The impact of purifying and background selection on the inference of population history: problems and prospects

Parul Johri[*, †], Kellen Riall[*], and Jeffrey D. Jensen[*, †]

[*]School of Life Sciences, Arizona State University, Tempe, AZ 85287

[†]Correspondence can be addressed to these authors

ABSTRACT

Current procedures for inferring population history are generally performed under the assumption of complete neutrality - that is, neglecting both direct selection and the effects of selection on linked sites. We here examine how the presence of purifying 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 (DFE) and the fraction of directly selected sites interacts with demographic parameter estimation. The results demonstrate that, even after masking functional genomic regions, background selection effects may result in the mis-inference of population growth under models of both constant population size as well as decline. This effect is amplified as the strength of purifying selection and the density of directly selected sites increases. We propose a potential solution to correct for this mis-inference. By treating the DFE as a nuisance parameter and averaging across all potential realizations, we demonstrate that even directly selected sites may be used to infer demographic histories with reasonable accuracy.

Keywords: demographic inference, background selection, MSMC, *fastsimcoal2*, approximate Bayesian computation (ABC)

INTRODUCTION

The characterization of population history is a central goal of genomic analysis - with applications spanning from anthropological to agricultural to clinical (see review by Beichman *et al.* 2018). Furthermore, use of an appropriate underlying demographic model provides a necessary null for assessing the impact of selective effects 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 relatedness (*e.g.*, Gutenkunst *et al.* 2009; Li and Durbin 2011; Lukic and Hey 2012; Harris and Nielsen 2013; Excoffier *et al.* 2013; Bhaskar *et al.* 2015; Sheehan and Song 2016; Ragsdale and Gutenkunst 2017; Steinrücken *et al.* 2019; Kelleher *et al.* 2019; Speidel *et al.* 2019). Although many methods perform well when evaluated under the stated assumption of neutrality, it is in fact difficult in practice to assure that the sites utilized for empirical analysis neither experience direct selection nor the effects of selection at linked sites. For example, inference is often performed using intergenic sites, or 4-fold degenerate and intronic sites in more compact genomes. While there is evidence for weak direct selection in all of these categories in multiple organisms (*e.g.*, Haddrill *et al.* 2005; Chamary and Hurst 2005; Andolfatto 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). These BGS effects are known to affect the local underlying effective population size, and alter both 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 light of this likely-to-be-violated model assumption. Yet it is not entirely uninvestigated. Rather than investigating existing software, Ewing and Jensen (2016) implemented an approximate Bayesian (ABC) approach to quantify the impact of unaccounted for 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 density of directly selected sites and the shape of the distribution of fitness effects (DFE) - parameters expected to be of great importance - have yet to be considered. Spanning the range of these potential parameter values is also important for understanding the expected impact on any given empirical application. As one example, the proportion of the genome experiencing direct selection can vary greatly between species, with estimates ranging from ~3-8% in humans, to ~12% in rice, to 37-53% in *D. melanogaster*, to 47-68% in *S. cerevisiae* (Siepel *et al.* 2005; Liang *et al.* 2018). Further, many organisms have highly compact genomes, with ~88% of the *E. coli* genome (Blattner *et al.* 1997) and effectively all of many virus genomes (*e.g.*, >95% of the SARS-CoV-2 genome; Wu *et al.* 2020).

While such estimates allow us to approximate the effects of BGS in model organisms in which recombination and mutation rates are well known, there are many non-model organisms in which it is difficult to predict these effects. Moreover, while 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; though see Pouyet *et al.* 2018), existing demographic inference approaches are similarly applied across organisms without a consideration of this important source of differences in levels of bias. Here, we examine the effects of 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. Simulating examples consistent with the structure of the *Drosophila* genome, we demonstrate greatly improved demographic inference even when using directly selected sites alone.

METHODS

Simulations: Simulations were performed using SLiM 3.1 (Haller and Messer 2019) with 10 replicates per evolutionary scenario. For every replicate, 22 chromosomes of 150Mb each were simulated, totaling ~3 Gb of information per individual genome (similar to the amount of information in a human genome). Within each chromosome, 3 different types of regions were simulated, representing non-coding intergenic, intronic, and exonic regions. Based on the NCBI RefSeq human genome annotation downloaded from the UCSC genome browser

(http://genome.ucsc.edu/; Kent *et al.* 2002) for hg19, 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 was comprised of 8 exons and 7 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 3000 bp and intergenic sequence of 31000 bp, genome10 of introns of 1500 bp and intergenic sequence of 15750 bp, while genome20 was comprised of introns of 600 bp and intergenic sequence of 6300 bp. The total chromosome sizes of these genomes were approximately 150 Mb (150,018,599 bp, 150,029,949 bp, 150,003,699 bp in genome5, 10, and 20, respectively). In order to be conservative and favorable with regards 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 at 1×10^{-8} /site / generation. Neither crossover interference (see discussion of Campos and Charlesworth 2019) nor gene conversion were here considered. Exonic regions in the genomes experienced direct purifying selection given by a discrete DFE comprised of 4 fixed classes: f_0 in which $0 \leq |2N_e s| < 1$ (*i.e.*, effectively neutral), f_1 in which $1 \leq |2N_e s| < 10$ (*i.e.*, weakly deleterious), f_2 in which $10 \leq |2N_e s| < 100$ (*i.e.*, moderately deleterious), and f_3 in which $100 \leq |2N_e s| < 2N_e$ (*i.e.*, strongly deleterious), where N_e is the effective population size and s is the decrease 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 semi-dominant. In all cases, the N_e corresponding to the DFE refers to the ancestral effective population size. Six different types of DFE were simulated: (1) a DFE skewed largely towards mildly deleterious mutations given by f_0=0.1, f_1=0.7, f_2=0.1, f_3=0.1, (2) a DFE skewed towards moderately deleterious mutations, f_0=0.1, f_1=0.1, f_2=0.7, f_3=0.1, (3) a DFE skewed towards strongly deleterious mutations f_0=0.1, f_1=0.1, f_2=0.1, f_3=0.7, (4) a DFE with equal proportions of all mutations given by f_0=0.25, f_1=0.25, f_2=0.25, f_3=0.25, (5) a DFE with equal proportions of neutral and strongly deleterious mutations given by f_0=0.5, f_1=0.0, f_2=0.0, f_3=0.5, and (6) a DFE with a majority of neutral mutations and a minority of strongly deleterious mutations given by f_0=0.7, f_1=0.0, f_2=0.0, f_3=0.3. In cases with

demographic non-equilibrium, the DFE shape was scaled with respect to the ancestral population size.

Three different demographic models were tested for each of these DFEs: 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). The precise parameters of these three scenarios are given in Supp. Table 1 and are also displayed in Figure 1a.

Running MSMC: In order to quantify the effect of purifying selection on demographic inference, we used the entire chromosomes generated by SLiM to generate input files for MSMC. For comparison, and 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 those results were the most accurate (Supp Figure 1). 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

msmc2_linux64bit -r 1.0 -t 5 -o output_genomeID input_chr1.tab input_chr2.tab … input_chr22.tab

Running Fastsimcoal2: In order to minimize the effects of linkage disequilibrium, only SNPs separated by 5 kb were used for inference, following Excoffier *et al.* (2013). Inference was also performed by including all SNPs in order to assess the impact of violating the assumption of no linkage disequilibrium on demographic inference. Site frequency spectra (SFS) were obtained for both sets of SNPs for all 10 replicates of every combination of demographic history and DFE. SNPs from all 22 chromosomes were pooled together to calculate the SFS. *Fastsimcoal2* was used to fit each SFS to 4 distinct models: (a) equilibrium, which estimates only a single population size parameter (N); (b) instantaneous size change (decline/growth), which fits 3 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 3 parameters - N_{anc}, N_{cur} and T; and (d) an instantaneous bottleneck model with 3 parameters – N_{anc}, intensity, and time of

bottleneck. 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 × number of parameters – 2 × ln(likelihood). The relative likelihoods (Akaike's weight of evidence) in favor of the i^{th} model was then calculated by:

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

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.

Performing inference by approximate Bayesian computation (ABC): ABC was performed using the R package "abc" (Csilléry *et al.* 2010) and non-linear 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. To infer posterior estimates, a tolerance of 0.1 was applied (*i.e.*, 10% of the total number of simulations were accepted by ABC 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, and (2) in the presence of direct purifying selection. In both cases only 2 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 $i = 0$ to 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.*, $f_i \in \{0.0, 0.05, 0.10, \ldots, 0.95, 1.0\}\ \forall\ i$), 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 comparison 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 10^5-10^7 diploid individuals,

while 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 of these parameters were scaled by the factor 320 in order to decrease computational time. The scaled population sizes thus ranged between ~300–30000 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), 100 haploid genomes were randomly sampled without replacement, and summary statistics were calculated using pylibseq 0.2.3 (Thornton 2003). Means and variances (between replicates) of the following summary statistics were calculated across the entire genomic region: 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). As opposed to the above examples, in this inference scheme only exonic data (*i.e*., directly selected sites) were utilized.

Data availability: The following data will be made publicly available at https://github.com/paruljohri upon acceptance of the manuscript: (1) Scripts used to perform simulations; (2) Input files used to run fastsimcoal2; (3) Scripts used for plotting; (4) Plotted results of MSMC and *fastsimcoal2* for all models and scenarios tested in this work.

RESULTS and DISCUSSION

Effect of SNP number and genome size on inference under neutrality

The accuracy and performance of demographic inference was 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 with 100 independent replicates each. Both *fastsimcoal2* and MSMC resulted in incorrect inferences for all segments smaller than 1 Gb (Supp Figure 1, Supp Figure 2). In addition, when 4 or 8 haploid genomes were used for inference, MSMC inferred a recent many-fold growth for all segment sizes, but performed well when using 2 genomes for large segments (Supp Figure 1). These results suggest exercising caution when performing inference on smaller regions or genomes, and when using more than 1 diploid individual in MSMC. All further analyses were restricted to characterizing demographic inference on data that roughly matched the structure and size of the human genome - for every diploid individual, 22 chromosomes 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 was performed using 1 and 50 diploid individuals for MSMC and *fastsimcoal2,* respectively.

Effect of the strength of purifying selection on demographic inference

In order to test 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 Methods) to vary the fraction of the genome under selection. Because deleterious mutations of differing severities result in background selection effects extending to variable distances, demographic inference was evaluated for different possible DFE realizations - summarized by four bins: $0 \leq |2N_e s| < 1$, $1 \leq |2N_e s| \leq 10$, $10 < |2N_e s| \leq 100$ and $100 < |2N_e s| \leq 2N_e$. Fitness effects of mutations were uniformly distributed within each bin and the DFE shape was altered by varying the proportion of mutations belonging to each fixed bin, given by f_0, f_1, f_2, and f_3, respectively. Three highly skewed DFEs were initially used to assess the impact of the strength of selection on demographic inference (with the remaining mutations equally distributed amongst the other three classes): (1) a DFE in which 70% of mutations have weakly deleterious fitness effects (*i.e.*, $f_1 = 0.7$), (2) a DFE in which 70% of mutations have moderately deleterious fitness effects (*i.e.*, $f_2 = 0.7$), and (3) a DFE in which 70% of mutations have strongly deleterious fitness effects (*i.e.*, $f_3 = 0.7$). In order to understand the effect of BGS, exonic sites were masked, and only linked neutral intergenic sites were used for demographic inference. Three demographic models were examined: (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 (Figure 1a). Although these models were parameterized using previous estimates of human demographic history (Supp Table 1; Gutenkunst *et al.* 2009) for the purpose of illustration, these basic demographic scenarios are applicable to many organisms.

Under demographic equilibrium, when 20% of the genome experiences direct selection (which is masked) we observed an underestimate of the true population size, as well as the mis-inference of recent population growth (Figure 1), a bias exacerbated in MSMC relative to *fastsimcoal2*. Stronger growth is inferred when the effects of mutations are more deleterious. Conversely, when the true demographic model is characterized by recent 30-fold growth, demographic inference is accurate and performs equally well for both MSMC and *fastsimcoal2*, with the exception of a slight under-estimation of the ancestral population size for all DFE types. When the true model is population decline, weakly deleterious mutations alone did not affect inference drastically and it was possible to recover the true model. However, moderately and strongly deleterious mutations resulted in an underestimation of population sizes and an inference of strong growth - to the extent that the decline model would be misinterpreted as a growth model. We further tested the effect of BGS on demographic inference when changes in population size were less severe, namely, when population growth and decline was only 2-fold - with qualitatively similar results (Supp Figure 3).

Finally, given strong evidence that most organisms have a bi-modal DFE with a significant proportion of strongly deleterious or lethal mutations (Sanjuán 2010; Kousathanas and Keightley 2013; Jacquier *et al.* 2013; Bank *et al.* 2014), 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 are strongly deleterious with fitness effects uniformly sampled between $100 \leq |2N_{anc}s| \leq 2N_{anc}$, with the remaining mutations being neutral. Similar to the above results, both equilibrium and decline models were falsely inferred as growth, with an order of magnitude under-estimation of the true population size (Figure 2).

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

Effect of the 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 will necessarily depend on the fraction of the genome experiencing direct selection. We therefore compared models in which genomes contained 5%, 10%, and 20% functional regions. For this comparison, an equal proportion of mutations in each DFE bin were assumed (*i.e.,* $f_0 = f_1 = f_2 = f_3 = 0.25$). As before, when the true model is exponential growth, inference was unbiased, with a slight underestimation of ancestral population size when 20% of the genome is under selection (Figure 3). When the true model is population decline, it is inferred reasonably well if less than 10% of the genome experiences direct selection, and may be mis-inferred as growth with greater functional density, as shown in Figure 3. Similarly, the extent to which population size is under-estimated at demographic equilibrium increases with the fraction of the genome under selection. Finally, it is noteworthy that many such changes in population size that were falsely inferred were of a magnitude greater than 2-fold - suggesting the need for great caution when interpreting these plots from empirical data.

Importantly, the results presented do not significantly differ between inference performed while including directly selected sites (*i.e.*, no masking of functional regions) versus inference performed using linked neutral sites (*i.e.*, masking functional regions) as shown in Supp Figure 4. These results suggest that the exclusion of exonic sites, which is often assumed to provide a sufficiently neutral dataset to enable accurate demographic inference, is not necessarily a satisfactory solution unless gene density is low.

A potential solution: averaging across all possible DFEs

As shown above, demographic inference can be strongly affected by unaccounted for background selection effects, as well as by direct purifying selection. A potential solution is thus to correct for these effects when performing inference of population history. A promising first approach to estimating direct selection effects, DFE-alpha, takes a step-wise approach to infer demography using a presumed neutral class (synonymous sites); conditional on that demography, it then estimates parameters associated with the shape of the DFE (Keightley and Eyre-Walker 2007; Eyre-Walker and Keightley 2009; Schneider *et al.* 2011). However, this approach does not include the effects of selection at linked sites, which can result an over-estimate of population growth, as well as general mis-inference of the DFE (Johri *et al.* 2020).

Building on this idea, Johri *et al.* (2020) recently proposed an approach to include both direct and background effects of purifying selection - simultaneously inferring 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, or in non-coding functional elements as well. However, this computationally-intensive approach is specifically concerned with jointly inferring the DFE with demography. As such, if an unbiased characterization of the population history is the sole aim, such a procedure may be needlessly involved. We thus here examine the possibility of rather 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 thus 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 generally in organisms with high gene densities.

In order to illustrate this solution, 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 Figure 4). 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 found in the other inference approaches examined, population sizes were under-estimated and a false inference of population growth was observed in almost all cases when selective effects are ignored (Figure 4). Indeed, the reduction in diversity resulting from BGS (B) was found to differ by demographic scenario, disproportionately amplifying mis-inference under certain models (Figure 5). Next, the assumption of neutrality was relaxed, and mutations were simulated with fitness effects characterized by a discrete DFE, with the same fitness classes given 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. 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 estimating only parameters related to population history, highly accurate inference of modern and ancestral population size

is possible (Figure 4). These results demonstrate that even if the true DFE of a population is unknown (as will always be the case empirically), it is possible to infer the demographic history with reasonable accuracy by approximately accounting for these selective effects.

CONCLUSION

While commonly-used approaches to infer demography assume neutrality and independence among segregating sites, empirically it is very difficult to verify those assumptions. In addition, there is considerable evidence for the wide-spread effects of selection on linked sites in many commonly studied organisms (Hernandez *et al.* 2011; Cutter and Payseur 2013; Elyashiv *et al.* 2016; Campos *et al.* 2017; Castellano *et al.* 2020). As such, we explored how violations of this assumption 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. Thus, when the true underlying model is in fact growth, demographic mis-inference is not particularly severe; when the true underlying model is of constant size or decline, the mis-inference can be extreme, with a many-fold under-estimation of population size.

It is important to note that background selection effects extend to 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, strongly deleterious mutations are likely to predominantly bias inference. In contrast, if synonymous sites are used to infer demographic history, mildly deleterious mutations will more predominantly influence inference. Thus, as we here focused on relatively sparsely-coding genomes (resembling human-like densities) and used intergenic sites to perform inference, moderately and strongly deleterious mutations resulted in more severe demographic mis-inference. These results suggest that demographic mis-inference will become stronger when such approaches are applied to smaller, and more compact, genomes with higher gene densities.

Comparing the two methods investigated here, it appears that *fastsimcoal2* is less prone to false fluctuations in population size caused by background selection. However, both methods falsely infer growth under alternative models, which is exacerbated as the density of coding regions increases. The correspondingly inferred times of population growth inferred by both methods appear to be relatively randomly affected. Further, we observed little difference in performance when using all SNPs relative to thinning SNPs to be separated by 5kb (Supp Figure 6), presumably because the models investigated do not generate strong LD. Overall, the degree of mis-inference caused by a neglect of BGS is largely similar between the two methods.

However, 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 as a series of size changes within MSMC. 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 (Supp Figure 5). Previous work has demonstrated that this resulting model does not in fact well-explain the observed site-frequency spectrum in the Yoruban population (Lapierre *et al.* 2017; Beichman *et al.* 2017). A similar shape has also been inferred in the vervet subspecies (Warren *et al.* 2015; Figure 4), in passenger pigeons (Hung *et al.* 2014; Figure 2), in elephants (Palkopoulou *et al.* 2018; Figure 4), in Arabidopsis (Fulgione *et al.* 2018; Figure 3), and in grapevines (Zhou *et al.* 2017; Figure 2A). Although under simulated neutrality the population size fluctuations 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 neutral, equilibrium performance of MSMC is concerning, and adds to the other previously published cautions related to 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, inference may suggest false size change (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 is utilized, or more than 2 haploid genomes are used to perform inference, MSMC falsely infers recent growth of varying magnitudes (as previously observed by Beichman *et al.* 2017).

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 realizations within an ABC framework appears to be highly 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 unlinked neutral sites either do not exist, or are difficult to identify and verify.

Acknowledgements

We would like to thank Brian Charlesworth and Susanne Pfeifer for helpful discussions related to this project, and feedback on the manuscript. This research was conducted using resources provided by Research Computing at Arizona State University (http://www.researchcomputing.asu.edu) and the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science. This work was funded by National Institutes of Health grant R01GM135899 to JDJ.

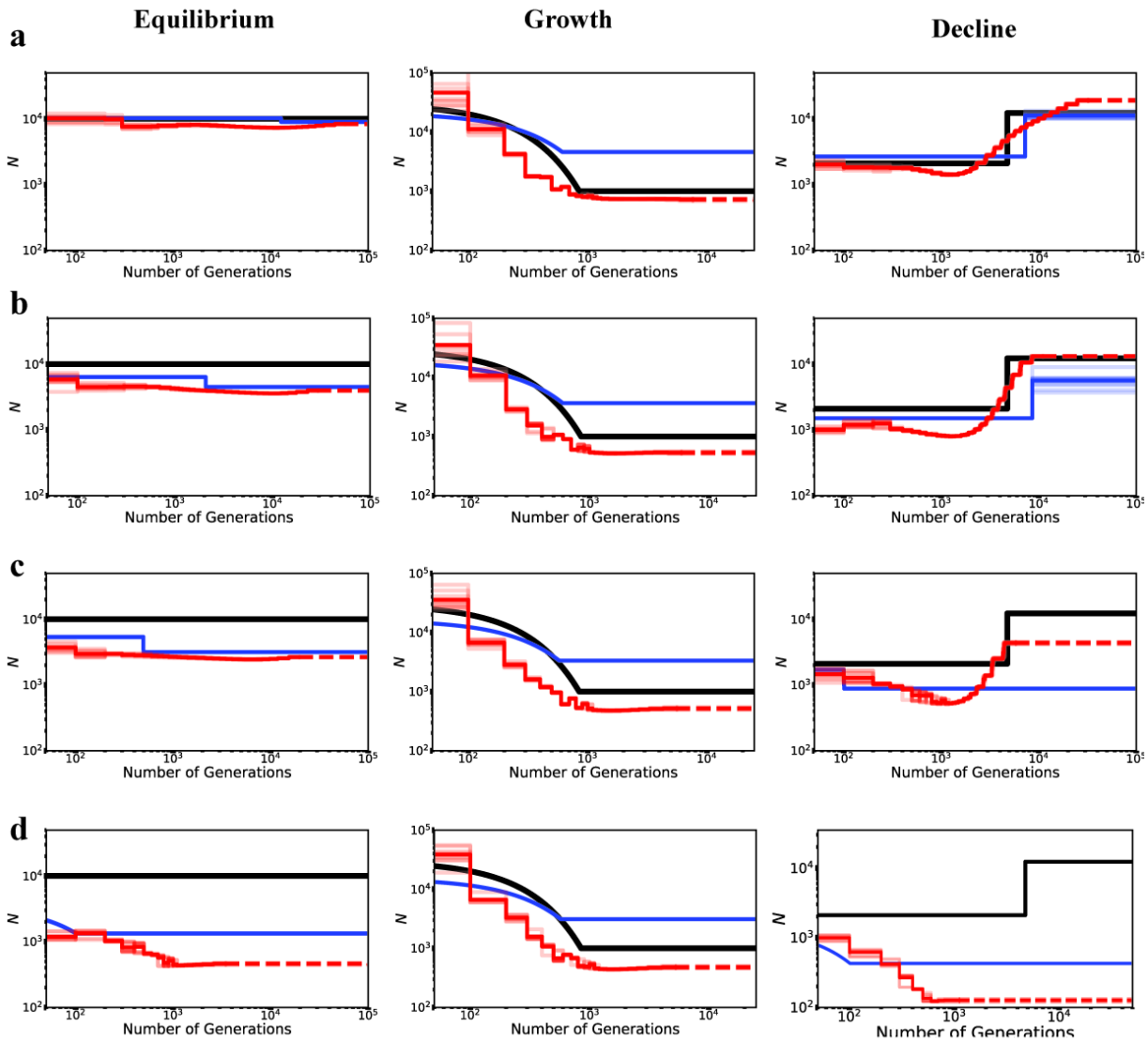


Figure 1: Inference of demography by MSMC (red lines; 10 replicates) and *fastsimcoal2* (blue lines; 10 replicates) with and without BGS, under demographic equilibrium (left column), 30-fold exponential growth (middle column), and ~6-fold instantaneous decline (right column). The true demographic model is depicted in black lines. (a) All genomic sites are strictly neutral. (b) Exonic sites experience purifying selection specified by a DFE comprised largely of weakly deleterious mutations ($f_0 = 0.1, f_1 = 0.7, f_2 = 0.1, f_3 = 0.1$). (c) Exonic sites experience purifying selection specified by a DFE comprised largely of moderately deleterious mutations ($f_0 = 0.1, f_1 = 0.1, f_2 = 0.7, f_3 = 0.1$). (d) Exonic sites experience purifying selection specified by a DFE comprised largely of strongly deleterious mutations ($f_0 = 0.1, f_1 = 0.1, f_2 = 0.1, f_3 = 0.7$). In this case, exons represent 20% of the genome. Here, exonic sites were masked/excluded when performing demographic inference - as such, this quantifies the effects of background selection alone.

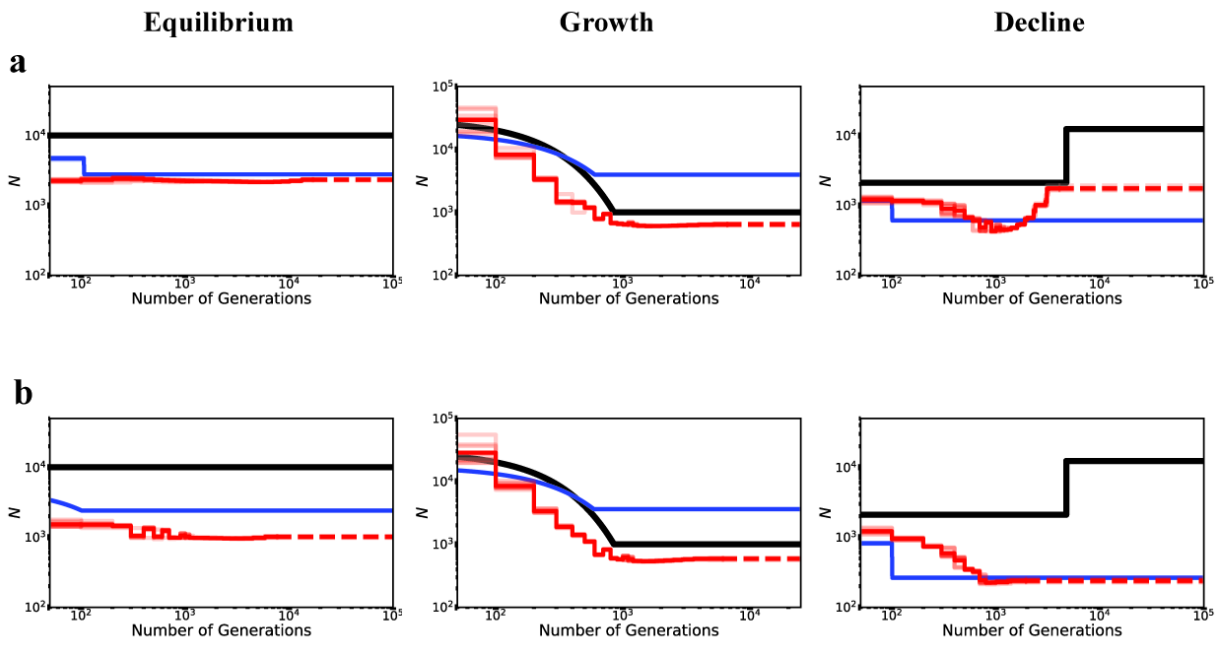


Figure 2: Inference of demography by MSMC (red lines; 10 replicates) and *fastsimcoal2* (blue lines; 10 replicates) in the presence of background selection generated by strongly deleterious mutations. Here, directly selected sites comprised 20% of the genome and were masked when performing demographic inference. (a) Exons experience purifying selection, with 30% of new mutations being strongly deleterious, and the remaining being neutral. (b) Exons experience purifying selection with 50% of new mutations being strongly deleterious, and the remaining being neutral. The true demographic model is given as black lines.

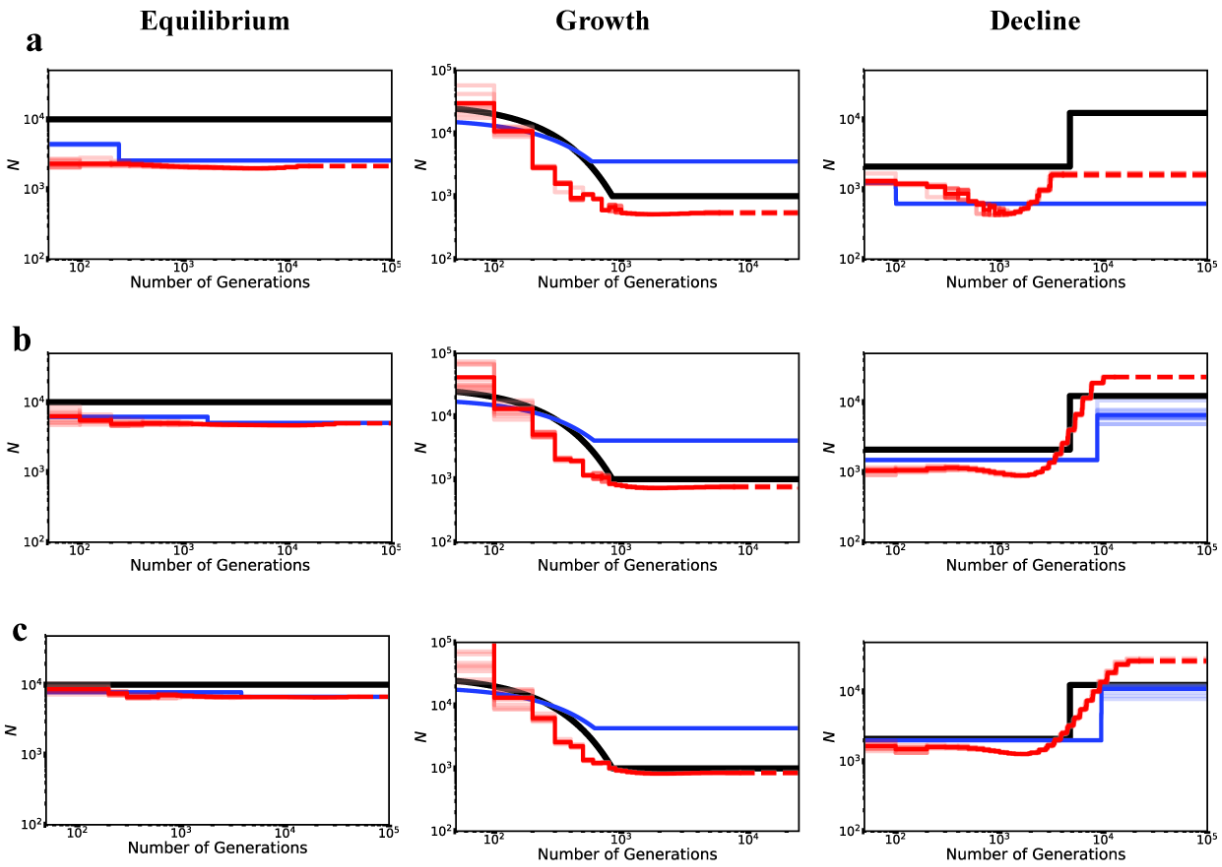


Figure 3: Inference of demography by MSMC (red lines; 10 replicates) and *fastsimcoal2* (blue lines; 10 replicates) in the presence of background selection with varying proportions of the genome under selection, for demographic equilibrium (left column), exponential growth (middle column), and instantaneous decline (right column). Exonic sites were simulated with purifying selection given by the DFE: f_0 =0.25, f_1 = 0.25, f_2 = 0.25, f_3 = 025 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 model is given by the black lines.

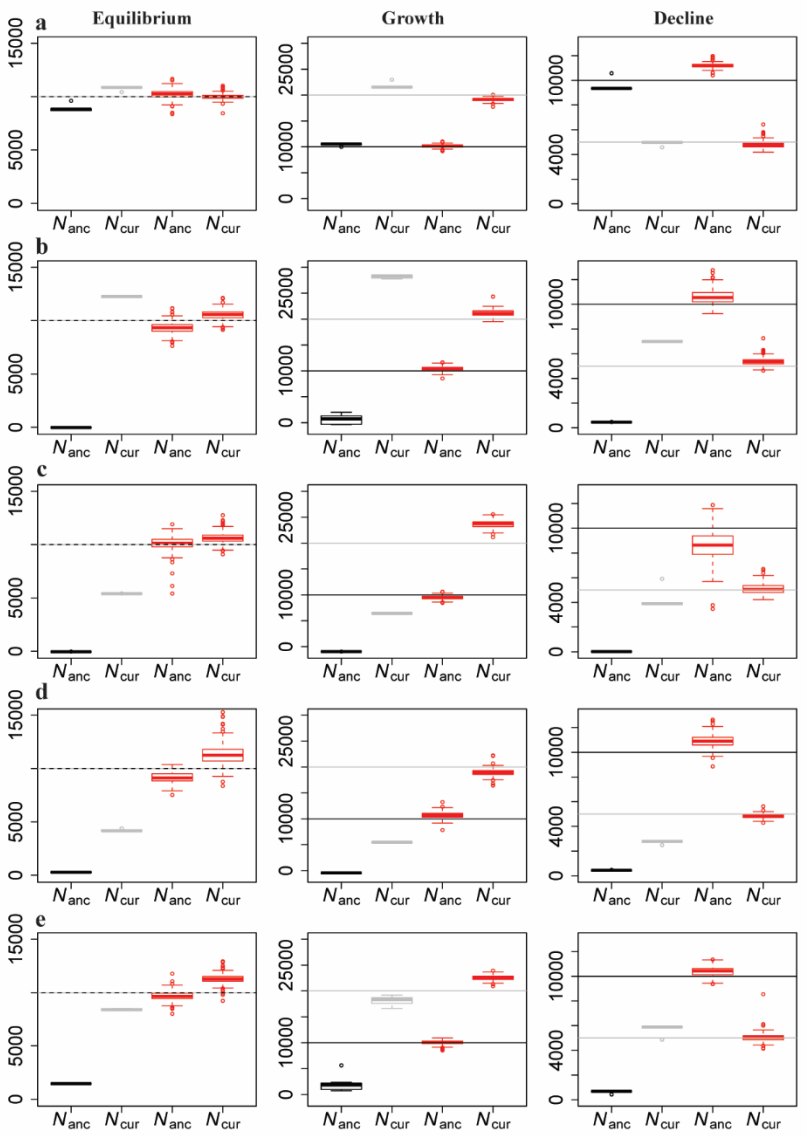


Figure 4: Comparison of inference of ancestral (N_{anc}) and current (N_{cur}) population sizes by assuming neutrality *vs* by varying the DFE shape as a nuisance parameter, using an ABC framework. Inference is shown for demographic equilibrium (left column), 2-fold exponential growth (middle column), 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 ($f_0 = 1, f_1 = 0, f_2 = 0, f_3 = 0$), (b) weak purifying selection ($f_0 = 0.1, f_1 = 0.7, f_2 = 0.1, f_3 = 0.1$), (c) moderately strong purifying selection ($f_0 = 0.1, f_1 = 0.1, f_2 = 0.7, f_3 = 0.1$), (d) strong purifying selection ($f_0 = 0.1, f_1 = 0.1, f_2 = 0.1, f_3 = 0.7$), and (e) a DFE in which all classes of mutations are equally present ($f_0 = f_1 = f_2 = f_3 = 0.25$). In each, the horizontal lines give the true value (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, while red boxes represent estimates when the DFE is treated as a nuisance parameter.

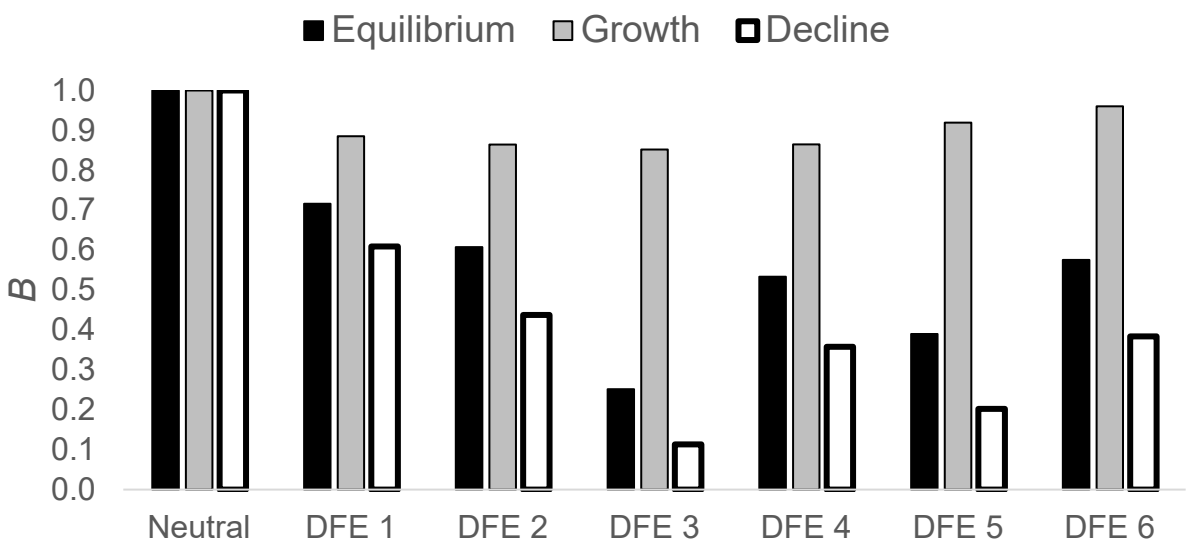


Figure 5: Reduction in nucleotide diversity (B) due to background selection under demographic equilibrium (black bars), population growth (gray bars), and population decline (white bars), under varying DFEs:
DFE 1: $f_0 = 0.1, f_1 = 0.7, f_2 = 0.1, f_3 = 0.1$;
DFE 2: $f_0 = 0.1, f_1 = 0.1, f_2 = 0.7, f_3 = 0.1$;
DFE 3: $f_0 = 0.1, f_1 = 0.1, f_2 = 0.1, f_3 = 0.7$;
DFE 4: $f_0 = f_1 = f_2 = f_3 = 0.25$;
DFE 5: $f_0 = 0.5, f_1 = 0.0, f_2 = 0.0, f_3 = 0.5$;
DFE 6: $f_0 = 0.7, f_1 = 0.0, f_2 = 0.0, f_3 = 0.3$.
Here, exonic sites comprise ~10% of the genome, roughly mimicking the density of the human genome. BGS effects differ strongly by demographic model. Note that here $B = \pi(D) / \pi_0(D)$, where $\pi(D)$ is the nucleotide diversity in neutral regions when exons experience purifying selection under demographic model D (*i.e.*, with BGS), while $\pi_0(D)$ is the nucleotide diversity under unlinked neutrality for the same demographic model D (*i.e.*, without BGS).

REFERENCES

Andolfatto P., 2005 Adaptive evolution of non-coding DNA in *Drosophila*. Nature 437: 1149–1152. https://doi.org/10.1038/nature04107

Bank C., G. B. Ewing, A. Ferrer-Admettla, M. Foll, and J. D. Jensen, 2014 Thinking too positive? Revisiting current methods of population genetic selection inference. Trends Genet. 30: 540–546. https://doi.org/10.1016/j.tig.2014.09.010

Beichman A. C., T. N. Phung, and K. E. Lohmueller, 2017 Comparison of single genome and allele frequency data reveals discordant demographic histories. G3 (Bethesda) 7: 3605–3620. https://doi.org/10.1534/g3.117.300259

Beichman A. C., E. Huerta-Sanchez, and K. E. Lohmueller, 2018 Using genomic data to infer historic population dynamics of nonmodel organisms. Annu. Rev. Ecol. Evol. Syst. 49: 433–456. https://doi.org/10.1146/annurev-ecolsys-110617-062431

Bhaskar A., Y. X. R. Wang, and Y. S. Song, 2015 Efficient inference of population size histories and locus-specific mutation rates from large-sample genomic variation data. Genome Res. 25: 268–279. https://doi.org/10.1101/gr.178756.114

Blattner F. R., G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, *et al.*, 1997 The complete genome sequence of *Escherichia coli* K-12. Science 277: 1453–1462. https://doi.org/10.1126/science.277.5331.1453

Bunnefeld L., L. A. F. Frantz, and K. Lohse, 2015 Inferring bottlenecks from genome-wide samples of short sequence blocks. Genetics 201: 1157–1169. https://doi.org/10.1534/genetics.115.179861

Campos J. L., L. Zhao, and B. Charlesworth, 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: E4762–E4771. https://doi.org/10.1073/pnas.1619434114

Campos J. L., and B. Charlesworth, 2019 The effects on neutral variability of recurrent selective sweeps and background selection. Genetics 212: 287–303. https://doi.org/10.1534/genetics.119.301951

Castellano D., A. Eyre-Walker, and K. Munch, 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: 3550–3561. https://doi.org/10.1093/gbe/evz215

Chamary J., and L. D. Hurst, 2005 Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. Genome Biol. 6: R75. https://doi.org/10.1186/gb-2005-6-9-r75

Charlesworth B., M. T. Morgan, and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics 134: 1289–1303.

544 Charlesworth B., 2013 Background selection 20 years on. The Wilhelmine E. Key 2012
545 invitational lecture. J. Hered. 104: 161–171. https://doi.org/10.1093/jhered/ess136

546 Chikhi L., W. Rodríguez, S. Grusea, P. Santos, S. Boitard, *et al.*, 2018 The IICR (inverse
547 instantaneous coalescence rate) as a summary of genomic diversity: insights into
548 demographic inference and model choice. Heredity 120: 13–24.
549 https://doi.org/10.1038/s41437-017-0005-6

550 Choi J. Y., and C. F. Aquadro, 2016 Recent and long term selection across synonymous sites in
551 *Drosophila ananassae*. J. Mol. Evol. 83: 50–60. https://doi.org/10.1007/s00239-016-
552 9753-9

553 Comeron J. M., R. Ratnappan, and S. Bailin, 2012 The many landscapes of recombination in
554 *Drosophila melanogaster*. PLoS Genet. 8: e1002905.
555 https://doi.org/10.1371/journal.pgen.1002905

556 Csilléry K., M. G. B. Blum, O. E. Gaggiotti, and O. François, 2010 Approximate Bayesian
557 Computation (ABC) in practice. Trends Ecol. Evol. 25: 410–418.
558 https://doi.org/10.1016/j.tree.2010.04.001

559 Cutter A. D., and B. A. Payseur, 2013 Genomic signatures of selection at linked sites: unifying
560 the disparity among species. Nat. Rev. Genet. 14: 262–274.
561 https://doi.org/10.1038/nrg3425

562 Elyashiv E., S. Sattath, T. T. Hu, A. Strutsovsky, G. McVicker, *et al.*, 2016 A genomic map of
563 the effects of linked selection in *Drosophila*. PLoS Genet. 12: e1006130.
564 https://doi.org/10.1371/journal.pgen.1006130

565 Ewing G. B., and J. D. Jensen, 2016 The consequences of not accounting for background
566 selection in demographic inference. Mol. Ecol. 25: 135–141.
567 https://doi.org/10.1111/mec.13390

568 Excoffier L., I. Dupanloup, E. Huerta-Sánchez, V. C. Sousa, and M. Foll, 2013 Robust
569 demographic inference from genomic and SNP data. PLoS Genet. 9: e1003905.
570 https://doi.org/10.1371/journal.pgen.1003905

571 Eyre-Walker A., and P. D. Keightley, 2009 Estimating the rate of adaptive molecular evolution
572 in the presence of slightly deleterious mutations and population size change. Mol. Biol.
573 Evol. 26: 2097–2108. https://doi.org/10.1093/molbev/msp119

574 Fiston-Lavier A.-S., N. D. Singh, M. Lipatov, and D. A. Petrov, 2010 *Drosophila melanogaster*
575 recombination rate calculator. Gene 463: 18–20.
576 https://doi.org/10.1016/j.gene.2010.04.015

577 Fulgione A., M. Koornneef, F. Roux, J. Hermisson, and A. M. Hancock, 2018 Madeiran
578 *Arabidopsis thaliana* reveals ancient long-range colonization and clarifies demography in
579 Eurasia. Mol. Biol. Evol. 35: 564–574. https://doi.org/10.1093/molbev/msx300

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. https://doi.org/10.1371/journal.pgen.1000695

Haddrill P. R., K. R. Thornton, B. Charlesworth, and P. Andolfatto, 2005 Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. Genome Res. 15: 790–799. https://doi.org/10.1101/gr.3541005

Haller B. C., and P. W. Messer, 2019 SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. Mol. Biol. Evol. 36: 632–637. https://doi.org/10.1093/molbev/msy228

Harris K., and R. Nielsen, 2013 Inferring demographic history from a spectrum of shared haplotype lengths. PLoS Genet. 9: e1003521. https://doi.org/10.1371/journal.pgen.1003521

Hernandez R. D., J. L. Kelley, E. Elyashiv, S. C. Melton, A. Auton, *et al.*, 2011 Classic selective sweeps were rare in recent human evolution. Science 331: 920–924. https://doi.org/10.1126/science.1198878

Hung C.-M., P.-J. L. Shaner, R. M. Zink, W.-C. Liu, T.-C. Chu, *et al.*, 2014 Drastic population fluctuations explain the rapid extinction of the passenger pigeon. Proc. Nat. Acad. Sci. U.S.A. 111: 10636–10641. https://doi.org/10.1073/pnas.1401526111

Jackson B. C., J. L. Campos, P. R. Haddrill, B. Charlesworth, and K. Zeng, 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: 102–123. https://doi.org/10.1093/gbe/evw291

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. U.S.A. 110: 13067–13072. https://doi.org/10.1073/pnas.1215206110

Jensen J. D., B. A. Payseur, W. Stephan, C. F. Aquadro, M. Lynch, *et al.*, 2019 The importance of the Neutral Theory in 1968 and 50 years on: A response to Kern and Hahn 2018. Evolution 73: 111–114. https://doi.org/10.1111/evo.13650

Johri P., B. Charlesworth, and J. D. Jensen, 2020 Towards an evolutionarily appropriate null model: jointly Inferring demography and purifying selection. Genetics. https://doi.org/10.1534/genetics.119.303002

Jones M. R., and J. M. Good, 2016 Targeted capture in evolutionary and ecological genomics. Mol. Ecol. 25: 185–202. https://doi.org/10.1111/mec.13304

Kaiser V. B., and B. Charlesworth, 2009 The effects of deleterious mutations on evolution in non-recombining genomes. Trends Genet. 25: 9–12. https://doi.org/10.1016/j.tig.2008.10.009

616 Keightley P. D., and A. Eyre-Walker, 2007 Joint inference of the distribution of fitness effects of
617 deleterious mutations and population demography based on nucleotide polymorphism
618 frequencies. Genetics 177: 2251–2261. https://doi.org/10.1534/genetics.107.080663

619 Keightley P. D., U. Trivedi, M. Thomson, F. Oliver, S. Kumar, *et al.*, 2009 Analysis of the
620 genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation
621 lines. Genome Res. 19: 1195–1201. https://doi.org/10.1101/gr.091231.109

622 Keightley P. D., R. W. Ness, D. L. Halligan, and P. R. Haddrill, 2014 Estimation of the
623 spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib
624 family. Genetics 196: 313–320. https://doi.org/10.1534/genetics.113.158758

625 Kelleher J., Y. Wong, A. W. Wohns, C. Fadil, P. K. Albers, *et al.*, 2019 Inferring whole-genome
626 histories in large population datasets. Nat. Genet. 51: 1330–1338.
627 https://doi.org/10.1038/s41588-019-0483-y

628 Kent W. J., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, *et al.*, 2002 The Human
629 Genome Browser at UCSC. Genome Res. 12: 996–1006.
630 https://doi.org/10.1101/gr.229102

631 Kousathanas A., and P. D. Keightley, 2013 A comparison of models to infer the distribution of
632 fitness effects of new mutations. Genetics 193: 1197–1208.
633 https://doi.org/10.1534/genetics.112.148023

634 Lander E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, *et al.*, 2001 Initial sequencing
635 and analysis of the human genome. Nature 409: 860–921.
636 https://doi.org/10.1038/35057062

637 Lapierre M., A. Lambert, and G. Achaz, 2017 Accuracy of demographic inferences from the site
638 frequency spectrum: the case of the Yoruba population. Genetics 206: 439–449.
639 https://doi.org/10.1534/genetics.116.192708

640 Li H., and R. Durbin, 2011 Inference of human population history from individual whole-
641 genome sequences. Nature 475: 493–496. https://doi.org/10.1038/nature10231

642 Liang P., H. S. A. Saqib, X. Zhang, L. Zhang, and H. Tang, 2018 Single-base resolution map of
643 evolutionary constraints and annotation of conserved elements across major grass
644 genomes. Genome Biol. Evol. 10: 473–488. https://doi.org/10.1093/gbe/evy006

645 Lukic S., and J. Hey, 2012 Demographic inference using spectral methods on SNP data, with an
646 analysis of the human out-of-Africa expansion. Genetics 192: 619–639.
647 https://doi.org/10.1534/genetics.112.141846

648 Lynch M., 2007 *The Origins of Genome Architecture*. Sinauer Associates, Sunderland,
649 Massachusetts.

650 Mazet O., W. Rodríguez, S. Grusea, S. Boitard, and L. Chikhi, 2016 On the importance of being
651 structured: instantaneous coalescence rates and human evolution--lessons for ancestral

652 population size inference? Heredity (Edinb) 116: 362–371.
653 https://doi.org/10.1038/hdy.2015.104

654 Nicolaisen L. E., and M. M. Desai, 2013 Distortions in genealogies due to purifying selection
655 and recombination. Genetics 195: 221–230. https://doi.org/10.1534/genetics.113.152983

656 O'Fallon B. D., J. Seger, and F. R. Adler, 2010 A continuous-state coalescent and the impact of
657 weak selection on the structure of gene genealogies. Mol. Biol. Evol. 27: 1162–1172.
658 https://doi.org/10.1093/molbev/msq006

659 Orozco-terWengel P., 2016 The devil is in the details: the effect of population structure on
660 demographic inference. Heredity (Edinb) 116: 349–350.
661 https://doi.org/10.1038/hdy.2016.9

662 Palkopoulou E., M. Lipson, S. Mallick, S. Nielsen, N. Rohland, *et al.*, 2018 A comprehensive
663 genomic history of extinct and living elephants. Proc. Nat. Acad. Sci. U.S.A. 115:
664 E2566–E2574. https://doi.org/10.1073/pnas.1720554115

665 Pouyet F., S. Aeschbacher, A. Thiéry, and L. Excoffier, 2018 Background selection and biased
666 gene conversion affect more than 95% of the human genome and bias demographic
667 inferences, (K. Veeramah, P. J. Wittkopp, and I. Gronau, Eds.). eLife 7: e36317.
668 https://doi.org/10.7554/eLife.36317

669 Ragsdale A. P., and R. N. Gutenkunst, 2017 Inferring demographic history using two-locus
670 statistics. Genetics 206: 1037–1048. https://doi.org/10.1534/genetics.117.201251

671 Sanjuán R., 2010 Mutational fitness effects in RNA and single-stranded DNA viruses: common
672 patterns revealed by site-directed mutagenesis studies. Philos. Trans. R. Soc. Lond., B,
673 Biol. Sci. 365: 1975–1982. https://doi.org/10.1098/rstb.2010.0063

674 Schiffels S., and R. Durbin, 2014 Inferring human population size and separation history from
675 multiple genome sequences. Nat. Genet. 46: 919–925. https://doi.org/10.1038/ng.3015

676 Schneider A., B. Charlesworth, A. Eyre-Walker, and P. D. Keightley, 2011 A method for
677 inferring the rate of occurrence and fitness effects of advantageous mutations. Genetics
678 189: 1427–1437. https://doi.org/10.1534/genetics.111.131730

679 Sheehan S., and Y. S. Song, 2016 Deep learning for population genetic inference. PLoS Comput.
680 Biol. 12: e1004845. https://doi.org/10.1371/journal.pcbi.1004845

681 Siepel A., G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, *et al.*, 2005 Evolutionarily
682 conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 15:
683 1034–1050. https://doi.org/10.1101/gr.3715005

684 Speidel L., M. Forest, S. Shi, and S. R. Myers, 2019 A method for genome-wide genealogy
685 estimation for thousands of samples. Nat. Genet. 51: 1321–1329.
686 https://doi.org/10.1038/s41588-019-0484-x

687 Steinrücken M., J. Kamm, J. P. Spence, and Y. S. Song, 2019 Inference of complex population
688 histories using whole-genome sequences from multiple populations. Proc. Natl. Acad.
689 Sci. U.S.A. 116: 17115–17120. https://doi.org/10.1073/pnas.1905060116

690 Teshima K. M., G. Coop, and M. Przeworski, 2006 How reliable are empirical genomic scans for
691 selective sweeps? Genome Res 16: 702–712. https://doi.org/10.1101/gr.5105206

692 Thornton K., 2003 Libsequence: a C++ class library for evolutionary genetic analysis.
693 Bioinformatics 19: 2325–2327. https://doi.org/10.1093/bioinformatics/btg316

694 Thornton K. R., and J. D. Jensen, 2007 Controlling the false-positive rate in multilocus genome
695 scans for selection. Genetics 175: 737–750. https://doi.org/10.1534/genetics.106.064642

696 Wang L., T. M. Beissinger, A. Lorant, C. Ross-Ibarra, J. Ross-Ibarra, *et al.*, 2017 The interplay
697 of demography and selection during maize domestication and expansion. Genome Biol.
698 18: 215. https://doi.org/10.1186/s13059-017-1346-4

699 Warren W. C., A. J. Jasinska, R. García-Pérez, H. Svardal, C. Tomlinson, *et al.*, 2015 The
700 genome of the vervet (*Chlorocebus aethiops sabaeus*). Genome Res. 25: 1921–1933.
701 https://doi.org/10.1101/gr.192922.115

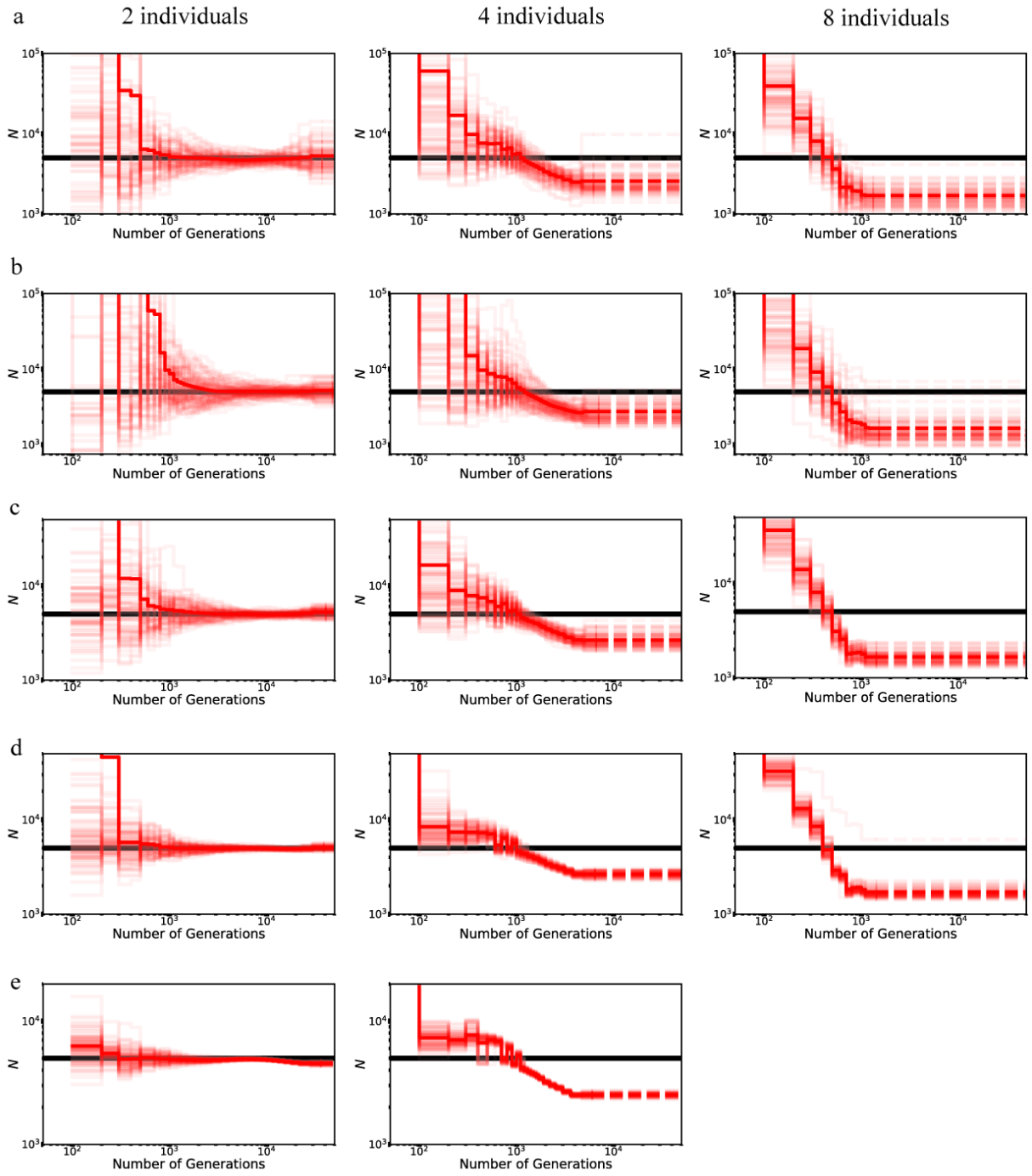
702 Wu F., S. Zhao, B. Yu, Y.-M. Chen, W. Wang, *et al.*, 2020 A new coronavirus associated with
703 human respiratory disease in China. Nature 579: 265–269.
704 https://doi.org/10.1038/s41586-020-2008-3

705 Zeng K., and B. Charlesworth, 2010 Studying patterns of recent evolution at synonymous sites
706 and intronic sites in *Drosophila melanogaster*. J. Mol. Evol. 70: 116–128.
707 https://doi.org/10.1007/s00239-009-9314-6

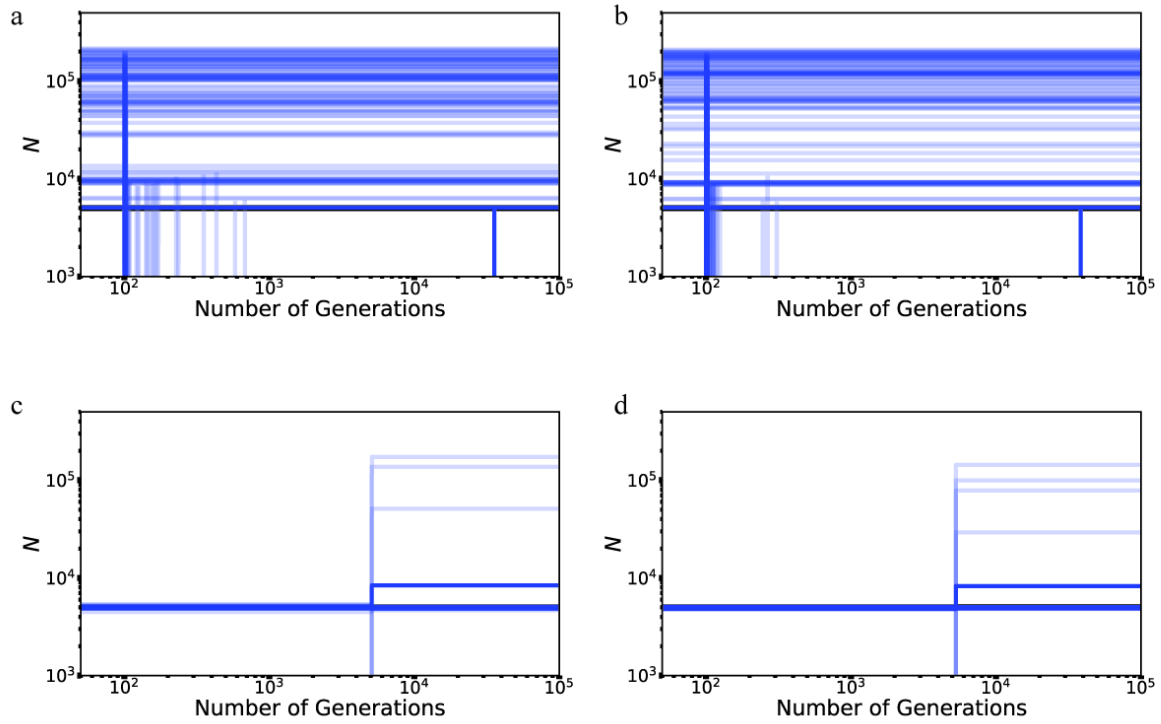
708 Zhou Y., M. Massonnet, J. S. Sanjak, D. Cantu, and B. S. Gaut, 2017 Evolutionary genomics of
709 grape (*Vitis vinifera ssp. vinifera*) domestication. Proc. Nat. Acad. Sci. U.S.A. 114:
710 11715–11720. https://doi.org/10.1073/pnas.1709257114

711

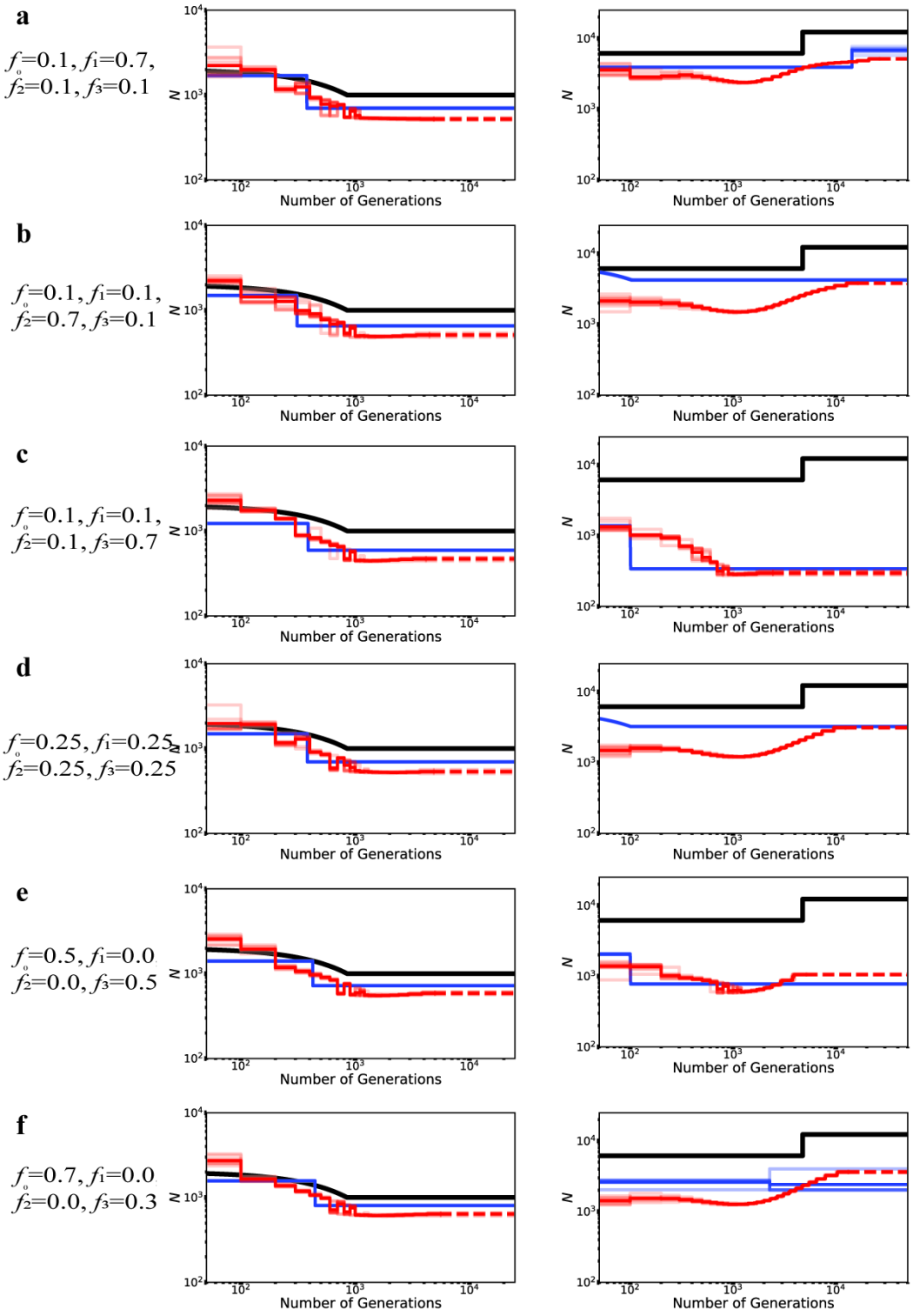
SUPPLEMENTARY



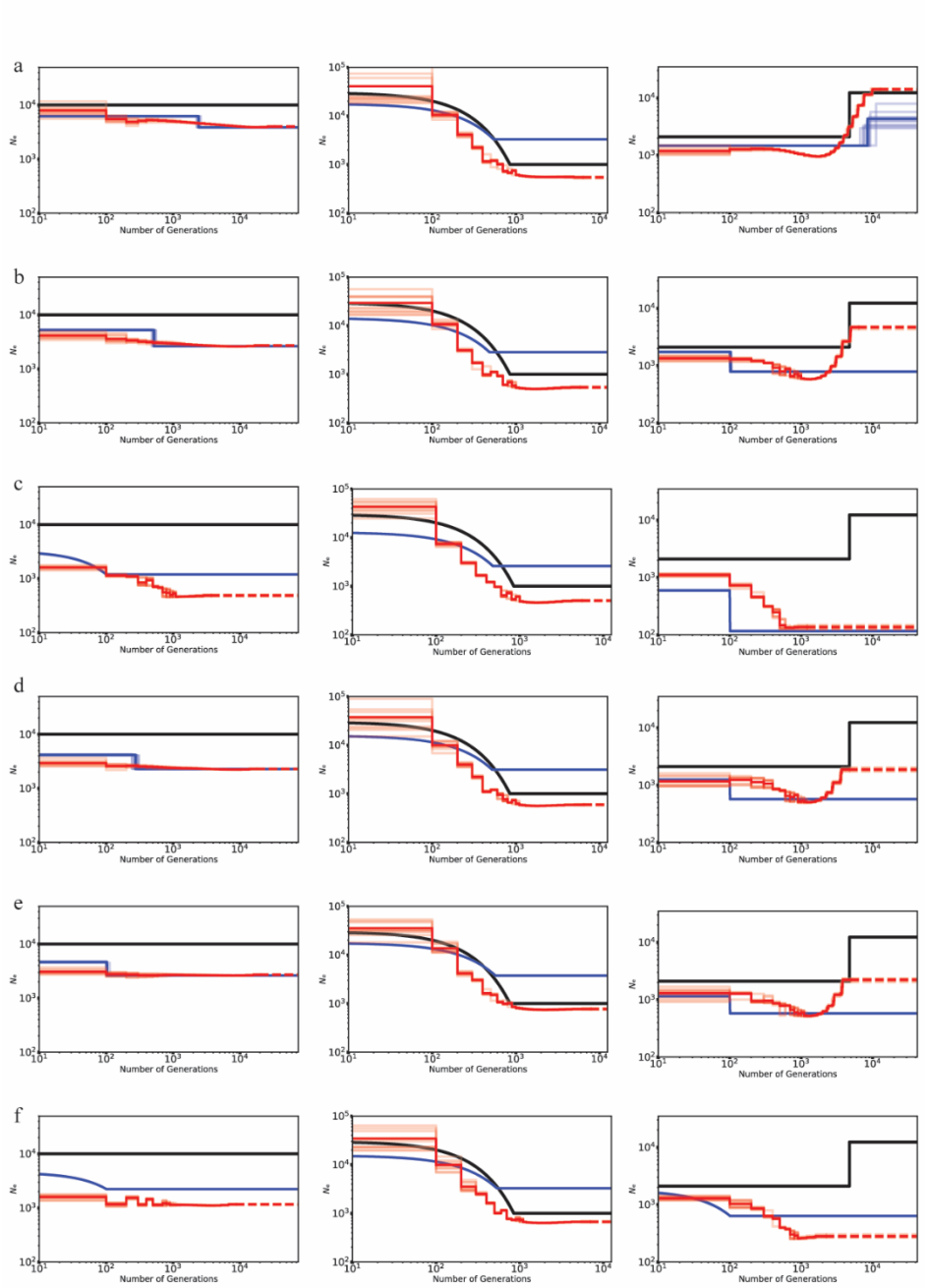
Supp Figure 1: Performance of MSMC under neutrality when using 2, 4 and 8 individuals for inference and on varying chromosome sizes – (a) 1 Mb, (b) 10 Mb, (c) 50 Mb, (d) 200 Mb, (e) 1 Gb. Approximately 100 replicates of each are shown above. Grey lines indicate 0.5N and 2N. As shown, MSMC has a tendency to take a common shape, often falsely indicating recent population growth.



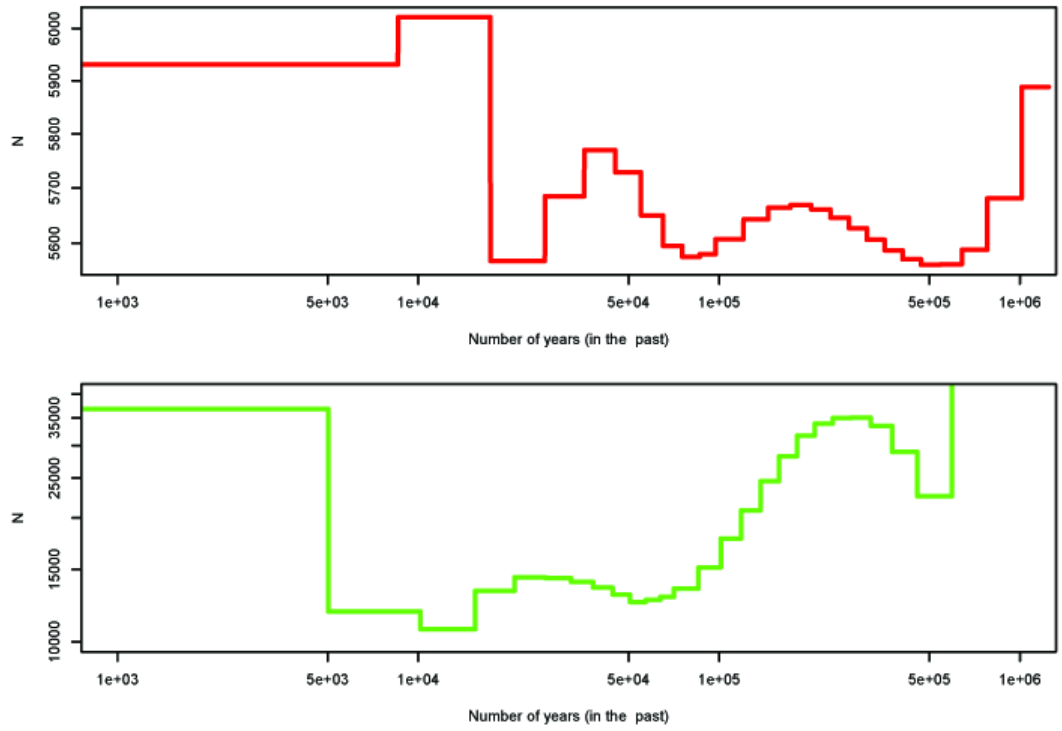
Supp Figure 2: Performance of *fastsimcoal2* under neutrality using 50 diploid individuals for inference when varying chromosome sizes – (a) 10 Mb, (b) 50 Mb, (c) 200 Mb, (d) 1 Gb. Approximately 100 replicates of each are shown above (blue lines) and the true model is shown in black.



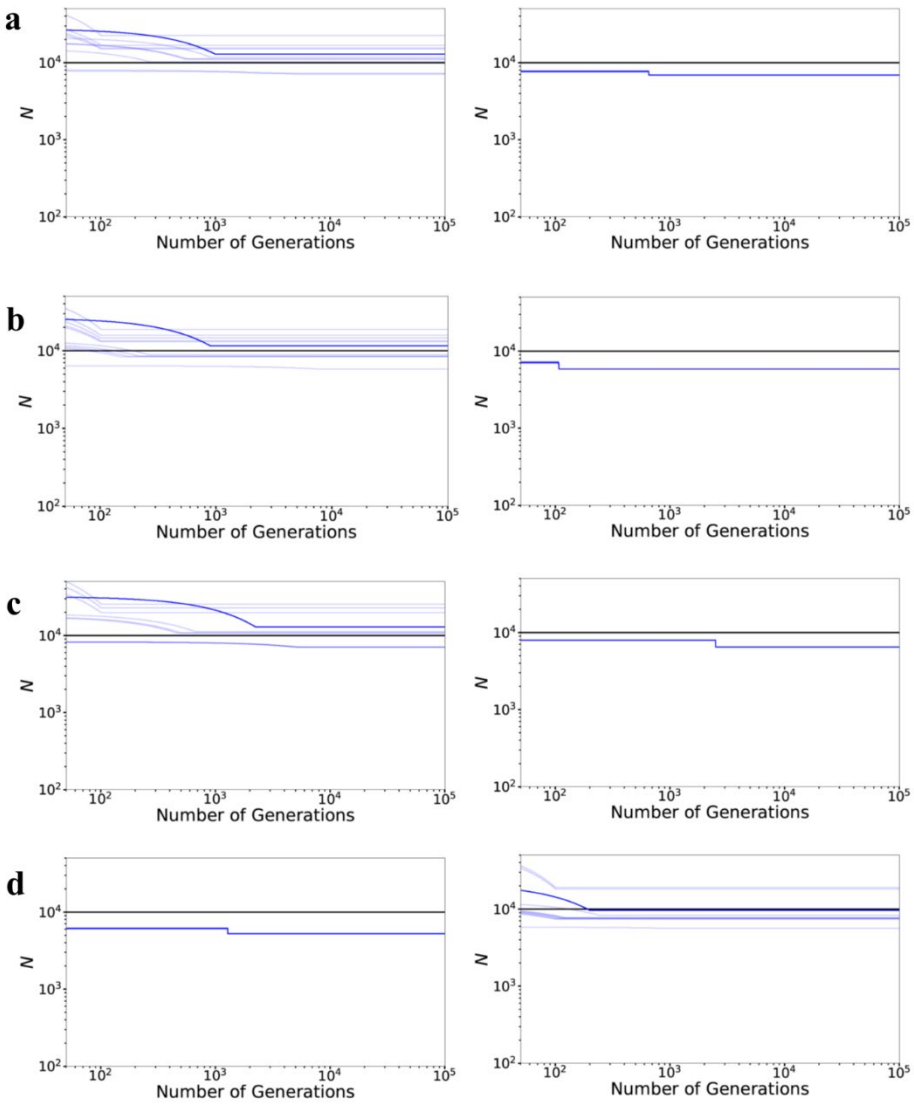
Supp Figure 3: Inferred demography by MSMC (red lines) and *fastsimcoal2* (blue lines) in the presence of background selection, with the true DFE shown to the left of the panel, for 2-fold instantaneous decline (right column) and 2-fold exponential growth (left column). In this case, 20% of the genome experiences direct selection. The true demographic model is shown as black lines.

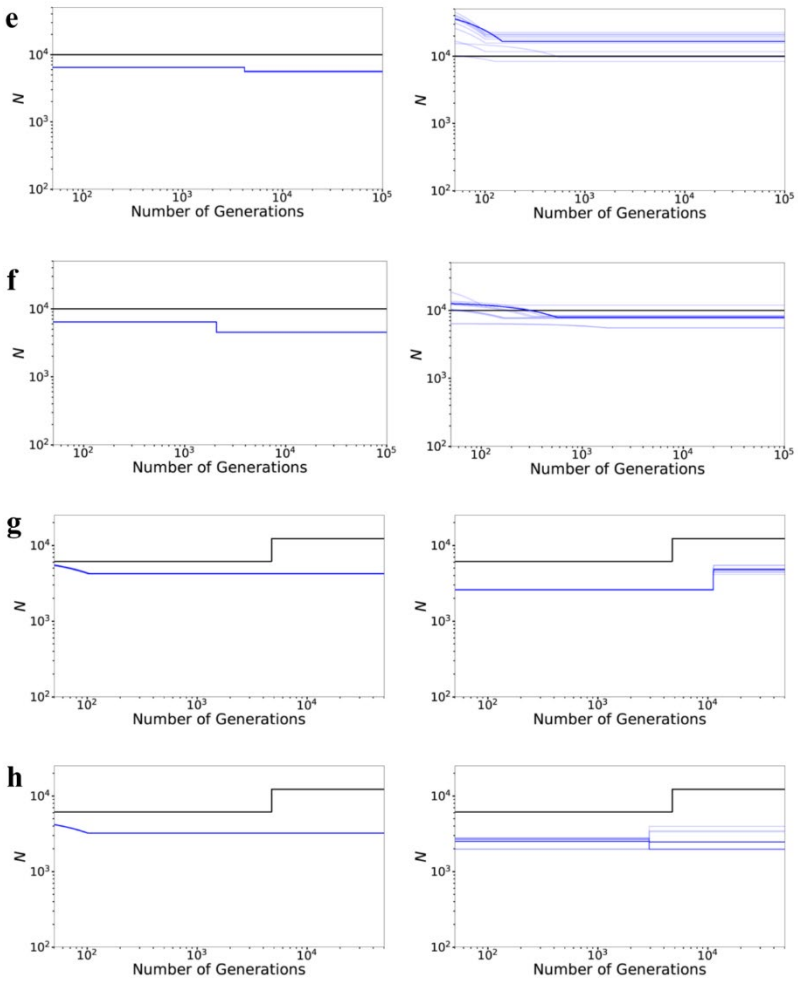


Supp Figure 4: Inference of demography by MSMC (red lines; 10 replicates) and *fastsimcoal2* (blue lines; 10 replicates) under demographic equilibrium (left column), 30-fold exponential growth (middle column), and ~6-fold instantaneous decline (right column) in the presence of direct purifying selection (*i.e.*, directly selected sites are not masked). The true demographic model is depicted in black lines. Exonic sites experience purifying selection specified by the following DFEs: (a) $f_0 = 0.1, f_1 = 0.7, f_2 = 0.1, f_3 = 0.1$, (b) $f_0 = 0.1, f_1 = 0.1, f_2 = 0.7, f_3 = 0.1$, (c) $f_0 = 0.1, f_1 = 0.1, f_2 = 0.1, f_3 = 0.7$, (d) $f_0 = 0.25, f_1 = 0.25, f_2 = 0.25, f_3 = 0.25$, (e) $f_0 = 0.7, f_1 = 0.0, f_2 = 0.0, f_3 = 0.3$, (f) $f_0 = 0.5, f_1 = 0.0, f_2 = 0.0, f_3 = 0.5$. In this case, 20% of the genome is composed of exons.



Supp Figure 5: Inference of demographic history by MSMC. Top panel / red line: simulations in which the true model is constant population size, and 50% of new mutations in exons are strongly deleterious with the remainder being neutral, where exons comprise of 5% of the genome. Bottom panel / green line: the empirical estimate of population history of the YRI population inferred with MSMC by Schiffels and Durbin (2014). The X-axis is in years (assuming a generation time of 30 years). Note that the Y-axes are on different scales, and the magnitude of change observed in the empirical data is considerably larger than that observed in the simulated data. Thus, this comparison is only meant to illustrate this common shape taken in MSMC plots (and see similar shapes in, for example, vervets (Warren *et al.* 2015; Figure 4) and passenger pigeons (Hung *et al.* 2014; Figure 2)).





Supp Figure 6: Demographic inference by *fastsimcoal2* when performed using all SNPs (right column) and when SNPs are thinned to be separated by 5 kb (left column). (a) Direct purifying selection (exonic sites are not masked) in 5% of the genome given by the DFE: $f_0 = f_1 = f_2 = f_3 = 0.25$. (b) Direct purifying selection (exonic sites are not masked) in 5% of the genome given by the DFE: $f_0 = 0.5, f_1 = 0.0, f_2 = 0.0, f_3 = 0.5$. (c) Direct purifying selection (exonic sites are not masked) in 10% of the genome given by the DFE: $f_0 = 0.1, f_1 = 0.7, f_2 = 0.1, f_3 = 0.1$. (d) Direct purifying selection (exonic sites are not masked) in 10% of the genome given by the DFE: $f_0 = 0.7, f_1 = 0.0, f_2 = 0.0, f_3 = 0.3$. (e) Background selection (*i.e.*, exonic sites are masked) in which 5% of the genome is exonic and given by the DFE: $f_0 = 0.5, f_1 = 0.0, f_2 = 0.0, f_3 = 0.5$. (f) Background selection (*i.e.*, exonic sites are masked) in which 20% of the genome is exonic and given by the DFE: $f_0 = 0.1, f_1 = 0.7, f_2 = 0.1, f_3 = 0.1$. (g) Background selection (*i.e.*, exonic sites are masked) in which 20% of the genome is exonic and given by the DFE: $f_0 = 0.1, f_1 = 0.1, f_2 = 0.7, f_3 = 0.1$. (h) Background selection (*i.e.*, exonic sites are masked) in which 20% of the genome is exonic and given by the DFE: $f_0 = f_1 = f_2 = f_3 = 0.25$.

Supp Table 1: Parameters underlying the human-like demographic models considered.

	Demographic models	**Ancestral population size**	**Current population size**	**Time of change in generations**
1	Equilibrium	10,000	10,000	NA
2	Exponential growth	1000	30,000	850
3	Instantaneous decline	12,300	2,100	4,750