_S = 2.5$. These simulations included sample sizes of $n = 24$ and $n = 2596$ chromosomes using a demographic model of constant size, a twofold instantaneous size change, and the demography inferred from the LuCamp data. We considered a variety of DFEs, which are described in more detail in specific instances below.

Because the PRF model makes several restrictive assumptions that are likely to not apply to real data sets, we performed an additional set of forward simulations violating these assumptions, and assessed the effect on inferences using Fit*dadi*. Specifically, we investigated the effects that unmodeled linkage, background selection, and population structure might have on our inference of the DFE. To do this, we simulated 100–150 Mb regions using the recombination rate and arrangement of functional elements on human chromosome 1 (Huber *et al.* 2016) using the forward simulation program SLiM (Messer 2013). We assumed a gamma DFE for both nonsynonymous ($\alpha = 0.2$, $\beta = 200$) and conserved noncoding sites ($\alpha = 0.0415$, $\beta = 50$; see Torgerson *et al.* 2009). We assume that 400 generations ago, the ancestral population expanded eightfold and split into eight genetically isolated populations. This population size increase reflects the Neolithic expansion into Europe under the demic diffusion model (Chikhi *et al.* 2002; Gazave *et al.* 2014). We then sampled 100 chromosomes equally across the eight populations, combined them together, and analyzed them as though they were from a single population. The ancestral population was simulated for a burn-in period of $10N$ generations. To avoid prohibitively slow forward simulations, we simulated with an ancestral effective population size of 1000 and scaled mutation rate, recombination rate, selection coefficients, and demographic parameters accordingly (Aberer and Stamatakis 2013). We then fit a single population demographic model (which is the incorrect model) to the synonymous SNPs in each simulated data set using *dadi*. Then, conditional on these demographic parameters, we inferred the DFE using Fit*dadi*. Our goal with these simulations is to mimic what researchers do in practice; where they do not know the true demographic model, but try to fit a simplified model to the data.

Data

We downloaded SNP data for 432 unrelated European (EUR) individuals from the 1000 Genomes Project phase 3 release (The 1000 Genomes Project Consortium 2015); 6503 individuals from the National Heart, Lung, and Blood Institute ESP6500SI-V2 European American (EUR) data set (Tennessen *et al.* 2012; Fu *et al.* 2013); and 2000 Danish individuals from the LuCamp project (Lohmueller *et al.* 2013). The 1000 Genomes phase 3 data were downloaded from the European Bioinformatics Institute file transfer protocol site <http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>. Related individuals were removed by sampling only mothers and fathers from trios or extended families. Only SNPs from the phase 3, exome-targeted sequencing which passed the strict

mask filter were used. The total length of sites considered in the analysis, $L_S + L_{NS}$, was computed by taking this filtering into account. Variants were annotated using the 1000 Genomes Project-filtered annotations. The Exome Sequencing Project (ESP) ESP6400 data set was downloaded from the Exome Variant Server (<http://evs.gs.washington.edu/EVS/>). Only sites with 1800 or more European individuals sequenced, according to the site-by-site annotations, were used for the analysis. The LuCamp data were obtained from Lohmueller *et al.* (2013). For computational tractability, a hypergeometric distribution was used to project the LuCamp and ESP data sets down to sample sizes of 1298 and 1300 diploids (Gutenkunst *et al.* 2009), respectively, after filtering problematic individuals. All 432 unrelated European individuals from the 1000 Genomes Project were used. From these data, we assembled the folded SFS of synonymous and nonsynonymous sites, which were used for subsequent inference. To examine the effect of a smaller sample size on inference of demography and selection, we also projected the data down to a sample size of 24 chromosomes.

Estimating s from 2Ns

The DFEs inferred using the approach described above were for the population-scaled selection coefficient, scaled by twice the ancestral population size ($\gamma = 2N_A s$). Because we were interested in the distribution of s , we needed to estimate N_A . We computed N_A from the value of θ_S inferred from synonymous sites (Table S1 in File S1) using the equation $\theta_S = 4N_A \mu L_S$. Detailed information about these parameters used for our analysis can be found in Table S2 in File S1. However, this value of N_A depends on assumptions about the per-base-pair mutation rate and the ratio of possible nonsynonymous to synonymous sites, L_{NS}/L_S , since these quantities are computed from the total number of coding sites, $L_S + L_{NS}$. We assumed the mutation rate to be $\mu = 1.5 \times 10^{-8}$ to reflect estimates of the mutation rate in the human exome (Ségurel *et al.* 2014). For comparison to results from Boyko *et al.* (2008), we assumed the mutation rate to be $\mu = 1.8 \times 10^{-8}$. For the re-analysis of the Boyko *et al.* (2008) data set, we assumed the same ancestral population size, $N = 7778$ diploids, instead of computing it. To compute the total number of coding sites, $L_S + L_{NS}$, in each data set, we intersected the coding exons from the University of California Santa Cruz canonical transcript with the relevant filters for each data set. For the 1000 Genomes data, we intersected the phase 3 strict mask, the exome targets, and the hg19 coding exons. For the analysis of the ESP data set, we used the intersection of the hg19 coding exons and the site-by-site annotations to count the total number of coding sites for which $n \geq 2600$ alleles had been sequenced. For the LuCamp data, we obtained the value of $L_S + L_{NS}$ from Lohmueller *et al.* (2013).

Data availability

This research uses previously published data sets obtained as previously described. The Fitdadi software is available at <https://github.com/LohmuellerLab/fitdadi>.

Results

Validation of Fitdadi by comparison to previous analyses

We first examined the performance of Fitdadi by fitting a demographic model and DFE to the African-American SFS from Boyko *et al.* (2008). Fitdadi produces similar estimates of the shape and scale parameters of the gamma distribution compared to Boyko *et al.* (2008) (Boyko: $\alpha = 0.184$, $\beta = 2488$; Fitdadi: $\alpha = 0.179$, $\beta = 3161$). Additionally, Fitdadi produced similar estimates of the proportions of mutations in different bins of the DFE (Table S3 in File S1).

Performance on simulated data

We further investigated the performance of Fitdadi by performing forward simulations under the PRF model (see *Materials and Methods*). We first considered the best-fit DFE of Boyko *et al.* (2008), rescaled to have an ancestral population size of $N = 10,085$ diploids ($\alpha = 0.184$, $\beta = 3226$). Fitdadi is able to accurately infer the DFE from our simulated data sets (Table 1). Predictably, the variance of our estimates of the most deleterious portion of the DFE ($|s| > 10^{-2}$) is five- to sixfold greater for the small ($n = 24$) samples. However, for the samples of size $n = 2596$, the variance in the estimates of this portion of the DFE is significantly reduced and overall estimates of the proportions of the DFE where $|s| > 10^{-3}$ are accurate. Therefore, this sample size should allow us to accurately infer the most deleterious portions of the DFE.

Because it is not certain that the DFE is truly gamma distributed, we simulated data sets of 2596 chromosomes with other DFEs. Again, we scaled these DFEs to an ancestral population size of 10,085 diploids. We considered the mixture distribution of Li *et al.* (2010), which consists of 20% neutral and 80% gamma-distributed ($\alpha = 4$, $\beta = 2.148$) selection coefficients (the neutral+gamma DFE). We also considered a mixture distribution consisting of five bins, (the discrete DFE) with breaks at $|s| = [0, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 1]$. Within each bin, the values of s were uniformly distributed. We examined three weighting schemes for this distribution. First, we computed the probability mass in each bin from the mixture distribution of Li *et al.* (2010). Then, we computed the probability mass in each bin from a gamma distribution with parameters $\alpha = 0.203$ and $\beta = 1082.1$, but where the mass in the $|s| = [10^{-2}, 1]$ bin was placed into the $|s| = [10^{-3}, 10^{-2}]$ bin, and the opposite case where the mass in the $|s| = [10^{-3}, 10^{-2}]$ bin was placed into the $|s| = [10^{-2}, 1]$ bin. This was done to evaluate whether we could distinguish between these discrete DFEs and to evaluate our ability to distinguish strongly deleterious mutations from moderately deleterious mutations in a large sample.

We find that if the true underlying DFE is distributed according to the Li *et al.* (2010) neutral+gamma distribution, the discrete DFE is able to estimate the true DFE, albeit with some limitations (Figure 2, A and B). For example, when the true DFE follows Li *et al.* (2010), our inference under the

Table 1 Performance of Fit $\partial a\partial i$ on simulated data sets

Demography	n_{chr}	α (shape)	β (scale)	$0 \leq s < 10^{-5}$	$10^{-5} \leq s < 10^{-4}$	$10^{-4} \leq s < 10^{-3}$	$10^{-3} \leq s < 10^{-2}$	$10^{-2} \leq s $
True ^a	—	0.184	3266 ^b	0.182	0.096	0.146	0.219	0.357
Constant size	2596	0.180 \pm 0.010	3712.2 \pm 980.2	0.186 \pm 0.009	0.095 \pm 0.002	0.144 \pm 0.006	0.213 \pm 0.013	0.363 \pm 0.016
	24	0.185 \pm 0.028	3613.1 \pm 4196.7	0.182 \pm 0.017	0.097 \pm 0.009	0.148 \pm 0.023	0.221 \pm 0.043	0.353 \pm 0.060
Two-fold expansion	2596	0.191 \pm 0.007	2606.0 \pm 410.7	0.178 \pm 0.008	0.098 \pm 0.002	0.152 \pm 0.004	0.230 \pm 0.009	0.341 \pm 0.010
	24	0.187 \pm 0.023	3259.8 \pm 2612.6	0.181 \pm 0.016	0.097 \pm 0.008	0.149 \pm 0.019	0.223 \pm 0.036	0.350 \pm 0.050
LuCamp	2596	0.182 \pm 0.008	3411.9 \pm 558.5	0.184 \pm 0.008	0.096 \pm 0.001	0.145 \pm 0.004	0.216 \pm 0.008	0.358 \pm 0.009
	24	0.186 \pm 0.027	3435.4 \pm 3249.9	0.182 \pm 0.017	0.097 \pm 0.009	0.148 \pm 0.022	0.222 \pm 0.042	0.351 \pm 0.060

95% intervals were estimated as ± 1.96 SD of 100 replicates in each simulation set. chr, chromosome.

^a Values show the true DFE used to simulate the data.

^b Here the simulation was scaled to an ancestral population size of $N = 10,085$ diploids.

discrete DFE correctly estimates a negligible fraction of moderately or strongly deleterious new mutations ($|s| > 10^{-3}$), and correctly infers a mode of weakly deleterious mutations ($10^{-4} \leq |s| < 10^{-3}$). However, estimates of the proportion of nearly neutral and neutral mutations ($|s| < 10^{-4}$) are less accurate (Figure 2A). When we simulate with the discretized distribution of Li *et al.* (2010), our estimates of the proportions of the discrete DFE are unbiased (Figure 2B). Additionally, we can distinguish between DFEs with varying proportions of moderately and strongly deleterious mutations (Figure 2, C and D). Although it is unlikely that the DFE of any natural population is discretized in this manner, these results show that the discrete DFE can help to approximate the general form of the underlying DFE, even if the true DFE is multimodal. This mimics what would be done in practice, where the precise functional form of the DFE is not known *a priori*. Therefore, fitting the discrete DFE should provide a general idea of the true DFE, especially if the true DFE is significantly multimodal. Notably, the discrete distribution can distinguish between strongly and moderately deleterious mutations at our sample size of 2596 chromosomes.

Simulations with linkage and population structure

The procedure of first inferring demography from the synonymous SFS and then selection from the nonsynonymous SFS provides unbiased estimates of selection, even in the presence of linkage (Boyko *et al.* 2008; Messer and Petrov 2013; Comeron 2014). In other words, this methodology controls for the effects of selection at linked sites. However, it is unclear what effect population structure might have on inference of the DFE. It is well known that such cryptic structure affects the SFS and may bias demographic inference (Ptak and Przeworski 2002; Gazave *et al.* 2014). Further, large human resequencing data sets may contain cryptic population structure (Novembre and Ramachandran 2011). For example, the 1000 Genomes European sample is comprised of five different subpopulations. To examine the performance of Fit $\partial a\partial i$ when applied to data sets where the assumptions of the PRF model and a single, unstructured population are violated, we performed forward simulations including background selection and population structure (see *Materials and Methods*). We fit a single population, single-size-change demographic model to synonymous sites; and then, conditional

on the size-change demographic model, a gamma DFE to nonsynonymous sites for each simulation replicate. Even when using the incorrect demographic model, we accurately infer selection from simulated data in the presence of linkage and population structure (Figure 3). Importantly, the single-size-change demographic model provides a reasonable approximation to the SFS when there are both population expansions and structure (Figure 3A). This in turn allows for the accurate estimation of both the shape and scale parameters of the gamma DFE (Figure 3B).

Therefore, simulations and a comparison to existing empirical data suggest that Fit $\partial a\partial i$ can reliably infer the DFE in the presence of complex demographic scenarios. Below we present additional simulation scenarios to examine the performance of Fit $\partial a\partial i$ with varying sample sizes and when the assumed demography and DFE are misspecified.

Demographic inference

We begin by fitting a demographic model to the synonymous SFS of each of the three data sets (LuCamp, ESP, and 1000 Genomes) using $\partial a\partial i$. Briefly, this demographic model incorporates an out-of-Africa bottleneck, a recovery period, and recent exponential population growth (Figures S3 and S4 in File S1). Our estimates of demography as well as the inferred population sizes are presented in Tables S1 and S2 in File S1. Predictably, the parameter describing the magnitude of recent population expansion is harder to infer when using a sample size of 24 chromosomes than when using the larger sample sizes ($n = 2596$ chromosomes). Although the demographic model we infer is biased by linked selection, this step controls for these effects when inferring selection (Boyko *et al.* 2008; Kousathanas and Keightley 2013; Messer and Petrov 2013; Huber *et al.* 2016).

Inference of the DFE from large data sets

Here we estimate the DFE for new nonsynonymous mutations using large samples. First, like previous studies, we fit a gamma distribution to the DFE (Table S4 in File S1). We infer a strongly leptokurtic distribution where there are many neutral and nearly neutral mutations (*i.e.*, 34–37% of new mutations have $|s| < 10^{-4}$), as well as a class of strongly deleterious mutations (*i.e.*, 15–22% of new mutations have $|s| > 10^{-2}$). Interestingly, the estimates from the three

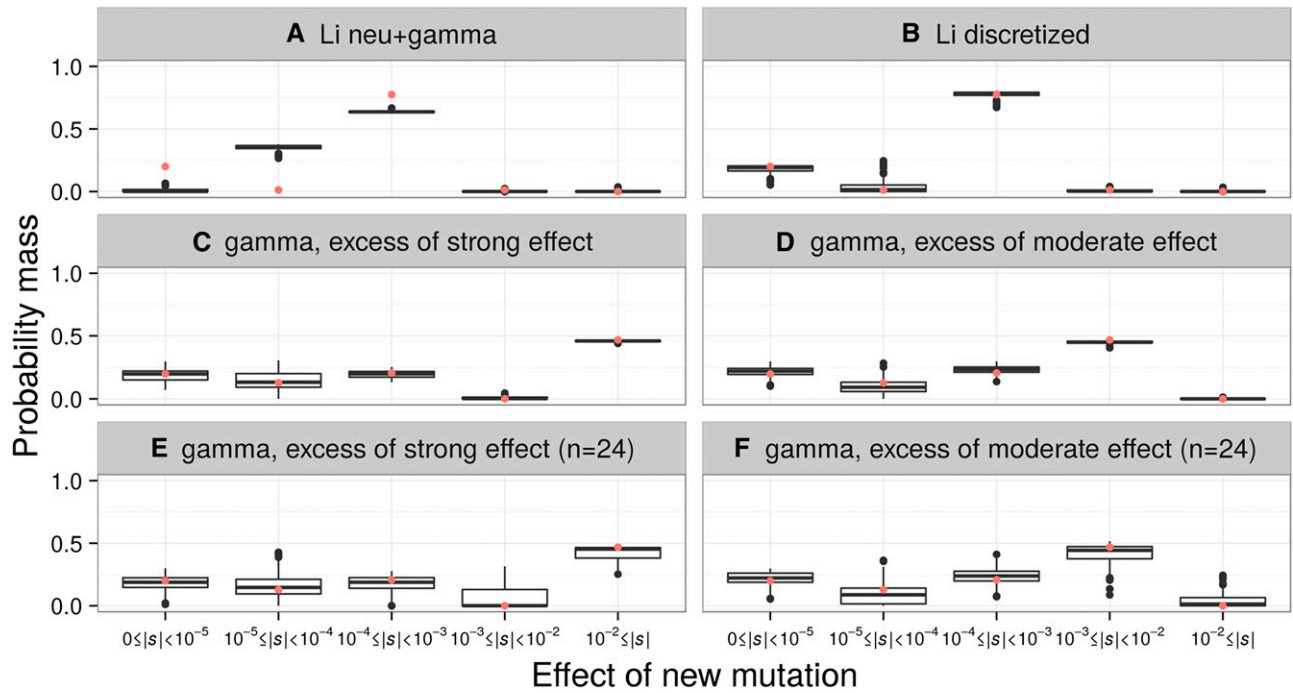


Figure 2 The discrete DFE can recover the approximate form of the DFE from simulated data. The distributions of the proportions of mutations with different selective effects, as inferred by the discrete DFE for 100 simulated data sets, are shown. Each simulation set assumed the demographic model fit to the LuCamp synonymous SFS. A red point depicts the true proportions of the simulated DFE. The true DFE for each set is: (A) the continuous neutral+gamma distribution of Li *et al.* (2010) ($p_{\text{neu}} = 0.2$, $\alpha = 4$, $\beta = 1.065 \times 10^{-4}$), (B) the discretized version of that distribution, (C–F) a gamma DFE ($\alpha = 0.215$, $\beta = 567.1$), but where (C and E) the mass of the $10^{-3} \leq |s| < 10^{-2}$ bin was added to the $10^{-2} \leq |s|$ bin, and (D and F) where the mass of the $10^{-2} \leq |s|$ bin was added to the $10^{-3} \leq |s| < 10^{-2}$ bin. The data sets simulated for (C) and (D) had sample sizes of $n = 2596$ chromosomes, while the data sets for (E) and (F) had sample sizes of $n = 24$ chromosomes.

different data sets are generally concordant, though the 95% C.I.'s sometimes do not overlap. While this may suggest that the differences cannot be attributable to limited amounts of data in the SFS, we caution that these C.I.'s are likely too narrow because they do not account for the nonindependence of SNPs or the uncertainty of demographic estimates.

When compared directly to Boyko *et al.* (2008), the best-fit gamma DFEs inferred from all three data sets are generally shifted toward neutrality (Table 2 and Tables S4 and S5 in File S1), even when matching the mutation rates to those of Boyko *et al.* (2008) ($\mu = 1.8 \times 10^{-8}$ and $L_{\text{NS}}/L_{\text{S}} = 2.5$). We infer 19.2–22.9% of new mutations have a selection coefficient $|s| < 10^{-5}$, compared to the 18.3% observed by Boyko

et al. (2008). This corresponds to a 1.05- to 1.25-fold increase. Additionally, we infer 24.5–29.8% of new mutations are strongly deleterious ($|s| > 10^{-2}$), which corresponds to a 0.69- to 0.84-fold decrease of the 35.5% inferred by Boyko *et al.* (2008). Taken together, when assuming a gamma distribution for the DFE, all three data sets suggest fewer strongly deleterious mutations than seen in Boyko *et al.* (2008).

Next, we explored the fit of complex DFEs to these large samples. Using the same combination of mutation rates as with the gamma, we fit the neutral+gamma mixture distribution; a mixture distribution of a point mass of neutral, a point mass of lethal, and exponentially distributed new

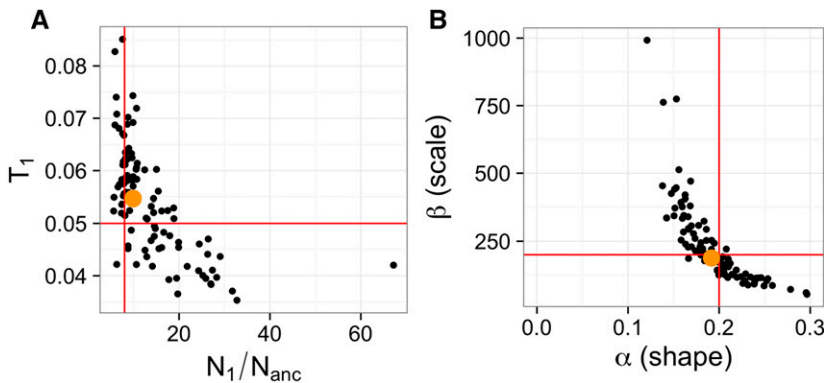


Figure 3 Inference of the DFE is robust to misspecification of the demographic model and background selection. Points show the MLEs of the (A) demographic parameters and (B) DFE parameters inferred from 100 simulated data sets with linkage and population structure. Red lines denote the true values and the yellow dots denote the median estimates across the 100 data sets. Estimates of time of expansion (T_1) and the ratio of current to ancestral population size (N_1/N_{anc}) tend to be biased because demography is incorrectly modeled due to background selection, but estimates of the DFE are unbiased.

Table 2 MLEs of various DFEs

Data set	DFE	Parameter MLEs	Log-likelihood	AIC	ΔAIC^a
LuCamp	Gamma	$\alpha = 0.215, \beta = 562.1$	-3334.7	6673.4	13
	Neu+gamma	$p_{\text{neu}} = 0.164, \alpha = 0.338, \beta = 367.7$	-3327.2	6660.4	0
	Neu+exp+let	$p_{\text{neu}} = 0.304, p_{\text{exp}} = 0.613, \lambda = 66.56$	-3337.8	6681.6	21.2
	Discrete ^b	$p_1 = 0.309, p_2 = 0.024, p_3 = 0.247, p_4 = 0.372$	-3334.1	6676.2	15.8
1kG EUR	Gamma	$\alpha = 0.186, \beta = 875.0$	-1450.5	2905.0	0
	Neu+gamma	$p_{\text{neu}} = 0.031, \alpha = 0.199, \beta = 820.6$	-1450.8	2907.6	2.6
	Neu+exp+let	$p_{\text{neu}} = 0.312, p_{\text{exp}} = 0.509, \lambda = 41.48$	-1472.0	2950.0	45
	Discrete ^b	$p_1 = 0.286, p_2 = 0.099, p_3 = 0.222, p_4 = 0.305$	-1453.4	2914.8	9.8
ESP EUR	Gamma	$\alpha = 0.169, \beta = 1327.4$	-3012.6	6029.2	2.6
	Neu+gamma	$p_{\text{neu}} = 0.092, \alpha = 0.207, \beta = 1082.3$	-3010.3	6026.6	0
	Neu+exp+let	$p_{\text{neu}} = 0.341, p_{\text{exp}} = 0.504, \lambda = 63.90$	-3071.6	6149.2	122.6
	Discrete ^b	$p_1 = 0.334, p_2 = 0.041, p_3 = 0.201, p_4 = 0.306$	-3029.5	6067.0	40.4

These results are reported assuming $L_{\text{NS}}/L_{\text{S}} = 2.31$ and $\mu = 1.5 \times 10^{-8}$. See Table S4 in File S1 for additional information. The shape and scale parameters of the gamma distribution are denoted with α and β , respectively, and the rate parameter of the exponential distribution is denoted with λ . Neu, neutral; exp, exponential; let, lethal; 1kG, 1000 Genomes.

^a Change in AIC relative to the model with the lowest AIC.

^b In terms of $|s|$, these parameters correspond to the ranges of $|s|$ corresponding to: $[0, 10^{-5})$, $[10^{-5}, 10^{-4})$, $[10^{-4}, 10^{-3})$, and $[10^{-3}, 10^{-2})$, respectively. The mass in the $[10^{-2}, 1]$ range is the complement of the total mass of the four aforementioned categories.

mutations (the “neutral+exp+lethal” DFE); and the discrete DFE described previously. The MLEs, as well as the proportion of mutations with varying selection coefficients predicted by these distributions, are depicted in Table 2, Table 3, and Table S4 in File S1.

When we assume $\mu = 1.5 \times 10^{-8}$ and $L_{\text{NS}}/L_{\text{S}} = 2.31$, the neutral+gamma DFE fit best to the LuCamp and ESP data sets as reflected by the highest log-likelihood and Akaike information criterion (AIC) score (Table 3 and Table S4 in File S1). The gamma still fit best to the 1000 Genomes data set. Compared to the gamma DFEs inferred previously for two data sets, our best-fitting DFEs predict slightly fewer (0.92–0.98 fold) new mutations with $|s| > 10^{-2}$, and slightly more (1.06–1.18 fold) new mutations of $|s| < 10^{-5}$. When we matched the mutation rates of Boyko *et al.* (2008) with $\mu = 1.8 \times 10^{-8}$ and $L_{\text{NS}}/L_{\text{S}} = 2.5$, the discrete DFE fit best to the LuCamp data set (Table 3 and Table S4 in File S1). However, the gamma DFE continued to fit best to the 1000 Genomes and ESP data sets under these assumptions. The best fitting DFEs are depicted in Figure 4 and Figure S5 in File S1, and a comparison of the model to the SFS of the data are shown in Figure S6 in File S1. When using the larger mutation rate, we find the discrete DFE to fit best to the LuCamp data set, which predicts significantly fewer (0.54 fold) new mutations of $|s| < 10^{-2}$ than the gamma DFE fit using the same mutation rate.

One concern is that biases in SNP calling may affect these inferences. One way to test for this is by masking the singletons in the analysis, since singletons may be enriched for false SNPs due to sequencing errors (Boyko *et al.* 2008). We fit the gamma and neutral+gamma DFEs while masking the singleton category in the SFS and find little difference in the inferred DFEs (Table S6 in File S1). This finding suggests our inferences are robust to potential errors in SNP calling in these data sets.

The DFEs we have inferred thus far differ from that inferred in Boyko *et al.* (2008). In that study, 35.5% of new nonsynon-

ymous mutations were inferred to be strongly deleterious in African-Americans, and 37.9% in Europeans. We infer fewer new strongly deleterious nonsynonymous mutations, even when matching the mutation rates used in Boyko *et al.* (2008) (Figure 4 and Figure S5 in File S1). Furthermore, the distribution of $2N_s$ also shows fewer strongly deleterious mutations (27.1–36.9% of mutations with $2N_s > 100$) than seen in Boyko *et al.* (2008) (40.4% of mutations with $2N_s > 100$; Figure S5 in File S1). Our results remain consistent across data sets and assumed forms of the DFE.

Additionally, our estimates of the DFE differ substantially from the estimates provided by Li *et al.* (2010). Specifically, Li *et al.* (2010) infer almost no strongly or moderately deleterious new nonsynonymous mutations. That is, 1% of new nonsynonymous mutations have selection coefficients of $10^{-3} < |s| < 10^{-2}$ and 0% have a selection coefficient $|s| > 10^{-2}$ (Figure 1). All of our estimates infer that at least ~30% of new nonsynonymous mutations have a selection coefficient $|s| > 10^{-3}$, even when the assumed mutation rate is small (Figure 4 and Table 3). Our simulations suggest if the true underlying DFE follows that suggested by Li *et al.* (2010), we should be able to estimate those proportions using both the neutral+gamma and the discrete DFE (Figure 2, A and B). The fact that our inferences did not show similar estimates to those inferred by Li *et al.* (2010) suggests that our data and analyses are not consistent with the distribution inferred by Li *et al.* (2010) (Table S5 in File S1). In the following sections, we explore several reasons why the different studies infer different DFEs.

Assessing the role of sample size and model misspecification using simulations

One possibility for the distinct estimates of the DFE is that the three studies used different sample sizes. Larger samples will have more moderately and strongly deleterious variants segregating than will smaller data sets (Figure S1 in File S1). To investigate the effect of sample size on our ability to infer the

Table 3 We infer more nearly neutral ($|s| < 10^{-5}$) and fewer strongly deleterious ($|s| \geq 10^{-2}$) new mutations than previous studies

Data set	μ , L_N/L_S	Best fit DFE	$0 \leq s < 10^{-5}$	$10^{-5} \leq s < 10^{-4}$	$10^{-4} \leq s < 10^{-3}$	$10^{-3} \leq s < 10^{-2}$	$10^{-2} \leq s $
Boyko <i>et al.</i> (2008) (AA) ^a	1.8×10^{-8} , 2.5 ^b	Gamma	0.183	0.096	0.147	0.219	0.355
1000 Genomes	1.8×10^{-8} , 2.5 ^b	Gamma	0.217 (0.212–0.223)	0.112 (0.111–0.113)	0.169 (0.165–0.172)	0.243 (0.235–0.249)	0.259 (0.252–0.266)
ESP		Gamma	0.229 (0.223–0.234)	0.105 (0.104–0.106)	0.152 (0.150–0.155)	0.216 (0.212–0.221)	0.298 (0.294–0.302)
LuCamp		Discrete	0.278 (0.221–0.303)	0.027 (0.001–0.110)	0.211 (0.167–0.234)	0.352 (0.330–0.373)	0.132 (0.124–0.142)
1000 Genomes	1.5×10^{-8} , 2.31	Gamma	0.237 (0.231–0.243)	0.127 (0.125–0.128)	0.192 (0.188–0.197)	0.266 (0.259–0.272)	0.178 (0.171–0.186)
ESP		Neu+gamma	0.263 (0.250–0.277)	0.104 (0.091–0.114)	0.167 (0.160–0.173)	0.249 (0.241–0.259)	0.217 (0.211–0.221)
LuCamp		Neu+gamma	0.242 (0.223–0.260)	0.091 (0.072–0.107)	0.194 (0.183–0.203)	0.332 (0.313–0.352)	0.141 (0.129–0.152)

Results in comparison with data from Boyko *et al.* (2008). C.I.'s were constructed by Poisson resampling the nonsynonymous SFS and fitting a DFE 200 times. See Table S5 in File S1 for additional information. Neu, neutral. ^a African-American.

^b The results presented with the assumptions $L_N/L_S = 2.5$ and $\mu = 1.8 \times 10^{-8}$ match the mutation rate assumptions of Boyko *et al.* (2008).

DFE, we simulated 200 data sets, without linkage, of sample sizes $n = 12, 24, 100, 250$, and 500 chromosomes. Each data set was simulated using the demographic model and gamma DFE inferred from the LuCamp data set.

First, we simulated neutral synonymous sites and inferred the demographic parameters from each data set. This was done in two ways. First, we estimated the parameters from the full demographic model that was used to generate the data (herein the “full” model). Second, we inferred the parameters in a demographic model with only three instantaneous size changes (herein the “three-epoch” model). This is meant to mimic the situation in Boyko *et al.* (2008), where the true demography of the European population was likely complex, but simpler three-epoch demographic models could accurately fit the synonymous SFS. Next, as done in our inference and in previous studies, we estimated the parameters of a gamma distribution for the DFE of nonsynonymous mutations, conditioning separately upon the two demographic models.

When the full demographic model was fit to the simulated data sets, we found the variance of our estimates, both of demography and selection, decreased as sample size increased (Figure S7 in File S1). We were unable to correctly infer the magnitude of recent population growth with small sample sizes, consistent with previous work (Keinan and Clark 2012; Nelson *et al.* 2012; Tennessen *et al.* 2012; Fu *et al.* 2013). However, this did not affect the inference of selection as long as the demographic model fit reasonably well to synonymous sites (Figure S7 in File S1). At small sample sizes, the three-epoch model could fit the synonymous SFS well and thus estimates of selection were also unbiased. However, we found that for sample sizes >100 chromosomes, the three-epoch model increasingly became unable to account for the excess of rare variants caused by recent growth. The inability to account for the rare variants in the sample then biased the estimates of both the shape and scale parameters of the gamma distribution. However, this effect seems to be negligible at a sample size of 24 chromosomes (Figure S7 in File S1).

As long as the demographic model fits the observed SFS of synonymous sites, small sample sizes can estimate the parameters in a gamma distributed DFE, even when the demographic model is not the correct one. The accuracy of the estimates increases with sample size, especially for the scale parameter, and notably provides better estimates of the strongly deleterious portion of the DFE (Figure S7B in File S1 and Table 1). Thus, the results of Boyko *et al.* (2008) are unlikely to be affected by misspecification of demography due to small sample size.

Another possibility for the varying estimates of the DFE is that the DFE itself may be misspecified. Although parametric distributions are convenient for approximating the DFE, the true form of the DFE is unknown. Additionally, we have shown that the neutral+gamma DFE and the discrete DFE can fit large data sets better than the gamma DFE. To investigate an example of what would happen if the DFE is misspecified, we

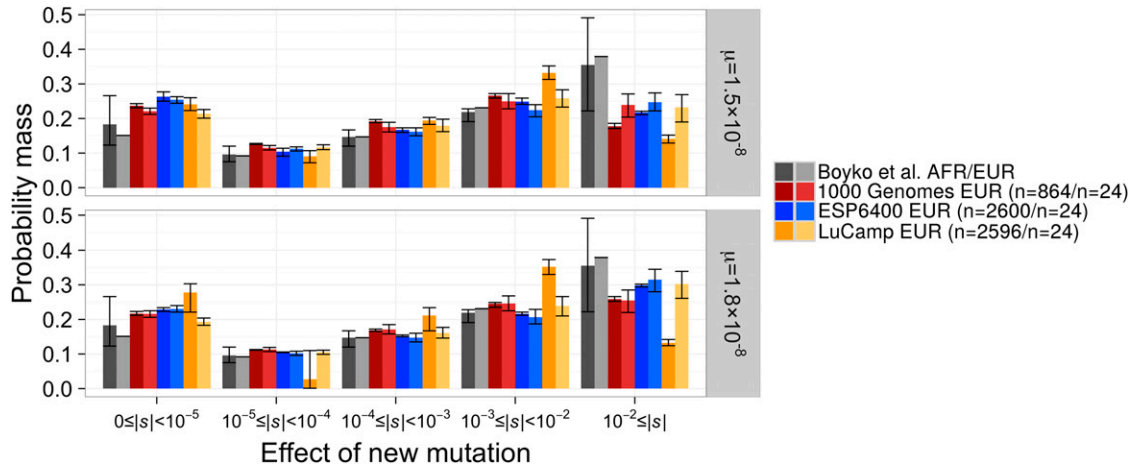


Figure 4 The distribution of selection coefficients of new mutations under our best-fit DFEs compared to Boyko *et al.* (2008). Results are presented for the best-fit DFE to each full data set and the best-fit DFE when the data were projected down to $n = 24$ chromosomes. C.I.'s were estimated by Poisson resampling the nonsynonymous SFS and fitting a DFE 200 times. C.I.'s for the DFE fit to the Boyko *et al.* (2008) European data set were unavailable. Note that our models predict more nearly neutral mutations ($0 \leq |s| < 10^{-5}$) and fewer strongly deleterious mutations ($10^{-2} \leq |s|$) than Boyko *et al.* (2008), across all mutation rates. Top panel denotes our favored mutation rate while the bottom panel denotes the mutation rate used by Boyko *et al.* (2008). See Figure S5 in File S1 for a comparison of the population-scaled selection coefficients ($2N_s$).

simulated 100 data sets without linkage for the best-fit neutral + gamma DFE inferred from the LuCamp data set, scaled to an ancestral population size of 10,085 diploids ($p_{\text{neu}} = 0.164$, $\alpha = 0.338$, $\beta = 358.8$). We also downsampled each data set to $n = 24$ chromosomes. Then, we fit a gamma and neutral + gamma DFE to each full and downsampled data set.

When the true DFE is neutral + gamma distributed, inference of the DFE from small samples overestimates the proportion of strongly deleterious mutations (Figure 5). When the DFE is correctly specified, we obtain unbiased estimates of the DFE even from small samples. However, at a sample size of $n = 24$ chromosomes, both the gamma and neutral + gamma distributions have a similar log-likelihood (Figure 5A). This was also observed in Boyko *et al.* (2008). Then, the extra parameter in the neutral + gamma distribution penalizes the true DFE when choosing the best-fit DFE by AIC score. This leads one to choose the gamma distribution as the best-fit DFE to the small sample, even when the true DFE follows a neutral + gamma distribution. Fitting the gamma distribution yields a DFE with more new mutations with $|s| > 10^{-2}$ and a decrease in the proportion of new mutations with $|s| < 10^{-5}$ compared to the true DFE (Figure 5B).

Assessing the role of sample size using real data

Next, we investigated the role that sample size has on inference of the DFE from real data. To do this, we projected our synonymous and nonsynonymous frequency spectra down to a sample size of $n = 24$ chromosomes to match the sample size of Boyko *et al.* (2008), then fit a demographic model and DFEs as previously described. Here we used the mutation rate assumptions $\mu = 1.5 \times 10^{-8}$ and $L_{\text{NS}}/L_{\text{S}} = 2.31$, but also matched the mutation rate of Boyko *et al.* (2008) ($\mu = 1.8 \times 10^{-8}$ and $L_{\text{NS}}/L_{\text{S}} = 2.5$). Then, we fit the gamma, neutral + gamma, and discrete DFEs—which were

the best-fitting distributions to the full data—to the downsampled data sets.

As predicted by our simulations, there is generally little difference in the fit of the different DFEs to the downsampled data sets in terms of log-likelihood (Table S5 in File S1). The neutral + gamma and discrete DFEs often fit better than the gamma, but the difference in log-likelihood is small (0.1–0.6 log-likelihood units). Thus, the gamma DFE is selected as the best-fit DFE for all downsampled data sets by AIC. These results mimic what was observed in our simulations, although the pattern is not wholly consistent across data sets and mutation rates. When we assume $\mu = 1.8 \times 10^{-5}$ and $L_{\text{NS}}/L_{\text{S}} = 2.5$, the gamma DFE fits best to both the full and downsampled 1000 Genomes and ESP data sets (Figure 4 and Table S5 in File S1). There also appears to be little difference between the gamma DFE fit to the full and downsampled data. By contrast, the discrete DFE fits best to the LuCamp data under these mutation rates. Additionally, the neutral + gamma fit best to the full ESP and LuCamp data when we assume $\mu = 1.5 \times 10^{-5}$ and $L_{\text{NS}}/L_{\text{S}} = 2.31$. The gamma DFE fit to the downsampled versions of these data sets predicts more strongly deleterious ($|s| > 10^{-2}$) and more nearly neutral ($|s| < 10^{-5}$) new mutations (Figure 4 and Table S5 in File S1). The DFE fit to the 1000 Genomes data using the lower mutation rates does not follow this pattern. The gamma DFE fits best to both versions of the data set, yet the estimates from the small data set still predict more strongly deleterious new mutations. These results seem to corroborate the results from our simulations. That is, fitting a DFE using a small sample may result in misspecification of the DFE, which, in turn, may cause inaccuracies in the inferred proportions of the DFE. We believe this may explain some of the differences between the findings of Boyko *et al.* (2008) and the findings in this study.

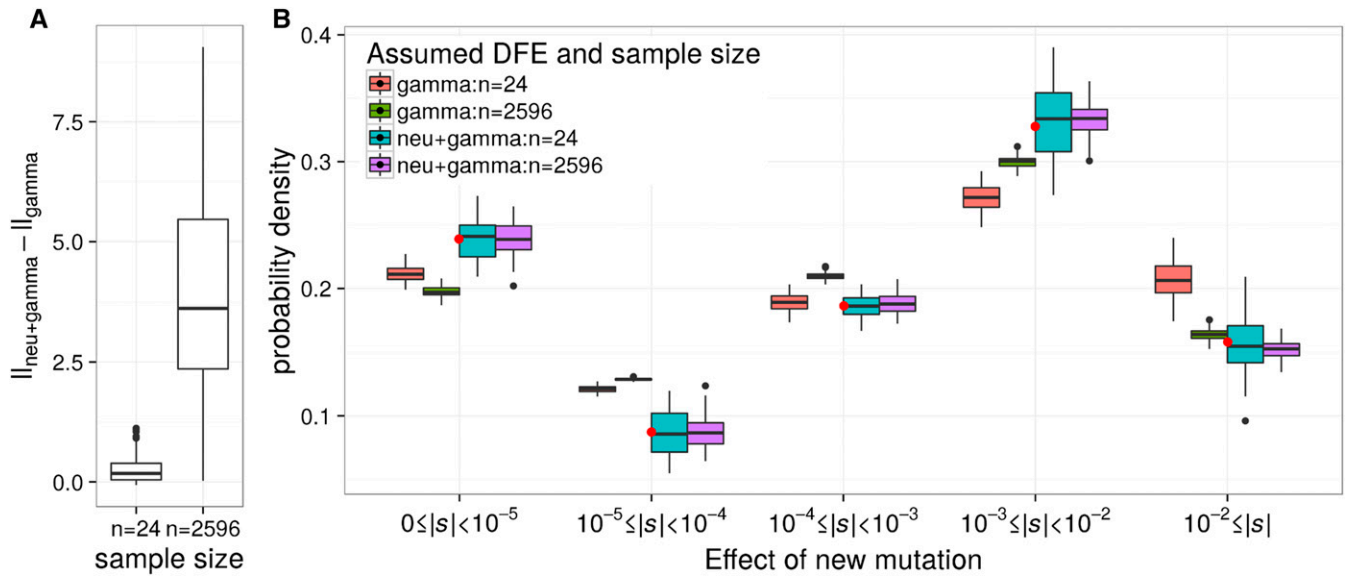


Figure 5 Small sample size and misspecification of the DFE can explain some of the differences between previous estimates and our estimates. Gamma and neutral+gamma DFEs were fit to 100 simulated data sets of sample sizes $n = 24$ and $n = 2596$ chromosomes, where the true DFE was neutral+gamma distributed ($p_{\text{neu}} = 0.164$, $\alpha = 0.338$, $\beta = 358.8$). (A) The distributions of the difference in log-likelihood between the gamma and neutral+gamma distributions. When the sample size is large ($n = 2596$) the neutral+gamma distribution has a higher log-likelihood than the gamma distribution. However, the small samples ($n = 24$) are unable to distinguish between the gamma and neutral+gamma distributions. (B) The estimated proportions of new mutations having different selective effects when fitting the gamma and neutral+gamma distributions. Note that when $n = 24$, the gamma distribution overpredicts the proportion of strongly deleterious mutations ($|s| \geq 0.01$). Red dots denote the true proportion of mutations in each bin. The boxes cover the first and third quartiles, and the band represents the median. The whiskers cover the highest and lowest datum within 1.5 times the interquartile range from the first and third quartiles. Lastly, any data outside that region are plotted as outlier points.

Assessing the role of the likelihood function using simulations

Next, we investigated the performance of the multinomial vs. Poisson likelihoods at inferring the DFE. In this study, as well as in Boyko *et al.* (2008), we fit the DFE using the Poisson likelihood, which uses an *a priori* estimate of the population-scaled mutation rate, θ , to fit the curvature of the SFS as well as the total number of SNPs. Too few segregating variants would suggest the presence of strongly deleterious variants that are not segregating in the sample (Boyko *et al.* 2008).

The multinomial likelihood fits the curvature of the SFS while conditioning on the total number of SNPs. As such, the number of SNPs provides no additional information. The multinomial inference is similar to how the DFE was inferred by Li *et al.* (2010) in that they only used information from the curvature of the SFS. Note, however, Li *et al.* (2010) fit the population frequency spectrum using a least-squares approach while the multinomial likelihood fits the sample frequency spectrum. As such, the multinomial likelihood function does not strictly mirror the procedure of Li *et al.* (2010). Using the multinomial likelihood is convenient because it does not require any prior assumptions about the population scaled mutation rate, θ , yet may be underpowered when trying to identify the proportion of strongly deleterious mutations, unless such variants are segregating in the sample.

To compare the two likelihood methods at varying sample sizes, we fit the full model to simulated data sets of $n = 12, 24, 50, 100, 150, 200$, and 250 chromosomes using both the

multinomial and Poisson likelihoods (Figure S8 in File S1). Again, we simulated 200 data sets at each sample size with the LuCamp demography and a gamma DFE with parameters $\alpha = 0.203$ and $\beta = 1082.1$. In general, the accuracy of our shape parameter estimate improves as the sample size increases, and we find the multinomial and Poisson likelihoods can both be used to reliably estimate the shape parameter. While this trend holds true for the scale parameter using the Poisson likelihood, we find that we are unable to accurately infer the scale parameter using the multinomial likelihood, even for a sample of $n = 250$ chromosomes. For example, nearly 50% of all the parameter estimates lie close to the maximum bound and 25% lie close to the minimum bound allowed during optimization. We found that this result can be explained by the likelihood surface being exceptionally flat with respect to the scale parameter. In other words, we cannot estimate the strength of purifying selection using only the curvature of the SFS with these sample sizes. Therefore, because Li *et al.* (2010) fit only the curvature of the SFS and excluded rare variants ($<2\%$ MAF) in a sample of size of $n = 400$ chromosomes, the power to detect strongly deleterious variants may be quite low, resulting in different parameter estimates from those in Boyko *et al.* (2008) and our present work.

Discussion

We developed a computational method to infer the DFE of new mutations from large data sets, and then estimated the DFEs

for nonsynonymous mutations using the SFS of 432 Europeans from the 1000 Genomes Project, 1300 Europeans from the ESP, and 1298 Europeans from the LuCamp project. The new DFEs suggest there are fewer strongly deleterious mutations than previously estimated (Figure 4). Although we find a neutral+gamma mixture DFE fits best to the ESP and LuCamp data sets, a gamma DFE seems to be a better fit to the 1000 Genomes data (Table 3). Nevertheless, our best-fit DFEs predict 0.38 to 0.84 as many strongly deleterious new mutations compared to the current, widely used estimates of Boyko *et al.* (2008). Additionally, these findings are robust to assumptions about the mutation rate or the assumed functional form of the DFE. We show small sample size can lead to incorrect estimates of the DFE, specifically when the DFE is approximated with a parametric distribution that is not the true distribution (Figure 5). Therefore, our estimates provide an important update to previous studies of the DFE that used smaller sample sizes. Our current estimates of the DFE, particularly the estimates of the proportion of moderately and strongly deleterious mutations, should be more reliable and precise than previous estimates. To facilitate their utility for future researchers, we provide scripts for implementing these models on GitHub (see *Materials and Methods*).

Our results suggest misspecification of the DFE may explain some of the differences in the DFEs we infer from small and large data sets. This is particularly relevant because the true DFE is almost certainly not a parametric distribution. At small sample sizes, different forms of the DFE tended to similarly fit the data in terms of log-likelihood (Table S5 in File S1). Therefore, the DFE that had fewer parameters (*i.e.*, gamma) was selected as the best-fit DFE. Additionally, we infer more strongly deleterious ($|s| > 10^{-2}$) new mutations from the downsampled data sets. We showed through simulations that even if the true DFE is neutral+gamma distributed, a gamma DFE is selected as the best fit to a small sample. Furthermore, this leads to inaccuracy in recovering the true proportions of the DFE (Figure 5). While the neutral+gamma distribution is also unlikely to be the true DFE, our simulations reproduce the patterns observed when downsampling the real data. Therefore, large sample size not only helps to improve the precision of the estimated DFEs, but also helps to approximate the correct form of the DFE. We expect this question to be better resolved as additional and larger sequencing data sets continue to be generated in the future.

A gamma or similarly shaped distribution of deleterious mutations is well supported by experimental estimates of the DFE in laboratory organisms (Martin and Lenormand 2006; Bataillon and Bailey 2014), although some studies suggest more complex distributions (Halligan and Keightley 2009; Hietpas *et al.* 2011; Jacquier *et al.* 2013). A number of theoretical models also predict the functional form of the DFE (Huber *et al.* 2016). Most progress in this regard comes from phenotype fitness-landscape models such as Fisher's geometric model (FGM) (Martin and Lenormand 2006; Chevin *et al.* 2010; Lourenço *et al.* 2011; Tenaillon 2014; Huber *et al.*

2016) and biophysical models of protein stability (Cherry 1998; Goldstein 2013; Serohijos and Shakhnovich 2014). Under fairly general assumptions, the predicted DFE under these models for a perfectly adapted population is gamma distributed (Martin and Lenormand 2006; Martin 2014; Serohijos and Shakhnovich 2014), and a strongly leptokurtic shape would suggest that most mutations have low pleiotropy (Martin and Lenormand 2006; Lourenço *et al.* 2011). However, our finding of a neutral+gamma distribution suggests that the general FGM is inadequate, since it does not predict the neutral point mass. Alternatively, our support for a neutral point mass might not reflect truly neutral mutations, but rather slightly beneficial mutations that behave effectively neutral under the historically small human population size (Huber *et al.* 2016). Since these mutations would segregate at frequencies similar to neutral mutations, and since we do not explicitly model the effect of beneficial mutations on the SFS, our method would place these mutations at the neutral point mass. Such slightly beneficial mutations are predicted under FGM when deleterious mutations fix and move the population away from the optimal phenotype (Lourenço *et al.* 2011). Slightly deleterious mutations can fix in the population when the effect of drift is large, *i.e.*, the effective population size is small. Thus, our support for the neutral+gamma distribution might be consistent with a large effect of drift in the relatively small human population (Huber *et al.* 2016). Alternatively, a recent change in the selective environment could have moved the human population away from the phenotypic optimum at many genes, leading to a similar increase of the proportion of slightly beneficial mutations (Martin and Lenormand 2006; Chevin *et al.* 2010; Lourenço *et al.* 2011; Bank *et al.* 2014).

Additionally, our results show that estimates of the DFE are sensitive to the mutation rate. For any given data set, assuming a higher nonsynonymous mutation rate will result in the inference of stronger purifying selection due to the increased number of SNPs expected (but not observed) relative to the nonsynonymous mutation rate. There are two assumptions that factor into the calculation of the nonsynonymous mutation rate: First, it is unclear what the true mutation rate is. Whole genome, pedigree-based estimates suggest a mutation rate of about 10^{-8} per base pair per generation, exome-based estimates suggest rates of 1.5×10^{-8} , and phylogenetic estimates suggest a mutation rate in the range of $2.0\text{--}2.5 \times 10^{-8}$ (Ségurel *et al.* 2014). Second, we infer a mutation rate from synonymous sites, but use that mutation rate to make an *a priori* assumption about the nonsynonymous mutation rate. Many studies have the nonsynonymous mutation rate at 2.5 times the synonymous mutation rate, but we believe 2.31 to be a more accurate estimate, as this takes into account the CpG mutational bias and a 3:1 transition:transversion ratio (Huber *et al.* 2016). These two factors combined can result in large differences in the DFE. For example, the gamma DFE fit to the LuCamp data predicts 15% of mutations to be strongly deleterious ($|s| > 10^{-2}$) when assuming $\theta_{NS}/\theta_S = 2.31$ and $\mu = 1.5 \times 10^{-8}$, but 25% of new mutations to be strongly

deleterious when assuming $\theta_{NS}/\theta_S = 2.5$ and $\mu = 2.5 \times 10^{-8}$. Although our results remain qualitatively consistent across the range of mutation rates, uncertainty about the true rate leads to uncertainty in estimating the DFE.

Another important aspect of our results is the consistency of our estimates of the DFE between data sets. Our estimates of the DFE suggest a skew toward neutrality compared to previous studies, and we infer a consistent range of neutral ($|s| < 10^{-5}$, 24–26%), moderately deleterious ($10^{-3} < |s| < 10^{-2}$, 25–33%), and strongly deleterious ($|s| < 10^{-2}$, 14–22%) new mutations between the three data sets. The consistency of our results across data sets suggests our inferences are accurate and robust to sampling from different populations, sequencing, bioinformatic processing, and sample size. This suggests the DFEs we have inferred are reliable updates to the DFEs inferred by Eyre-Walker *et al.* (2006) and Boyko *et al.* (2008).

It is also worth noting that our methodology has key differences from that of Li *et al.* (2010). Li *et al.* (2010) estimated the DFE using the population frequency spectrum excluding rare variants (MAF < 2%), under a constant size demographic model using a least-squares method, and fit the curvature of the SFS while not considering the total number of SNPs in the sample. The extent to which these methodological differences as well as differences in sequencing or bioinformatic processing of the data between their study and our present study contribute to the different estimated DFEs remains unclear. However, we have shown that for small and moderately sized samples, fitting only the curvature of the SFS is insufficient for estimating the scale parameter of the DFE. In other words, for smaller samples, the number of SNPs in the data must be considered to estimate the proportions of moderately and strongly deleterious new mutations, since moderately to strongly deleterious mutations are unlikely to be found in the sample.

Fit*dadi* infers the DFE for new mutations rather than segregating variants. Interestingly, even inference using the multinomial likelihood function, which only uses the frequencies of segregating variants, still infers the DFE for new mutations. We used simulations to compare the DFE of new mutations to that of segregating variants (Figure S9 and Table S7 in File S1). The DFE of segregating variants depends on the sample size and is shifted to be more neutral than that of new mutations. Approximating the DFE of segregating variants using a gamma distribution reveals a different shape parameter than that of new mutations (Table S7 in File S1); confirming that Fit*dadi*, when applied to segregating variants, estimates the DFE of new mutations. While long-tailed distributions such as the gamma do not directly capture the mode of very strongly deleterious new mutations observed by many experimental studies (Eyre-Walker and Keightley 2007), the proportion could be extrapolated from the DFEs we inferred. For example, for a gamma-distributed DFE, the proportion of new mutations that is lethal would be computed as the mass of the distribution greater than $|s| > 1$.

Our results argue that, provided the demographic model fit to the data can adequately match the SFS of neutral synonymous sites, inference of the DFE should be robust to cryptic, unmodeled, population structure. In other words, the skew in the frequency spectrum due to demography must be reproduced accurately, but inference of selection is relatively robust to the way the skew is modeled. This result is consistent with the work of Ma *et al.* (2013). Alternately, other studies rescale the entries of the nonsynonymous SFS independently to match the skew of the synonymous SFS from the standard neutral model (Eyre-Walker *et al.* 2006; Galtier 2016; Tataru *et al.* 2016). However, this method is not always accurate for demographies including recent, rapid expansions since the skew on neutral and selected sites may differ (Eyre-Walker *et al.* 2006). Further, fitting many independent scaling parameters to large samples can be problematic (Tataru *et al.* 2016). Thus, Fit*dadi* offers an advantage over the rescaling methods in these contexts.

Although Fit*dadi* was developed to work with large sequencing data sets, it still has several limitations. The inference framework we use becomes increasingly slower for larger samples and requires significant computational resources and time to initially generate the SFS for the range of selection coefficients. Additionally, the frequency spectrum becomes difficult to compute for larger selection coefficients ($2N_s > 10,000$). This is mainly because finer integration grids must be used to accurately estimate low frequency variants. Also, like the method of Boyko *et al.* (2008), our method assumes additive selective effects and should be interpreted as averaging of selection over all heterozygotes and genetic backgrounds. Nevertheless, we anticipate that our method will be useful for estimating the DFE across the tree of life as polymorphism data sets from different species continue to accumulate.

Our results suggest that there may be more weakly and moderately deleterious nonsynonymous mutations than previously appreciated. This has a number of important implications for medical genetic studies. These variants could possibly contribute to disease risk. However, these mutations could also confound statistical tests that compare observed levels of variation to those predicted by population genetic models. For example, using the DFE of Boyko *et al.* (2008) would predict fewer segregating deleterious variants because more new mutations were estimated to be strongly deleterious and would not segregate in the sample. However, if those mutations were instead only moderately deleterious, some could drift up in frequency and actually segregate in the sample. Further, a common approach to modeling how deleterious variants affect complex traits (Eyre-Walker 2010) assigns mutational effects on a trait as a function of their effects on fitness. This approach has been widely used to quantify the architecture of complex traits (Morris *et al.* 2012; Mancuso *et al.* 2016), to study the effects of demography on traits (Lohmueller 2014a; Simons *et al.* 2014), and to assess the power of rare variant association tests (Uricchio *et al.* 2016). The accuracy and realism

of these models depend on having accurate estimates of the DFE.

Additionally, the DFE determines the extent to which background selection affects patterns of neutral variation. Accurately characterizing the reduction in diversity (*i.e.*, effective population size N_e) should reduce bias when trying to learn the true demography of a population using sites linked to selected variants (Ewing and Jensen 2016; Schrider *et al.* 2016). We used a deterministic approximation (Nicolaisen and Desai 2013) of the models of Zeng and Charlesworth (2011) to contrast the effects of background selection predicted from the DFE of Boyko *et al.* (2008) and the DFEs we inferred in our study (Figure S10 in File S1). We computed the reduction in N_e (*i.e.*, increase in the rate of coalescence) as a function of time due to background selection for the two different mutation rates used in our inferences ($\mu = 1.5 \times 10^{-8}$, $L_{NS}/L_S = 2.31$; $\mu = 1.8 \times 10^{-8}$ and $L_{NS}/L_S = 2.5$) as well as the higher deleterious mutation rate of McVicker *et al.* (2009): $\mu = 7.4 \times 10^{-8}$, also assuming $L_{NS}/L_S = 2.5$. Importantly, all the DFEs predict that background selection will reduce diversity and skew the SFS toward an excess of rare variants compared to models of constant population size (Figure S10 in File S1). However, DFEs with fewer strongly deleterious mutations, like the best fit DFEs to the ESP EUR and LuCamp data sets, predict less of an overall reduction in neutral diversity compared to Boyko *et al.* (2008). Further, the change in coalescent rates over time varies across DFEs, suggesting that the degree to which background selection affects the curvature of the SFS does depend on the specific DFE.

More broadly, our results have important implications for understanding and quantifying deleterious variants across human populations (Lohmueller *et al.* 2008; Lohmueller 2014b; Simons *et al.* 2014; Do *et al.* 2015). Specifically, the fate of strongly deleterious mutations is relatively insensitive to population demography. The fate of weakly and moderately deleterious mutations, however, is linked more closely with effective population size (Henn *et al.* 2016). Human evolution in particular is influenced by nearly neutral processes due to relatively small effective population sizes. Then, a DFE containing fewer strongly deleterious new mutations suggests the nature of purifying selection in humans may be different from what is currently understood. For example, larger proportions of moderately and weakly deleterious mutations may suggest greater differences in the proportion of segregating deleterious mutations and genetic load between human populations (Henn *et al.* 2016). Accurate inferences of the DFE are critical in this regard as researchers begin to use explicit models of demography and selection to quantify differences in the amounts of deleterious variants across populations (Brandvain and Wright 2016; Gravel 2016).

Acknowledgments

We thank Emilia Huerta-Sanchez, Diego Ortega-Del Vecchyo, and Noah Rosenberg, as well as two reviewers, for constructive

comments on the manuscript. This work was supported by a Searle Scholars Fellowship, an Alfred P. Sloan Research Fellowship in Computational and Molecular Biology, and National Institutes of Health grant R35 GM-119856 to K.E.L.

Literature Cited

- Aberer, A. J., and A. Stamatakis, 2013 Rapid forward-in-time simulation at the chromosome and genome level. *BMC Bioinformatics* 14: 216.
- Acevedo, A., L. Brodsky, and R. Andino, 2014 Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* 505: 686–690.
- Bank, C., R. T. Hietpas, A. Wong, D. N. Bolon, and J. D. Jensen, 2014 A Bayesian MCMC approach to assess the complete distribution of fitness effects of new mutations: uncovering the potential for adaptive walks in challenging environments. *Genetics* 196: 841–852.
- Bataillon, T., and S. F. Bailey, 2014 Effects of new mutations on fitness: insights from models and data. *Ann. N. Y. Acad. Sci.* 1320: 76–92.
- Boucher, J. I., P. Cote, J. Flynn, L. Jiang, A. Laban *et al.*, 2014 Viewing protein fitness landscapes through a next-gen lens. *Genetics* 198: 461–471.
- Boyko, A. R., S. H. Williamson, A. R. Indap, J. D. Degenhardt, R. D. Hernandez *et al.*, 2008 Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet.* 4: e1000083.
- Brandvain, Y., and S. I. Wright, 2016 The limits of natural selection in a nonequilibrium world. *Trends Genet.* 32: 201–210.
- Castellano, D., M. Coronado-Zamora, J. L. Campos, A. Barbadilla, and A. Eyre-Walker, 2016 Adaptive evolution is substantially impeded by Hill–Robertson interference in *Drosophila*. *Mol. Biol. Evol.* 33: 442–455.
- Cherry, J. L., 1998 Should we expect substitution rate to depend on population size? *Genetics* 150: 911–919.
- Chevin, L.-M., R. Lande, and G. M. Mace, 2010 Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *PLoS Biol.* 8: e1000357.
- Chikhi, L., R. A. Nichols, G. Barbujani, and M. A. Beaumont, 2002 Y genetic data support the Neolithic demic diffusion model. *Proc. Natl. Acad. Sci. USA* 99: 11008–11013.
- Comeron, J. M., 2014 Background selection as baseline for nucleotide variation across the *Drosophila* genome. *PLoS Genet.* 10: e1004434.
- Do, R., D. Balick, H. Li, I. Adzhubei, S. Sunyaev *et al.*, 2015 No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. *Nat. Genet.* 47: 126–131.
- Ewing, G. B., and J. D. Jensen, 2016 The consequences of not accounting for background selection in demographic inference. *Mol. Ecol.* 25: 135–141.
- Eyre-Walker, A., 2010 Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proc. Natl. Acad. Sci. USA* 107: 1752–1756.
- Eyre-Walker, A., and P. D. Keightley, 2007 The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* 8: 610–618.
- Eyre-Walker, A., M. Woolfit, and T. Phelps, 2006 The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* 173: 891–900.
- Fowler, D. M., C. L. Araya, S. J. Fleishman, E. H. Kellogg, J. J. Stephany *et al.*, 2010 High-resolution mapping of protein sequence-function relationships. *Nat. Methods* 7: 741–746.

- Fu, W., T. D. O'Connor, G. Jun, H. M. Kang, G. Abecasis *et al.*, 2013 Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 493: 216–220.
- Galtier, N., 2016 Adaptive protein evolution in animals and the effective population size hypothesis. *PLoS Genet.* 12: e1005774.
- Gazave, E., L. Ma, D. Chang, A. Coventry, F. Gao *et al.*, 2014 Neutral genomic regions refine models of recent rapid human population growth. *Proc. Natl. Acad. Sci. USA* 111: 757–762.
- Goldstein, R. A., 2013 Population size dependence of fitness effect distribution and substitution rate probed by biophysical model of protein thermostability. *Genome Biol. Evol.* 5: 1584–1593.
- Gossmann, T. I., P. D. Keightley, and A. Eyre-Walker, 2012 The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. *Genome Biol. Evol.* 4: 658–667.
- Gravel, S., 2016 When is selection effective? *Genetics* 203: 451–462.
- Gutenkunst, R. N., R. D. Hernandez, S. H. Williamson, and C. D. Bustamante, 2009 Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 5: e1000695.
- Halligan, D. L., and P. D. Keightley, 2009 Spontaneous mutation accumulation studies in evolutionary genetics. *Annu. Rev. Ecol. Evol. Syst.* 40: 151–172.
- Halligan, D. L., A. Kousathanas, R. W. Ness, B. Harr, L. Eöry *et al.*, 2013 Contributions of protein-coding and regulatory change to adaptive molecular evolution in murid rodents. *PLoS Genet.* 9: e1003995.
- Harris, K., and R. Nielsen, 2016 The genetic cost of Neanderthal introgression. *Genetics* 203: 881–891.
- Hartl, D. L., E. N. Moriyama, and S. A. Sawyer, 1994 Selection intensity for codon bias. *Genetics* 138: 227–234.
- Henn, B. M., L. R. Botigué, S. Peischl, I. Dupanloup, M. Lipatov *et al.*, 2016 Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. *Proc. Natl. Acad. Sci. USA* 113: E440–E449.
- Hernandez, R. D., 2008 A flexible forward simulator for populations subject to selection and demography. *Bioinformatics* 24: 2786–2787.
- Hernandez, R. D., S. H. Williamson, and C. D. Bustamante, 2007 Context dependence, ancestral misidentification, and spurious signatures of natural selection. *Mol. Biol. Evol.* 24: 1792–1800.
- Hietpas, R. T., J. D. Jensen, and D. N. A. Bolon, 2011 Experimental illumination of a fitness landscape. *Proc. Natl. Acad. Sci. USA* 108: 7896–7901.
- Huber, C. D., B. Y. Kim, C. D. Marsden, and K. E. Lohmueller, 2016 Determining the factors driving selective effects of new nonsynonymous mutations. *bioRxiv* Available at: <http://www.biorxiv.org/content/early/2016/08/23/071209>.
- Jacquier, H., A. Birgy, H. L. Nagard, Y. Mechulam, E. Schmitt *et al.*, 2013 Capturing the mutational landscape of the beta-lactamase TEM-1. *Proc. Natl. Acad. Sci. USA* 110: 13067–13072.
- Keightley, P. D., and A. Eyre-Walker, 2007 Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics* 177: 2251–2261.
- Keightley, P. D., and S. P. Otto, 2006 Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* 443: 89–92.
- Keinan, A., and A. G. Clark, 2012 Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* 336: 740–743.
- Koufopanou, V., S. Lomas, I. J. Tsai, and A. Burt, 2015 Estimating the fitness effects of new mutations in the wild yeast *Saccharomyces paradoxus*. *Genome Biol. Evol.* 7: 1887–1895.
- Kousathanas, A., and P. D. Keightley, 2013 A comparison of models to infer the distribution of fitness effects of new mutations. *Genetics* 193: 1197–1208.
- Kraft, D., 1988 *A Software Package for Sequential Quadratic Programming*. Wiss. Berichtswesen d, DFVLR, Cologne.
- Li, Y., N. Vinckenbosch, G. Tian, E. Huerta-Sanchez, T. Jiang *et al.*, 2010 Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat. Genet.* 42: 969–972.
- Lohmueller, K. E., 2014a The impact of population demography and selection on the genetic architecture of complex traits. *PLoS Genet.* 10: e1004379.
- Lohmueller, K. E., 2014b The distribution of deleterious genetic variation in human populations. *Curr. Opin. Genet. Dev.* 29: 139–146.
- Lohmueller, K. E., A. R. Indap, S. Schmidt, A. R. Boyko, R. D. Hernandez *et al.*, 2008 Proportionally more deleterious genetic variation in European than in African populations. *Nature* 451: 994–997.
- Lohmueller, K. E., T. Sparsø, Q. Li, E. Andersson, T. Korneliussen *et al.*, 2013 Whole-exome sequencing of 2,000 Danish individuals and the role of rare coding variants in type 2 diabetes. *Am. J. Hum. Genet.* 93: 1072–1086 (erratum: *Am. J. Hum. Genet.* 94: 479).
- Lourenço, J., N. Galtier, and S. Glémin, 2011 Complexity, pleiotropy, and the fitness effect of mutations. *Evolution* 65: 1559–1571.
- Ma, X., J. L. Kelley, K. Eilertson, S. Musharoff, J. D. Degenhardt *et al.*, 2013 Population genomic analysis reveals a rich speciation and demographic history of orang-utans (*Pongo pygmaeus* and *Pongo abelii*). *PLoS One* 8: e77175.
- Mancuso, N., N. Rohland, K. A. Rand, A. Tandon, A. Allen *et al.*, 2016 The contribution of rare variation to prostate cancer heritability. *Nat. Genet.* 48: 30–35.
- Martin, G., 2014 Fisher's geometrical model emerges as a property of complex integrated phenotypic networks. *Genetics* 197: 237–255.
- Martin, G., and T. Lenormand, 2006 A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution* 60: 893–907.
- McManus, K. F., J. L. Kelley, S. Song, K. R. Veeramah, A. E. Woerner *et al.*, 2015 Inference of gorilla demographic and selective history from whole-genome sequence data. *Mol. Biol. Evol.* 32: 600–612.
- McVicker, G., D. Gordon, C. Davis, and P. Green, 2009 Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet.* 5: e1000471.
- Messer, P. W., 2013 SLiM: simulating evolution with selection and linkage. *Genetics* 194: 1037–1039.
- Messer, P. W., and D. A. Petrov, 2013 Frequent adaptation and the McDonald–Kreitman test. *Proc. Natl. Acad. Sci. USA* 110: 8615–8620.
- Moon, S., and J. M. Akey, 2016 A flexible method for estimating the fraction of fitness influencing mutations from large sequencing data sets. *Genome Res.* 26: 834–843.
- Morris, A. P., B. F. Voight, T. M. Teslovich, T. Ferreira, A. V. Segrè *et al.*, 2012 Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* 44: 981–990.
- Nelson, M. R., D. Wegmann, M. G. Ehm, D. Kessner, P. St Jean *et al.*, 2012 An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* 337: 100–104.
- Nicolaisen, L. E., and M. M. Desai, 2013 Distortions in genealogies due to purifying selection and recombination. *Genetics* 195: 221–230.
- Novembre, J., and S. Ramachandran, 2011 Perspectives on human population structure at the cusp of the sequencing era. *Annu. Rev. Genomics Hum. Genet.* 12: 245–274.

- Ortega-Del Vecchio, D., C. D. Marsden, and K. E. Lohmueller, 2016 PReFerSim: fast simulation of demography and selection under the Poisson random field model. *Bioinformatics* 32: 3516–3518.
- Ptak, S. E., and M. Przeworski, 2002 Evidence for population growth in humans is confounded by fine-scale population structure. *Trends Genet.* 18: 559–563.
- Ragsdale, A. P., A. J. Coffman, P. Hsieh, T. J. Struck, and R. N. Gutenkunst, 2016 Triallelic population genomics for inferring correlated fitness effects of same site nonsynonymous mutations. *Genetics* 203: 513–523.
- Sawyer, S. A., and D. L. Hartl, 1992 Population genetics of polymorphism and divergence. *Genetics* 132: 1161–1176.
- Schrider, D. R., A. G. Shanku, and A. D. Kern, 2016 Effects of linked selective sweeps on demographic inference and model selection. *Genetics* 204: 1207–1223.
- Ségurel, L., M. J. Wyman, and M. Przeworski, 2014 Determinants of mutation rate variation in the human germline. *Annu. Rev. Genomics Hum. Genet.* 15: 47–70.
- Serohijos, A. W. R., and E. I. Shakhnovich, 2014 Contribution of selection for protein folding stability in shaping the patterns of polymorphisms in coding regions. *Mol. Biol. Evol.* 31: 165–176.
- Simons, Y. B., M. C. Turchin, J. K. Pritchard, and G. Sella, 2014 The deleterious mutation load is insensitive to recent population history. *Nat. Genet.* 46: 220–224.
- Tataru, P., M. Mollion, S. Glemin, and T. Bataillon, 2016 Inference of distribution of fitness effects and proportion of adaptive substitutions from polymorphism data. *bioRxiv* Available at: <http://biorxiv.org/content/early/2016/07/05/062216>.
- Tenaillon, O., 2014 The utility of Fisher's geometric model in evolutionary genetics. *Annu. Rev. Ecol. Evol. Syst.* 45: 179–201.
- Tennessen, J. A., A. W. Bigham, T. D. O'Connor, W. Fu, E. E. Kenny *et al.*, 2012 Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337: 64–69.
- The 1000 Genomes Project Consortium, 2015 A global reference for human genetic variation. *Nature* 526: 68–74.
- Torgerson, D. G., A. R. Boyko, R. D. Hernandez, A. Indap, X. Hu *et al.*, 2009 Evolutionary processes acting on candidate cis-regulatory regions in humans inferred from patterns of polymorphism and divergence. *PLoS Genet.* 5: e1000592.
- Uricchio, L. H., N. A. Zaitlen, C. J. Ye, J. S. Witte, and R. D. Hernandez, 2016 Selection and explosive growth alter genetic architecture and hamper the detection of causal rare variants. *Genome Res.* 26: 863–873.
- Williamson, S. H., R. Hernandez, A. Fedel-Alon, L. Zhu, R. Nielsen *et al.*, 2005 Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proc. Natl. Acad. Sci. USA* 102: 7882–7887.
- Wilson Sayres, M. A., K. E. Lohmueller, and R. Nielsen, 2014 Natural selection reduced diversity on human Y chromosomes. *PLoS Genet.* 10: e1004064.
- Zeng, K., and B. Charlesworth, 2011 The joint effects of background selection and genetic recombination on local gene genealogies. *Genetics* 189: 251–266.

Communicating editor: N. A. Rosenberg