The Impact of Purifying and Background Selection on the Inference of Population History: Problems and Prospects

Parul Johri,*[,1] Kellen Riall,[1] Hannes Becher,[2] Laurent Excoffier,[3,4] Brian Charlesworth,[2] and Jeffrey D. Jensen*[,1]

[1]School of Life Sciences, Arizona State University, Tempe, AZ, USA
[2]Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom
[3]Institute of Ecology and Evolution, University of Berne, Berne, Switzerland
[4]Swiss Institute of Bioinformatics, Lausanne, Switzerland

***Corresponding authors**: E-mails: pjohri1@asu.edu; jeffrey.d.jensen@asu.edu.
Associate editor: Arndt von Haeseler

Abstract

Current procedures for inferring population history generally assume complete neutrality—that is, they neglect both direct selection and the effects of selection on linked sites. We here examine how the presence of direct purifying selection and background selection may bias demographic inference by evaluating two commonly-used methods (MSMC and *fastsimcoal2*), specifically studying how the underlying shape of the distribution of fitness effects and the fraction of directly selected sites interact with demographic parameter estimation. The results show that, even after masking functional genomic regions, background selection may cause the mis-inference of population growth under models of both constant population size and decline. This effect is amplified as the strength of purifying selection and the density of directly selected sites increases, as indicated by the distortion of the site frequency spectrum and levels of nucleotide diversity at linked neutral sites. We also show how simulated changes in background selection effects caused by population size changes can be predicted analytically. We propose a potential method for correcting for the mis-inference of population growth caused by selection. By treating the distribution of fitness effect as a nuisance parameter and averaging across all potential realizations, we demonstrate that even directly selected sites can be used to infer demographic histories with reasonable accuracy.

Key words: demographic inference, background selection, distribution of fitness effects, MSMC, *fastsimcoal2*, approximate Bayesian computation (ABC).

Introduction

The characterization of past population size change is a central goal of population genomic analysis, with applications ranging from anthropological to agricultural to clinical (see review by Beichman et al. 2018). Furthermore, use of an appropriate demographic model provides a necessary null model for assessing the impact of selection across the genome (e.g., Teshima et al. 2006; Thornton and Jensen 2007; Jensen et al. 2019). Multiple strategies have been proposed for performing demographic inference, utilizing expectations related to levels of variation, the site frequency spectrum, linkage disequilibrium, and within- and between-population divergence (e.g., Gutenkunst et al. 2009; Li and Durbin 2011; Lukic and Hey 2012; Excoffier et al. 2013; Harris and Nielsen 2013; Bhaskar et al. 2015; Boitard et al. 2016; Sheehan and Song 2016; Ragsdale and Gutenkunst 2017; Kelleher et al. 2019; Speidel et al. 2019; Steinrücken et al. 2019).

Although many methods perform well when evaluated under the standard assumption of neutrality, it is difficult in practice to assure that the nucleotide sites used in empirical analyses experience neither direct selection nor the effects of selection at linked sites. For example, inference is often performed using intergenic, 4-fold degenerate, or intronic sites. Although there is evidence for weak direct selection on all of these categories in multiple organisms (e.g., Andolfatto 2005; Chamary and Hurst 2005; Haddrill et al. 2005; Lynch 2007; Zeng and Charlesworth 2010; Choi and Aquadro 2016; Jackson et al. 2017), it is also clear that such sites near or in coding regions will also experience background selection (BGS; Charlesworth et al. 1993; Charlesworth 2013), and may periodically be affected by selective sweeps as well (Messer and Petrov 2013; Schrider et al. 2016). These effects are known to affect the local underlying effective population size, and alter both the levels and patterns of variation and linkage disequilibrium (Charlesworth et al. 1993; Kaiser and Charlesworth 2009; O'Fallon et al. 2010; Charlesworth 2013; Nicolaisen and Desai 2013; Ewing and Jensen 2016; Johri et al. 2020).

However, commonly-used approaches for performing demographic inference that assume complete neutrality, including *fastsimcoal2* (Excoffier et al. 2013) and MSMC/PSMC (Li and Durbin 2011; Schiffels and Durbin 2014), have yet to

be thoroughly evaluated in the light of this assumption, which is likely to be violated in practice. There are, however, some exceptions, as well as subsequent suggestions on how best to choose the least-affected genomic data for analysis (Pouyet et al. 2018). Rather than investigating existing software, Ewing and Jensen (2016) implemented an approximate Bayesian (ABC) approach quantifying the impact of BGS effects, demonstrating that weak purifying selection can generate a skew towards rare alleles that would be mis-interpreted as population growth. Under certain scenarios, this resulted in a many-fold mis-inference of population size change. However, the effects of the density of directly selected sites and the shape of the distribution of fitness effects (DFE), which are probably of great importance, have yet to be fully considered. Spanning the range of these potential parameter values is important for understanding the implications for empirical applications. For example, the proportion of the genome experiencing direct purifying selection can vary greatly between species, with estimates ranging from ∼3% to 8% in humans, ∼12% in rice, 37% to 53% in *Drosophila melanogaster*, and 47% to 68% in *Saccharomyces cerevisiae* (Siepel et al. 2005; Liang et al. 2018). Furthermore, many organisms have highly compact genomes, with ∼88% of the *Escherichia coli* genome (Blattner et al. 1997), and effectively all of many virus genomes, being functional (e.g., >95% of the SARS-CoV-2 genome, Wu et al. 2020).

Although such estimates allow us to approximate the effects of BGS in some model organisms, in which recombination and mutation rates are well known, it is difficult to predict these effects in the vast majority of study systems. Moreover, although the genome-wide mean of *B*, a widely-used measure of BGS effects that measures the level of variability relative to neutral expectation, can range from ∼0.45 in *D. melanogaster* to ∼0.94 in humans (Charlesworth 2013; but see Pouyet et al. 2018), existing demographic inference approaches are usually applied across organisms without considering this important source of differences in levels of bias. Here, we examine the effects of the DFE shape and functional density on two common demographic inference approaches—the multiple sequentially Markovian coalescent (MSMC) and *fastsimcoal2*. Finally, we propose an extension within the approximate Bayesian computation (ABC) framework to address this issue, treating the DFE as a nuisance parameter and demonstrating greatly improved demographic inference even when using directly selected sites alone.

Results and Discussion

Effects of SNP Numbers, Density, and Genome Size on Inference under Neutral Equilibrium

The accuracy and performance of demographic inference were evaluated using two popular methods, MSMC (Schiffels and Durbin 2014) and *fastsimcoal2* (Excoffier et al. 2013). In order to assess performance, it was first necessary to determine how much genomic information is required to make accurate inference when the assumptions of neutrality are met. Chromosomal segments of varying sizes (1 Mb,

10 Mb, 50 Mb, 200 Mb, and 1 Gb) were simulated under neutrality and demographic equilibrium (i.e., a constant population size of 5,000 diploid individuals) with 100 independent replicates each. For each replicate, this amounted to the mean [SD] number of segregating sites for each diploid individual being 1,944 [283], 9,996 [418], 40,046 [957], and 200,245 [1,887]; for 50 diploid individuals, these values were 10,354 [225], 51,863 [567], 207,118 [1,139], and 1,035,393 [2,476] for 10 Mb, 50 Mb, 200 Mb, and 1 Gb, respectively. Use of MSMC resulted in incorrect inferences for all segments smaller than 1 Gb (supplementary figs. 1 and 2, Supplementary Material online). Specifically, very strong recent growth was inferred instead of demographic equilibrium, although ancestral population sizes were correctly estimated. In addition, when two or four diploid genomes were used for inference, MSMC again inferred a recent many-fold growth for all segment sizes even when the true model was equilibrium, but performed well when using one diploid genome with large segments (supplementary figs. 1 and 2, Supplementary Material online). These results suggest caution when performing inference with MSMC on smaller regions or genomes, specifically when the number of SNPs is less than ∼200,000 per single diploid individual. Extra caution should be used when interpreting population size changes inferred by MSMC when using more than one diploid individual.

When using *fastsimcoal2* to perform demographic inference, parameters were accurately estimated for all chromosomal segment sizes when the correct model (i.e., equilibrium) was specified (supplementary table 1, Supplementary Material online). However, when model selection was performed using a choice of four models (equilibrium, instantaneous size change, exponential size change, and instantaneous bottleneck), the correct model was chosen more often (∼30% of replicates) when the simulated chromosome sizes were small (1 and 10 Mb), whereas an alternative model of either instantaneous size change or instant bottleneck was increasingly preferred for larger regions (supplementary tables 2 and 3, Supplementary Material online), although the estimates of ancestral sizes were correct. This finding suggests that the nonindependence of SNPs may result in model mis-identification. Indeed, since the model choice procedure assumes that SNPs are independent, the true number of independent SNPs is overestimated, which results in an overestimation in the confidence of the model choice with an increasing amount of data. However, it is interesting to note that the parameter values underlying the non-constant size preferred model were often pointing towards a constant population size (see below). When model selection was performed using sparser SNP densities (i.e., 1 SNP per 5, 50, or 100 kb), the correct model was recovered for longer chromosomes up to 200 Mb (supplementary tables 2 and 3 and figs. 3 and 4, Supplementary Material online), although model selection was slightly less accurate for smaller chromosomes due to the decrease in the total amount of data. As suspected, the biases introduced by the nonindependence of SNPs were found to be concordant with the level of linkage disequilibrium among SNPs used for the analysis

(for ten SNP windows, in which SNPs were separated by 50 kb [100 kb], mean $r^2 = 0.027$ (0.020), compared with the all-SNP mean r^2 of 0.118, and to the completely unlinked SNPs mean r^2 of 0.010; supplementary table 4, Supplementary Material online). Additionally, AIC performed on partially linked SNPs may impose an insufficient penalty on a larger number of parameters, resulting in an undesirable preference for parameter-rich models. We found that implementing a more severe penalty improved inference considerably, even for 1-Gb chromosome sizes (supplementary tables 5 and 6, Supplementary Material online). This model selection performance, the potential corrections related to increased penalties, as well as the total number of SNPs and SNP thinning, should be investigated on a case-by-case basis in empirical applications, owing to the contribution of multiple underlying parameters (e.g., chromosome length, recombination rates, and SNP densities).

In the light of this performance assessment, all further analyses were restricted to characterizing demographic inference on data that far exceeded 1 Gb and roughly matched the structure and size of the human genome—for every diploid individual, 22 chromosomes (autosomes) of size 150 Mb each were simulated, which amounted to roughly 3 Gb of total sequence. Ten independent replicates of each parameter combination were performed throughout, and inference utilized one and fifty diploid individuals for MSMC and *fastsimcoal2*, respectively.

Effect of the Strength of Purifying Selection on Demographic Inference

In order to test the accuracy of demographic inference in the presence of BGS, all 22 chromosomes were simulated with exons of size 350 bp each, with varying sizes of introns and intergenic regions (see Materials and Methods) in order to vary the fraction (5%, 10%, and 20%) of the genome under selection. Because the strength of selection acting on deleterious mutations affects the distance over which the effects of BGS extend, demographic inference was evaluated for various DFEs (table 1). The DFE was modeled as a discrete distribution with four fixed classes: $0 \leq 2N_{anc}s < 1$, $1 \leq 2N_{anc}s < 10$, $10 \leq 2N_{anc}s < 100$, and $100 \leq 2N_{anc}s < 2N_{anc}$, where N_{anc} is the ancestral effective population size and s is the reduction in the fitness of the homozygous mutant relative to wildtype. The fitness effects of mutations were uniformly distributed within each bin, and assumed to be semidominant, following a multiplicative fitness model for multiple loci; the DFE shape was altered by varying the proportion of mutations belonging to each class, given by f_0, f_1, f_2, and f_3, respectively (see Materials and Methods). Three DFEs highly skewed towards a particular class were initially used to assess the impact of the strength of selection on demographic inference (with the remaining mutations equally distributed among the other three classes): DFE1: a DFE in which 70% of mutations have weakly deleterious fitness effects (i.e., $f_1 = 0.7$); DFE2: a DFE in which 70% of mutations have moderately deleterious fitness effects (i.e., $f_2 = 0.7$); and DFE3: a DFE in which 70% of mutations have strongly deleterious fitness effects (i.e., $f_3 = 0.7$). A DFE with

Table 1. Proportion (f_i) of Mutations in Each Class of the Discrete Distribution of Fitness Effects (DFE) Simulated in This Study.

	f_0	f_1	f_2	f_3
DFE1	0.1	0.7	0.1	0.1
DFE2	0.1	0.1	0.7	0.1
DFE3	0.1	0.1	0.1	0.7
DFE4	0.25	0.25	0.25	0.25
DFE5	0.5	0.0	0.0	0.5
DFE6	0.7	0.0	0.0	0.3

equal proportions of all deleterious classes (i.e., DFE4: $f_0 = f_1 = f_2 = f_3 = 0.25$) was also simulated to evaluate the combined effect of different selective strengths. In addition, two bimodal DFEs consisting of only the neutral and the strongly deleterious class of mutations were simulated to characterize the role of strongly deleterious mutations (DFE5: a DFE in which 50% of mutations have strongly deleterious effects (i.e., $f_3 = 0.5$) with the remaining being neutral; and DFE6: a DFE in which 30% of mutations were strongly deleterious (i.e., $f_3 = 0.3$) with the remaining being neutral).

In order to understand the effects of BGS, exonic sites were masked, and only linked neutral intergenic and intronic sites were used for demographic inference by both MSMC and *fastsimcoal2* (although comparisons are presented under certain models to analyses based on nonmasked data sets). The three demographic models examined were: 1) demographic equilibrium, 2) a 30-fold exponential growth, mimicking the recent growth experienced by European human populations, and 3) ∼6-fold instantaneous decline, mimicking the out-of-Africa bottleneck in human populations (fig. 1a). Although these models were parameterized using previous estimates of human demographic history (supplementary table 7, Supplementary Material online; Gutenkunst et al. 2009), these basic demographic scenarios are applicable to many organisms, although the magnitudes of population size changes in this case may represent an extreme. Under neutrality, inference of parameters of all three simulated demographic models was highly accurate with both MSMC and *fastsimcoal2* (fig. 1a and supplementary table 8, Supplementary Material online). However, when inferring parameters using *fastsimcoal2*, the time of change in case of the population decline model was consistently overestimated when SNPs separated by 5 kb were used, whereas the time was accurately inferred when using all SNPs (supplementary table 8, Supplementary Material online). We therefore present our results using all SNPs throughout (with comparisons to one SNP per 5 kb and one SNP per 100 kb thinning, under certain models), and recommend caution when implementing thinning procedures.

Under demographic equilibrium, when 20% of the genome experiences direct selection (with masking of the directly selected sites), we found the true population size to be underestimated as expected, and recent population growth misinferred (fig. 1 and supplementary fig. 5, Supplementary Material online), even when only one SNP per 100 kb was used and a higher AIC penalty was employed (supplementary fig. 6, Supplementary Material online). Conversely, when the
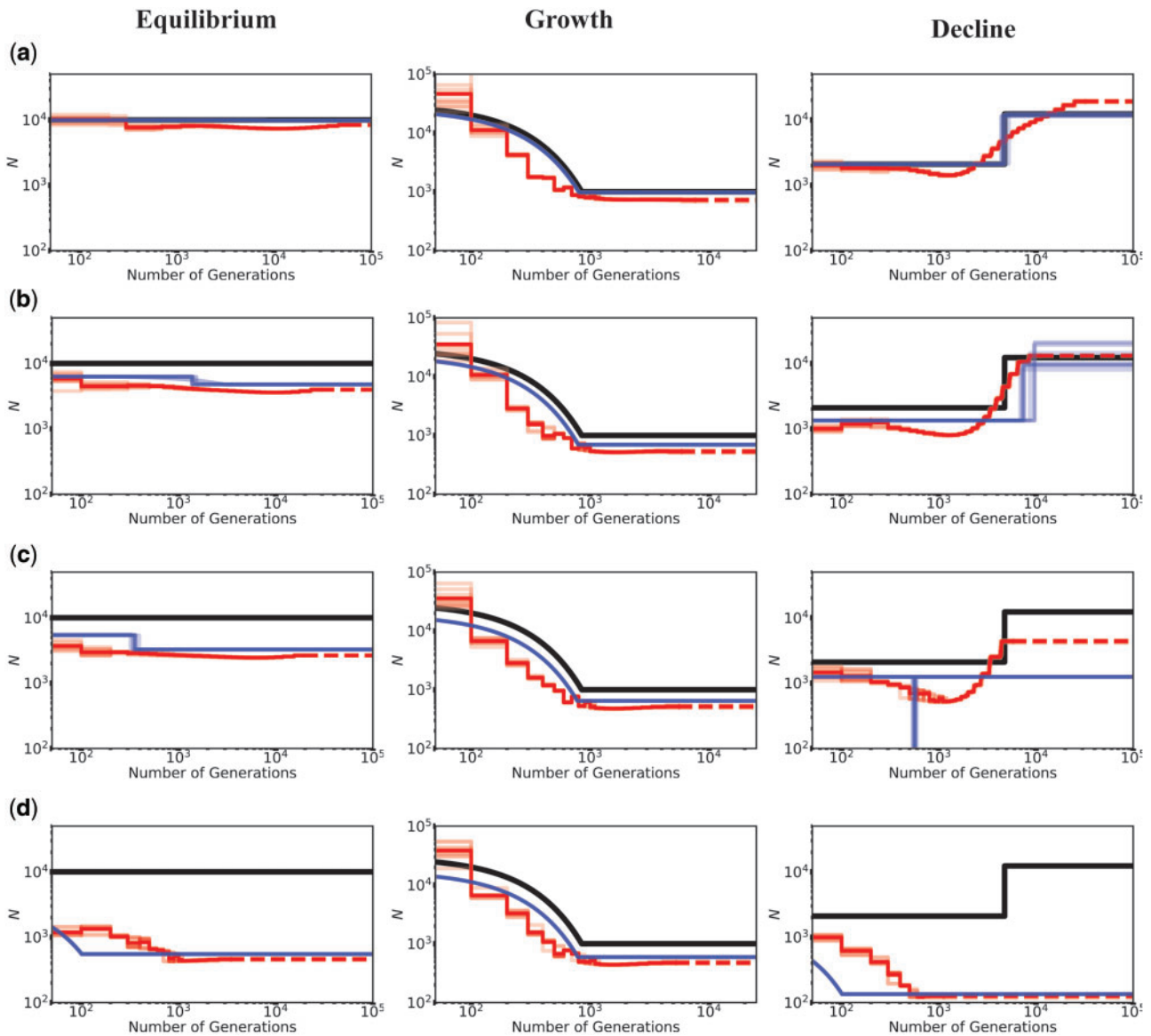
FIG. 1. Inference of demography by MSMC (red lines; ten replicates) and *fastsimcoal2* (blue lines; ten replicates) with and without BGS, under demographic equilibrium (left column), 30-fold exponential growth (middle column), and ∼6-fold instantaneous decline (right column). The true demographic models are depicted as black lines, with the *x*-axis origin representing the present day. (*a*) All genomic sites are strictly neutral. Exonic sites experience purifying selection specified by (*b*) DFE1, (*c*) DFE2, and (*d*) DFE3 (see table 1). Exons represent 20% of the genome, and exonic sites were masked (i.e., excluded) when performing demographic inference, quantifying the effects of BGS alone. The dashed lines represent indefinite extensions of the ancestral population sizes. Detailed methods including command lines can be found at: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/Figure1.txt.

true demographic model was characterized by a recent 30-fold growth, demographic inference was accurate and performed equally well for both MSMC and *fastsimcoal2*, with the exception of a slight underestimation of the ancestral population size for all DFE types. When the true model was population decline, weakly deleterious mutations alone did not affect inference drastically with either method, and it was possible to recover the true model (i.e., decline vs. growth) by *fastsimcoal2* in all replicates (supplementary fig. 7, Supplementary Material online). However, moderately and strongly deleterious mutations resulted in an underestimation of population size and the inference of an instantaneous bottleneck and strong recent growth,

respectively, to the extent that population decline was misinterpreted as a bottleneck/growth in all replicates (supplementary figs. 5 and 7, Supplementary Material online). Strong recent growth was inferred (in the presence of moderately and strongly deleterious mutations) even when SNPs separated by 100 kb were used, and an increased penalty was employed against parameter-rich models (supplementary fig. 6, Supplementary Material online). We further tested the effect of BGS on demographic inference when changes in population size were less severe, namely, when population growth and decline were only 2-fold, with qualitatively similar results (supplementary fig. 8, Supplementary Material online).
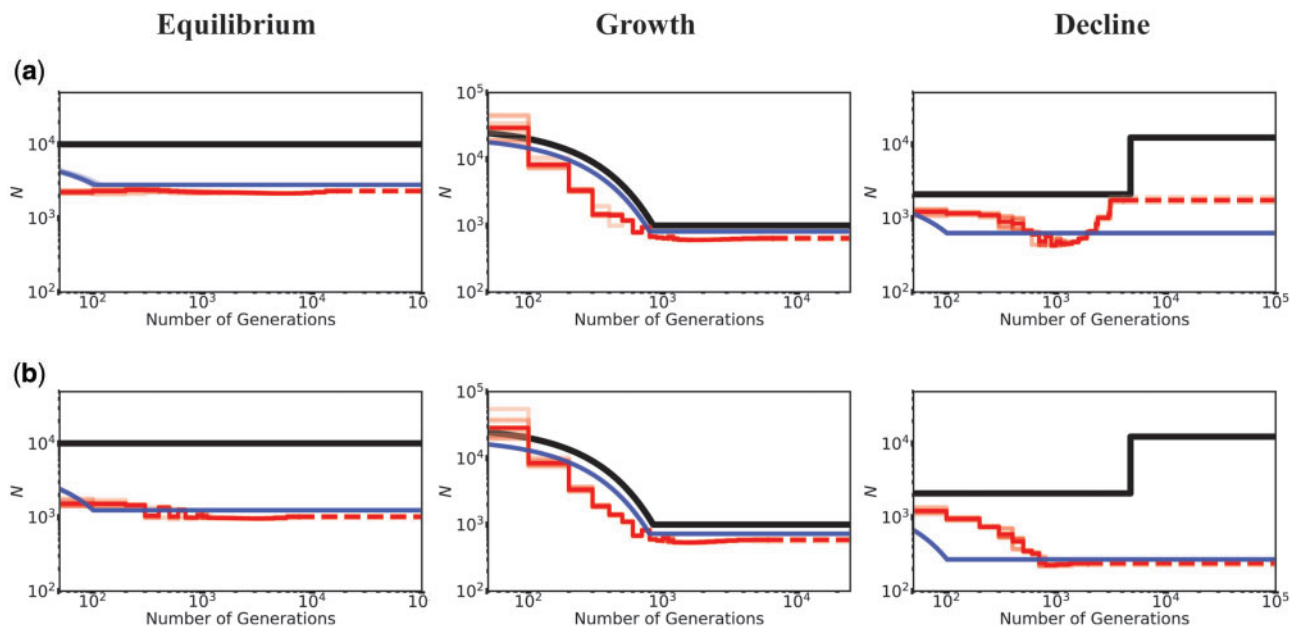
FIG. 2. Inference of demography by MSMC (red lines; ten replicates) and *fastsimcoal2* (blue lines; ten replicates) in the presence of BGS generated by strongly deleterious mutations. Directly selected sites comprised 20% of the genome and were masked when performing demographic inference. Exons experience purifying selection specified by (*a*) DFE6 and (*b*) DFE5 (see table 1). The true demographic models are given as black lines, with the *x*-axis origin representing the present day. The dashed lines represent indefinite extensions of the ancestral population sizes. Detailed methods including command lines can be found at: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/Figure2.txt.

Finally, given the strong evidence that most organisms have a bimodal DFE with a significant proportion of strongly deleterious or lethal mutations (Sanjuán 2010; Jacquier et al. 2013; Kousathanas and Keightley 2013; Bank et al. 2014; Charlesworth 2015; Galtier and Rousselle 2020), we investigated the effect of this strongly deleterious class further. Thus, for comparison with the above, we simulated a rather extreme case in which 30% or 50% of exonic mutations were strongly deleterious with fitness effects uniformly sampled between $100 \leq 2N_{anc}s < 2N_{anc}$, with the remaining mutations being neutral (i.e., DFE5 and DFE6; see table 1). As with the above results, both equilibrium and decline models were falsely inferred as growth, with an order of magnitude underestimation of the true population size (fig. 2).

In sum, neglecting BGS frequently results in the inference of population growth, almost regardless of the true underlying demographic model.

Effects of Density and Inclusion/Exclusion of Directly Selected Sites on Inference

Although we have shown that the presence of purifying selection biases demographic inference, the extent of mis-inference necessarily depends on the fraction of the genome experiencing direct selection. We therefore compared models in which 5%, 10%, or 20% of the genome was functional. For this comparison, equal proportions of mutations in each DFE bin were assumed corresponding to DFE4 (table 1). As before, when the true model was growth, inference was unbiased, with a slight underestimation of ancestral population size when 20% of the genome experienced selection (fig. 3). Population decline was inferred reasonably well if less than

10% of the genome experienced direct selection, but could be mis-inferred as growth with greater functional density, as shown in figure 3. Similarly, the extent to which population size was underestimated at demographic equilibrium increased with the fraction of the genome under selection. Finally, it is noteworthy that many changes in population size that were falsely inferred were greater than 2-fold in size, suggesting the need for great caution when inferring such changes from real data.

Importantly, the results presented do not significantly differ between inference performed while including directly selected sites (i.e., no masking of functional regions; supplementary fig. 9, Supplementary Material online) versus inference performed using linked neutral sites (i.e., masking functional regions; figs. 1–3). These results suggest that the exclusion of exonic sites, which is often assumed to provide a sufficiently neutral data set to enable accurate demographic inference, is not necessarily a satisfactory solution unless gene density is low. For example, demographic inference would naturally be expected to be less biased by BGS for human-like genomes with a relatively low functional density, and more biased in genomes with higher functional density like *D. melanogaster*.

Effect of BGS on Model Selection and Inferred Time of Size Change Using *fastsimcoal2*

In order to quantify the effects of BGS on model selection, four competing models were used for inference: equilibrium, instantaneous size change (growth/decline), exponential size change (growth/decline), and an instantaneous bottleneck. Although demographic equilibrium was almost always
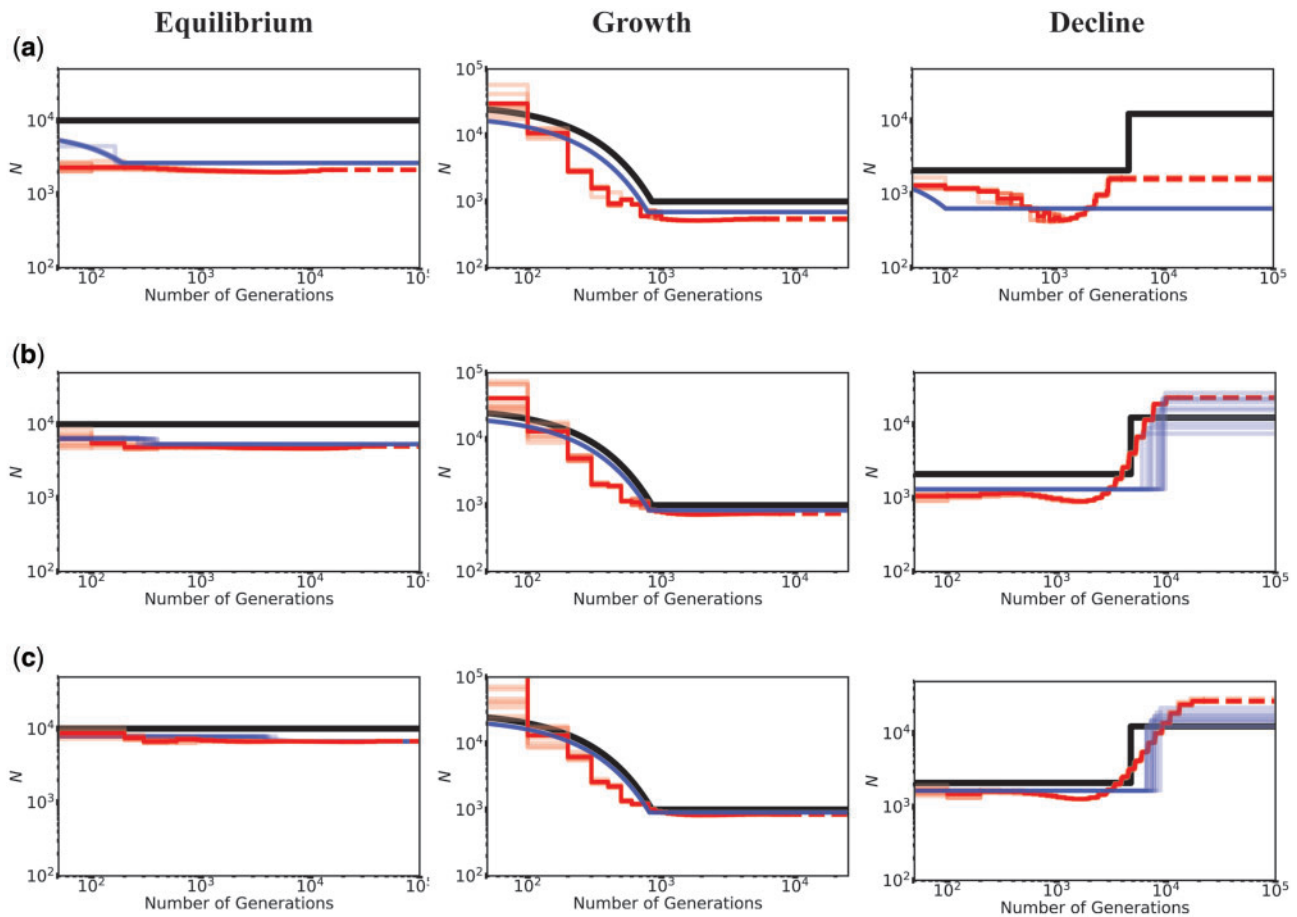
Fig. 3. Inference of demography by MSMC (red lines; ten replicates) and *fastsimcoal2* (blue lines; ten replicates) in the presence of BGS with varying proportions of the genome under selection, for demographic equilibrium (left column), exponential growth (middle column), and instantaneous decline (right column). Exonic sites were simulated with purifying selection with all f_i values equal to 0.25 (DFE4; see table 1), and were masked when performing inference. Directly selected sites comprise (*a*) 20% of the simulated genome, (*b*) 10% of the simulated genome, and (*c*) 5% of the simulated genome. The true demographic models are given by the black lines, with the *x*-axis origin representing the present day. The dashed lines represent indefinite extensions of the ancestral population sizes. Detailed methods including command lines can be found at: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/Figure3.txt.

inferred as an instantaneous size change (70–100% of replicates), the fitted parameters of the size change model were nearly indistinguishable from the correct model (fig. 1*a*). In other words, the inferred size change was so inconsequential so as to be nearly a constant-size model, suggesting that parameter estimation is usually more reliable than model selection. When there was a substantial proportion of highly deleterious mutations (DFE3 and DFE5), exponential growth was generally inferred. However, when there was a true size change, *fastsimcoal2* performed well in distinguishing between exponential versus instantaneous change models even in the presence of BGS (supplementary figs. 5 and 6, Supplementary Material online), provided that the magnitude of size change was large. When size changes were on the order of 2-fold, exponential growth was consistently inferred to be instantaneous.

With respect to model choice between growth and decline in the presence of BGS (irrespective of instantaneous vs. exponential change), as the density of selected sites and strength of purifying selection increased, both equilibrium and decline models were more likely to be inferred as growth

and occasionally as instantaneous bottlenecks (supplementary fig. 7, Supplementary Material online), whereas true growth models were generally chosen correctly. It should be added that with such large chromosome sizes (3 Gb of total sequence data), model selection was not observed to vary between replicates using *fastsimcoal2* for any given parameter combination. Thus, in the presence of BGS, high-confidence calls of an incorrect underlying demographic model appear likely.

With regard to the time of inferred size change, when the true model was exponential growth, the model was always correctly identified and inference of the time of change was slightly underestimated in the presence of BGS (supplementary figs. 10 and 11, Supplementary Material online), consistent with the fact that BGS will further skew the site frequency spectrum towards rare alleles. When the true model was decline, and the model was correctly identified as such, the time of change was modestly overestimated (supplementary figs. 12 and 13, Supplementary Material online)—up to ∼2-fold for 6× growth and 2.5-fold for 2× growth (when 20% of the genome was exonic).

Effect of Heterogeneity in Recombination Rates, Mutation Rates, and Repeat Masking

Variation in recombination and mutation rates, as well as the masking of repeat regions, may also affect demographic inference procedures. We evaluated this issue by simulating heterogeneity in both mutation and recombination rates (based on estimated human genome maps, as described in Materials and Methods), and masking 10% of each simulated segment drawing from the empirical distribution of repeat lengths in the human genome (supplementary fig. 14, Supplementary Material online). In general, inferences under neutrality (supplementary figs. 15–17, Supplementary Material online) as well as under BGS (supplementary figs. 18–20, Supplementary Material online) were not affected to a great extent, suggesting such heterogeneity to have a comparatively minor role for the parameter space considered in this study. Thus, serious mis-inference is more likely to be caused by selection. These observations also suggest that simulations performed with mean rates of recombination and mutation, as in this study, are sufficient to evaluate biases caused by BGS.

Effects of BGS on Diversity and the SFS under Various Demographic Models: Theoretical Expectations versus Simulation Results

To better understand how BGS can lead to different biases in the inference of population history, we investigated the extent of BGS effects under all three demographic models, with respect to both the expected diversity in the presence of BGS relative to neutrality (B), as well as the shape of the SFS at linked neutral sites. First, we found that B differed among demographic scenarios, with much lower values in the case of equilibrium and decline, concordant with stronger demographic mis-inference (fig. 4). After a population decline, B was lower than that before the size change; whereas after population expansion, B increased relative to that in the ancestral population, sometimes approaching 1 (fig. 4). This may seem paradoxical, given that the magnitude of the scaled selection coefficient ($2N_e s$) decreases with decreasing N_e (i.e., the efficacy of purifying selection decreases, and could thus be expected to result in larger values of B under population decline). Conversely, with increasing N_e, B should be expected to reduce.

However, these expectations apply only once a population has maintained a given N_e for sufficient time such that mutation-drift-equilibrium has been approached. During the initial stages of population size change, and shortly afterwards, the dynamics of B tend to show a trend opposing these long-term expectations (see also Figure 5 of Torres et al. 2020). This is because differences in N_e caused by different initial levels of BGS cause differences in the rates of response to changes in population size—a small value of N_e (corresponding to low B) results in a faster response compared with a high value (Fay and Wu 1999; Hey and Harris 1999; Pool and Nielsen 2007, 2009; Campos et al. 2014; Torres et al. 2020). In other words, diversity in a growing population will increase more rapidly in regions experiencing stronger

BGS than in completely neutral regions, whereas diversity in a declining population will decrease at a faster rate in regions with BGS relative to those with neutrality, resulting in temporarily higher and lower B, respectively. The relative diversity values observed with different initial equilibrium B values after a short period of population size change may thus be very different from both the initial and final equilibrium values. The overall effect is that there is an apparent increase in B immediately following a population decline, and a decrease immediately following an expansion. Analytical models describing these effects are presented in Appendix. These models used the simulated values of B at equilibrium before the population size changes to predict the apparent B values at the ends of the periods of size change (see Materials and Methods and Appendix). It can be seen from figure 4 that there is good agreement between these predictions and the simulation results.

Because several demographic estimation methods are based on fitting a demographic model to the SFS, it is also of interest to determine whether BGS can skew the SFS to different extents under different demographic models. Although it is well known that BGS causes a skew of the SFS towards rare variants under equilibrium models (Charlesworth et al. 1995; Nicolaisen and Desai 2013), the effect of BGS on the SFS with population size change has not been much explored (but see Johri et al. 2020; Torres et al. 2020). As shown in figure 5, with a population size decline, the SFS of derived alleles is more skewed towards rare variants when BGS is operating, especially when B is initially small, since the effects of BGS work in opposition to the effects of the population size reduction. This difference in the left skew of the SFS with and without BGS is much less noticeable in the case of population expansion, since here the effects of BGS and the expansion act in a similar direction.

As with the estimates of the apparent B values discussed above, analytical predictions of the expected SFS after an instantaneous/exponential change in population size can be made, using the values of B and the SFS at equilibrium in the ancestral population before the population size change using the formulae of Polanski and Kimmel (2003) and Polanski et al. (2003) for the purely neutral case, as described in Materials and Methods. Importantly, the use of the B parameter does not in itself cause a skew in the SFS, it merely affects overall diversity values. Figure 5 shows that the overall shape of the SFS is predicted reasonably well by the analytical results, although deviations are to be expected for the rare allele classes, which are the most sensitive to demographic change and selection. Overall, the results imply that BGS is more likely to bias demographic inference post-decline compared with post-expansion, consistent with the performance of the methods described above. Although it is notable that the SFS can be reasonably well predicted by correcting for the rescaling effects of BGS if the effects of BGS in the ancestral population are accurately known, the exact allele frequency patterns observed will depend on the timing of population size changes relative to the time of sampling, as well as the value of B prior to the size change.
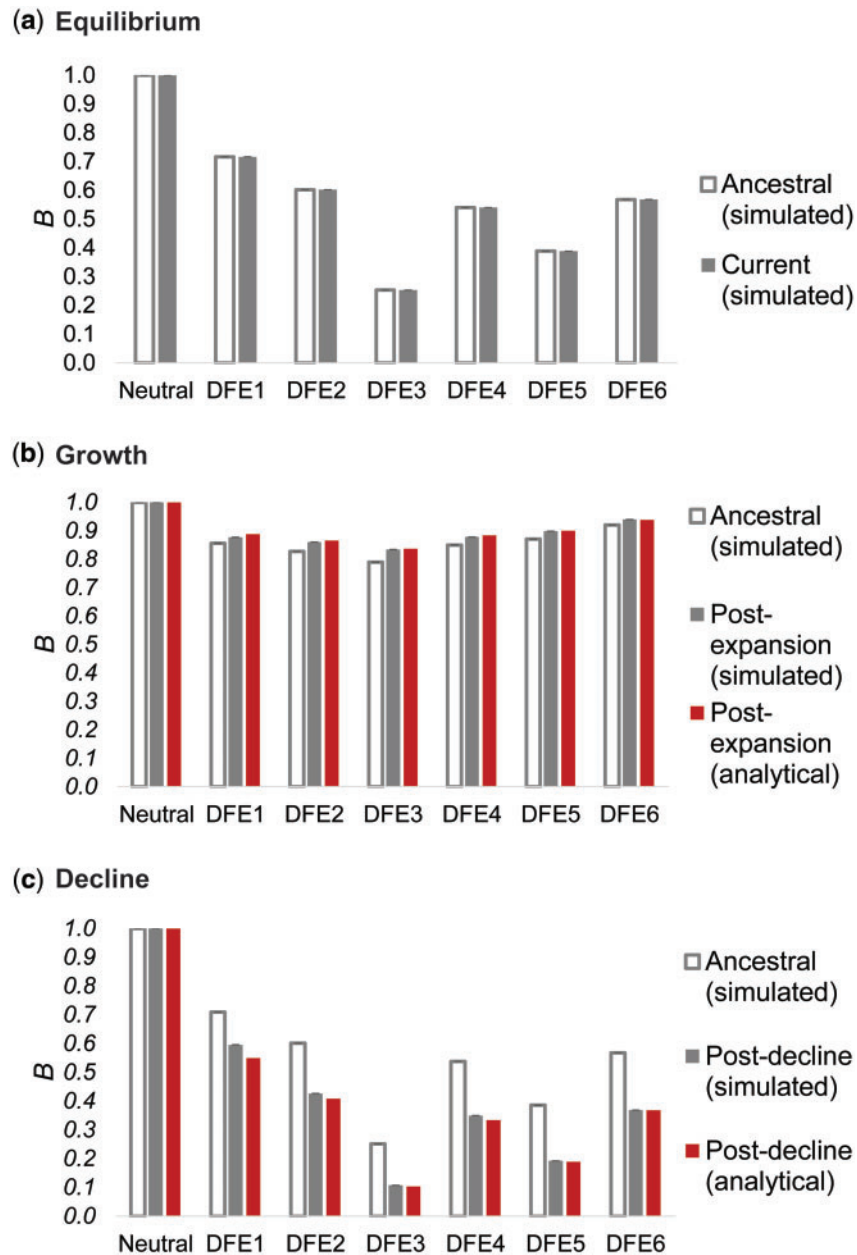
FIG. 4. Nucleotide site diversity with BGS (*B*) relative to its purely neutral expectation (π_0) for varying DFEs (specified in table 1) and demographic scenarios. The results are shown for (*a*) demographic equilibrium, (*b*) population growth, and (*c*) population decline. All cases refer to size changes forward in time, the ancestral *B* (i.e., *B* pre-change in population size) is shown in white bars, *B* post-change in population size is shown in solid gray bars, and the analytical expectations for the post-size change *B* are shown as red bars. Exonic sites comprised ~10% of the genome, roughly mimicking the density of the human genome. Detailed methods including command lines can be found at: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/Figure4.txt.

The patterns described here thus represent only a small subset of the possibilities.

A Potential Solution: Averaging across All Possible DFEs

As shown above, demographic inference can be strongly affected by BGS effects that have not been taken into account, as well as by direct purifying selection. A potential solution is thus to correct for these effects when performing inferences of population history. A widely-used approach to estimating direct selection effects, DFE-alpha, takes a stepwise approach to inferring demography, by using a presumed neutral class

(synonymous sites); conditional on that demography, it then estimates the parameters of the DFE (Keightley and Eyre-Walker 2007; Eyre-Walker and Keightley 2009; Schneider et al. 2011; Kousathanas and Keightley 2013). However, this approach does not include the possibility of effects of selection at linked sites, which can result an overestimate of population growth, and while the DFE may not be mis-inferred strongly (Kim et al. 2017), there is substantial mis-inference of the DFE if synonymous sites experience direct selection (Johri et al. 2020).

Building on this idea, Johri et al. (2020) recently proposed an approach that includes both direct and background effects
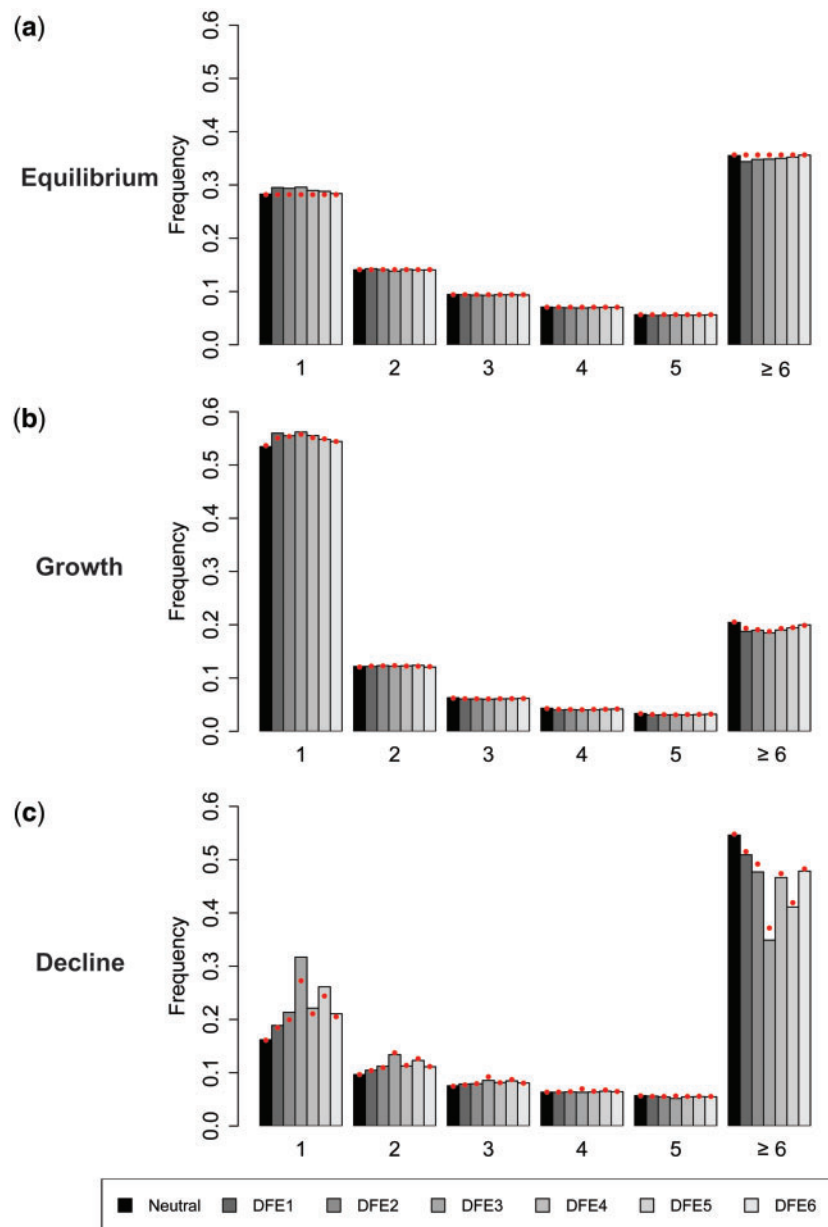
FIG. 5. The site frequency spectrum (SFS) of derived allele frequencies at neutral sites from ten diploid genomes under (a) demographic equilibrium, (b) population growth, and (c) population decline, under the same DFEs as shown in figure 4. The x-axis indicates the number of sampled alleles (out of 20) carrying the derived variant. Exonic sites comprised ~10% of the genome, roughly mimicking the density of the human genome. The red solid circles give the values predicted analytically with a purely neutral model, but correcting for BGS by using the B values of the ancestral population (i.e., pre-change in population size) obtained from simulations, in order to quantify the effective population size. Detailed methods including command lines can be found at: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/Figure5.txt.

of purifying selection, and simultaneously infers the deleterious DFE and demography. By utilizing the decay of BGS effects around functional regions, they demonstrated high accuracy under the simple demographic models examined. Moreover, the method makes no assumptions about the neutrality of synonymous sites, and can thus be used to estimate selection acting on these sites, as well as in noncoding functional elements. However, this computationally intensive approach is specifically concerned with jointly inferring the DFE and demographic

parameters. As such, if an unbiased characterization of the population history is the sole aim, this procedure may be needlessly involved. We thus here examine the possibility of instead treating the DFE as an unknown nuisance parameter, averaging across all possible DFE shapes, in order to assess whether demographic inference may be improved simply by correcting for these selection effects without inferring their underlying parameter values. This approach utilizes functional (i.e., directly selected) regions, a potential advantage in populations for

which only coding data may be available (e.g., exome-capture data; see Jones and Good 2016), or more generally in organisms with largely functional genomes.

In order to illustrate this approach, a functional genomic element was simulated under demographic equilibrium, 2-fold exponential population growth and 2-fold exponential population decline with four different DFE shapes (as described previously, and shown in fig. 6). A number of summary statistics were calculated (see Materials and Methods) for the entire functional region. Inference was first performed assuming strict neutrality, and inferring a one-epoch size change (thus estimating the ancestral [N_{anc}] and current population sizes [N_{cur}]). As was found with the other inference approaches examined, population sizes were underestimated and a false inference of population growth was observed in almost all cases when selective effects are ignored (fig. 6).

Next, the assumption of neutrality was relaxed, and mutations were simulated with fitness effects characterized by a discrete DFE, with the fitness classes used above (f_0, f_1, f_2, f_3). Values for f_i were drawn from a uniform prior between 0 and 1, such that $\sum f_i = 1$. Note that no assumptions were made about which sites in the genomic region were functionally important, or regarding the presence/absence of a neutral class. These directly selected sites were then used to infer demographic parameters. We found that, by varying the shape of the DFE, averaging across all realizations, and only estimating parameters related to population history, highly accurate inference of modern and ancestral population sizes is possible (fig. 6). These results demonstrate that, even if the true DFE of a population is unknown (as will always be the case in reality), it is possible to infer demographic history with reasonable accuracy by approximately correcting for these selective effects.

This proposed method is most applicable to organisms in which recombination rates are reasonably well known. If the assumed recombination rate is 2-fold lower than the true rate, the ABC approach infers growth by overestimating the current population size; correspondingly, if the assumed recombination rate is higher than the true rate, the current population size is underestimated (supplementary fig. 21, Supplementary Material online). Interestingly, in both cases the ancestral population sizes are correctly inferred, consistent with previous results (Johri et al. 2020).

Conclusions

Although commonly used approaches for inferring demography assume neutrality and independence among segregating sites, these assumptions are likely to be violated in practice. In addition, there is considerable evidence for wide-spread effects of selection at linked sites in many commonly studied organisms (Hernandez et al. 2011; Cutter and Payseur 2013; Williamson et al. 2014; Elyashiv et al. 2016; Campos et al. 2017; Booker and Keightley 2018; Pouyet et al. 2018; Ragsdale et al. 2018; Torres et al. 2018; Castellano et al. 2020). Accordingly, we have explored how violations of the assumption of neutrality may affect demographic inference, particularly with regard to the underlying strength of purifying selection and

the genomic density of directly selected sites. Generally speaking, the neglect of these effects (i.e., background selection) results in an inference of population growth, with the severity of the growth model roughly scaling with selection strength and density, as well as the inference of historical bottlenecks with some frequency. Thus, when the true underlying model is in fact growth, demographic mis-inference is not particularly severe; when the true underlying model is constant size or decline, the mis-inference can be extreme, with a many-fold underestimation of population size.

However, given that BGS will lead to the false inference of recent growth nearly regardless of the true history, it would be difficult in practice to determine the accuracy of this model without independent information on any given empirical application. Moreover, as the two very different methods investigated here result in highly similar mis-inference, we propose that this performance is unlikely to be a feature of these specific approaches, but rather a quantification of the fact that the underlying genealogies are distorted in the presence of BGS. Thus, these problems are likely to be common to all demographic inference based on polymorphism data.

It is important to note that BGS effects extend over genomic distances in a way that is positively related to the strength of purifying selection. For instance, strongly and moderately deleterious mutations affect patterns of diversity at large genomic distances, whereas mildly deleterious mutations primarily skew allele frequencies at adjacent sites. Thus, if intergenic regions further away from exons are used to perform demographic inference, it is predominantly moderately deleterious mutations that are likely to bias inferences; if these are relatively rare, they may not cause significant problems. In contrast, if synonymous sites are used to infer demographic history, mildly deleterious mutations arising in the coding sequences to which they belong may have significant effects. As we have focused here on relatively sparsely-coding genomes (with human-like gene densities) and used intergenic sites for inference, moderately deleterious mutations resulted in more severe mis-inference. The effect of the decay with distance of BGS due to mildly deleterious mutations depends on multiple parameters. For instance, with an exon of length 500 bp and Drosophila-like parameters (e.g., $N_e = 10^6$; recombination rate = mutation rate = 10^{-8}/site/generation), B increases from 0.53 (at ten bases from the end of the exon) to 0.94 at a distance of 1,000 bases. On the other hand, with human-like parameters ($N_e = 10^4$; recombination rate = mutation rate = 10^{-8}/site/generation) the corresponding change in B is only from 0.981 to 0.982 (supplementary table 9, Supplementary Material online).

Thus, mildly deleterious mutations have drastically different effects, depending on the underlying population parameters. Although these results certainly suggest that demographic inference ought to be less biased by BGS in neutral regions very distant from functional elements (for species with sufficiently high recombination/functionally sparse genomes), it is noteworthy that purifying selection on moderately and strongly deleterious mutations can have long-range effects, and that the complex interaction of population history with purifying/background selection
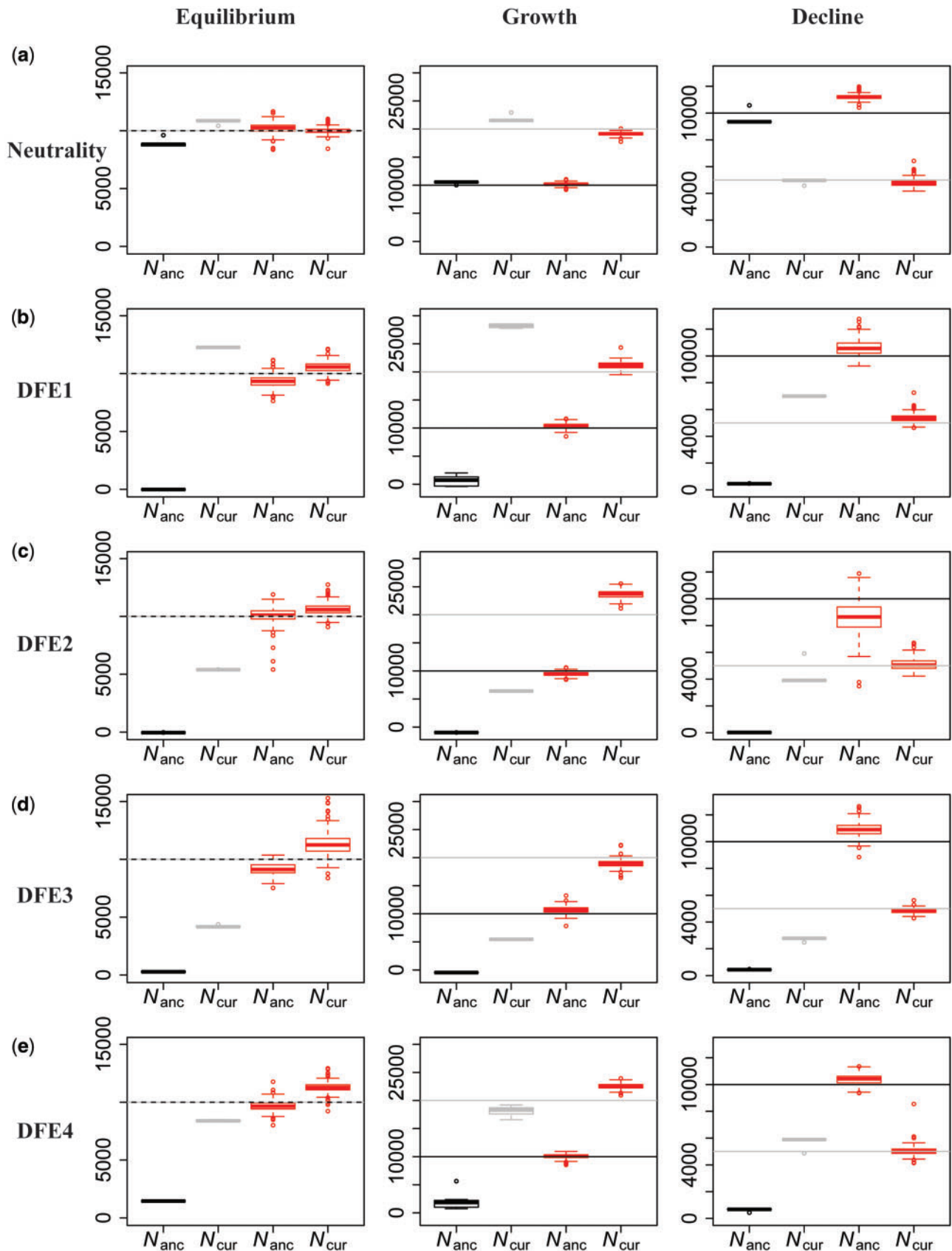
Fig. 6. Comparison of estimates of ancestral (N_{anc}) and current (N_{cur}) population sizes when assuming neutrality versus when varying the DFE shape as a nuisance parameter, using an ABC framework. Inference is shown for demographic equilibrium (left column), 2-fold exponential growth (middle column), and 2-fold population decline (right column), for five separate DFE shapes that define the extent of direct purifying selection acting on the genomic segment for which demographic inference is performed: (a) neutrality, (b) DFE1, (c) DFE2, (d) DFE3, and (e) DFE4 (see table 1). In each case, the horizontal lines give the true values (black for N_{anc} and gray for N_{cur}) and the box-plots give the estimated values. Black and gray boxes represent estimates when assuming neutrality, whereas red boxes represent estimates when the DFE is treated as a nuisance parameter. Detailed methods including command lines can be found at: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/Figure6.txt.

necessitates a consideration of this topic in any given empirical application.

Comparing the two inference methods investigated here, it appears that *fastsimcoal2* is less prone to inferring false fluctuations in population size. However, both methods falsely infer growth in the presence of BGS, with increasing severity as the density of coding regions increases. The times of population growth inferred by both methods appear to be affected in unpredictable ways when the inferred model is incorrect. When the general model is correctly identified, BGS leads to inference of more recent growth, and more ancient decline, than the reality. In addition, although variation in mutation and recombination rates across the genome alone did not strongly affect demographic inference, our evaluations in the current study are restricted to a specific parameter space resembling those of human populations. The effects of this variation on organisms with more extreme rate fluctuations remain in need of investigation.

It is noteworthy that, even when all sites are strictly neutral or only 5% of the genome experiences direct selection, demographic equilibrium is mis-estimated by MSMC as a series of size changes. The pattern of these erroneous size changes lend a characteristic shape to the MSMC curve (i.e., ancient decline and recent growth) which appears to resemble the demographic history previously inferred for the Yoruban population (Schiffels and Durbin 2014), including the time at which changes in population size occurred (supplementary fig. 22). Previous work has demonstrated that the resulting demographic model does not in fact fit the observed SFS in the Yoruban population (Beichman et al. 2017; Lapierre et al. 2017). A similar shape has also been inferred in the vervet subspecies (Warren et al. 2015; fig. 4), in passenger pigeons (Hung et al. 2014; fig. 2), in elephants (Palkopoulou et al. 2018; fig. 4), in *Arabidopsis* (Fulgione et al. 2018; fig. 3), and in grapevines (Zhou et al. 2017; fig. 2A).

Although the inferred population size fluctuations under simulated neutrality are only ∼1.2-fold, in most empirical applications the fluctuations are of a somewhat larger magnitude (∼2-fold in pigeons, *Arabidopsis*, and grapevines). Nonetheless, this performance of MSMC under neutral demographic equilibrium is concerning, and adds to the other previously published cautions concerning the interpretation of MSMC results. For example, Mazet et al. (2016) and Chikhi et al. (2018) demonstrated that, under constant population size with hidden structure, MSMC may suggest false size changes (see also Orozco-terWengel 2016). In addition, MSMC has been reported to falsely infer growth prior to instantaneous bottlenecks (Bunnefeld et al. 2015). In addition, we observed that, if insufficient genomic data are used, or more than one diploid genome is used to perform inference, MSMC falsely infers recent growth of varying magnitudes, the latter having been previously observed by Beichman et al. (2017) and Adrion et al. (2020).

In sum, we find that the effects of purifying and background selection result in similar demographic mis-inference across approaches, and that masking functional sites does not yield accurate parameter estimates. In order to side-step many of these difficulties, our proposed approach

of inferring demography by averaging selection effects across all possible DFE shapes within an ABC framework appears to be promising. Utilizing only functional regions, we found a great improvement in accuracy, without making any assumptions regarding the true underlying shape of the DFE or the neutrality of particular classes of sites. As such, this approach represents a more computationally efficient avenue if only demographic parameters are of interest, and ought to be particularly useful in the great majority of organisms in which independent neutral sites either do not exist, or are difficult to identify and verify.

Materials and Methods

Simulations of Chromosomal Segments under Neutral Equilibrium

When assessing the amount of genomic information required for accurate demographic inference, chromosomal segments of varying sizes (1 Mb, 10 Mb, 50 Mb, 200 Mb, and 1 Gb) were simulated under neutral equilibrium. In all cases, the effective population size (N_e) simulated was 5,000, and mutation and recombination rates were both 1×10^{-8} per site per generation. Simulations were performed with both SLiM 3.1 (Haller and Messer 2019) for a $10N_e$ generation burn-in, and with msprime 0.7.3 (Kelleher et al. 2016). In all cases 100 replicates were simulated, with the exception of 1-Gb chromosomes simulated by SLiM, in which only ten replicates were obtained.

Simulations of Human-Like Chromosomes (with and without Selection)

Simulations were performed using SLiM 3.1 (Haller and Messer 2019) for a burn-in of $10N_{anc}$ generations, with ten replicates per evolutionary scenario. For every replicate, 22 chromosomes of 150 Mb each were simulated, totaling ∼3 Gb of information per individual genome (similar to the amount of information in a human genome). Within each chromosome, three different types of regions were simulated, representing noncoding intergenic, intronic, and exonic regions. Based on the NCBI RefSeq human genome annotation, downloaded from the UCSC genome browser for hg19 (http://genome.ucsc.edu/; Kent et al. 2002), mean values of exon sizes and intron numbers per gene were calculated. To represent mean values for the human genome (Lander et al. 2001), each gene comprised eight exons and seven introns, and exon lengths were fixed at 350 bp. By varying the lengths of the intergenic and intronic regions, three different genomic configurations with varying densities of functional elements were simulated and compared—with 5%, 10%, and 20% of the genome being under direct selection—hereafter referred to as genome5, genome10, and genome20, respectively. Genome5 was comprised of introns of 3,000 bp and intergenic sequence of 31,000 bp, genome10 of introns of 1,500 bp and intergenic sequence of 15,750 bp, whereas genome20 was comprised of introns of 600 bp and intergenic sequence of 6,300 bp. The total chromosome sizes of these genomes were approximately 150 Mb (150,018,599, 150,029,949, and 150,003,699 bp) with 2737, 5164, and 11278 genes per

chromosome in genome5, genome10, and genome20, respectively. In order to be conservative with respect to the performance of existing demographic estimators, intronic and intergenic regions were assumed to be neutral.

Recombination and mutation rates were assumed to be equal to 1×10^{-8}/site/generation. Neither crossover interference nor gene conversion was modeled (see the Discussion in Campos and Charlesworth 2019). Exonic regions in the genomes experienced direct purifying selection given by a discrete DFE comprised of four fixed classes (Johri et al. 2020), whose frequencies are denoted by f_i: f_0, with $0 \leq 2N_es < 1$ (i.e., effectively neutral mutations), f_1, with $1 \leq 2N_es < 10$ (i.e., weakly deleterious mutations), f_2, with $10 \leq 2N_es < 100$ (i.e., moderately deleterious mutations), and f_3, with $100 \leq 2N_es < 2N_e$ (i.e., strongly deleterious mutations), where N_e is the effective population size and s is the reduction in fitness of the mutant homozygote relative to wild-type. Within each bin, the distribution of s was assumed to be uniform. All mutations were assumed to be semidominant. In all cases, the N_e corresponding to the DFE refers to the ancestral effective population size.

Six different types of DFE were simulated, described by the parameters provided in table 1. Three different demographic models were tested for each of these DFEs (supplementary table 7, Supplementary Material online): 1) demographic equilibrium, 2) recent exponential 30-fold growth, resembling that estimated for the human CEU population (Gutenkunst et al. 2009), and 3) ~6-fold instantaneous decline, resembling the out-of-Africa bottleneck in humans (Gutenkunst et al. 2009). For simulations of demographic equilibrium and decline, population sizes and time of change were scaled down by a factor of 10 (with corresponding scaling of the recombination rate, mutation rate, and selection coefficients), whereas simulations of growth were not scaled.

Running MSMC

In order to quantify the effect of purifying selection on demographic inference, we used entire chromosomes generated by SLiM to generate input files for MSMC. For comparison, and in order to quantify the effect of BGS alone on demographic inference, we masked the exonic regions to generate input files. For all parameters, MSMC was performed on a single diploid genome, as the results for this case were the most accurate (supplementary figs. 1 and 2, Supplementary Material online). Input files were made using the script ms2multihetsep.py provided in the msmc-tools-Repository downloaded from https://github.com/stschiff/msmc-tools. MSMC1 and 2 were run as follows:

> msmc_1.1.0_linux64bit -t 5 -r 1.0 -o output_
> genomeID input_chr1.tab input_chr2.tab
> . . . input_chr22.tab.

Population sizes obtained from MSMC were plotted up to the maximum number of generations obtained from MSMC, and the final value of the ancestral population size was extended indefinitely as a dashed line.

Running fastsimcoal2

Inference was performed by masking all exonic SNPs and using all intronic and intergenic SNPs in order to obtain the most accurate estimates. In order to minimize the effects of linkage disequilibrium (LD), SNPs separated by 5 or 100 kb were also used for inference in some cases to assess the impact of violating the assumption of independence. When choosing SNPs separated by a particular distance, the first SNP from each chromosome was chosen and if the distance to the next consecutive SNP was greater than or equal to 5 kb/100 kb, that SNP was included, otherwise the next downstream SNP was evaluated. Site frequency spectra (SFS) were obtained for all sets of SNPs for all ten replicates of every combination of demographic history and DFE. SNPs from all 22 chromosomes were pooled together to calculate the SFS. In the case of SNPs separated by 5 kb/100 kb, the "0" class of the SFS was scaled down by the same extent as the decrease in the total number of SNPs. Fastsimcoal2 was used to fit each SFS to four distinct models: (a) equilibrium, which estimates only a single population size parameter (N); (b) instantaneous size change (decline/growth), which fits three parameters—ancestral population size (N_{anc}), current population size (N_{cur}), and time of change (T); (c) exponential size change (decline/growth), which also estimates three parameters—N_{anc}, N_{cur}, and T; and (d) an instantaneous bottleneck model with three parameters—N_{anc}, intensity, and time of bottleneck. The parameter search ranges for both ancestral and current population sizes in all cases were specified to be uniformly distributed between 100 and 500,000 individuals, whereas the parameter range for time of change was specified to be uniform between 100 and 10,000 generations in all models. The intensity of the bottleneck was sampled from a log-uniform distribution between 10^{-5} and 2. The following command line was used to run fastsimcoal2:

> fsc26 -t demographic_model.tpl -n 150000 -d -e demographic_model.est -M -L 50 -q.

Model selection was performed as recommended by Excoffier et al. (2013). For each demographic model, the maximum of maximum likelihoods from all replicates was used to calculate the Akaike Information Criterion (AIC) = $2 \times$ number of parameters—$2 \times \ln(\text{likelihood}) = 2 \times$ number of parameters—$2 \times \ln(10) \times L_{10}$, where L_{10} is the logarithm (with respect to base 10) of the best likelihood provided by fastsimcoal2. For model choice comparison, we also implemented a stricter penalty of $25\times$ (see supplementary tables 5 and 6, Supplementary Material online), in which case AIC = $25 \times$ number of parameters—$2 \times \ln(\text{likelihood})$. The relative likelihoods (Akaike's weight of evidence) in favor of the ith model were then calculated as:

$$w(i) = \frac{e^{-0.5\Delta_i}}{\sum_{j=1}^{4} e^{-0.5\Delta_j}},$$

where $\Delta_i = AIC_i - AIC_{\min}$. The model with the highest relative likelihood was selected as the best model, and the parameters estimated using that model were used to plot the final inferred demography.

Simulations of Variable Recombination and Mutation Rates, and Repeat Masking

In order to simulate variation in recombination and mutation rates, all 22 chromosomes were simulated by mimicking chromosome 6 ($\sim$171 Mb) of the human genome. Recombination rates (HapMap) obtained from Yoruban populations ([McVean et al. 2004](); [Myers et al. 2005]()) were obtained from the UCSC genome browser, whereas the mutation rate map (https://molgenis26.target.rug.nl/downloads/gonl_public/mutation_rate_map/release2/) was assumed to correspond to estimates obtained from de novo mutations ([Francioli et al. 2015]()), as in [Castellano et al. (2020)](). Absolute values of mutation rates were normalized in order to maintain the mean mutation rate across the genome at $\sim 1.0 \times 10^{-8}$ per site per generation. Recombination and mutation rate estimates were taken from positions of $\sim$10–160 Mb, with the recombination map starting at 10,010,063 bp and the mutation map starting at 10,010,001 bp. Regions with missing data for either of the two estimates were simulated with rates corresponding to the previous window, except for the case of centromeres in which no recombination was assumed. In order to understand the effect of excluding centromeric regions in empirical studies, the 4Mb region corresponding to the centromere was masked, corresponding to 48.5–52.5 Mb of the simulated 150Mb chromosomes. In order to evaluate the effect of masking repeat regions, random segments comprising 10% of each chromosome were masked. The lengths of these segments were drawn from the lengths of repeat regions found in the human genome ([supplementary fig. 14](), [Supplementary Material]() online), as obtained from the repeat regions in the hg19 assembly of the human genome from the UCSC genome browser.

Performing Inference by ABC

ABC was performed using the R package "abc" ([Csilléry et al. 2010]()), and nonlinear regression aided by a neural net (used with default parameters as provided by the package) was used to correct for the relationship between parameters and statistics ([Johri et al. 2020]()). To infer posterior estimates, a tolerance of 0.1 was applied (i.e., 10% of the total number of simulations were accepted by ABC in order to estimate the posterior probability of each parameter). The weighted medians of the posterior estimates for each parameter were used as point estimates. ABC inference was performed under two conditions: 1) complete neutrality, or 2) the presence of direct purifying selection. In both cases, only two parameters were inferred—ancestral (N_{anc}) and current (N_{cur}) population sizes. However, in scenario 2, the shape of the DFE was also varied. Specifically, the parameters f_0, f_1, f_2, and f_3 were treated as nuisance parameters and were sampled such that $0 \leq f_i \leq 1$, and $\Sigma_i f_i = 1$, for $i = 0$–3. In addition, in order to limit the computational complexity involved in the ABC framework, values of f_i were restricted to multiples of 0.05 (i.e., $f_i \epsilon \{0.0, 0.05, 0.10, \ldots, 0.95, 1.0\} \; \forall \; i$), which allowed us to sample 1,771 different DFE realizations. Simulations were performed with functional genomic regions, and the demographic model was characterized by 1-epoch changes in

which the population either grows or declines exponentially from ancestral to current size, beginning at a fixed time in the past.

For the purpose of illustration, and for a contrast with the human-like parameter set above, parameters for ABC testing were selected to resemble those of *D. melanogaster* African populations. Priors on ancestral and current population sizes were drawn from a uniform distribution between 10^5 and 10^7 diploid individuals, whereas the time of change was fixed at 10^6 ($\sim N_e$) generations. In order to simulate functional regions, 94 single-exon genes, as described in [Johri et al. (2020)]() and provided in https://github.com/paruljohri/BGS_Demography_DFE/blob/master/DPGP3_data.zip, were simulated with recombination rates specific to those exons (https://petrov.stanford.edu/cgi-bin/recombination-rates_updateR5.pl) ([Fiston-Lavier et al. 2010](); [Comeron et al. 2012]()). Mutation rates were assumed to be fixed at 3×10^{-9} per site per generation ([Keightley et al. 2009](), [2014]()).

All parameters were scaled by the factor 320 in order to decrease computational time, using the principle first described by [Hill and Robertson (1966)](), and subsequently employed by others ([Comeron and Kreitman 2002](); [Hoggart et al. 2007](); [Kim and Wiehe 2008](); [Kaiser and Charlesworth 2009](); [Uricchio and Hernandez 2014](); [Campos and Charlesworth 2019]()). The scaled population sizes thus ranged between $\sim$300 and 30,000 and were reported as scaled values in the main text. One thousand replicate simulations were performed for every parameter combination (N_{anc}, N_{cur}, f_0, f_1, f_2, f_3); for performing ABC inference, 50 diploid genomes were randomly sampled without replacement, and summary statistics were calculated using pylibseq 0.2.3 ([Thornton 2003]()). The following summary statistics were calculated across the entire exonic region for every exon: nucleotide site diversity (π), Watterson's θ, Tajima's D, Fay and Wu's H (both absolute and normalized), number of singletons, haplotype diversity, LD-based statistics (r^2, D, D′), and divergence (i.e., number of fixed mutations per site per generation after the burn-in period). Means and variances (between exons) of all of the above (a total of 22) were used as final summary statistics to perform ABC. As opposed to the above examples, in this inference scheme only exonic data (i.e., directly selected sites) were utilized. Test data sets were generated in exactly the same fashion as described above.

Analytical Expectations for the Relative Site Frequencies

To compute the expected relative frequencies of site frequency classes, the approach of [Polanski and Kimmel (2003)]() was followed. They describe a method for computing the "probability that a SNP has *b* mutant bases," which is equivalent to the expected site frequency spectrum (SFS) of derived variants. This method (their eqs. 3–10) allows for the specification of arbitrary population size histories and sample sizes. For reasons of computational precision, a sample size of ten diploid genomes was chosen. The demographic scenarios were implemented as piecewise functions of the effective population size (counting haploid genomes), and the effect of BGS was included by scaling these functions by values of B

before population size change as obtained from the forwards-in-time simulations described above. A Mathematica notebook detailing these results is available online (see Data Availability statement). In addition, analytical expressions can be obtained for pairwise diversity values when there are step changes or exponential growth in population size, as described in the Appendix and in an example program that calculates diversity values after exponential growth.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Data Availability

The following data are publicly available: 1) the general workflow for simulating and performing demographic inference; 2) scripts used for performing simulations, masking genomes, calculating the SFS, performing model selection, and plotting the final results; 3) input files used to run *fastsimcoal2*, including for the calculated SFS; 4) output files of MSMC and *fastsimcoal2* for all simulated scenarios; 5) a Mathematica (version 12.1) notebook detailing the calculations of analytical expectations for the relative SFS; and 6) an example program (Fortran script) demonstrating how to obtain analytical expressions for values of B after exponential growth. All supplementary files, scripts, command lines, and descriptions may be found at: https://github.com/paruljohri/demographic_inference_with_selection

Appendix

There are two scenarios of population size change for which simple explicit expressions for the expected pairwise coalescent time or diversity can be obtained, without using the methodology of Polanski and Kimmel (2003) and Polanski et al. (2003)—a step change in N or an exponential growth in N. First consider the coalescent process for a step change, where the current and initial effective population sizes are denoted by N_{e1} and N_{e0}, respectively. Let B be the background selection parameter at the start of the process of change, corresponding to effective size N_{e0}. For convenience, time is scaled in units of $2N_{e1}$ generations, and the time of the change in population size on this scale is denoted by T_0, counting back from the present time, $T = 0$. T_0 is assumed to be sufficiently small that B remains

approximately constant during the period since the change in size. Denote the ratio N_{e0}/N_{e1} by R. The derivation for the case of a step change in population size is similar to that given by Pool and Nielsen (2009) for the purpose of comparing X chromosomes and autosomes.

Between times T and T_0, coalescence occurs at a rate B^{-1} on the chosen timescale, so that the contribution from this period to the net coalescent time for a pair of alleles sampled at $T = 0$ is:

$$B^{-1} \int_0^{T_0} T \, \exp(-B^{-1}T) \, dT = B - B \, \exp(-B^{-1}T_0) - T_0 \, \exp(-B^{-1}T_0).$$

There is a probability of $\exp(-B^{-1}T_0)$ that there is no coalescence when T lies between 0 and T_0, after which coalescence occurs at a rate $1/BR$, giving a net contribution to the coalescence time of:

$$(BR + T_0) \, \exp(-B^{-1}T_0).$$

The net coalescence time for the stepwise change with BGS is given by the sum of these two expressions:

$$B[1 + (R - 1) \, \exp(-B^{-1}T_0)]. \tag{1a}$$

If this expression is compared with the corresponding equation with $B = 1$, the apparent value of B at the time of sampling of the pair of alleles is given by:

$$B_s = \frac{B[1 + (R - 1) \, \exp(-B^{-1}T_0)]}{[1 + (R - 1) \, \exp(-T_0)]}. \tag{1b}$$

Next, consider a process of exponential change in population size, starting at an initial effective size of N_{e0} at t_0 generations in the past and ending at size N_{e1}, such that the instantaneous growth rate r per generation is $r = \ln(N_{e1}/N_{e0})/t_0$. The effective population size at time t in the past is $N_e(t) = N_{e1}\exp(-rt)$; with BGS, the rate of coalescence at time t is $1/BN_e(t)$. As before, the BGS parameter is assumed to remain constant over the period of population size change. It follows that the probability of no coalescence by generation t in the past is:

$$P_{nc}(t) = \exp[-\int_0^t (2BN_{e1})^{-1} \, \exp(rt) \, dt]$$
$$= \exp[c^{-1}(1 - e^{rt})], \tag{2}$$

where $c = 2BN_{e1}r$.

The pre-growth period with $t > t_0$ contributes an expected coalescent time of $(2BN_{e0} + t_0)P_{nc}(t_0)$, on the scale of generations.

Following Slatkin and Hudson (1991), to obtain the contribution from the period with $t > t_0$, it is convenient to measure time as $\tau = rt$. The probability of coalescence between τ and $\tau + d\tau$ is then given by:

$$P_c(\tau) = c^{-1}e^{\tau}\exp[c^{-1}(1 - e^{\tau})]d\tau. \tag{3}$$

The contribution from this period to the expected coalescent time is given by the integral of $\tau P_c(\tau)$ between 0 and

τ_0. Following Slatkin and Hudson (1991), by transforming to $u = \exp(\tau)$, this contribution can be expressed as the following integral:

$$\bar{\tau}_1 = c^{-1} e^{c^{-1}} \int_1^{u_0} \ln(u) e^{-uc^{-1}} du. \qquad (4)$$

This integral can easily be evaluated numerically. The corresponding mean coalescent time on the scale of generations is obtained by division by r, and the result can be added to $(2BN_{e0} + t_0)P_{nc}(t_0)$, yielding the net expected coalescent time. By dividing the resulting expression by the corresponding expression with $B = 1$, the apparent BGS effect at the time of sampling can be obtained, in the same way as for the step change model.

References

Adrion JR, Cole CB, Dukler N, Galloway JG, Gladstein AL, Gower G, Kyriazis CC, Ragsdale AP, Tsambos G, Baumdicker F, et al. 2020. A community-maintained standard library of population genetic models. *eLife* 9:e54967.

Andolfatto P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* 437(7062):1149–1152.

Bank C, Ewing GB, Ferrer-Admettla A, Foll M, Jensen JD. 2014. Thinking too positive? Revisiting current methods of population genetic selection inference. *Trends Genet*. 30(12):540–546.

Beichman AC, Huerta-Sanchez E, Lohmueller KE. 2018. Using genomic data to infer historic population dynamics of nonmodel organisms. *Annu Rev Ecol Evol Syst*. 49(1):433–456.

Beichman AC, Phung TN, Lohmueller KE. 2017. Comparison of single genome and allele frequency data reveals discordant demographic histories. *G3 (Bethesda)* 7:3605–3620.

Bhaskar A, Wang YXR, Song YS. 2015. Efficient inference of population size histories and locus-specific mutation rates from large-sample genomic variation data. *Genome Res*. 25(2):268–279.

Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, et al. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 277(5331):1453–1462.

Boitard S, Rodríguez W, Jay F, Mona S, Austerlitz F. 2016. Inferring population size history from large samples of genome-wide molecular data – an approximate Bayesian computation approach. *PLoS Genet*. 12(3):e1005877.

Booker TR, Keightley PD. 2018. Understanding the factors that shape patterns of nucleotide diversity in the house mouse genome. *Mol Biol Evol*. 35(12):2971–2988.

Bunnefeld L, Frantz LAF, Lohse K. 2015. Inferring bottlenecks from genome-wide samples of short sequence blocks. *Genetics* 201(3):1157–1169.

Campos JL, Charlesworth B. 2019. The effects on neutral variability of recurrent selective sweeps and background selection. *Genetics* 212(1):287–303.

Campos JL, Halligan DL, Haddrill PR, Charlesworth B. 2014. The relation between recombination rate and patterns of molecular evolution and variation in *Drosophila melanogaster*. *Mol Biol Evol*. 31(4):1010–1028.

Campos JL, Zhao L, Charlesworth B. 2017. Estimating the parameters of background selection and selective sweeps in *Drosophila* in the presence of gene conversion. *Proc Natl Acad Sci U S A*. 114(24):E4762–E4771.

Castellano D, Eyre-Walker A, Munch K. 2020. Impact of mutation rate and selection at linked sites on DNA variation across the genomes of humans and other Homininae. *Genome Biol Evol*. 12(1):3550–3561.

Chamary J, Hurst LD. 2005. Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *Genome Biol*. 6(9):R75.

Charlesworth B. 2013. Background selection 20 years on. The Wilhelmine E. Key 2012 invitational lecture. *J Hered*. 104(2):161–171.

Charlesworth B. 2015. Causes of natural variation in fitness: evidence from studies of *Drosophila* populations. *Proc Natl Acad Sci U S A*. 112(6):1662–1669.

Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134(4):1289–1303.

Charlesworth D, Charlesworth B, Morgan MT. 1995. The pattern of neutral molecular variation under the background selection model. *Genetics* 141(4):1619–1632.

Chikhi L, Rodríguez W, Grusea S, Santos P, Boitard S, Mazet O. 2018. The IICR (inverse instantaneous coalescence rate) as a summary of genomic diversity: insights into demographic inference and model choice. *Heredity* 120(1):13–24.

Choi JY, Aquadro CF. 2016. Recent and long term selection across synonymous sites in *Drosophila ananassae*. *J Mol Evol*. 83(1–2):50–60.

Comeron JM, Kreitman M. 2002. Population, evolutionary and genomic consequences of interference selection. *Genetics* 161(1):389–410.

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

Csilléry K, Blum MGB, Gaggiotti OE, François O. 2010. Approximate Bayesian Computation (ABC) in practice. *Trends Ecol Evol*. 25(7):410–418.

Cutter AD, Payseur BA. 2013. Genomic signatures of selection at linked sites: unifying the disparity among species. *Nat Rev Genet*. 14(4):262–274.

Elyashiv E, Sattath S, Hu TT, Strutsovsky A, McVicker G, Andolfatto P, Coop G, Sella G. 2016. A genomic map of the effects of linked selection in *Drosophila*. *PLoS Genet*. 12(8):e1006130.

Ewing GB, Jensen JD. 2016. The consequences of not accounting for background selection in demographic inference. *Mol Ecol*. 25(1):135–141.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genet*. 9(10):e1003905.

Eyre-Walker A, Keightley PD. 2009. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol Biol Evol*. 26(9):2097–2108.

Fay JC, Wu CI. 1999. A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. *Mol Biol Evol*. 16(7):1003–1005.

Fiston-Lavier A-S, Singh ND, Lipatov M, Petrov DA. 2010. *Drosophila melanogaster* recombination rate calculator. *Gene* 463(1–2):18–20.

Francioli LC, Polak PP, Koren A, Menelaou A, Chun S, Renkens I, van Duijn CM, Swertz M, Wijmenga C, van Ommen G, et al. 2015. Genome-wide patterns and properties of de novo mutations in humans. *Nat Genet*. 47(7):822–826.

Fulgione A, Koornneef M, Roux F, Hermisson J, Hancock AM. 2018. Madeiran Arabidopsis thaliana reveals ancient long-range colonization and clarifies demography in Eurasia. *Mol Biol Evol*. 35(3):564–574.

Galtier N, Rousselle M. 2020. How much does N_e vary among species? *Genetics* 216(2):559–572.

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

Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P. 2005. Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Res*. 15(6):790–799.

Haller BC, Messer PW. 2019. SLiM 3: forward genetic simulations beyond the Wright–Fisher model. *Mol Biol Evol*. 36(3):632–637.

Harris K, Nielsen R. 2013. Inferring demographic history from a spectrum of shared haplotype lengths. *PLoS Genet*. 9(6):e1003521.

Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G, Sella G, Przeworski M. 2011. Classic selective sweeps were rare in recent human evolution. *Science* 331(6019):920–924.

Hey J, Harris E. 1999. Population bottlenecks and patterns of human polymorphism. *Mol Biol Evol.* 16(10):1423–1426.

Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genet Res.* 8(3):269–294.

Hoggart CJ, Chadeau-Hyam M, Clark TG, Lampariello R, Whittaker JC, Iorio MD, Balding DJ. 2007. Sequence-level population simulations over large genomic regions. *Genetics* 177(3):1725–1731.

Hung C-M, Shaner P-JL, Zink RM, Liu W-C, Chu T-C, Huang W-S, Li S-H. 2014. Drastic population fluctuations explain the rapid extinction of the passenger pigeon. *Proc Natl Acad Sci U S A.* 111(29):10636–10641.

Jackson BC, Campos JL, Haddrill PR, Charlesworth B, Zeng K. 2017. Variation in the intensity of selection on codon bias over time causes contrasting patterns of base composition evolution in *Drosophila*. *Genome Biol Evol.* 9(1):102–123.

Jacquier H, Birgy A, Nagard HL, Mechulam Y, Schmitt E, Glodt J, Bercot B, Petit E, Poulain J, Barnaud G, et al. 2013. Capturing the mutational landscape of the beta-lactamase TEM-1. *Proc Natl Acad Sci U S A.* 110(32):13067–13072.

Jensen JD, Payseur BA, Stephan W, Aquadro CF, Lynch M, Charlesworth D, Charlesworth B. 2019. The importance of the Neutral Theory in 1968 and 50 years on: a response to Kern and Hahn 2018. *Evolution* 73(1):111–114.

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

Jones MR, Good JM. 2016. Targeted capture in evolutionary and ecological genomics. *Mol Ecol.* 25(1):185–202.

Kaiser VB, Charlesworth B. 2009. The effects of deleterious mutations on evolution in non-recombining genomes. *Trends Genet.* 25(1):9–12.

Keightley PD, Eyre-Walker A. 2007. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics* 177(4):2251–2261.

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

Keightley PD, Trivedi U, Thomson M, Oliver F, Kumar S, Blaxter ML. 2009. Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome Res.* 19(7):1195–1201.

Kelleher J, Etheridge AM, McVean G. 2016. Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Comput Biol.* 12(5):e1004842.

Kelleher J, Wong Y, Wohns AW, Fadil C, Albers PK, McVean G. 2019. Inferring whole-genome histories in large population datasets. *Nat Genet.* 51(9):1330–1338.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The Human Genome Browser at UCSC. *Genome Res.* 12(6):996–1006.

Kim BY, Huber CD, Lohmueller KE. 2017. Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. *Genetics* 206(1):345–361.

Kim Y, Wiehe T. 2008. Simulation of DNA sequence evolution under models of recent directional selection. *Brief Bioinformatics.* 10(1):84–96.

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

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409(6822):860–921.

Lapierre M, Lambert A, Achaz G. 2017. Accuracy of demographic inferences from the site frequency spectrum: the case of the Yoruba population. *Genetics* 206(1):439–449.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475(7357):493–496.

Liang P, Saqib HSA, Zhang X, Zhang L, Tang H. 2018. Single-base resolution map of evolutionary constraints and annotation of conserved elements across major grass genomes. *Genome Biol Evol.* 10(2):473–488.

Lukic S, Hey J. 2012. Demographic inference using spectral methods on SNP data, with an analysis of the human out-of-Africa expansion. *Genetics* 192(2):619–639.

Lynch M. 2007. The origins of genome architecture. Sunderland (MA): Sinauer Associates.

Mazet O, Rodríguez W, Grusea S, Boitard S, Chikhi L. 2016. On the importance of being structured: instantaneous coalescence rates and human evolution–lessons for ancestral population size inference? *Heredity* 116(4):362–371.

McVean GAT, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. 2004. The fine-scale structure of recombination rate variation in the human genome. *Science* 304(5670):581–584.

Messer PW, Petrov DA. 2013. Frequent adaptation and the McDonald–Kreitman test. *Proc Natl Acad Sci U S A.* 110(21):8615–8620.

Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. 2005. A fine-scale map of recombination rates and hotspots across the human genome. *Science* 310(5746):321–324.

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

O'Fallon BD, Seger J, Adler FR. 2010. A continuous-state coalescent and the impact of weak selection on the structure of gene genealogies. *Mol Biol Evol.* 27(5):1162–1172.

Orozco-terWengel P. 2016. The devil is in the details: the effect of population structure on demographic inference. *Heredity* 116(4):349–350.

Palkopoulou E, Lipson M, Mallick S, Nielsen S, Rohland N, Baleka S, Karpinski E, Ivancevic AM, To T-H, Kortschak RD, et al. 2018. A comprehensive genomic history of extinct and living elephants. *Proc Natl Acad Sci U S A.* 115(11):E2566–E2574.

Polanski A, Bobrowski A, Kimmel M. 2003. A note on distributions of times to coalescence, under time-dependent population size. *Theor Popul Biol.* 63(1):33–40.

Polanski A, Kimmel M. 2003. New explicit expressions for relative frequencies of single-nucleotide polymorphisms with application to statistical inference on population growth. *Genetics* 165(1):427–436.

Pool JE, Nielsen R. 2007. Population size changes reshape genomic patterns of diversity. *Evolution* 61(12):3001–3006.

Pool JE, Nielsen R. 2009. Correction for Pool and Nielsen (2007). *Evolution* 63:1671.

Pouyet F, Aeschbacher S, Thiéry A, Excoffier L. 2018. Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. *eLife* 7:e36317.

Ragsdale AP, Gutenkunst RN. 2017. Inferring demographic history using two-locus statistics. *Genetics* 206(2):1037–1048.

Ragsdale AP, Moreau C, Gravel S. 2018. Genomic inference using diffusion models and the allele frequency spectrum. *Curr Opin Genet Dev.* 53:140–147.

Sanjuán R. 2010. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. *Philos Trans R Soc B.* 365(1548):1975–1982.

Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nat Genet.* 46(8):919–925.

Schneider A, Charlesworth B, Eyre-Walker A, Keightley PD. 2011. A method for inferring the rate of occurrence and fitness effects of advantageous mutations. *Genetics* 189(4):1427–1437.

Schrider DR, Shanku AG, Kern AD. 2016. Effects of linked selective sweeps on demographic inference and model selection. *Genetics* 204(3):1207–1223.

Sheehan S, Song YS. 2016. Deep learning for population genetic inference. *PLoS Comput Biol.* 12(3):e1004845.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, et al. 2005. Evolutionarily

conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15(8):1034–1050.

Slatkin M, Hudson RR. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129(2):555–562.

Speidel L, Forest M, Shi S, Myers SR. 2019. A method for genome-wide genealogy estimation for thousands of samples. *Nat Genet.* 51(9):1321–1329.

Steinrücken M, Kamm J, Spence JP, Song YS. 2019. Inference of complex population histories using whole-genome sequences from multiple populations. *Proc Natl Acad Sci U S A.* 116(34):17115–17120.

Teshima KM, Coop G, Przeworski M. 2006. How reliable are empirical genomic scans for selective sweeps? *Genome Res.* 16(6):702–712.

Thornton K. 2003. Libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics* 19(17):2325–2327.

Thornton KR, Jensen JD. 2007. Controlling the false-positive rate in multilocus genome scans for selection. *Genetics* 175(2):737–750.

Torres R, Stetter MG, Hernandez RD, Ross-Ibarra J. 2020. The temporal dynamics of background selection in nonequilibrium populations. *Genetics* 214(4):1019–1030.

Torres R, Szpiech ZA, Hernandez RD. 2018. Human demographic history has amplified the effects of background selection across the genome. *PLoS Genet.* 14(6):e1007387.

Uricchio LH, Hernandez RD. 2014. Robust forward simulations of recurrent hitchhiking. *Genetics* 197(1):221–236.

Warren WC, Jasinska AJ, García-Pérez R, Svardal H, Tomlinson C, Rocchi M, Archidiacono N, Capozzi O, Minx P, Montague MJ, et al. 2015. The genome of the vervet (*Chlorocebus aethiops sabaeus*). *Genome Res.* 25(12):1921–1933.

Williamson RJ, Josephs EB, Platts AE, Hazzouri KM, Haudry A, Blanchette M, Wright SI. 2014. Evidence for widespread positive and negative selection in coding and conserved noncoding regions of *Capsella grandiflora*. *PLoS Genet.* 10(9):e1004622.

Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, Hu Y, Tao Z-W, Tian J-H, Pei Y-Y, et al. 2020. A new coronavirus associated with human respiratory disease in China. *Nature* 579(7798):265–269.

Zeng K, Charlesworth B. 2010. Studying patterns of recent evolution at synonymous sites and intronic sites in *Drosophila melanogaster*. *J Mol Evol.* 70(1):116–128.

Zhou Y, Massonnet M, Sanjak JS, Cantu D, Gaut BS. 2017. Evolutionary genomics of grape (*Vitis vinifera ssp. vinifera*) domestication. *Proc Natl Acad Sci U S A.* 114(44):11715–11720.