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 85287, USA

[2]Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, EH9 3FL, United Kingdom

[3]Institute of Ecology and Evolution, University of Berne, Berne 3012, Switzerland

[4]Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland

[*]Correspondence can be addressed to these authors

Corresponding authors:

Parul Johri

pjohri1@asu.edu

Jeffrey D. Jensen

Jeffrey.D.Jensen@asu.edu

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 (DFE) 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 DFE 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.

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

Running title: Demographic inference with selection

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. While there is evidence for weak direct selection on 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), 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 ~3-8% in humans, ~12% in rice, 37-53% in *D. melanogaster*, and 47-68% in *S. cerevisiae* (Siepel et al. 2005; Liang et al. 2018). Furthermore, many organisms have highly compact genomes, with ~88% of the *E. 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).

While 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, 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; 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 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 and demographic equilibrium (*i.e.,* a constant population size of 5000 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 [1887]; for 50 diploid individuals, these values were 10,354 [225], 51,863 [567], 207,118 [1139] and 1,035,393 [2476] 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 (Supp Figures 1, 2). 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 1 diploid genome with large segments (Supp Figures 1, 2). 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 1 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 (Supp Table 1). 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), while an alternative model of either instantaneous size change or instant bottleneck was increasingly preferred for larger regions (Supp Tables 2, 3),

155 although the estimates of ancestral sizes were correct. This finding suggests that the non-
156 independence of SNPs may result in model mis-identification. Indeed, since the model choice
157 procedure assumes that SNPs are independent, the true number of independent SNPs is
158 overestimated, which results in an overestimation in the confidence of the model choice with
159 increasing amount of data. However, it is interesting to note that the parameter values underlying
160 the non-constant size preferred model were often pointing towards a constant-population size
161 (see below). When model selection was performed using sparser SNP densities (*i.e.*, 1 SNP per 5
162 kb, 50 kb or 100 kb), the correct model was recovered for longer chromosomes up to 200 Mb
163 (Supp Tables 2, 3; Supp Figures 3, 4), although model selection was slightly less accurate for
164 smaller chromosomes due to the decrease in the total amount of data. As suspected, the biases
165 introduced by the non-independence of SNPs were found to be concordant with the level of
166 linkage disequilibrium amongst SNPs used for the analysis (for 10 SNP windows, in which SNPs
167 were separated by 50 kb (100 kb), mean r^2 = 0.027 (0.020), compared to the all-SNP mean r^2 of
168 0.118, and to the completely unlinked SNPs mean r^2 of 0.010; Supp Table 4). Additionally, AIC
169 performed on partially linked SNPs may impose an insufficient penalty on larger number of
170 parameters, resulting in an undesirable preference for parameter-rich models. We found that
171 implementing a more severe penalty improved inference considerably, even for 1 Gb
172 chromosome sizes (Supp Table 5, 6). This model selection performance, the potential corrections
173 related to increased penalties, as well as the total number of SNPs and SNP thinning, should be
174 investigated on a case-by-case basis in empirical applications, owing to the contribution of
175 multiple underlying parameters (*e.g.*, chromosome length, recombination rates, and SNP
176 densities).
177 In the light of this performance assessment, all further analyses were restricted to
178 characterizing demographic inference on data that far exceeded 1 Gb and roughly matched the
179 structure and size of the human genome - for every diploid individual, 22 chromosomes
180 (autosomes) of size 150 Mb each were simulated, which amounted to roughly 3 Gb of total
181 sequence. Ten independent replicates of each parameter combination were performed
182 throughout, and inference utilized one and fifty diploid individuals for MSMC and *fastsimcoal2,*
183 respectively.
184
185

186 **Effect of the strength of purifying selection on demographic inference**

187 In order to test the accuracy of demographic inference in the presence of BGS, all 22
188 chromosomes were simulated with exons of size 350 bp each, with varying sizes of introns and
189 intergenic regions (see Methods) in order to vary the fraction (5%, 10% and 20%) of the genome
190 under selection. Because the strength of selection acting on deleterious mutations affects the
191 distance over which the effects of BGS extend, demographic inference was evaluated for various
192 DFEs (Table 1). The DFE was modelled as a discrete distribution with four fixed classes: $0 \leq$
193 $2N_{anc}s < 1$, $1 \leq 2N_{anc}s < 10$, $10 \leq 2N_{anc}s < 100$ and $100 < 2N_{anc}s < 2N_{anc}$, where N_{anc}
194 is the ancestral effective population size and s is the reduction in the fitness of the homozygous
195 mutant relative to wildtype. The fitness effects of mutations were uniformly distributed within
196 each bin, and assumed to be semi-dominant, following a multiplicative fitness model for multiple
197 loci; the DFE shape was altered by varying the proportion of mutations belonging to each class,
198 given by f_0, f_1, f_2, and f_3, respectively (see Methods). Three DFEs highly skewed towards a
199 particular class were initially used to assess the impact of the strength of selection on
200 demographic inference (with the remaining mutations equally distributed amongst the other three
201 classes): DFE1: a DFE in which 70% of mutations have weakly deleterious fitness effects (*i.e.*, f_1
202 $= 0.7$); DFE2: a DFE in which 70% of mutations have moderately deleterious fitness effects
203 (*i.e.*, $f_2 = 0.7$); and DFE3: a DFE in which 70% of mutations have strongly deleterious fitness
204 effects (*i.e.*, $f_3 = 0.7$). A DFE with equal proportions of all deleterious classes (*i.e.*, DFE4: $f_0 =$
205 $f_1 = f_2 = f_3 = 0.25$) was also simulated to evaluate the combined effect of different selective
206 strengths. In addition, two bimodal DFEs consisting of only the neutral and the strongly
207 deleterious class of mutations were simulated to characterize the role of strongly deleterious
208 mutations (DFE5: a DFE in which 50% of mutations have strongly deleterious effects (*i.e.*, $f_3 =$
209 0.5) with the remaining being neutral; and DFE6: a DFE in which 30% of mutations were
210 strongly deleterious (*i.e.*, $f_3 = 0.3$) with the remaining being neutral).

211 In order to understand the effects of BGS, exonic sites were masked, and only linked
212 neutral intergenic and intronic sites were used for demographic inference by both MSMC and
213 *fastsimcoal2* (although comparisons are presented under certain models to analyses based on
214 non-masked datasets). The three demographic models examined were (1) demographic
215 equilibrium, (2) a 30-fold exponential growth, mimicking the recent growth experienced by
216 European human populations, and (3) ~6-fold instantaneous decline, mimicking the out-of-

217 Africa bottleneck in human populations (Figure 1a). Although these models were parameterized
218 using previous estimates of human demographic history (Supp Table 7; Gutenkunst *et al.* 2009),
219 these basic demographic scenarios are applicable to many organisms, although the magnitudes of
220 population size changes in this case may represent an extreme. Under neutrality, inference of
221 parameters of all three simulated demographic models was highly accurate with both MSMC and
222 *fastsimcoal2* (Figure 1a; Supp Table 8). However, when inferring parameters using *fastsimcoal2*,
223 the time of change in case of the population decline model was consistently over-estimated when
224 SNPs separated by 5 kb were used, while the time was accurately inferred when using all SNPs
225 (Supp Table 8). We therefore present our results using all SNPs throughout (with comparisons to
226 1 SNP per 5 kb and 1 SNP per 100 kb thinning, under certain models), and recommend caution
227 when implementing thinning procedures.
228 Under demographic equilibrium, when 20% of the genome experiences direct selection
229 (with masking of the directly selected sites), we found the true population size to be
230 underestimated as expected, and recent population growth mis-inferred (Figure 1, Supp Figure
231 5), even when only 1 SNP per 100kb was used and a higher AIC penalty was employed (Supp
232 Figure 6). Conversely, when the true demographic model was characterized by a recent 30-fold
233 growth, demographic inference was accurate and performed equally well for both MSMC and
234 *fastsimcoal2*, with the exception of a slight underestimation of the ancestral population size for
235 all DFE types. When the true model was population decline, weakly deleterious mutations alone
236 did not affect inference drastically with either method, and it was possible to recover the true
237 model (*i.e.*, decline vs growth) by *fastsimcoal2* in all replicates (Supp Figure 7). However,
238 moderately and strongly deleterious mutations resulted in an underestimation of population size
239 and the inference of an instantaneous bottleneck and strong recent growth respectively, to the
240 extent that population decline was misinterpreted as a bottleneck/growth in all replicates (Supp
241 Figures 5,7). Strong recent growth was inferred (in the presence of moderately and strongly
242 deleterious mutations) even when SNPs separated by 100 kb were used, and an increased penalty
243 was employed against parameter-rich models (Supp Figure 6). We further tested the effect of
244 BGS on demographic inference when changes in population size were less severe, namely, when
245 population growth and decline were only 2-fold, with qualitatively similar results (Supp Figure
246 8).

247 Finally, given the strong evidence that most organisms have a bi-modal DFE with a
248 significant proportion of strongly deleterious or lethal mutations (Sanjuán 2010; Jacquier et al.
249 2013; Kousathanas and Keightley 2013; Bank et al. 2014; Charlesworth 2015; Galtier and
250 Rousselle 2020), we investigated the effect of this strongly deleterious class further. Thus, for
251 comparison with the above, we simulated a rather extreme case in which 30% or 50% of exonic
252 mutations were strongly deleterious with fitness effects uniformly sampled between $100 \leq$
253 $2N_{\text{anc}}s \leq 2N_{\text{anc}}$, with the remaining mutations being neutral (*i.e.*, DFE5 and DFE6; see Table 1).
254 As with the above results, both equilibrium and decline models were falsely inferred as growth,
255 with an order of magnitude underestimation of the true population size (Figure 2).
256 In sum, neglecting BGS frequently results in the inference of population growth, almost
257 regardless of the true underlying demographic model.
258
259 **Effects of density and inclusion/exclusion of directly selected sites on inference**
260 Although we have shown that the presence of purifying selection biases demographic inference,
261 the extent of mis-inference necessarily depends on the fraction of the genome experiencing direct
262 selection. We therefore compared models in which 5%, 10% or 20% of the genome was
263 functional. For this comparison, equal proportions of mutations in each DFE bin were assumed
264 corresponding to DFE4 (Table 1). As before, when the true model was growth, inference was
265 unbiased, with a slight underestimation of ancestral population size when 20% of the genome
266 experienced selection (Figure 3). Population decline was inferred reasonably well if less than
267 10% of the genome experienced direct selection, but could be mis-inferred as growth with
268 greater functional density, as shown in Figure 3. Similarly, the extent to which population size
269 was under-estimated at demographic equilibrium increased with the fraction of the genome under
270 selection. Finally, it is noteworthy that many changes in population size that were falsely inferred
271 were greater than 2-fold in size, suggesting the need for great caution when inferring such
272 changes from real data.
273 Importantly, the results presented do not significantly differ between inference performed
274 while including directly selected sites (*i.e.*, no masking of functional regions; Supp Figure 9)
275 versus inference performed using linked neutral sites (*i.e.*, masking functional regions; Figures 1-
276 3). These results suggest that the exclusion of exonic sites, which is often assumed to provide a
277 sufficiently neutral dataset to enable accurate demographic inference, is not necessarily a

278 satisfactory solution unless gene density is low. For example, demographic inference would
279 naturally be expected to be less biased by BGS for human-like genomes with a relatively low
280 functional density, and more biased in genomes with higher functional density like *D.*
281 *melanogaster.*
282
283 **Effect of BGS on model selection and inferred time of size change using *fastsimcoal2***
284 In order to quantify the effects of BGS on model selection, four competing models were used for
285 inference: equilibrium, instantaneous size change (growth/decline), exponential size change
286 (growth/decline), and an instantaneous bottleneck. Although demographic equilibrium was
287 almost always inferred as an instantaneous size change (70-100% of replicates), the fitted
288 parameters of the size change model were nearly indistinguishable from the correct model
289 (Figure 1a). In other words, the inferred size change was so inconsequential so as to be nearly a
290 constant-size model, suggesting that parameter estimation is usually more reliable than model
291 selection. When there was a substantial proportion of highly deleterious mutations (DFE3 and
292 DFE5), exponential growth was generally inferred. However, when there was a true size change,
293 *fastsimcoal2* performed well in distinguishing between exponential *vs.* instantaneous change
294 models even in the presence of BGS (Supp Figures 5, 6), provided that the magnitude of size
295 change was large. When size changes were on the order of 2-fold, exponential growth was
296 consistently inferred to be instantaneous.
297 With respect to model choice between growth and decline in the presence of BGS
298 (irrespective of instantaneous *vs.* exponential change), as the density of selected sites and
299 strength of purifying selection increased, both equilibrium and decline models were more likely
300 to be inferred as growth and occasionally as instantaneous bottlenecks (Supp Figure 7), while
301 true growth models were generally chosen correctly. It should be added that with such large
302 chromosome sizes (3 Gb of total sequence data), model selection was not observed to vary
303 between replicates using *fastsimcoal2* for any given parameter combination. Thus, in the
304 presence of BGS, high-confidence calls of an incorrect underlying demographic model appear
305 likely.
306 With regard to the time of inferred size change, when the true model was exponential
307 growth, the model was always correctly identified and inference of the time of change was
308 slightly under-estimated in the presence of BGS (Supp Figures 10, 11), consistent with the fact

309 that BGS will further skew the site frequency spectrum towards rare alleles. When the true model
310 was decline, and the model was correctly identified as such, the time of change was modestly
311 over-estimated (Supp Figures 12, 13) – up to ~2-fold for 6× growth and 2.5-fold for 2× growth
312 (when 20% of the genome was exonic).
313
314 **Effect of heterogeneity in recombination rates, mutation rates, and repeat masking**
315 Variation in recombination and mutation rates, as well as the masking of repeat regions, may
316 also affect demographic inference procedures. We evaluated this issue by simulating
317 heterogeneity in both mutation and recombination rates (based on estimated human genome
318 maps, as described in the Methods section), and masking 10% of each simulated segment
319 drawing from the empirical distribution of repeat lengths in the human genome (Supp Figure 14).
320 In general, inferences under neutrality (Supp Figures 15-17) as well as under BGS (Supp Figures
321 18-20) were not affected to a great extent, suggesting such heterogeneity to have a comparatively
322 minor role for the parameter space considered in this study. Thus, serious mis-inference is more
323 likely to be caused by selection. These observations also suggest that simulations performed with
324 mean rates of recombination and mutation, as in this study, are sufficient to evaluate biases
325 caused by BGS.
326
327 **Effects of BGS on diversity and the SFS under various demographic models: theoretical**
328 **expectations versus simulation results**
329 To better understand how BGS can lead to different biases in the inference of population history,
330 we investigated the extent of BGS effects under all three demographic models, with respect to
331 both the expected diversity in the presence of BGS relative to neutrality (B), as well as the shape
332 of the SFS at linked neutral sites. First, we found that B differed among demographic scenarios,
333 with much lower values in the case of equilibrium and decline, concordant with stronger
334 demographic mis-inference (Figure 4). After a population decline, B was lower than that before
335 the size change; while after population expansion, B increased relative to that in the ancestral
336 population, sometimes approaching 1 (Figure 4). This may seem paradoxical, given that the
337 magnitude of the scaled selection coefficient ($2N_es$) decreases with decreasing N_e (*i.e.*, the
338 efficacy of purifying selection decreases, and could thus be expected to result in larger values of
339 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; Pool and Nielsen 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, while 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 the 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 the 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.

371 As with the estimates of the apparent *B* values discussed above, analytical predictions of
372 the expected SFS after an instantaneous / exponential change in population size can be made,
373 using the values of *B* and the SFS at equilibrium in the ancestral population before the population
374 size change using the formulae of Polanski and Kimmel (2003) and Polanski *et al.* (2003) for the
375 purely neutral case, as described in the Methods section. Importantly, the use of the *B* parameter
376 does not in itself cause a skew in the SFS, it merely affects overall diversity values. Figure 5
377 shows that the overall shape of the SFS is predicted reasonably well by the analytical results,
378 although deviations are to be expected for the rare allele classes, which are the most sensitive to
379 demographic change and selection. Overall, the results imply that BGS is more likely to bias
380 demographic inference post-decline compared with post-expansion, consistent with the
381 performance of the methods described above. Although it is notable that the SFS can be
382 reasonably well predicted by correcting for the re-scaling effects of BGS if the effects of BGS in
383 the ancestral population are accurately known, the exact allele frequency patterns observed will
384 depend on the timing of population size changes relative to the time of sampling, as well as the
385 value of *B* prior to the size change. The patterns described here thus represent only a small subset
386 of the possibilities.

387

388 **A potential solution: averaging across all possible DFEs**

389 As shown above, demographic inference can be strongly affected by BGS effects that have not
390 been taken into account, as well as by direct purifying selection. A potential solution is thus to
391 correct for these effects when performing inferences of population history. A widely-used
392 approach to estimating direct selection effects, DFE-alpha, takes a stepwise approach to inferring
393 demography, by using a presumed neutral class (synonymous sites); conditional on that
394 demography, it then estimates the parameters of the DFE (Keightley and Eyre-Walker 2007;
395 Eyre-Walker and Keightley 2009; Schneider et al. 2011; Kousathanas and Keightley 2013).
396 However, this approach does not include the possibility of effects of selection at linked sites,
397 which can result an over-estimate of population growth, and while the DFE may not be mis-
398 inferred strongly (Kim et al. 2017), there is substantial mis-inference of the DFE if synonymous
399 sites experience direct selection (Johri et al. 2020).

400 Building on this idea, Johri et al. (2020) recently proposed an approach that includes both
401 direct and background effects of purifying selection, and simultaneously infers the deleterious

402 DFE and demography. By utilizing the decay of BGS effects around functional regions, they
403 demonstrated high accuracy under the simple demographic models examined. Moreover, the
404 method makes no assumptions about the neutrality of synonymous sites, and can thus be used to
405 estimate selection acting on these sites, as well as in non-coding functional elements. However,
406 this computationally-intensive approach is specifically concerned with jointly inferring the DFE
407 and demographic parameters. As such, if an unbiased characterization of the population history
408 is the sole aim, this procedure may be needlessly involved. We thus here examine the possibility
409 of instead treating the DFE as an unknown nuisance parameter, averaging across all possible
410 DFE shapes, in order to assess whether demographic inference may be improved simply by
411 correcting for these selection effects without inferring their underlying parameter values. This
412 approach utilizes functional (*i.e.*, directly selected) regions, a potential advantage in populations
413 for which only coding data may be available (*e.g.*, exome-capture data; see Jones and Good
414 2016), or more generally in organisms with largely functional genomes.

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

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

433 reality), it is possible to infer demographic history with reasonable accuracy by approximately
434 correcting for these selective effects.
435 This proposed method is most applicable to organisms in which recombination rates are
436 reasonably well known. If the assumed recombination rate is 2-fold lower than the true rate, the
437 ABC approach infers growth by over-estimating the current population size; correspondingly, if
438 the assumed recombination rate is higher than the true rate, the current population size is under-
439 estimated (Supp Figure 21). Interestingly, in both cases the ancestral population sizes are
440 correctly inferred, consistent with previous results (Johri *et al*. 2020).
441
442
443 **CONCLUSIONS**
444
445 While commonly used approaches for inferring demography assume neutrality and independence
446 among segregating sites, these assumptions are likely to be violated in practice. In addition,
447 there is considerable evidence for wide-spread effects of selection at linked sites in many
448 commonly studied organisms (Hernandez et al. 2011; Cutter and Payseur 2013; Williamson et al.
449 2014; Elyashiv et al. 2016; Campos et al. 2017; Booker and Keightley 2018; Pouyet et al. 2018;
450 Ragsdale et al. 2018; Torres et al. 2018; Castellano et al. 2020). Accordingly, we have explored
451 how violations of the assumption of neutrality may affect demographic inference, particularly
452 with regard to the underlying strength of purifying selection and the genomic density of directly
453 selected sites. Generally speaking, the neglect of these effects (*i.e.*, background selection) results
454 in an inference of population growth, with the severity of the growth model roughly scaling with
455 selection strength and density, as well as the inference of historical bottlenecks with some
456 frequency. Thus, when the true underlying model is in fact growth, demographic mis-inference is
457 not particularly severe; when the true underlying model is constant size or decline, the mis-
458 inference can be extreme, with a many-fold underestimation of population size.
459 However, given that BGS will lead to the false inference of recent growth nearly
460 regardless of the true history, it would be difficult in practice to determine the accuracy of this
461 model without independent information on any given empirical application. Moreover, as the
462 two very different methods investigated here result in highly similar mis-inference, we propose
463 that this performance is unlikely to be a feature of these specific approaches, but rather a

464 quantification of the fact that the underlying genealogies are distorted in the presence of BGS.
465 Thus, these problems are likely to be common to all demographic inference based on
466 polymorphism data.

467 It is important to note that BGS effects extend over genomic distances in a way that is
468 positively related to the strength of purifying selection. For instance, strongly and moderately
469 deleterious mutations affect patterns of diversity at large genomic distances, whereas mildly
470 deleterious mutations primarily skew allele frequencies at adjacent sites. Thus, if intergenic
471 regions further away from exons are used to perform demographic inference, it is predominantly
472 moderately deleterious mutations that are likely to bias inferences; if these are relatively rare,
473 they may not cause significant problems. In contrast, if synonymous sites are used to infer
474 demographic history, mildly deleterious mutations arising in the coding sequences to which they
475 belong may have significant effects. As we have focused here on relatively sparsely-coding
476 genomes (with human-like gene densities) and used intergenic sites for inference, moderately
477 deleterious mutations resulted in more severe mis-inference. The effect of the decay with
478 distance of BGS due to mildly deleterious mutations depends on multiple parameters. For
479 instance, with an exon of length 500 bp and *Drosophila*-like parameters (*e.g.,* $N_e = 10^6$;
480 recombination rate = mutation rate = 10^{-8} / site / generation), B increases from 0.53 (at 10 bases
481 from the end of the exon) to 0.94 at a distance of 1000 bases. On the other hand, with human-like
482 parameters ($N_e = 10^4$; recombination rate = mutation rate = 10^{-8} / site / generation) the
483 corresponding change in B is only from 0.981 to 0.982 (Supp Table 9).

484 Thus, mildly deleterious mutations have drastically different effects, depending on the
485 underlying population parameters. While these results certainly suggest that demographic
486 inference ought to be less biased by BGS in neutral regions very distant from functional elements
487 (for species with sufficiently high recombination / functionally sparse genomes), it is noteworthy
488 that purifying selection on moderately and strongly deleterious mutations can have long-range
489 effects, and that the complex interaction of population history with purifying / background
490 selection necessitates a consideration of this topic in any given empirical application.

491 Comparing the two inference methods investigated here, it appears that *fastsimcoal2* is
492 less prone to inferring false fluctuations in population size. However, both methods falsely infer
493 growth in the presence of BGS, with increasing severity as the density of coding regions
494 increases. The times of population growth inferred by both methods appear to be affected in

495 unpredictable ways when the inferred model is incorrect. When the general model is correctly
496 identified, BGS leads to inference of more recent growth, and more ancient decline, than the
497 reality. In addition, although variation in mutation and recombination rates across the genome
498 alone did not strongly affect demographic inference, our evaluations in the current study are
499 restricted to a specific parameter space resembling those of human populations. The effects of
500 this variation on organisms with more extreme rate fluctuations remain in need of investigation.
501 It is noteworthy that, even when all sites are strictly neutral or only 5% of the genome
502 experiences direct selection, demographic equilibrium is mis-estimated by MSMC as a series of
503 size changes. The pattern of these erroneous size changes lend a characteristic shape to the
504 MSMC curve (*i.e.*, ancient decline and recent growth) which appears to resemble the
505 demographic history previously inferred for the Yoruban population (Schiffels and Durbin
506 2014), including the time at which changes in population size occurred (Supp Figure 22).
507 Previous work has demonstrated that the resulting demographic model does not in fact fit the
508 observed SFS in the Yoruban population (Beichman et al. 2017; Lapierre et al. 2017). A similar
509 shape has also been inferred in the vervet subspecies (Warren *et al.* 2015; Figure 4), in passenger
510 pigeons (Hung *et al.* 2014; Figure 2), in elephants (Palkopoulou *et al.* 2018; Figure 4), in
511 *Arabidopsis* (Fulgione *et al.* 2018; Figure 3), and in grapevines (Zhou *et al.* 2017; Figure 2A).
512 Although the inferred population size fluctuations under simulated neutrality are only
513 ~1.2-fold, in most empirical applications the fluctuations are of a somewhat larger magnitude (~
514 2-fold in pigeons, *Arabidopsis*, and grapevines). Nonetheless, this performance of MSMC under
515 neutral demographic equilibrium is concerning, and adds to the other previously published
516 cautions concerning the interpretation of MSMC results. For example, Mazet *et al*. (2016) and
517 Chikhi *et al*. (2018) demonstrated that, under constant population size with hidden structure,
518 MSMC may suggest false size changes (see also Orozco-terWengel 2016). In addition, MSMC
519 has been reported to falsely infer growth prior to instantaneous bottlenecks (Bunnefeld et al.
520 2015). In addition, we observed that, if insufficient genomic data are used, or more than one
521 diploid genome is used to perform inference, MSMC falsely infers recent growth of varying
522 magnitudes, the latter having been previously observed by Beichman et al. (2017) and Adrion et
523 al. (2020).
524 In sum, we find that the effects of purifying and background selection result in similar
525 demographic mis-inference across approaches, and that masking functional sites does not yield

526 accurate parameter estimates. In order to side-step many of these difficulties, our proposed
527 approach of inferring demography by averaging selection effects across all possible DFE shapes
528 within an ABC framework appears to be promising. Utilizing only functional regions, we found
529 a great improvement in accuracy, without making any assumptions regarding the true underlying
530 shape of the DFE or the neutrality of particular classes of sites. As such, this approach represents
531 a more computationally efficient avenue if only demographic parameters are of interest, and
532 ought to be particularly useful in the great majority of organisms in which independent neutral
533 sites either do not exist, or are difficult to identify and verify.
534
535
536
537 **METHODS**
538
539 ***Simulations of chromosomal segments under neutral equilibrium*:** When assessing the amount
540 of genomic information required for accurate demographic inference, chromosomal segments of
541 varying sizes (1 Mb, 10 Mb, 50 Mb, 200 Mb and 1 Gb) were simulated under neutral
542 equilibrium. In all cases, the effective population size (N_e) simulated was 5000, and mutation
543 and recombination rates were both 1×10^{-8} per site per generation. Simulations were
544 performed with both SLiM 3.1 (Haller and Messer 2019) for a $10N_e$ generation burn-in, and with
545 msprime 0.7.3 (Kelleher et al. 2016). In all cases 100 replicates were simulated, with the
546 exception of 1 Gb chromosomes simulated by SLiM, in which only 10 replicates were obtained.
547
548 ***Simulations of human-like chromosomes (with and without selection)***: Simulations were
549 performed using SLiM 3.1 (Haller and Messer 2019) for a burn-in of $10N_{anc}$ generations, with
550 10 replicates per evolutionary scenario. For every replicate, 22 chromosomes of 150Mb each
551 were simulated, totaling ~3 Gb of information per individual genome (similar to the amount of
552 information in a human genome). Within each chromosome, 3 different types of regions were
553 simulated, representing non-coding intergenic, intronic, and exonic regions. Based on the NCBI
554 RefSeq human genome annotation, downloaded from the UCSC genome browser for hg19
555 (http://genome.ucsc.edu/; Kent *et al.* 2002), mean values of exon sizes and intron numbers per
556 gene were calculated. To represent mean values for the human genome (Lander et al. 2001), each

557 gene comprised 8 exons and 7 introns, and exon lengths were fixed at 350 bp. By varying the
558 lengths of the intergenic and intronic regions, three different genomic configurations with
559 varying densities of functional elements were simulated and compared - with 5%, 10% and 20%
560 of the genome being under direct selection - hereafter referred to as genome5, genome10, and
561 genome20, respectively. Genome5 was comprised of introns of 3000 bp and intergenic sequence
562 of 31000 bp, genome10 of introns of 1500 bp and intergenic sequence of 15750 bp, while
563 genome20 was comprised of introns of 600 bp and intergenic sequence of 6300 bp. The total
564 chromosome sizes of these genomes were approximately 150 Mb (150,018,599 bp, 150,029,949
565 bp, and 150,003,699 bp) with 2737, 5164, and 11278 genes per chromosome in genome5,
566 genome10, and genome20, respectively. In order to be conservative with respect to the
567 performance of existing demographic estimators, intronic and intergenic regions were assumed
568 to be neutral.

569 Recombination and mutation rates were assumed to be equal to 1×10^{-8} /site / generation.
570 Neither crossover interference nor gene conversion were modeled (see the discussion in Campos
571 and Charlesworth 2019). Exonic regions in the genomes experienced direct purifying selection
572 given by a discrete DFE comprised of 4 fixed classes (Johri et al. 2020), whose frequencies are
573 denoted by f_i: f_0, with $0 \leq 2N_e s < 1$ (*i.e.*, effectively neutral mutations), f_1, with $1 \leq 2N_e s < 10$
574 (*i.e.*, weakly deleterious mutations), f_2, with $10 \leq 2N_e s < 100$ (*i.e.*, moderately deleterious
575 mutations), and f_3, with $100 \leq 2N_e s < 2N_e$ (*i.e.*, strongly deleterious mutations), where N_e is the
576 effective population size and s is the reduction in fitness of the mutant homozygote relative to
577 wild-type. Within each bin, the distribution of s was assumed to be uniform. All mutations were
578 assumed to be semi-dominant. In all cases, the N_e corresponding to the DFE refers to the
579 ancestral effective population size.

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

588

589 ***Running MSMC***: In order to quantify the effect of purifying selection on demographic
590 inference, we used entire chromosomes generated by SLiM to generate input files for MSMC.
591 For comparison, and in order to quantify the effect of BGS alone on demographic inference, we
592 masked the exonic regions to generate input files. For all parameters, MSMC was performed on
593 a single diploid genome, as the results for this case were the most accurate (Supp Figure 1, 2).
594 Input files were made using the script ms2multihetsep.py provided in the msmc-tools-Repository
595 downloaded from https://github.com/stschiff/msmc-tools. MSMC1 and 2 were run as follows:
596 *msmc_1.1.0_linux64bit -t 5 -r 1.0 -o output_genomeID input_chr1.tab input_chr2.tab …*
597 *input_chr22.tab.* Population sizes obtained from MSMC were plotted up to the maximum
598 number of generations obtained from MSMC, and the final value of the ancestral population size
599 was extended indefinitely as a dashed line.
600

601 ***Running fastsimcoal2***: Inference was performed by masking all exonic SNPs and using all
602 intronic and intergenic SNPs in order to obtain the most accurate estimates. In order to minimize
603 the effects of linkage disequilibrium (LD), SNPs separated by 5 kb or 100 kb were also used for
604 inference in some cases to assess the impact of violating the assumption of independence. When
605 choosing SNPs separated by a particular distance, the first SNP from each chromosome was
606 chosen and if the distance to the next consecutive SNP was greater than or equal to 5 kb/100 kb,
607 that SNP was included, otherwise the next downstream SNP was evaluated. Site frequency
608 spectra (SFS) were obtained for all sets of SNPs for all 10 replicates of every combination of
609 demographic history and DFE. SNPs from all 22 chromosomes were pooled together to calculate
610 the SFS. In the case of SNPs separated by 5 kb/100 kb, the "0" class of the SFS was scaled down
611 by the same extent as the decrease in the total number of SNPs. *Fastsimcoal2* was used to fit
612 each SFS to 4 distinct models: (a) equilibrium, which estimates only a single population size
613 parameter (N); (b) instantaneous size change (decline/growth), which fits 3 parameters -
614 ancestral population size (N_{anc}), current population size (N_{cur}), and time of change (T); (c)
615 exponential size change (decline/growth), which also estimates 3 parameters - N_{anc}, N_{cur} and T;
616 and (d) an instantaneous bottleneck model with 3 parameters – N_{anc}, intensity, and time of
617 bottleneck. The parameter search ranges for both ancestral and current population sizes in all
618 cases were specified to be uniformly distributed between 100-500000 individuals, while the

619 parameter range for time of change was specified to be uniform between 100-10000 generations
620 in all models. The intensity of the bottleneck was sampled from a log-uniform distribution
621 between 10^{-5} and 2. The following command line was used to run *fastsimcoal2*:
622 *fsc26 -t demographic_model.tpl -n 150000 -d -e demographic_model.est -M -L 50 -q,*
623 Model selection was performed as recommended by Excoffier *et al*. (2013). For each
624 demographic model, the maximum of maximum likelihoods from all replicates was used to
625 calculate the Akaike Information Criterion (AIC) = 2 × number of parameters – 2 ×
626 ln(likelihood) = 2 × number of parameters – 2 × ln(10) × L_{10}, where L_{10} is the logarithm (with
627 respect to base 10) of the best likelihood provided by *fastsimcoal2*. For model choice
628 comparison, we also implemented a stricter penalty of 25× (see Supp Table 5, 6), in which case
629 AIC = 25 × number of parameters – 2 × ln(likelihood). The relative likelihoods (Akaike's weight
630 of evidence) in favor of the ith model were then calculated as:

631 $$w(i) = \frac{e^{-0.5\Delta_i}}{\sum_{j=1}^{4} e^{-0.5\Delta_j}}$$

632 where $\Delta_i = AIC_i - AIC_{min}$. The model with the highest relative likelihood was selected as the
633 best model, and the parameters estimated using that model were used to plot the final inferred
634 demography.
635
636 ***Simulations of variable recombination and mutation rates, and repeat masking***: In order to
637 simulate variation in recombination and mutation rates, all 22 chromosomes were simulated by
638 mimicking chromosome 6 (~171Mb) of the human genome. Recombination rates (HapMap)
639 obtained from Yoruban populations (McVean et al. 2004; Myers et al. 2005) were obtained from
640 the UCSC genome browser, while the mutation rate map
641 (https://molgenis26.target.rug.nl/downloads/gonl_public/mutation_rate_map/release2/) was
642 assumed to correspond to estimates obtained from *de novo* mutations (Francioli et al. 2015), as in
643 Castellano *et al.* (2020). Absolute values of mutation rates were normalized in order to maintain
644 the mean mutation rate across the genome at ~ 1.0×10^{-8} per site per generation. Recombination
645 and mutation rate estimates were taken from positions of approximately 10 Mb to 160 Mb, with
646 the recombination map starting at 10010063 bp and the mutation map starting at 10010001 bp.
647 Regions with missing data for either of the two estimates were simulated with rates
648 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 to 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 (Supp Figure 14), as obtained from the repeat regions in the *hg19* assembly of the human genome from the UCSC genome browser.

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 (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 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 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 10^5-10^7 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

680 https://github.com/paruljohri/BGS_Demography_DFE/blob/master/DPGP3_data.zip, were
681 simulated with recombination rates specific to those exons (https://petrov.stanford.edu/cgi-
682 bin/recombination-rates_updateR5.pl) (Fiston-Lavier et al. 2010; Comeron et al. 2012). Mutation
683 rates were assumed to be fixed at 3×10^{-9} per site per generation (Keightley et al. 2009;
684 Keightley et al. 2014).
685 All parameters were scaled by the factor 320 in order to decrease computational time,
686 using the principle first described by Hill and Robertson (1966), and subsequently employed by
687 others (Comeron and Kreitman 2002; Hoggart et al. 2007; Kaiser and Charlesworth 2009; Kim
688 and Wiehe 2009; Uricchio and Hernandez 2014; Campos and Charlesworth 2019). The scaled
689 population sizes thus ranged between ~300–30000 and were reported as scaled values in the
690 main text. One thousand replicate simulations were performed for every parameter combination
691 (N_{anc}, N_{cur}, f_0, f_1, f_2, f_3); for performing ABC inference, 50 diploid genomes were randomly
692 sampled without replacement, and summary statistics were calculated using pylibseq 0.2.3
693 (Thornton 2003). The following summary statistics were calculated across the entire exonic
694 region for every exon: nucleotide site diversity (π), Watterson's θ, Tajima's D, Fay and Wu's H
695 (both absolute and normalized), number of singletons, haplotype diversity, LD-based statistics
696 (r^2, D, D'), and divergence (*i.e.*, number of fixed mutations per site per generation after the burn-
697 in period). Means and variances (between exons) of all of the above (a total of 22) were used as
698 final summary statistics to perform ABC. As opposed to the above examples, in this inference
699 scheme only exonic data (*i.e.*, directly selected sites) were utilized. Test datasets were generated
700 in exactly the same fashion as described above.
701
702 ***Analytical expectations for the relative site frequencies***: To compute the expected relative
703 frequencies of site frequency classes, the approach of Polanski and Kimmel (2003) was
704 followed. They describe a method for computing the "probability that a SNP has *b* mutant
705 bases", which is equivalent to the expected site frequency spectrum (SFS) of derived variants.
706 This method (their equations 3-10) allows for the specification of arbitrary population size
707 histories and sample sizes. For reasons of computational precision, a sample size of 10 diploid
708 genomes was chosen. The demographic scenarios were implemented as piecewise functions of
709 the effective population size (counting haploid genomes), and the effect of BGS was included by
710 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.

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; (6) An example program (Fortran script) demonstrating how to obtain analytical expressions for values of *B* after exponential growth. All supplemental files, scripts, command lines, and descriptions may be found at: https://github.com/paruljohri/demographic_inference_with_selection

ACKNOWLEDGEMENTS

We would like to thank Susanne Pfeifer for helpful discussions related to this project, and for 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 grants R01GM135899 and 1R35GM139383-01 to JDJ.

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:

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

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:

$$(BR + T_0)\exp\left(-B^{-1}T_0\right)$$

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

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

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

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

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:

$$P_{nc}(t) = \exp\left[-\int_0^t (2BN_{e1})^{-1}\exp(rt)\,\mathrm{d}t\right] = \exp[c^{-1}(1-e^{rt})] \qquad (2)$$

where $c = 2BNe_1r$.

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 $\tau + \mathrm{d}\tau$ is then given by:

$$P_c(\tau) = c^{-1}e^{\tau}\exp[c^{-1}(1-e^{\tau})]\,\mathrm{d}\tau \qquad (3)$$

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.

827 **FIGURES AND TABLES**

828

829 **Table 1**: Proportion (f_i) of mutations in each class of the discrete distribution of fitness effects

830 (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

831

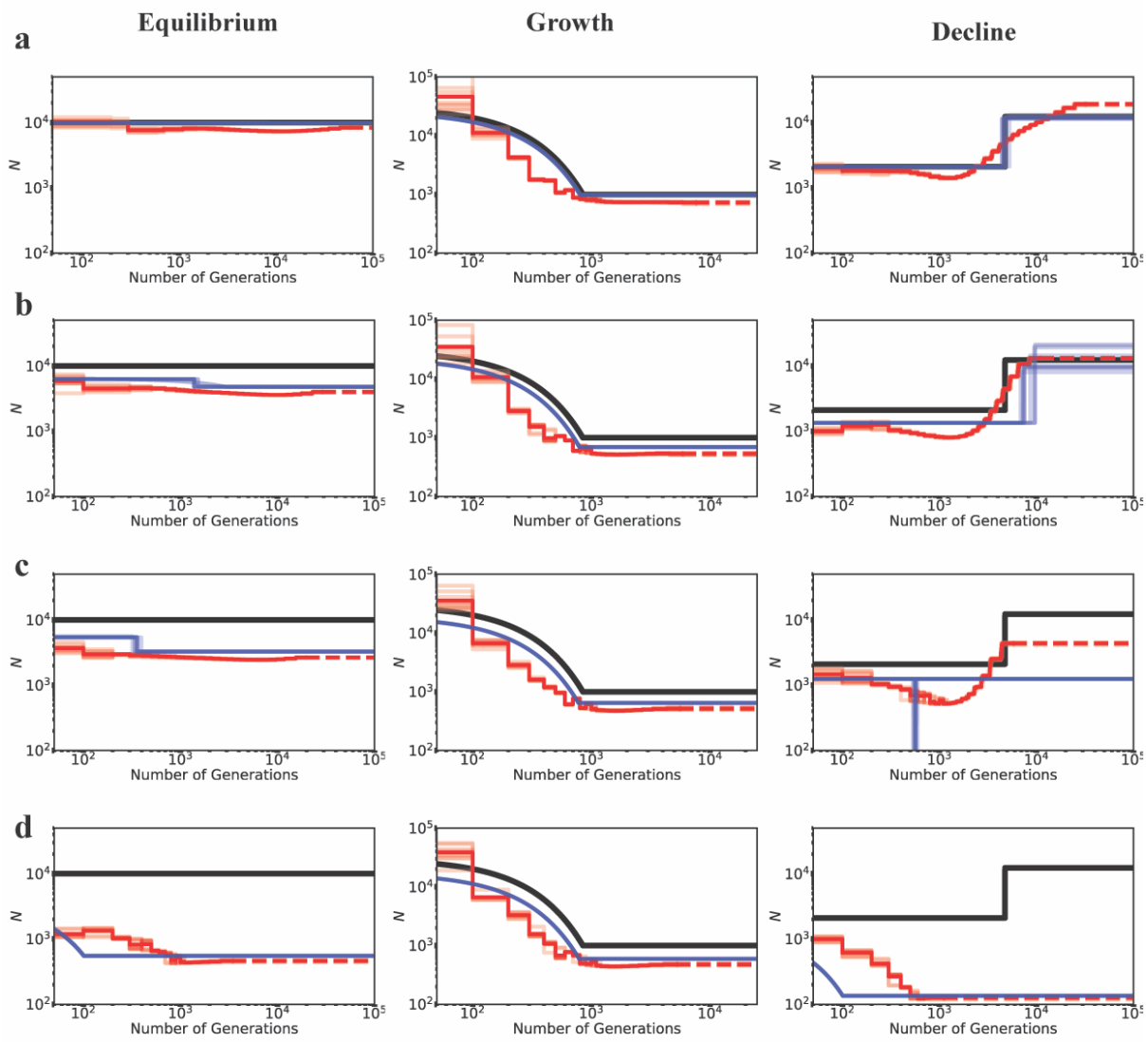


Figure 1: Inference of demography by MSMC (red lines; 10 replicates) and *fastsimcoal2* (blue lines; 10 replicates) with and without BGS, under demographic equilibrium (left column), 30-fold exponential growth (middle column), and ~6-fold instantaneous decline (right column). The true demographic models 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/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

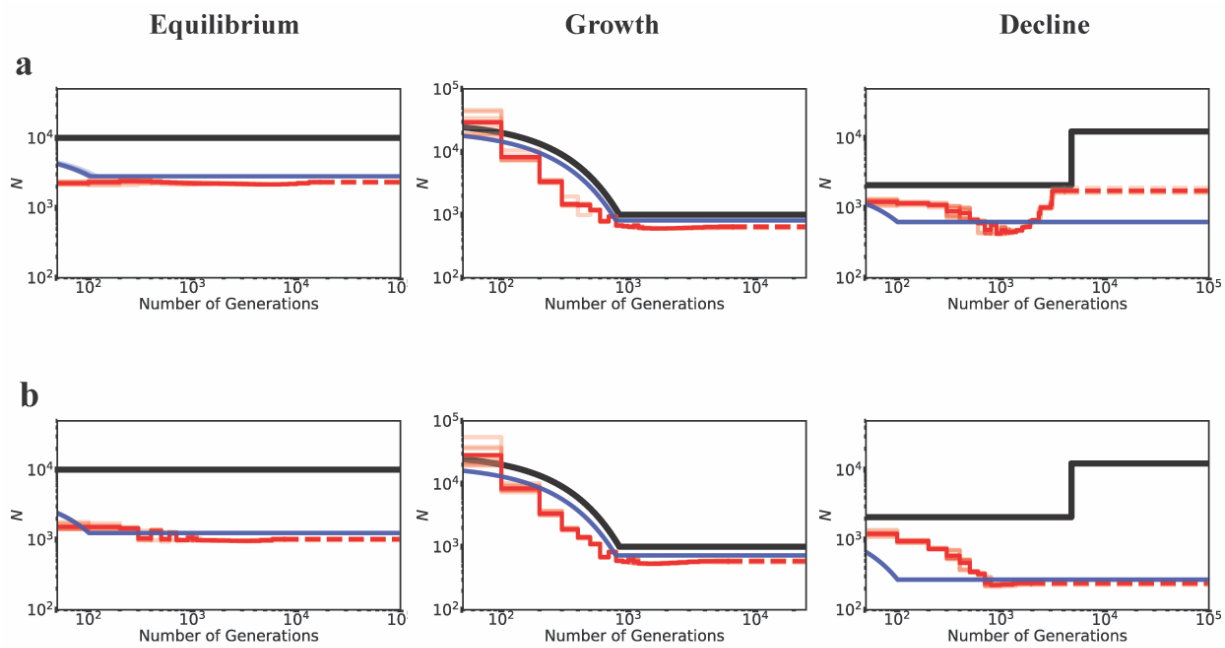


Figure 2: Inference of demography by MSMC (red lines; 10 replicates) and *fastsimcoal2* (blue lines; 10 replicates) in the presence of BGS generated by strongly deleterious mutations. Directly selected sites comprised 20% of the genome and were masked when performing demographic inference. Exons experience 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

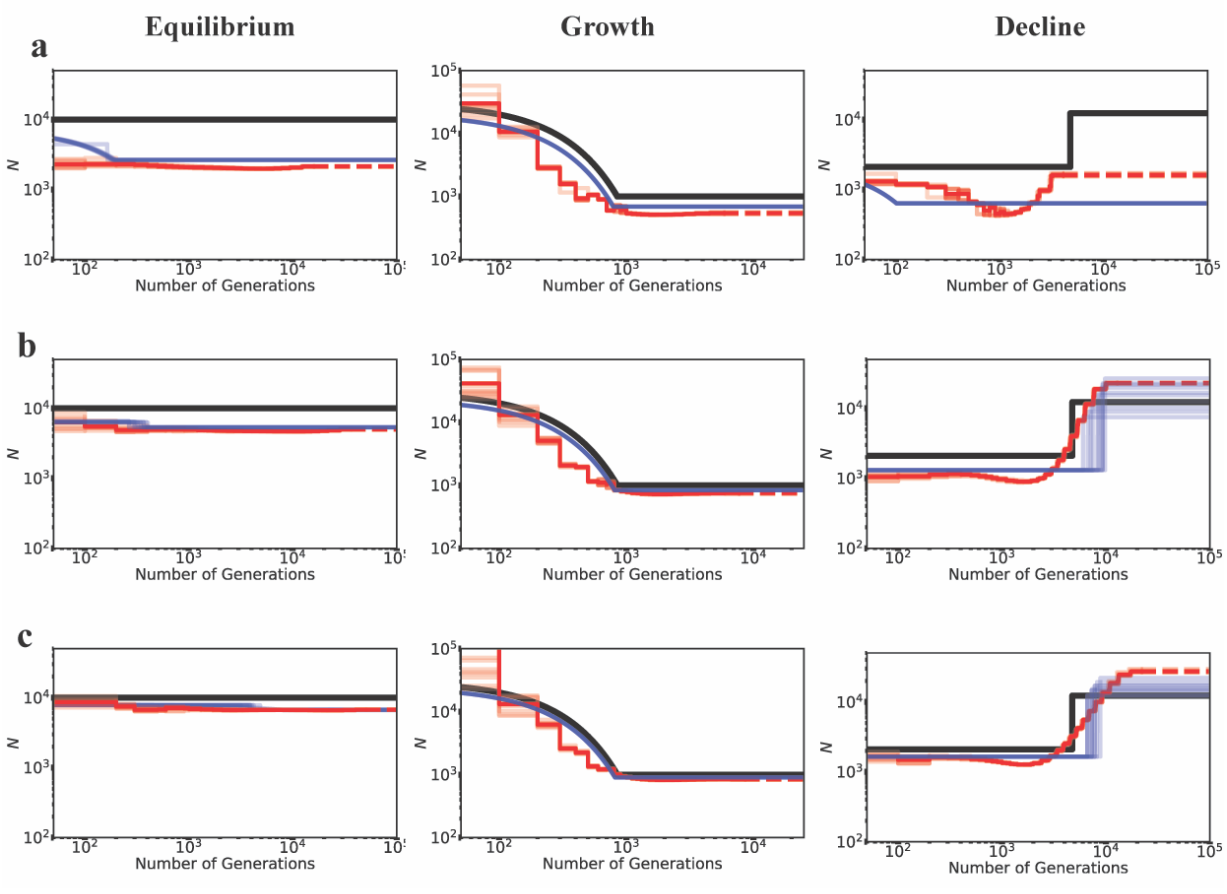


Figure 3: Inference of demography by MSMC (red lines; 10 replicates) and *fastsimcoal2* (blue lines; 10 replicates) in the presence of BGS with varying proportions of the genome under selection, for demographic equilibrium (left column), exponential growth (middle column), and instantaneous decline (right 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

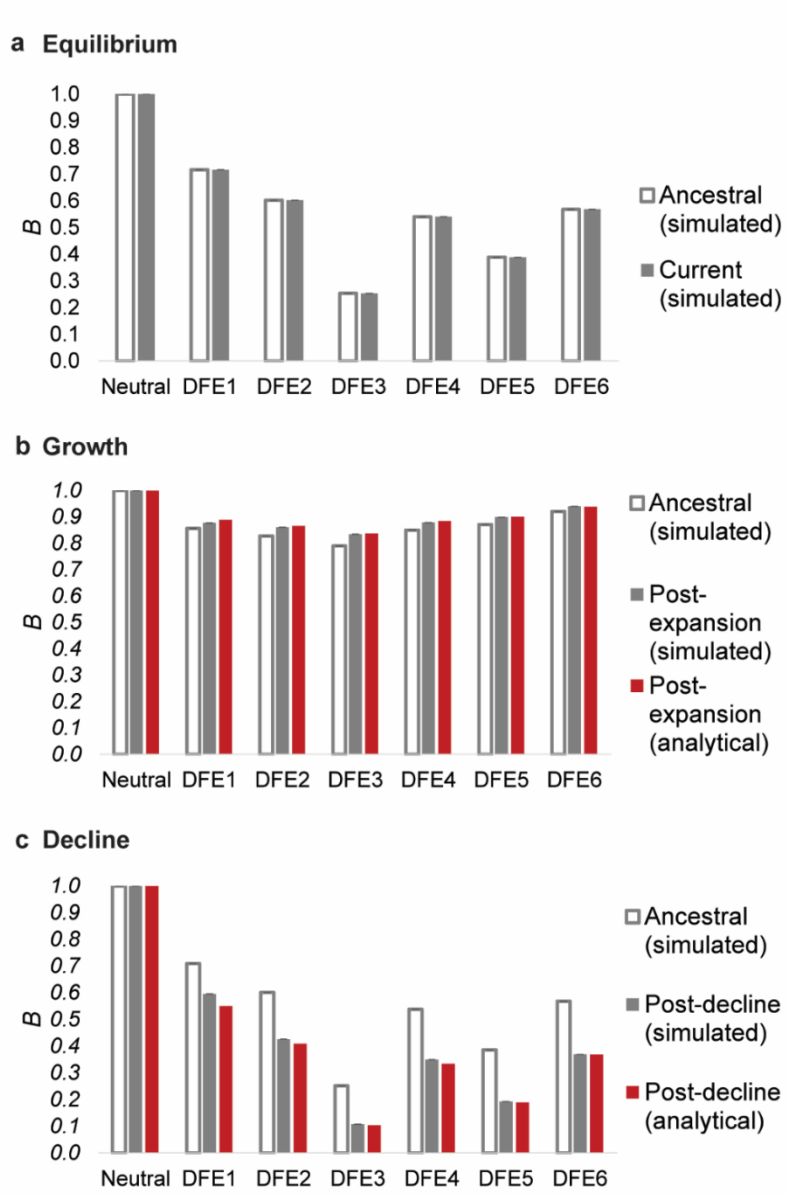


Figure 4: Nucleotide site diversity with BGS (*B*) relative to its purely neutral expectation (π_0) for varying DFEs (specified in Table 1) and demographic scenarios. The results are shown for (a) demographic equilibrium, (b) population growth, and (c) population decline. All cases refer to size changes 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* is 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

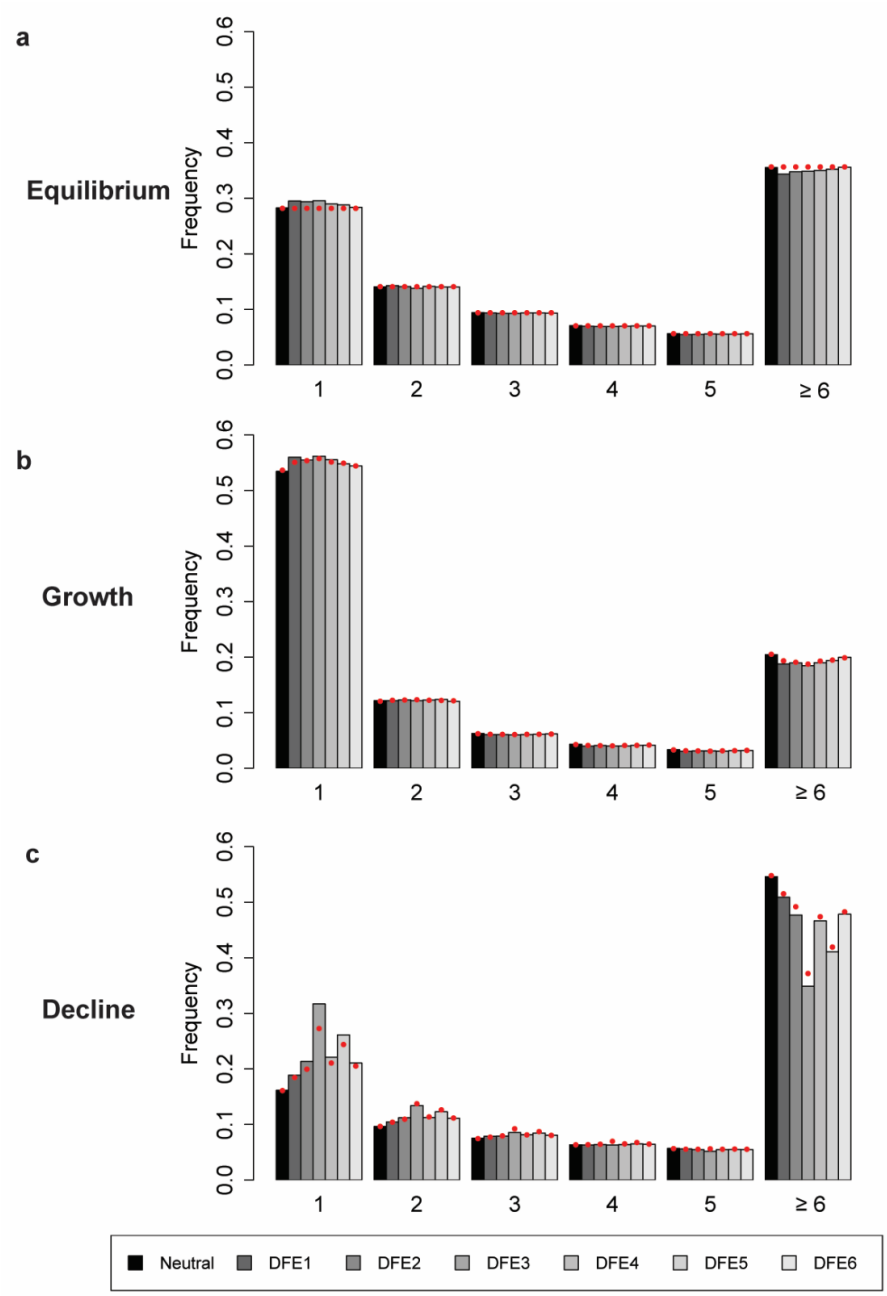


Figure 5: The site frequency spectrum (SFS) of derived allele frequencies at neutral sites from 10 diploid genomes under (a) demographic equilibrium, (b) population growth, and (c) population decline, under the same DFEs as shown in Figure 4. The x-axis indicates the number of sample alleles (out of 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

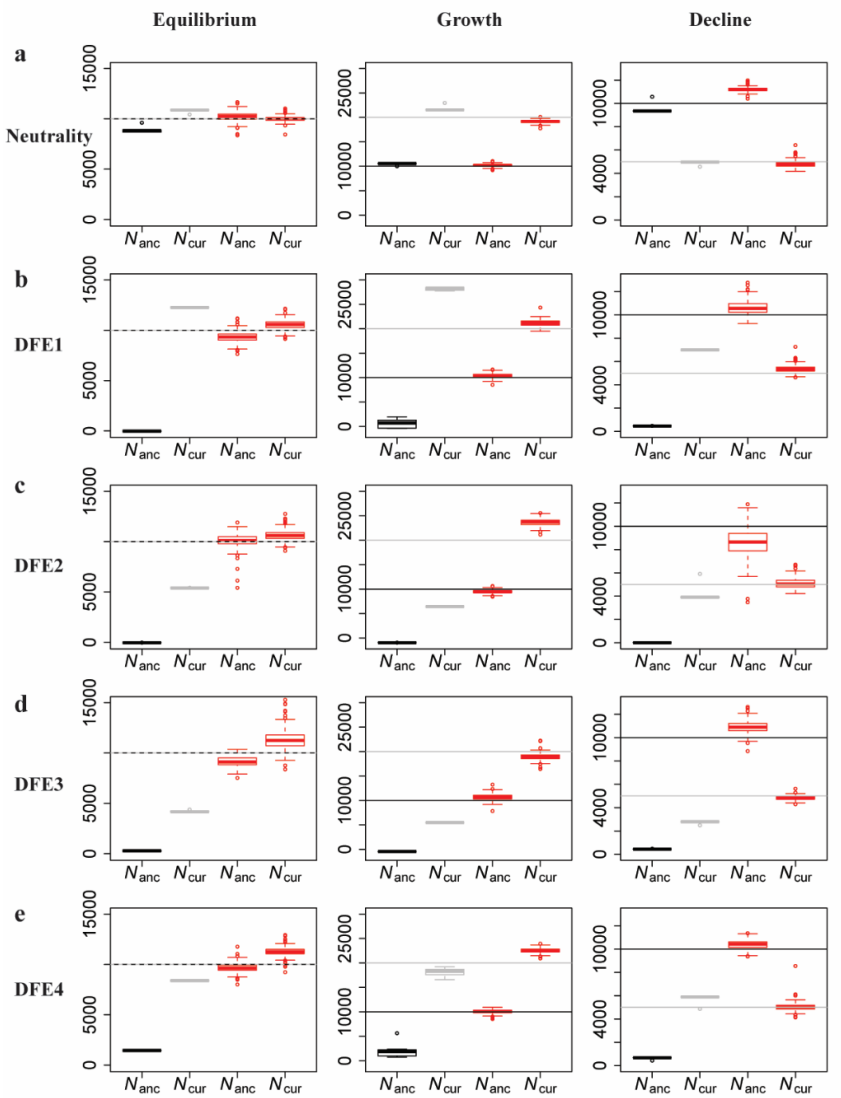


Figure 6: Comparison of estimates of ancestral (N_{anc}) and current (N_{cur}) population sizes when assuming neutrality *vs* when varying the DFE shape as a nuisance parameter, using an ABC framework. Inference is shown for demographic equilibrium (left column), 2-fold exponential growth (middle column), 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, while 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

914 **REFERENCES**

915

916 Adrion JR, Cole CB, Dukler N, Galloway JG, Gladstein AL, Gower G, Kyriazis CC, Ragsdale
917 AP, Tsambos G, Baumdicker F, et al. 2020. A community-maintained standard library of
918 population genetic models.Coop G, Wittkopp PJ, Novembre J, Sethuraman A, Mathieson
919 S, editors. *eLife* 9:e54967.

920 Andolfatto P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* 437:1149–
921 1152.

922 Bank C, Ewing GB, Ferrer-Admettla A, Foll M, Jensen JD. 2014. Thinking too positive?
923 Revisiting current methods of population genetic selection inference. *Trends Genet.*
924 30:540–546.

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

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

929 Bhaskar A, Wang YXR, Song YS. 2015. Efficient inference of population size histories and
930 locus-specific mutation rates from large-sample genomic variation data. *Genome Res.*
931 25:268–279.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

976 Elyashiv E, Sattath S, Hu TT, Strutsovsky A, McVicker G, Andolfatto P, Coop G, Sella G. 2016.
977 A genomic map of the effects of linked selection in *Drosophila*. *PLoS Genet.*
978 12:e1006130.

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

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

983 Eyre-Walker A, Keightley PD. 2009. Estimating the rate of adaptive molecular evolution in the
984 presence of slightly deleterious mutations and population size change. *Mol. Biol. Evol.*
985 26:2097–2108.

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

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

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

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

996 Galtier N, Rousselle M. 2020. How much does Ne vary among species? *Genetics* 216:559–572.

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

1000 Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P. 2005. Multilocus patterns of
1001 nucleotide variability and the demographic and selection history of *Drosophila*
1002 *melanogaster* populations. *Genome Res.* 15:790–799.

1003 Haller BC, Messer PW. 2019. SLiM 3: Forward genetic simulations beyond the Wright–Fisher
1004 model. *Mol. Biol. Evol.* 36:632–637.

1005 Harris K, Nielsen R. 2013. Inferring demographic history from a spectrum of shared haplotype
1006 lengths. *PLoS Genet.* 9:e1003521.

1007 Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G, Project 1000 Genomes,
1008 Sella G, Przeworski M. 2011. Classic selective sweeps were rare in recent human
1009 evolution. *Science* 331:920–924.

1010 Hey J, Harris E. 1999. Population bottlenecks and patterns of human polymorphism. *Mol Biol*
1011 *Evol* 16:1423–1426.

1012 Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genet. Res.*
1013 8:269–294.

1014 Hoggart CJ, Chadeau-Hyam M, Clark TG, Lampariello R, Whittaker JC, Iorio MD, Balding DJ.
1015 2007. Sequence-level population simulations over large genomic regions. *Genetics*
1016 177:1725–1731.

1017 Hung C-M, Shaner P-JL, Zink RM, Liu W-C, Chu T-C, Huang W-S, Li S-H. 2014. Drastic
1018 population fluctuations explain the rapid extinction of the passenger pigeon. *Proc. Nat.*
1019 *Acad. Sci. U.S.A.* 111:10636–10641.

1020 Jackson BC, Campos JL, Haddrill PR, Charlesworth B, Zeng K. 2017. Variation in the intensity
1021 of selection on codon bias over time causes contrasting patterns of base composition
1022 evolution in *Drosophila*. *Genome Biol. Evol.* 9:102–123.

1023 Jacquier H, Birgy A, Nagard HL, Mechulam Y, Schmitt E, Glodt J, Bercot B, Petit E, Poulain J,
1024 Barnaud G, et al. 2013. Capturing the mutational landscape of the beta-lactamase TEM-1.
1025 *Proc. Natl. Acad. Sci. U.S.A.* 110:13067–13072.

1026 Jensen JD, Payseur BA, Stephan W, Aquadro CF, Lynch M, Charlesworth D, Charlesworth B.
1027 2019. The importance of the Neutral Theory in 1968 and 50 years on: A response to Kern
1028 and Hahn 2018. *Evolution* 73:111–114.

1029 Johri P, Charlesworth B, Jensen JD. 2020. Toward an evolutionarily appropriate null model:
1030 jointly inferring demography and purifying selection. *Genetics* 215:173–192.

1031 Jones MR, Good JM. 2016. Targeted capture in evolutionary and ecological genomics. *Mol.*
1032 *Ecol.* 25:185–202.

1033 Kaiser VB, Charlesworth B. 2009. The effects of deleterious mutations on evolution in non-
1034 recombining genomes. *Trends Genet.* 25:9–12.

1035 Keightley PD, Eyre-Walker A. 2007. Joint inference of the distribution of fitness effects of
1036 deleterious mutations and population demography based on nucleotide polymorphism
1037 frequencies. *Genetics* 177:2251–2261.

1038 Keightley PD, Ness RW, Halligan DL, Haddrill PR. 2014. Estimation of the spontaneous
1039 mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. *Genetics*
1040 196:313–320.

1041 Keightley PD, Trivedi U, Thomson M, Oliver F, Kumar S, Blaxter ML. 2009. Analysis of the
1042 genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation
1043 lines. *Genome Res.* 19:1195–1201.

1044 Kelleher J, Etheridge AM, McVean G. 2016. Efficient coalescent simulation and genealogical
1045 analysis for large sample sizes. *PLoS Comput. Biol.* 12:e1004842.

1046 Kelleher J, Wong Y, Wohns AW, Fadil C, Albers PK, McVean G. 2019. Inferring whole-
1047 genome histories in large population datasets. *Nat. Genet.* 51:1330–1338.

1048 Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler and D. 2002.
1049 The Human Genome Browser at UCSC. *Genome Res.* 12:996–1006.

1050 Kim BY, Huber CD, Lohmueller KE. 2017. Inference of the distribution of selection coefficients
1051 for new nonsynonymous mutations using large samples. *Genetics* 206:345–361.

1052 Kim Y, Wiehe T. 2009. Simulation of DNA sequence evolution under models of recent
1053 directional selection. *Brief. Bioinformatics* 10:84–96.

1054 Kousathanas A, Keightley PD. 2013. A comparison of models to infer the distribution of fitness
1055 effects of new mutations. *Genetics* 193:1197–1208.

1056 Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle
1057 M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome.
1058 *Nature* 409:860–921.

1059 Lapierre M, Lambert A, Achaz G. 2017. Accuracy of demographic inferences from the site
1060 frequency spectrum: the case of the Yoruba population. *Genetics* 206:439–449.

1061 Li H, Durbin R. 2011. Inference of human population history from individual whole-genome
1062 sequences. *Nature* 475:493–496.

1063 Liang P, Saqib HSA, Zhang X, Zhang L, Tang H. 2018. Single-base resolution map of
1064 evolutionary constraints and annotation of conserved elements across major grass
1065 genomes. *Genome Biol. Evol.* 10:473–488.

1066 Lukic S, Hey J. 2012. Demographic inference using spectral methods on SNP data, with an
1067 analysis of the human out-of-Africa expansion. *Genetics* 192:619–639.

1068 Lynch M. 2007. The Origins of Genome Architecture. Sunderland, Massachusetts: Sinauer
1069 Associates

1070 Mazet O, Rodríguez W, Grusea S, Boitard S, Chikhi L. 2016. On the importance of being
1071 structured: instantaneous coalescence rates and human evolution--lessons for ancestral
1072 population size inference? *Heredity* 116:362–371.

1073 McVean GAT, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. 2004. The fine-scale
1074 structure of recombination rate variation in the human genome. *Science* 304:581–584.

1075 Messer PW, Petrov DA. 2013. Frequent adaptation and the McDonald–Kreitman test. *Proc. Natl.*
1076 *Acad. Sci. U.S.A.* 110:8615–8620.

1077 Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. 2005. A fine-scale map of
1078 recombination rates and hotspots across the human genome. *Science* 310:321–324.

1079 Nicolaisen LE, Desai MM. 2013. Distortions in genealogies due to purifying selection and
1080 recombination. *Genetics* 195:221–230.

1081 O'Fallon BD, Seger J, Adler FR. 2010. A continuous-state coalescent and the impact of weak
1082 selection on the structure of gene genealogies. *Mol. Biol. Evol.* 27:1162–1172.

1083 Orozco-terWengel P. 2016. The devil is in the details: the effect of population structure on
1084 demographic inference. *Heredity* 116:349–350.

1085 Palkopoulou E, Lipson M, Mallick S, Nielsen S, Rohland N, Baleka S, Karpinski E, Ivancevic
1086 AM, To T-H, Kortschak RD, et al. 2018. A comprehensive genomic history of extinct
1087 and living elephants. *Proc. Nat. Acad. Sci. U.S.A.* 115:E2566–E2574.

1088 Polanski A, Bobrowski A, Kimmel M. 2003. A note on distributions of times to coalescence,
1089 under time-dependent population size. *Theor Popul Biol* 63:33–40.

1090 Polanski A, Kimmel M. 2003. New explicit expressions for relative frequencies of single-
1091 nucleotide polymorphisms with application to statistical inference on population growth.
1092 *Genetics* 165:427–436.

1093 Pool JE, Nielsen R. 2007. Population size changes reshape genomic patterns of diversity.
1094 *Evolution* 61:3001–3006.

1095 Pool JE, Nielsen R. 2009. Correction for Pool and Nielsen (2007). *Evolution* 63:1671.

1096 Pouyet F, Aeschbacher S, Thiéry A, Excoffier L. 2018. Background selection and biased gene
1097 conversion affect more than 95% of the human genome and bias demographic inferences.
1098 *eLife* 7:e36317.

1099 Ragsdale AP, Gutenkunst RN. 2017. Inferring demographic history using two-locus statistics.
1100 *Genetics* 206:1037–1048.

1101 Ragsdale AP, Moreau C, Gravel S. 2018. Genomic inference using diffusion models and the
1102 allele frequency spectrum. *Current Opinion in Genetics & Development* 53:140–147.

1103 Sanjuán R. 2010. Mutational fitness effects in RNA and single-stranded DNA viruses: common
1104 patterns revealed by site-directed mutagenesis studies. *Phil. Trans. R. Soc. B.* 365:1975–
1105 1982.

1106 Schiffels S, Durbin R. 2014. Inferring human population size and separation history from
1107 multiple genome sequences. *Nat. Genet.* 46:919–925.

1108 Schneider A, Charlesworth B, Eyre-Walker A, Keightley PD. 2011. A method for inferring the
1109 rate of occurrence and fitness effects of advantageous mutations. *Genetics* 189:1427–
1110 1437.

1111 Schrider DR, Shanku AG, Kern AD. 2016. Effects of linked selective sweeps on demographic
1112 inference and model selection. *Genetics* 204:1207–1223.

1113 Sheehan S, Song YS. 2016. Deep learning for population genetic inference. *PLoS Comput. Biol.*
1114 12:e1004845.

1115 Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J,
1116 Hillier LW, Richards S, et al. 2005. Evolutionarily conserved elements in vertebrate,
1117 insect, worm, and yeast genomes. *Genome Res.* 15:1034–1050.

1118 Slatkin M, Hudson RR. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable
1119 and exponentially growing populations. *Genetics* 129:555–562.

1120 Speidel L, Forest M, Shi S, Myers SR. 2019. A method for genome-wide genealogy estimation
1121 for thousands of samples. *Nat. Genet.* 51:1321–1329.

1122 Steinrücken M, Kamm J, Spence JP, Song YS. 2019. Inference of complex population histories
1123 using whole-genome sequences from multiple populations. *Proc. Natl. Acad. Sci. U.S.A.*
1124 116:17115–17120.

1125 Teshima KM, Coop G, Przeworski M. 2006. How reliable are empirical genomic scans for
1126 selective sweeps? *Genome Res* 16:702–712.

1127 Thornton K. 2003. Libsequence: a C++ class library for evolutionary genetic analysis.
1128 *Bioinformatics* 19:2325–2327.

1129 Thornton KR, Jensen JD. 2007. Controlling the false-positive rate in multilocus genome scans
1130 for selection. *Genetics* 175:737–750.

1131 Torres R, Stetter MG, Hernandez RD, Ross-Ibarra J. 2020. The temporal dynamics of
1132 background selection in nonequilibrium populations. *Genetics* 214:1019–1030.

1133 Torres R, Szpiech ZA, Hernandez RD. 2018. Human demographic history has amplified the
1134 effects of background selection across the genome. *PLoS Genet.* 14:e1007387.

1135 Uricchio LH, Hernandez RD. 2014. Robust forward simulations of recurrent hitchhiking.
1136 *Genetics* 197:221–236.

1137 Warren WC, Jasinska AJ, García-Pérez R, Svardal H, Tomlinson C, Rocchi M, Archidiacono N,
1138 Capozzi O, Minx P, Montague MJ, et al. 2015. The genome of the vervet (*Chlorocebus*
1139 *aethiops sabaeus*). *Genome Res.* 25:1921–1933.

1140 Williamson RJ, Josephs EB, Platts AE, Hazzouri KM, Haudry A, Blanchette M, Wright SI.
1141 2014. Evidence for widespread positive and negative selection in coding and conserved
1142 noncoding regions of *Capsella grandiflora*. *PLoS Genet.* 10:e1004622.

1143 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.
1144 2020. A new coronavirus associated with human respiratory disease in China. *Nature*
1145 579:265–269.

1146 Zeng K, Charlesworth B. 2010. Studying patterns of recent evolution at synonymous sites and
1147 intronic sites in *Drosophila melanogaster*. *J. Mol. Evol.* 70:116–128.

1148 Zhou Y, Massonnet M, Sanjak JS, Cantu D, Gaut BS. 2017. Evolutionary genomics of grape
1149 (*Vitis vinifera ssp. vinifera*) domestication. *Proc. Nat. Acad. Sci. U.S.A.* 114:11715–
1150 11720.

1151
1152
1153

Supplementary Information

Supp Table 1: Performance of *fastsimcoal2* under neutrality and demographic equilibrium with varying chromosome sizes, when the correct model was specified. Inference was performed using 50 diploid individuals and all SNPs, and a comparison is given between simulations performed in SLiM (3.1) as well as msprime (0.7.3). The number of replicates for each chromosome size was set to 100, except for 1Gb chromosomes simulated in SLiM for which we report 10 replicates. Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppTable1.txt.

Simulated in SLiM 3.1			
Chromosome size	**True population size**	**Inferred population size**	
		Mean	**SD**
1 Mb	5000	5010	320
10 Mb	5000	5005	118
50 Mb	5000	4995	65
200 Mb	5000	4992	30
1 Gb	5000	4988	18
Simulated in msprime 0.7.3			
1 Mb	5000	5043	345
10 Mb	5000	5004	109
50 Mb	5000	5015	57
200 Mb	5000	5003	29
1 Gb	5000	5000	17

Supp Table 2: Model selection in *fastsimcoal2* for 100 replicates, in which the true model was neutral equilibrium and selection was performed with 4 models: equilibrium, instantaneous size change, exponential size change, and instantaneous bottleneck. Varying densities of SNPs (all SNPs, 1 per 5 kb, 1 per 50 kb and 1 per 100 kb) were used to perform inference. Simulations were performed using SLiM (3.1) for 10*N* generations. Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure3_SuppTable2_5.txt.

Density of SNPs	Genome length	Equilibrium	Exponential change	Instantaneous change	Instantaneous Bottleneck	Total number of replicates
all	1Mb	33	11	18	38	100
	10Mb	32	17	27	24	100
	50Mb	16	25	29	30	100
	200Mb	5	10	41	44	100
	1Gb	0	4	3	3	10
1 per 5 kb	1Mb	69	7	9	15	100
	10Mb	49	8	18	25	100
	50Mb	10	4	35	51	100
	200Mb	0	0	48	52	100
	1Gb	0	0	5	5	10
1 per 50 kb	1Mb	83	7	4	6	100
	10Mb	83	5	4	8	100
	50Mb	59	10	10	21	100
	200Mb	20	6	40	34	100
	1Gb	0	0	4	6	10
1 per 100 kb	1Mb	86	4	3	7	100
	10Mb	78	5	7	10	100
	50Mb	64	6	14	16	100
	200Mb	43	6	30	21	100
	1Gb	0	1	5	4	10

Supp Table 3: Model selection in *fastsimcoal2* for 100 replicates, in which the true model was neutral equilibrium and selection was performed with 4 models: equilibrium, instantaneous size change, exponential size change, and instantaneous bottleneck. Varying densities of SNPs (all SNPs, 1 per 5 kb, 1 per 50 kb and 1 per 100 kb) were used to perform inference. Simulations were performed using msprime (0.7.3). Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure4_SuppTable3_6.txt.

Density of SNPs	Genome length	Equilibrium	Exponential change	Instantaneous change	Instantaneous Bottleneck	Total number of replicates
all	1Mb	29	16	18	37	100
	10Mb	22	12	31	35	100
	50Mb	13	18	38	31	100
	200Mb	11	8	31	40	100
	1Gb	5	12	26	57	100
1 per 5 kb	1Mb	63	3	12	22	100
	10Mb	42	10	19	29	100
	50Mb	4	7	42	47	100
	200Mb	1	2	55	42	100
	1Gb	0	1	66	33	100
1 per 50 kb	1Mb	83	4	6	7	100
	10Mb	80	4	6	10	100
	50Mb	61	9	12	18	100
	200Mb	19	6	29	46	100
	1Gb	0	2	48	50	100
1 per 100 kb	1Mb	89	3	3	5	100
	10Mb	86	4	4	6	100
	50Mb	70	3	11	16	100
	200Mb	47	5	20	28	100
	1Gb	0	3	47	50	100

Supp Table 4: Linkage disequilibrium (r^2) summarized in varying chromosomal sizes, and sampled in different densities, calculated using *Pylibseq* 0.2.3 in non-overlapping sliding windows across chromosomes with ~10 SNPs per window. SNPs from separate chromosomes represent completely unlinked SNPs – 1 SNP was randomly sampled from each of 100 separate replicate chromosomes simulated, and this random sampling was performed 100 times. Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppTable4.txt.

		Number of SNPs per window		r^2	
Chromosome size	**SNP density**	**mean**	**SD**	**mean**	**SD**
10Mb	all	10.7	3.6	0.1181	0.0885
	1 per 5 kb	8.3	0.7	0.0731	0.0504
	1 per 50 kb	9.8	0.4	0.0271	0.0141
	1 per 100 kb	10.0	0.2	0.0203	0.0101
	separate chromosomes	10.5	0.5	0.0103	0.0047
1 Mb	all	10.7	3.6	0.1174	0.0889
	1 per 5 kb	8.4	0.7	0.0725	0.0503
	1 per 50 kb	10.0	0.0	0.0252	0.0108
	1 per 100 kb	10.0	0.0	0.0193	0.0080
	separate chromosomes	10.5	0.5	0.0102	0.0047
50Mb	all	10.7	3.6	0.1181	0.0888
	1 per 5 kb	8.3	0.7	0.0729	0.0502
	1 per 50 kb	9.8	0.4	0.0269	0.0136
	1 per 100 kb	9.9	0.3	0.0203	0.0099
	separate chromosomes	10.5	0.5	0.0101	0.0046
200 Mb	all	10.6	3.6	0.1180	0.0888
	1 per 5 kb	8.3	0.7	0.0728	0.0499
	1 per 50 kb	9.8	0.4	0.0269	0.0137
	1 per 100 kb	9.9	0.3	0.0202	0.0098
	separate chromosomes	10.5	0.5	0.0101	0.0046
1 Gb	all	10.6	3.6	0.1181	0.0891
	1 per 5 kb	8.4	0.7	0.0727	0.0495
	1 per 50 kb	9.8	0.4	0.0269	0.0134
	1 per 100 kb	9.9	0.3	0.0204	0.0102
	separate chromosomes	6.5	0.5	0.0092	0.0046

Supp Table 5: Model selection with a higher AIC penalty (=25×the number of parameters) in which the true model was neutral equilibrium, and model selection was performed with 4 models: equilibrium, instantaneous size change, exponential size change, and instantaneous bottleneck. Varying densities of SNPs (all SNPs, 1 per 5 kb, 1 per 50 kb and 1 per 100 kb) were used to perform inference. Simulations were performed in SLiM (3.1) for 10*N* generations**.** Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure3_SuppTable2_5.txt.

Density of SNPs	Genome length	Equilibrium	Exponential change	Instantaneous change	Instantaneous Bottleneck	Total number of replicates
all	1Mb	98	2	0	0	100
	10Mb	100	0	0	0	100
	50Mb	98	0	2	0	100
	200Mb	99	0	0	1	100
	1Gb	5	2	2	1	10
1 per 5 kb	1Mb	100	0	0	0	100
	10Mb	100	0	0	0	100
	50Mb	99	0	1	0	100
	200Mb	46	0	27	27	100
	1Gb	0	0	5	5	10
1 per 50 kb	1Mb	100	0	0	0	100
	10Mb	100	0	0	0	100
	50Mb	100	0	0	0	100
	200Mb	100	0	0	0	100
	1Gb	9	0	0	1	10
1 per 100 kb	1Mb	100	0	0	0	100
	10Mb	100	0	0	0	100
	50Mb	100	0	0	0	100
	200Mb	100	0	0	0	100
	1Gb	10	0	0	0	10

Supp Table 6: Model selection with a higher AIC penalty (=25×the number of parameters) in which the true model was neutral equilibrium, and model selection was performed with 4 models: equilibrium, instantaneous size change, exponential size change, and instantaneous bottleneck. Varying densities of SNPs (all SNPs, 1 per 5 kb, 1 per 50 kb and 1 per 100 kb) were used to perform inference. Simulations were performed using msprime (0.7.3). Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure4_SuppTable3_6.txt.

Density of SNPs	Genome length	Equilibrium	Exponential change	Instantaneous change	Instantaneous Bottleneck	Total number of replicates
all	1Mb	100	0	0	0	100
	10Mb	100	0	0	0	100
	50Mb	97	0	2	1	100
	200Mb	96	0	4	0	100
	1Gb	93	1	3	3	100
1 per 5 kb	1Mb	100	0	0	0	100
	10Mb	100	0	0	0	100
	50Mb	97	1	0	2	100
	200Mb	45	1	30	24	100
	1Gb	0	1	66	33	100
1 per 50 kb	1Mb	100	0	0	0	100
	10Mb	100	0	0	0	100
	50Mb	100	0	0	0	100
	200Mb	100	0	0	0	100
	1Gb	89	1	4	6	100
1 per 100 kb	1Mb	100	0	0	0	100
	10Mb	100	0	0	0	100
	50Mb	100	0	0	0	100
	200Mb	100	0	0	0	100
	1Gb	100	0	0	0	100

Supp Table 7: Parameters underlying the human-like demographic models considered.

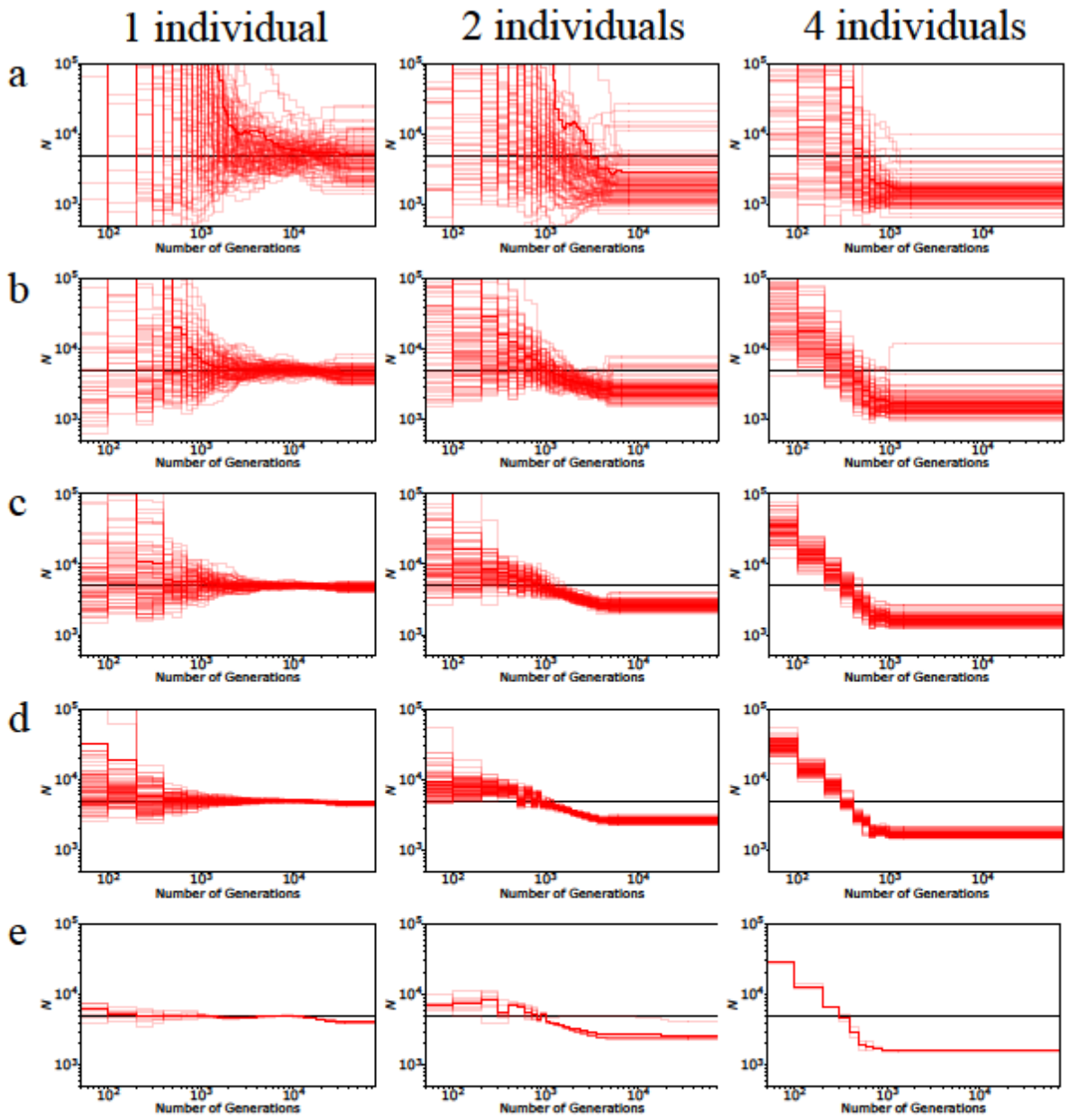
	Demographic models	**Ancestral population size (N_{anc})**	**Current population size (N_{cur})**	**Time of change in generations**
1	Equilibrium	10,000	10,000	NA
2	Exponential growth	1000	30,000	850
3	Instantaneous decline	12,300	2,100	4,750

Supp Table 8: Comparison of inference of parameters using *fastsimcoal2* under neutrality when inference is performed using all SNPs *vs*. using sparser SNP sampling (1 per 5 kb). In this example, 20% of the genome is exonic, and these exonic regions were masked when performing inference.

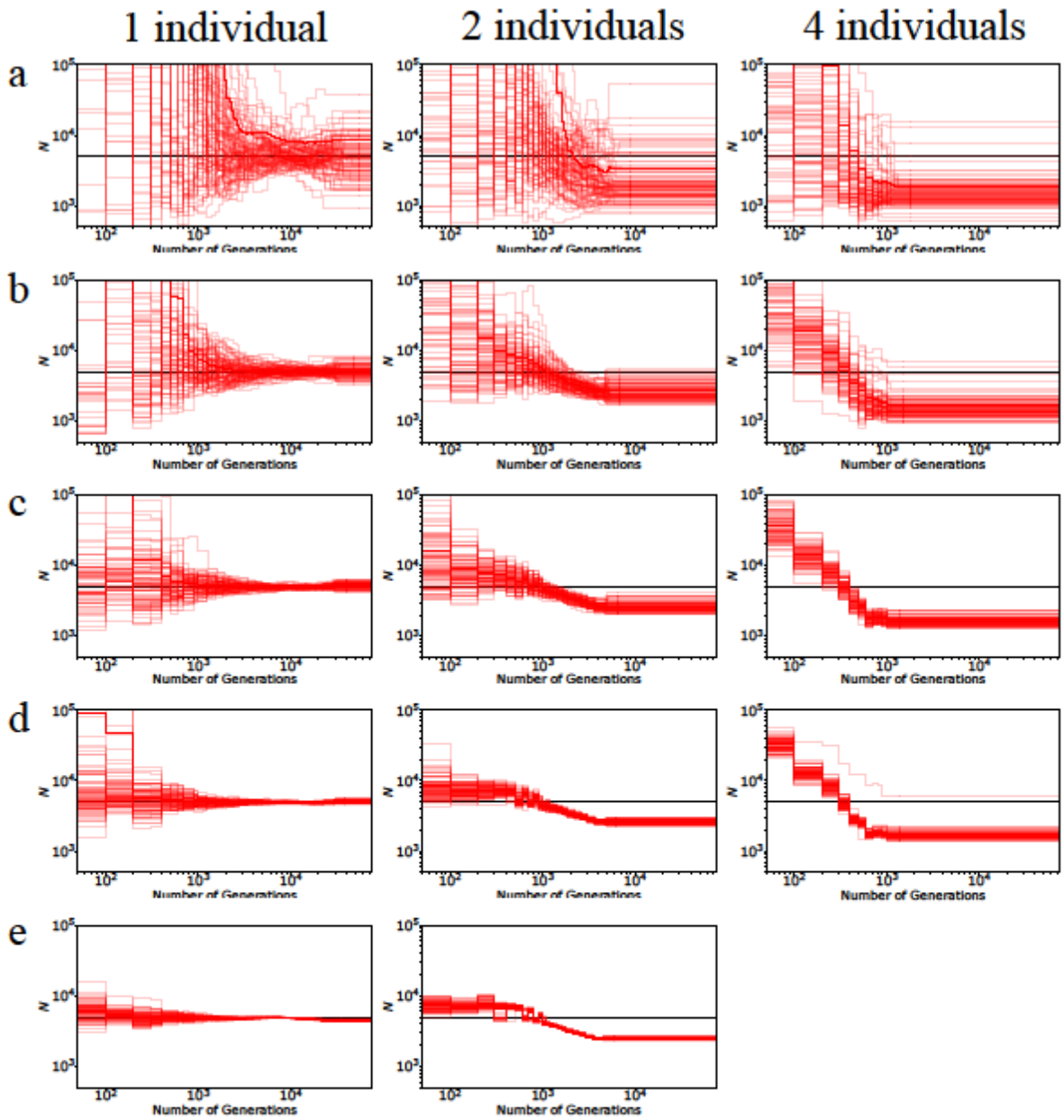
Demographic model	**Parameter**	**True value**	**Inferred value using all SNPs (Mean±SD)**	**Inferred value using 1 SNP per 5 kb (Mean±SD)**
Equilibrium	N_{anc}	10,000	9,781 ± 23	9,378 ± 567
Exponential growth	N_{anc}	1000	977 ± 3	963 ± 4
	N_{cur}	30,000	25,489 ± 123	25,641 ± 184
	Time of change	850	808 ± 3	816 ± 4
Instantaneous decline	N_{anc}	12,300	11,978 ± 463	11,513 ± 894
	N_{cur}	2,100	2,097 ± 15	2,681 ± 12
	Time of change	4,750	4,703 ± 246	7,458 ± 596

Supp Table 9: Nucleotide diversity in the presence of BGS relative to that under neutrality (*B*), calculated for a neutral site distance *y* bases from the end of a gene/exon of length 500 bp. The exon experiences purifying selection with strength $2N_e s=10$, where N_e is the effective population size and *s* is the reduction in fitness. Shown below is $t=hs$ where *h* is the dominance coefficient (assumed to be 0.5 here), *r* is the recombination rate per site per generation, and *u* is the mutation rate per site per generation. *B* was calculated using Equation 2 of Johri *et al*. (2020).

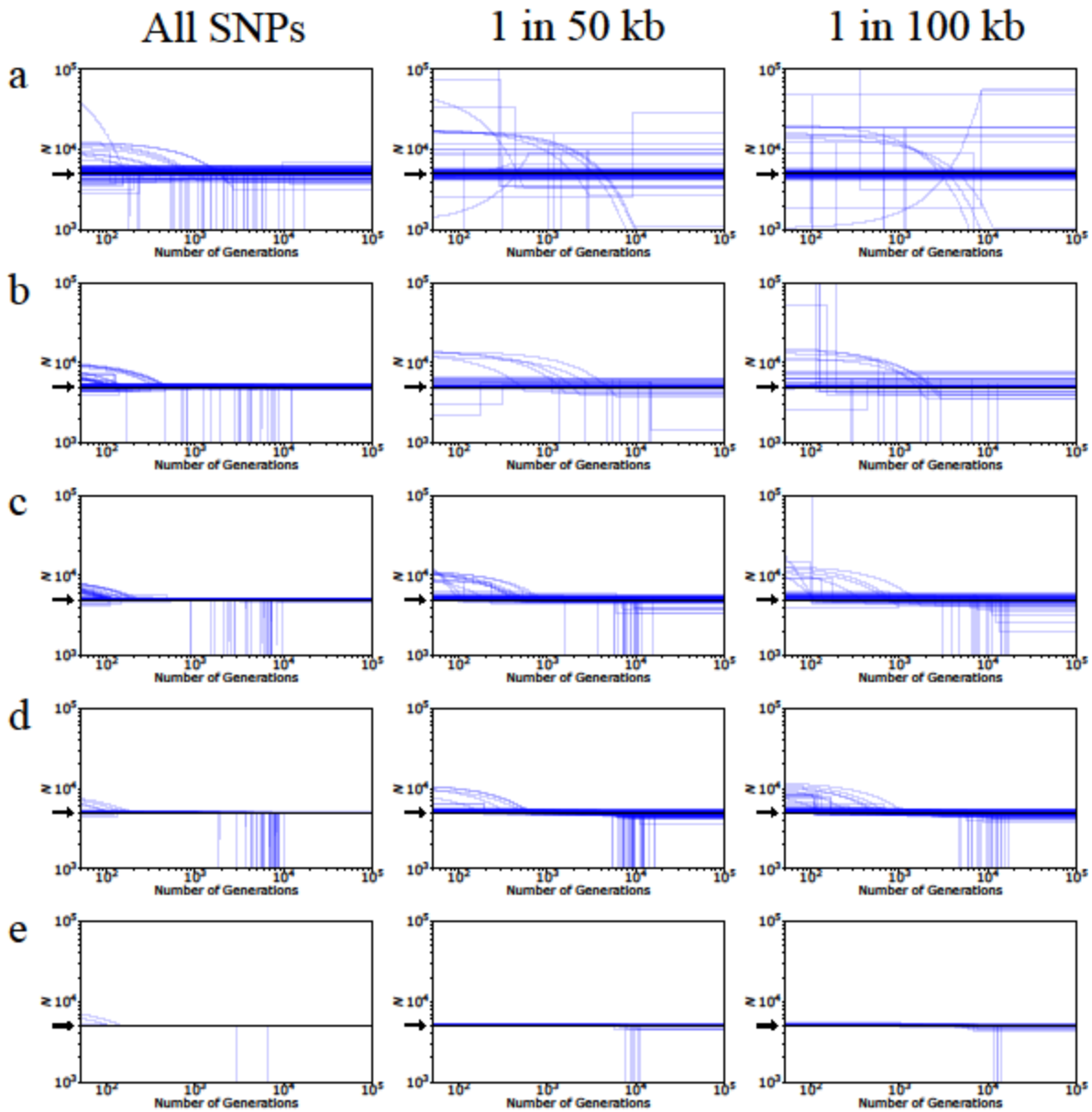
N_e	$t\,(2N_e s=10)$	r	u	$B\,(y=1)$	$B\,(y=10)$	$B\,(y=1000)$
10^4	0.00025	1.00×10^{-8}	1.00×10^{-8}	0.9805	0.9806	0.9820
		1.00×10^{-6}	1.00×10^{-6}	0.5152	0.5312	0.9444
10^6	2.5×10^{-6}	1.00×10^{-8}	1.00×10^{-8}	0.5152	0.5312	0.9445
		1.00×10^{-6}	1.00×10^{-6}	0.4920	0.8227	0.9992



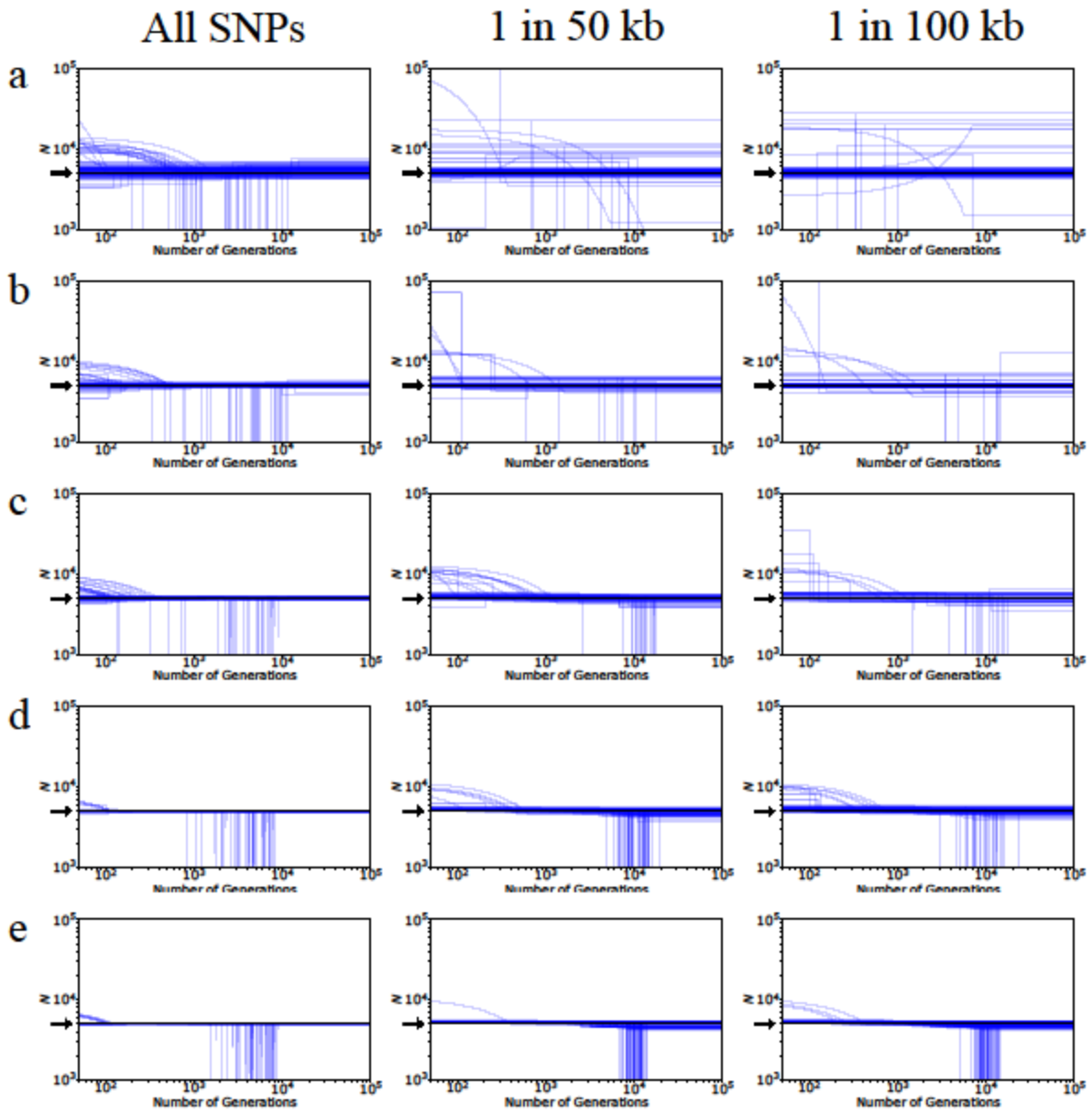
Supp Figure 1: Performance of MSMC under neutrality and demographic equilibrium when using 1, 2, and 4 diploid individuals for inference, for varying chromosome sizes: (a) 1Mb, (b) 10 Mb, (c) 50 Mb, (d) 200 Mb, (e) 1 Gb. Simulations were performed using SLiM (3.1) for 10*N* generations with 100 replicates for panels (a-d) and 10 for panel (e). Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure1.txt.



Supp Figure 2: Performance of MSMC under neutrality and demographic equilibrium when using 1, 2, and 4 diploid individuals for inference, for varying chromosome sizes: (a) 1 Mb, (b) 10 Mb, (c) 50 Mb, (d) 200 Mb, (e) 1 Gb. Simulations were performed using msprime (0.7.3) with 100 replicates for each panel. MSMC runs with 4 diploid individuals at 1Gb could not be obtained due to the long computational times required. Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure2.txt.



Supp Figure 3: Performance of *fastsimcoal2* under neutrality and demographic equilibrium when simulations were performed using SLiM (3.1) for varying chromosome sizes: (a) 1 Mb, (b) 10 Mb, (c) 50 Mb, (d) 200 Mb, (e) 1 Gb. Model selection was performed with 4 models – equilibrium, instantaneous size change, exponential size change, and instantaneous bottleneck. Inference was performed using 50 diploid individuals using all SNPs (left column), 1 SNP per 50 kb (middle column) and 1 SNP per 100 kb (right column). The inferred population size estimates of the best model are plotted (blue lines). The numbers of replicates for (a)-(d) were 100, while for (e) were 10. The true model is shown in black and indicated by the position of the arrow. Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure3_SuppTable2_5.txt.



Supp Figure 4: Performance of *fastsimcoal2* under neutrality and demographic equilibrium when simulations were performed using msprime (0.7.3) for varying chromosome sizes: (a) 1 Mb, (b) 10 Mb, (c) 50 Mb, (d) 200 Mb, (e) 1 Gb. Model selection was performed with 4 models – equilibrium, instantaneous size change, exponential size change, and instantaneous bottleneck. Inference was performed using 50 diploid individuals using all SNPs (left column), 1 SNP per 50 kb (middle column) and 1 SNP per 100 kb (right column). The inferred population size estimates of the best model are plotted (blue lines), with 100 replicates for each panel. The true model is shown in black and indicated by the position of the arrow. Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure4_SuppTable3_6.txt.

	Functional Density	5%				10%				20%			
True Model	Test Model / DFE Shape	Equilibrium	Exponential Change	Instantaneous change	Instant Bottleneck	Equilibrium	Exponential Change	Instantaneous change	Instant Bottleneck	Equilibrium	Exponential Change	Instantaneous change	Instant Bottleneck
Equilibrium	DFE0	0	1	9	0	0	0	10	0	0	0	10	0
	DFE1	0	1	9	0	0	2	8	0	0	2	8	0
	DFE2	0	0	10	0	0	2	8	0	0	0	10	0
	DFE3	0	6	4	0	0	10	0	0	0	10	0	0
	DFE4	0	1	9	0	0	0	10	0	0	9	1	0
	DFE5	0	6	4	0	0	10	0	0	0	10	0	0
	DFE6	0	6	4	0	0	6	4	0	0	10	0	0
Instantaneous Decline (6x)	DFE0	0	0	10	0	0	0	10	0	0	0	10	0
	DFE1	0	0	10	0	0	0	10	0	0	0	10	0
	DFE2	0	0	10	0	0	0	10	0	0	0	0	10
	DFE3	0	0	10	0	0	10	0	0	0	10	0	0
	DFE4	0	0	10	0	0	0	10	0	0	10	0	0
	DFE5	0	0	10	0	0	0	0	10	0	10	0	0
	DFE6	0	0	10	0	0	0	10	0	0	10	0	0
Instantaneous Decline (2x)	DFE0	0	0	10	0	0	0	10	0	0	0	10	0
	DFE1	0	0	10	0	0	0	10	0	0	0	10	0
	DFE2	0	0	10	0	0	0	10	0	0	0	0	10
	DFE3	0	0	10	0	0	0	0	10	0	10	0	0
	DFE4	0	0	10	0	0	0	10	0	0	0	0	10
	DFE5	0	0	10	0	0	0	10	0	0	10	0	0
	DFE6	0	0	10	0	0	0	10	0	0	0	0	10
Exponential Growth (30x)	DFE0	0	10	0	0	0	10	0	0	0	10	0	0
	DFE1	0	10	0	0	0	10	0	0	0	10	0	0
	DFE2	0	10	0	0	0	10	0	0	0	10	0	0
	DFE3	0	10	0	0	0	10	0	0	0	10	0	0
	DFE4	0	10	0	0	0	10	0	0	0	10	0	0
	DFE5	0	10	0	0	0	10	0	0	0	10	0	0
	DFE6	0	10	0	0	0	10	0	0	0	10	0	0
Exponential Growth (2x)	DFE0	0	0	10	0	0	0	10	0	0	0	10	0
	DFE1	0	0	10	0	0	0	10	0	0	0	10	0
	DFE2	0	0	10	0	0	0	10	0	0	0	10	0
	DFE3	0	0	10	0	0	0	10	0	0	0	10	0
	DFE4	0	0	10	0	0	0	10	0	0	0	10	0
	DFE5	0	0	10	0	0	0	10	0	0	0	10	0
	DFE6	0	0	10	0	0	0	10	0	0	0	10	0

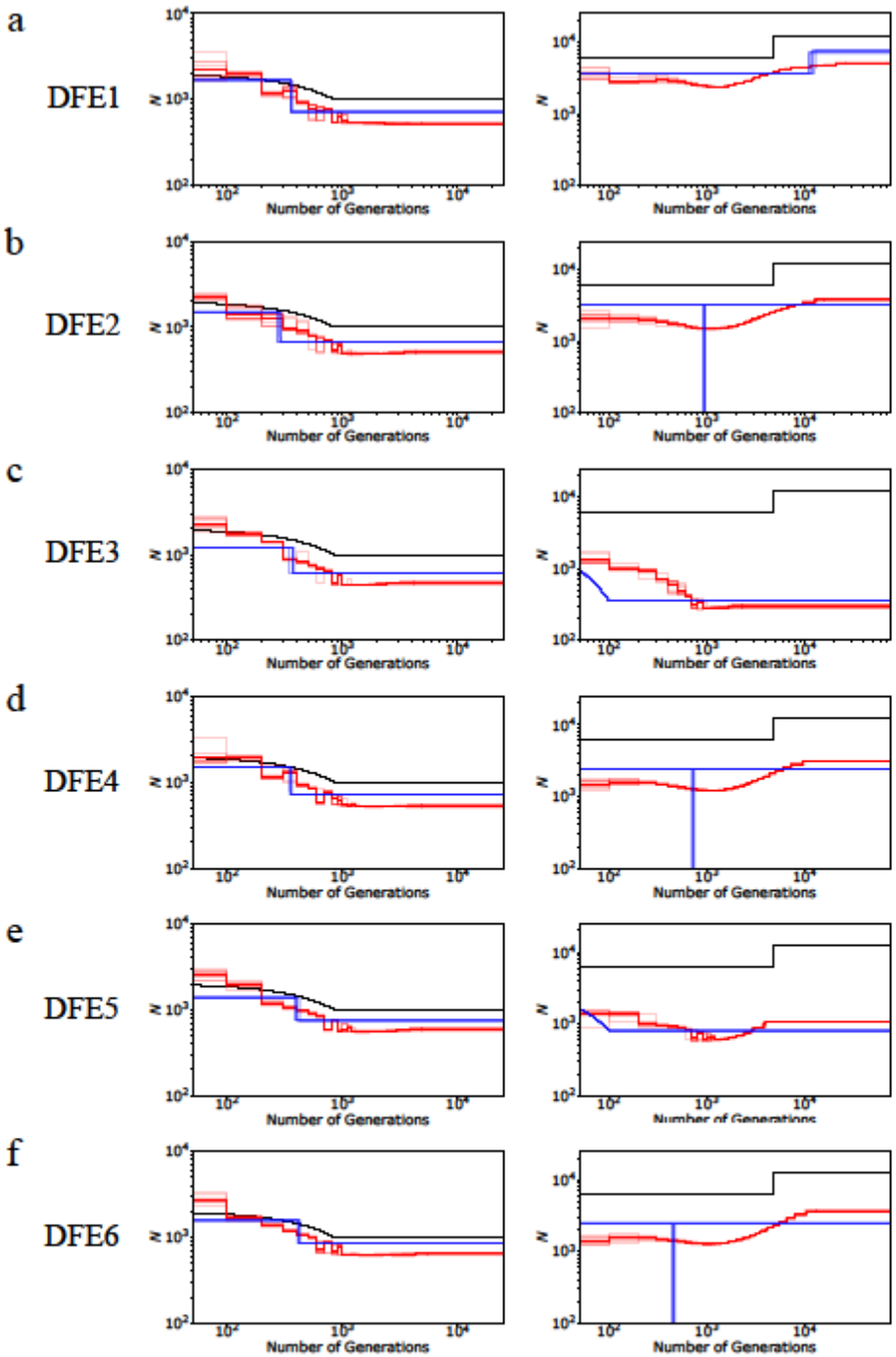
Supp Figure 5: Model selection by *fastsimcoal2* in the presence of BGS, when chosen from four possible models: equilibrium, instantaneous size change, exponential size change, and instantaneous bottleneck. The DFEs are specified in Table 1, and DFE0 refers to neutrality. Results are shown when all SNPs were used for inference, when 5%, 10% or 20% of the genomes were exonic (with exonic sites masked), and the standard AIC penalty was used. Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure5.txt.

AIC penalty	True Model	Test Model / DFE Shape	Equilibrium	Exponential Change	Instantaneous change	Instant Bottleneck
2x	Equilibrium	DFE0	0	0	8	2
		DFE1	0	0	10	0
		DFE2	0	1	9	0
		DFE3	0	10	0	0
		DFE4	0	3	7	0
		DFE5	0	10	0	0
		DFE6	0	6	4	0
	Instantaneous Decline (6x)	DFE0	0	0	10	0
		DFE1	0	0	0	10
		DFE2	0	10	0	0
		DFE3	0	10	0	0
		DFE4	0	10	0	0
		DFE5	0	10	0	0
		DFE6	0	10	0	0
	Exponential Growth (30x)	DFE0	0	10	0	0
		DFE1	0	10	0	0
		DFE2	0	10	0	0
		DFE3	0	10	0	0
		DFE4	0	10	0	0
		DFE5	0	10	0	0
		DFE6	0	10	0	0
25x	Equilibrium	DFE0	10	0	0	0
		DFE1	0	0	10	0
		DFE2	0	1	9	0
		DFE3	0	10	0	0
		DFE4	0	3	7	0
		DFE5	0	10	0	0
		DFE6	0	6	4	0
	Instantaneous Decline (6x)	DFE0	0	0	10	0
		DFE1	0	0	0	10
		DFE2	0	10	0	0
		DFE3	0	10	0	0
		DFE4	0	10	0	0
		DFE5	0	10	0	0
		DFE6	0	10	0	0
	Exponential Growth (30x)	DFE0	0	10	0	0
		DFE1	0	10	0	0
		DFE2	0	10	0	0
		DFE3	0	10	0	0
		DFE4	0	10	0	0
		DFE5	0	10	0	0
		DFE6	0	10	0	0

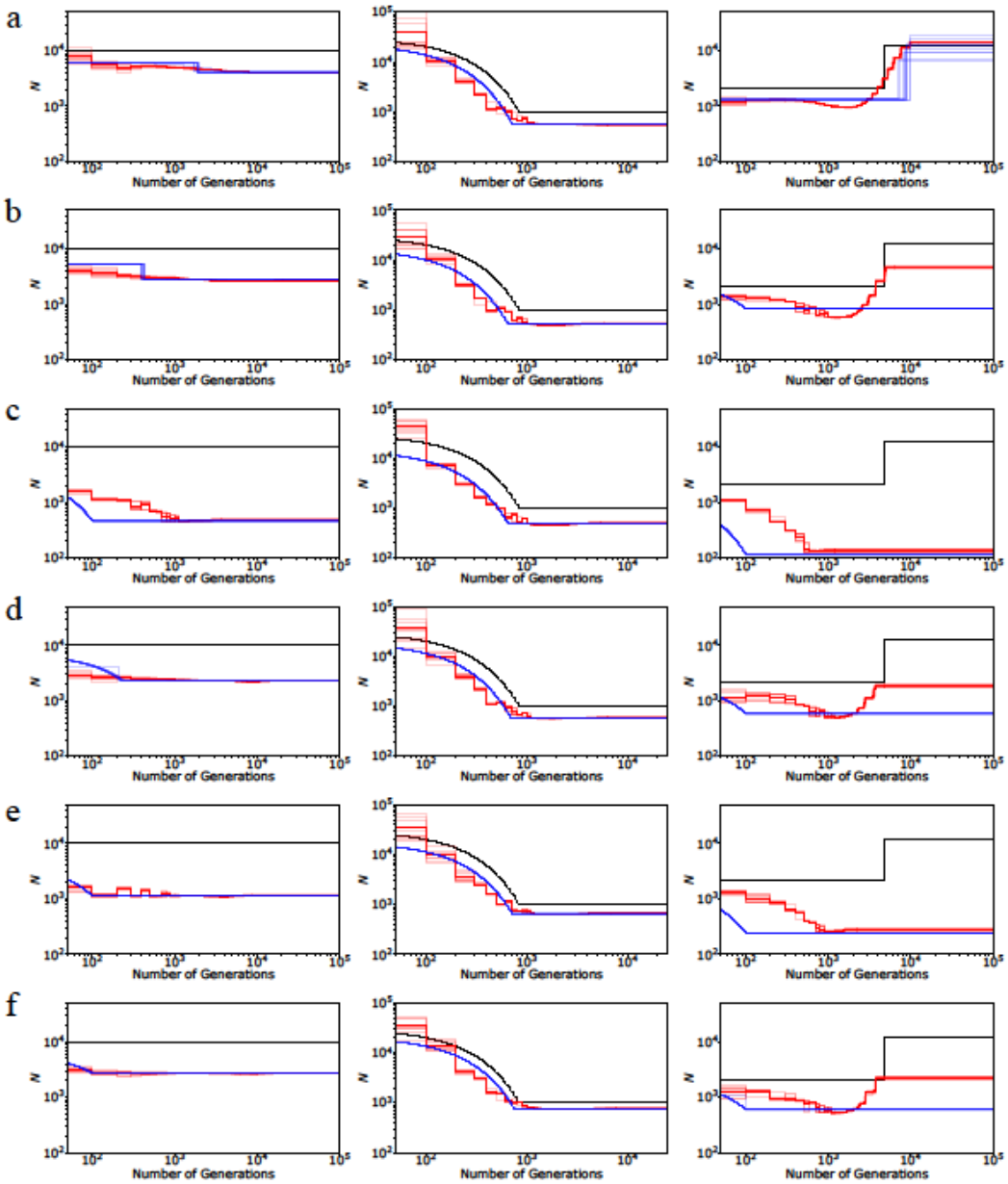
Supp Figure 6: Model selection by *fastsimcoal2* in the presence of BGS, when chosen from four possible models: equilibrium, instantaneous size change, exponential size change, and instantaneous bottleneck. The DFEs are specified in Table 1, and DFE0 refers to neutrality. Results are shown when SNPs used for inference were separated at a distance of 100 kb, 20% of the genomes were exonic, and the AIC penalty was 2× (standard) or 25× the number of parameters. Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure6.txt.

	Functional Density	5%			10%			20%		
True Model	Behavior / DFE Shape	Growth	Decline	Bottleneck	Growth	Decline	Bottleneck	Growth	Decline	Bottleneck
Equilibrium	DFE1	10	0	0	10	0	0	10	0	0
	DFE2	10	0	0	10	0	0	10	0	0
	DFE3	10	0	0	10	0	0	10	0	0
	DFE4	10	0	0	10	0	0	10	0	0
	DFE5	10	0	0	10	0	0	10	0	0
	DFE6	10	0	0	10	0	0	10	0	0
Instantaneous Decline (6x)	DFE1	0	10	0	0	10	0	0	10	0
	DFE2	0	10	0	0	10	0	0	0	10
	DFE3	0	10	0	10	0	0	10	0	0
	DFE4	0	10	0	0	10	0	10	0	0
	DFE5	0	10	0	0	0	10	10	0	0
	DFE6	0	10	0	0	10	0	10	0	0
Instantaneous Decline (2x)	DFE1	0	10	0	0	10	0	0	10	0
	DFE2	0	10	0	0	10	0	0	0	10
	DFE3	0	10	0	0	0	10	10	0	0
	DFE4	0	10	0	0	10	0	0	0	10
	DFE5	0	10	0	0	10	0	10	0	0
	DFE6	0	10	0	0	10	0	0	0	10
Exponential Growth (30x)	DFE1	10	0	0	10	0	0	10	0	0
	DFE2	10	0	0	10	0	0	10	0	0
	DFE3	10	0	0	10	0	0	10	0	0
	DFE4	10	0	0	10	0	0	10	0	0
	DFE5	10	0	0	10	0	0	10	0	0
	DFE6	10	0	0	10	0	0	10	0	0
Exponential Growth (2x)	DFE1	10	0	0	10	0	0	10	0	0
	DFE2	10	0	0	10	0	0	10	0	0
	DFE3	10	0	0	10	0	0	10	0	0
	DFE4	10	0	0	10	0	0	10	0	0
	DFE5	10	0	0	10	0	0	10	0	0
	DFE6	10	0	0	10	0	0	10	0	0

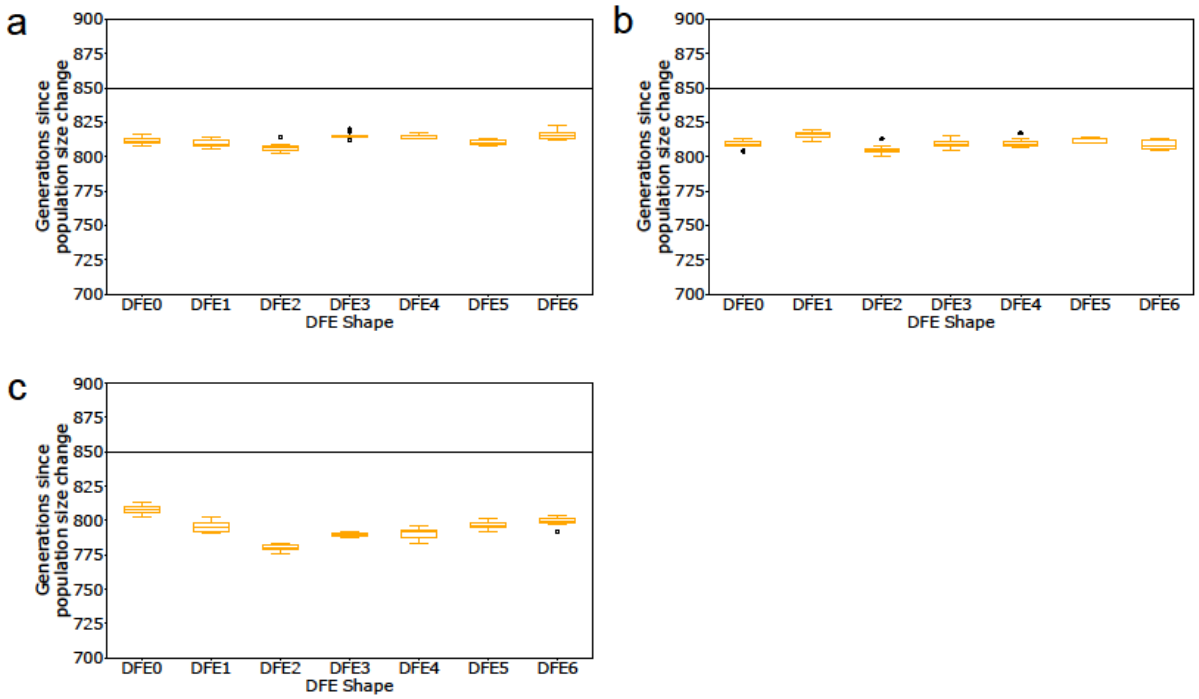
Supp Figure 7: Effects of BGS on inference of growth or decline by *fastsimcoal2*. The inferred model was classified as growth if $N_{anc} < N_{cur}$ and as decline if $N_{anc} > N_{cur}$. The DFEs are specified in Table 1. Results are shown for all SNPs, when 5%, 10% and 20% of the genomes were exonic (with exonic sites masked), and a standard AIC penalty was used. Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure7.txt.



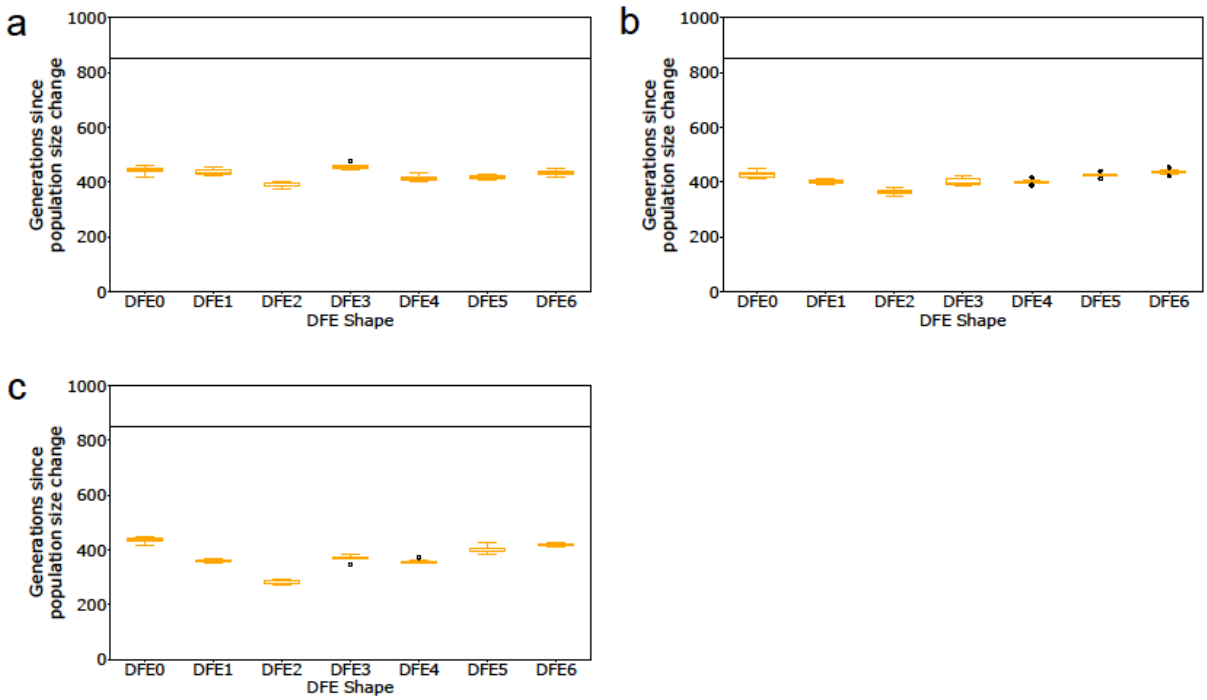
Supp Figure 8: Inferred demography from MSMC (red lines) and *fastsimcoal2* (blue lines) in the presence of background selection, with the true DFE shown to the left of the panel, for 2-fold instantaneous decline (right column) and 2-fold exponential growth (left column). In this case, 20% of the genome was exonic (and exonic sites were masked). The true demographic models are shown as black lines. Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure8.txt.



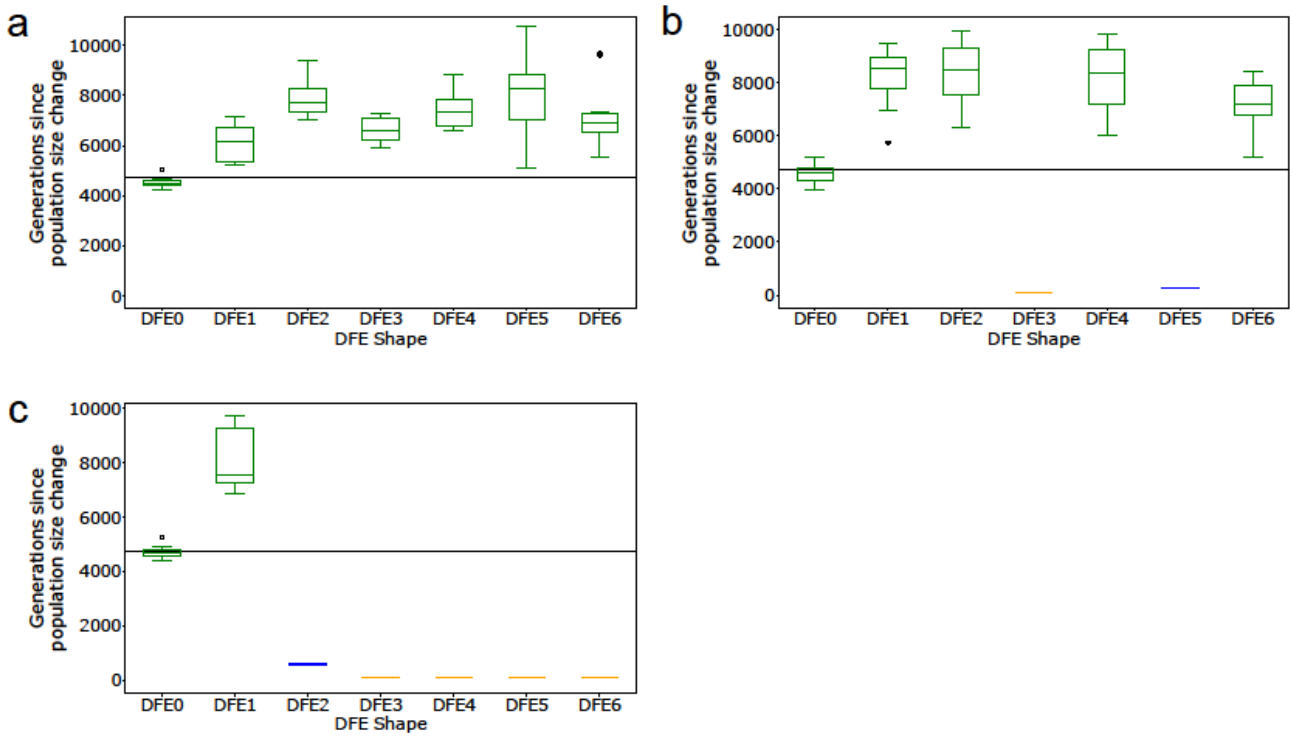
Supp Figure 9: Inference of demography by MSMC (red lines; 10 replicates) and *fastsimcoal2* (blue lines; 10 replicates) under demographic equilibrium (left column), 30-fold exponential growth (middle column), and ~6-fold instantaneous decline (right column) in the presence of direct purifying selection (*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 (defined in Table 1): (a) DFE1, (b) DFE2, (c) DFE3, (d) DFE4, (e) DFE5, (f) DFE6. In this case, 20% of the genome was exonic and all SNPs were used for inference, including exonic sites. Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure9.txt.



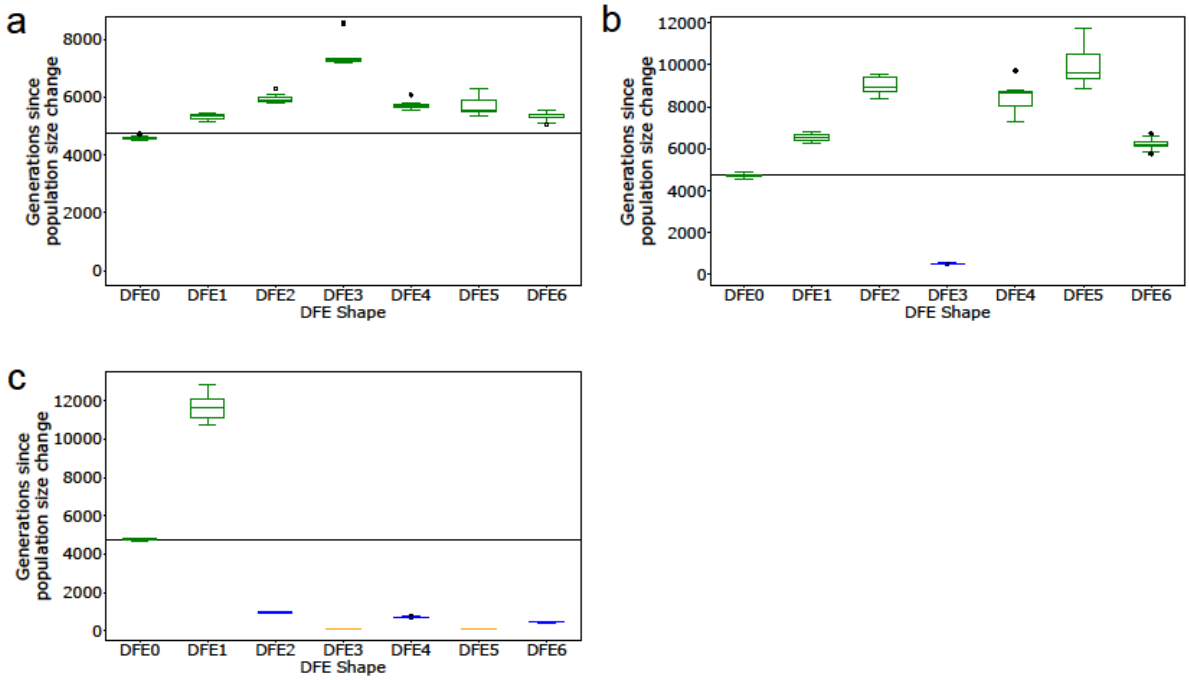
Supp Figure 10: Inference of the timing of 30-fold exponential growth by *fastsimcoal2* in the presence of BGS compared to neutrality (DFE0). The DFEs are specified in Table 1. Results are shown for the case where (a) 5%, (b) 10%, and (c) 20% of the genome is exonic (with exonic sites masked), and all non-exonic SNPs are used for inference. The black horizontal line represents the true timing.



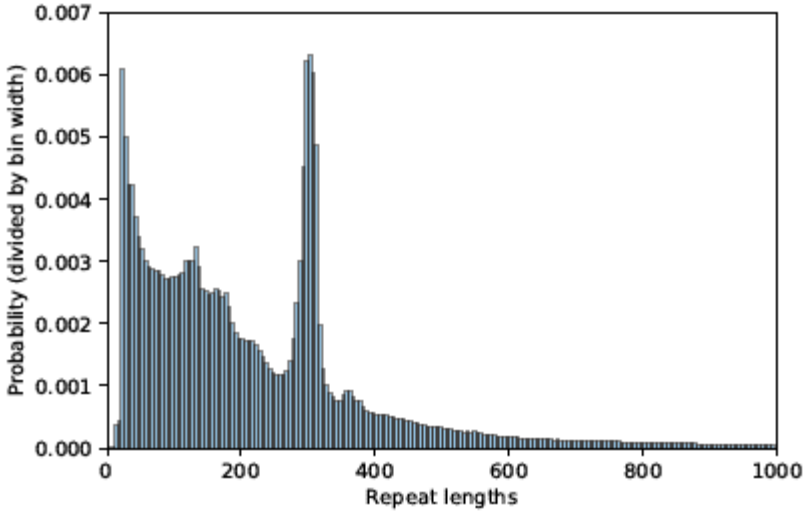
Supp Figure 11: Inference of the timing of 2-fold exponential growth by *fastsimcoal2* in the presence of BGS compared to neutrality (DFE0). Results are shown for the case where (a) 5%, (b) 10%, and (c) 20% of the genome is exonic (with exonic sites masked), and all non-exonic SNPs are used for inference. The black horizontal line represents the true timing.



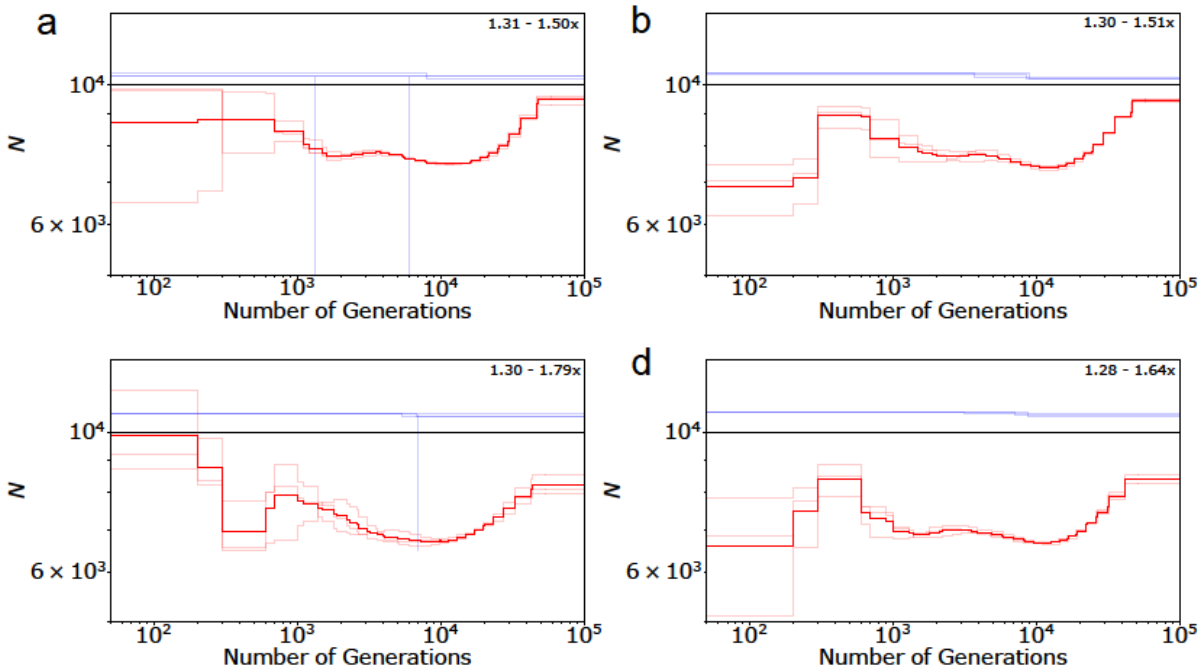
Supp Figure 12: Inference of the timing of 6-fold instantaneous decline by *fastsimcoal2* in the presence of BGS compared to neutrality (DFE0). The DFEs are specified in Table 1. Results are shown for the case in which (a) 5%, (b) 10%, and (c) 20% of the genome is exonic (with exonic sites masked), and all non-exonic SNPs are used for inference. The black horizontal line represents the true timing. Boxplots are presented in green if decline was inferred, in yellow if growth was inferred, and in blue if a bottleneck was inferred.



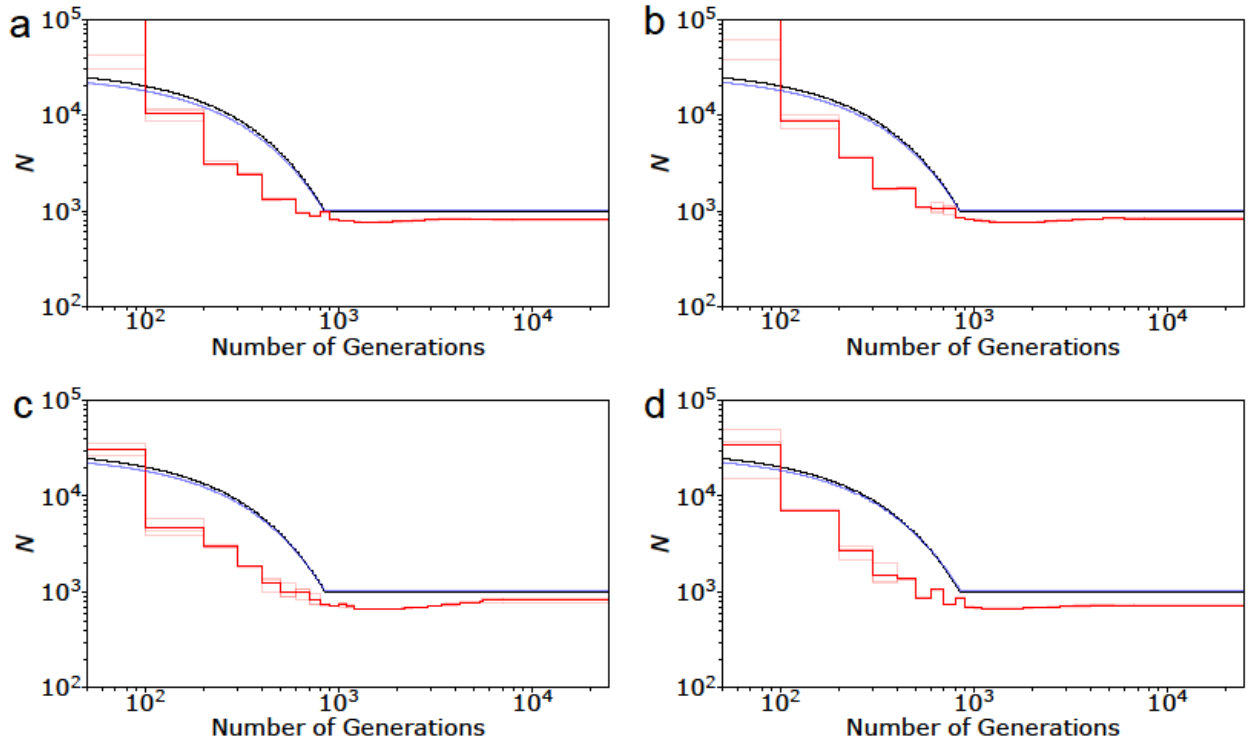
Supp Figure 13: Inference of the timing of 2-fold instantaneous decline by *fastsimcoal2* in the presence of BGS compared to neutrality (DFE0). The DFEs are specified in Table 1. Results are shown for the case where (a) 5%, (b) 10%, and (c) 20% of the genome is exonic (with exonic sites masked), and all non-exonic SNPs are used for inference. The black horizontal line represents the true timing. Boxplots are presented in green if decline was inferred, in yellow if growth was inferred, and in blue if a bottleneck was inferred.



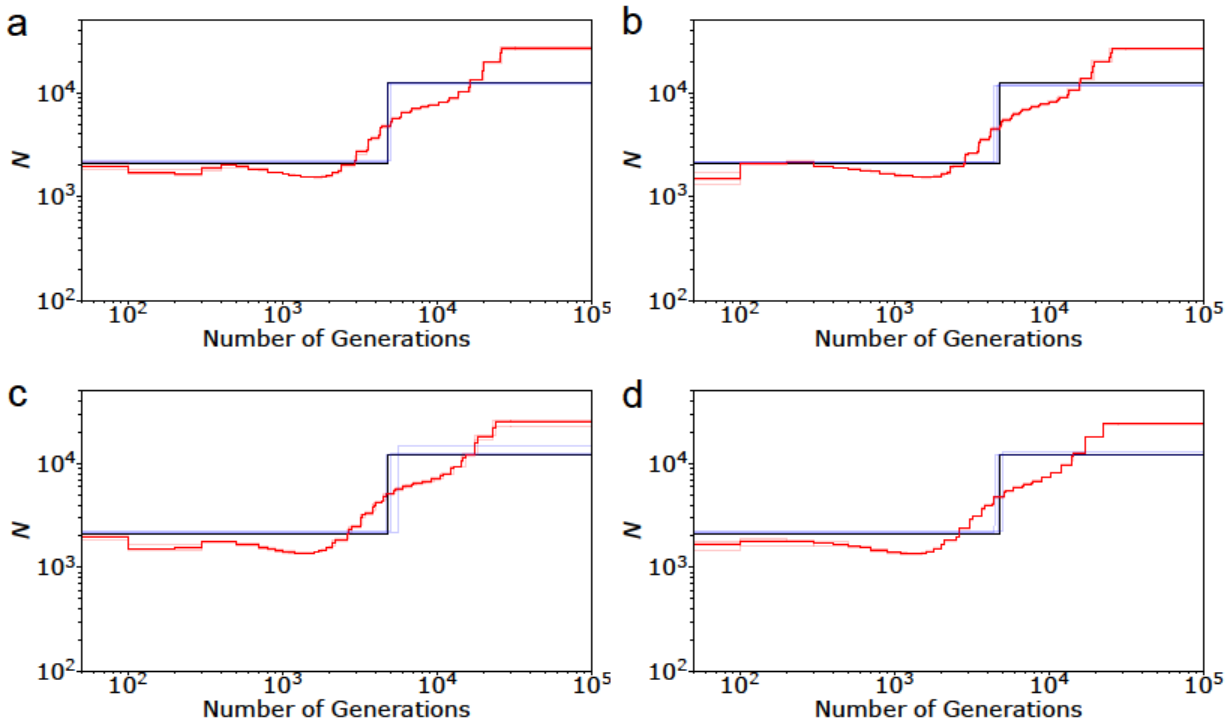
Supp Figure 14: Distribution of lengths of repeat regions in the human genomes (*hg19*). Shown above is the distribution of lengths up to 1000 bp, although lengths of repeat regions range between 6 to 160602 bp. Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure14.txt.



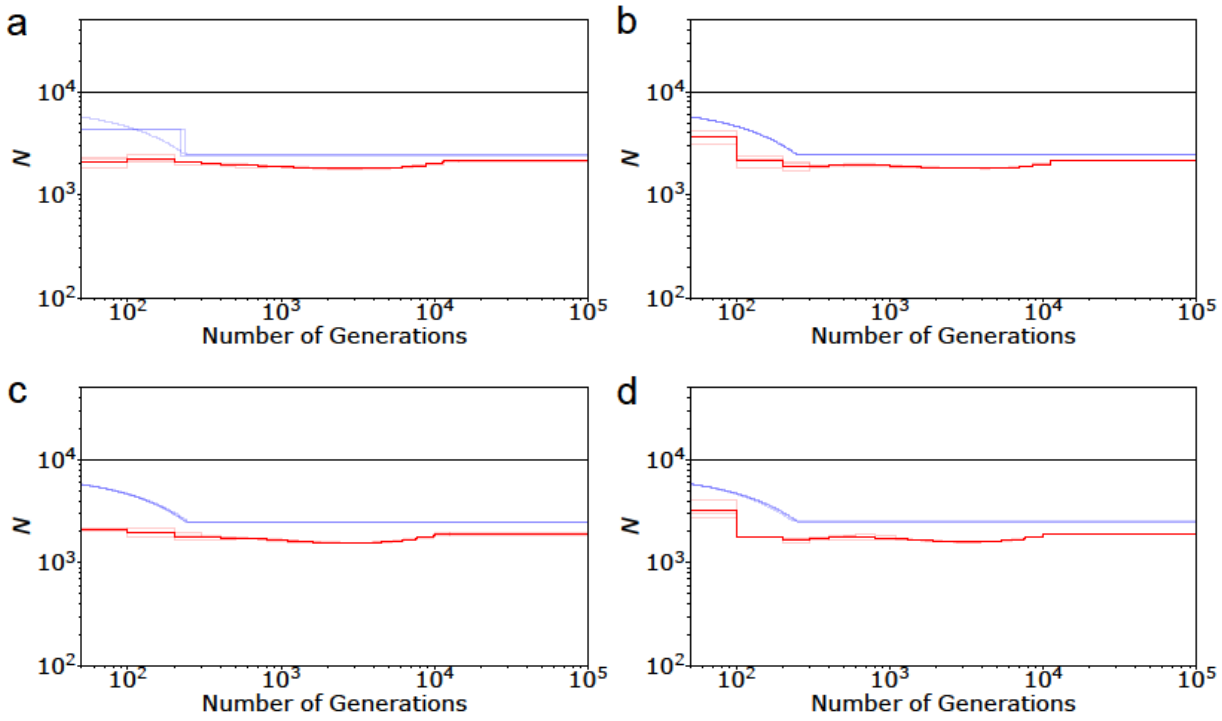
Supp Figure 15: Performance of demographic inference by MSMC (red lines) and *fastsimcoal2* (blue lines) under different scenarios of neutrality, when the true model is equilibrium: (a) there is variation in recombination and mutation rates, (b) there is variation in recombination and mutation rates and the centromeric region is masked, (c) there is variation in recombination and mutation rates, and short regions resembling repeats (comprising 10% of each chromosome) are randomly masked across the genome, and (d) there is variation in recombination and mutation rates, and the centromere as well as small-sized repeats are randomly masked across the genome. The maximum and minimum fold change detected in every scenario is indicated on the upper right corner. Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure15-20.txt.



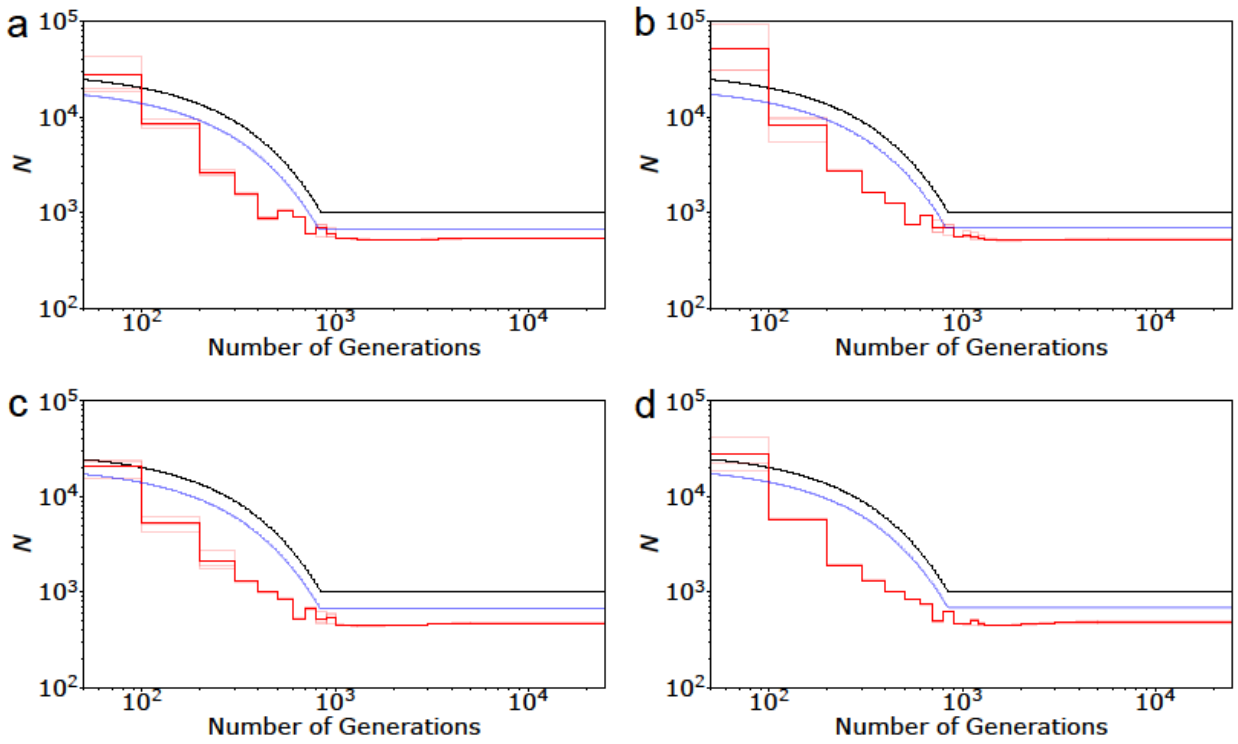
Supp Figure 16 : Performance of demographic inference by MSMC (red lines) and *fastsimcoal2* (blue lines) under different scenarios of neutrality, when the true model is 30-fold exponential growth: (a) there is variation in recombination and mutation rates across the genome, (b) there is variation in recombination and mutation rates, and the centromeric region is masked, (c) there is variation in recombination and mutation rates, and short regions resembling repeats are randomly masked across the genome (comprising of 10% of each chromosome), and (d) there is variation in recombination and mutation rates, and the centromere as well as small-sized repeats are randomly masked across the genome. Detailed methods including command can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure15-20.txt.



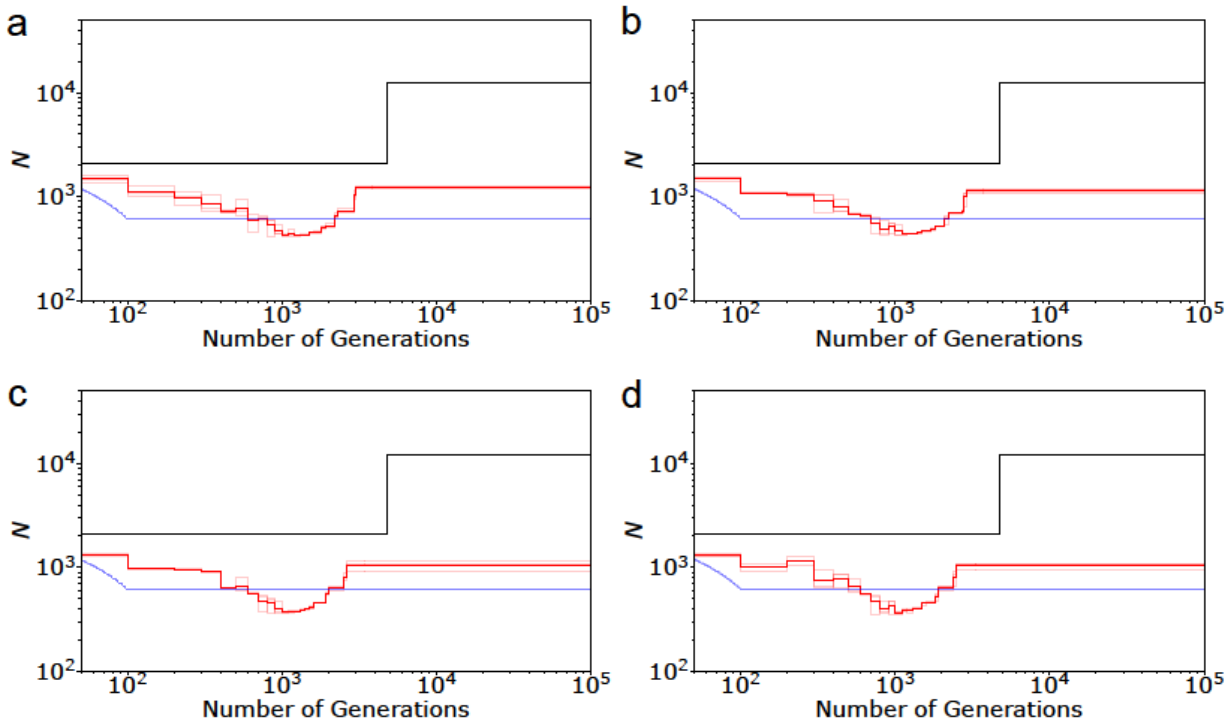
Supp Figure 17: Performance of demographic inference by MSMC (red lines) and *fastsimcoal2* (blue lines) under different scenarios of neutrality, when the true model is 6-fold instantaneous decline: (a) there is variation in recombination and mutation rates across the genome, (b) there is variation in recombination and mutation rates, and the centromeric region is masked, (c) there is variation in recombination sand mutation rates, and short regions resembling repeats are randomly masked across the genome (comprising of 10% of each chromosome), and (d) there is variation in recombination and mutation rates, and the centromere as well as small-sized repeats are randomly masked across the genome. Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure15-20.txt.



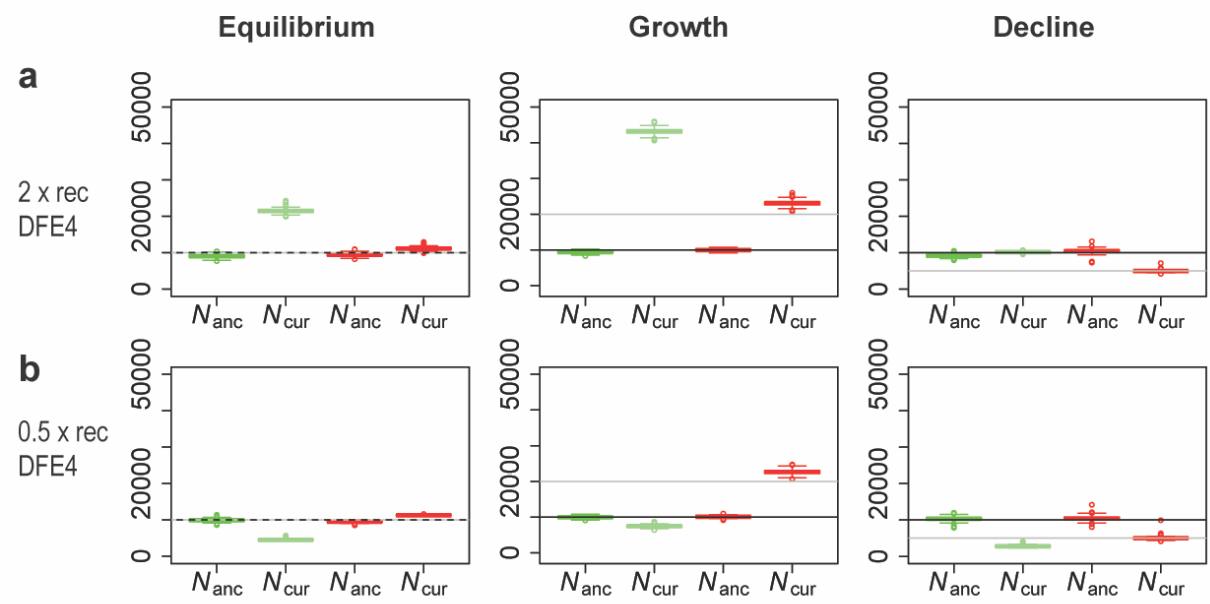
Supp Figure 18: Performance of demographic inference by MSMC (red lines) and *fastsimcoal2* (blue lines) in the presence of background selection, under different scenarios when the true model is equilibrium: (a) there is variation in recombination and mutation rates, (b) there is variation in recombination and mutation rates and the centromeric region is masked, (c) there is variation in recombination and mutation rates, and short regions resembling repeats (comprising 10% of each chromosome) are randomly masked across the genome, and (d) there is variation in recombination and mutation rates, and the centromere as well as small-sized repeats are randomly masked across the genome. Exons comprise of 20% of the genome, experience purifying selection given by DFE4 ($f_0 = f_1 = f_2 = f_3 = 0.25$), and are masked when performing inference. Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure15-20.txt.



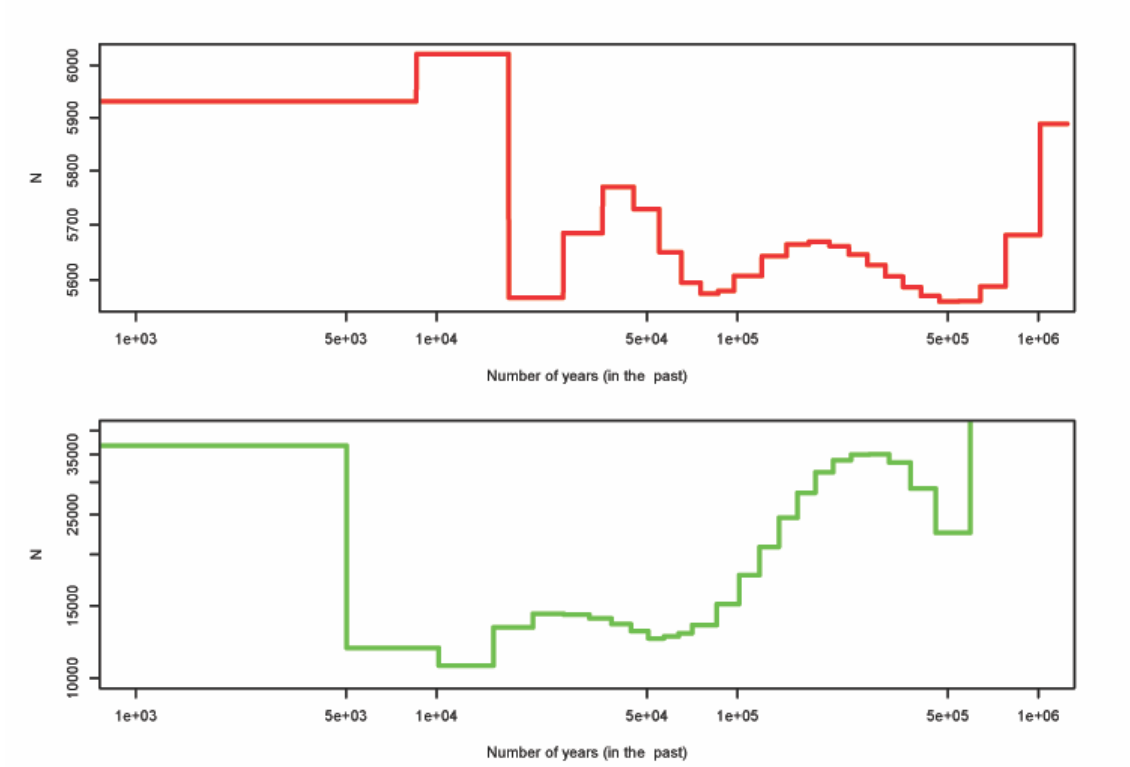
Supp Figure 19: Performance of demographic inference by MSMC (red lines) and *fastsimcoal2* (blue lines) in the presence of background selection, under different scenarios when the true model is 30-fold exponential growth: (a) there is variation in recombination and mutation rates, (b) there is variation in recombination and mutation rates and the centromeric region is masked, (c) there is variation in recombination and mutation rates, and short regions resembling repeats (comprising 10% of each chromosome) are randomly masked across the genome, and (d) there is variation in recombination and mutation rates, and the centromere as well as small-sized repeats are randomly masked across the genome. Exons comprise of 20% of the genome, experience purifying selection given by DFE4 ($f_0 = f_1 = f_2 = f_3 = 0.25$), and are masked when performing inference. Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure15-20.txt.



Supp Figure 20: Performance of demographic inference by MSMC (red lines) and *fastsimcoal2* (blue lines) in the presence of background selection, under different scenarios when the true model is a 6-fold instantaneous decline: (a) there is variation in recombination and mutation rates, (b) there is variation in recombination and mutation rates and the centromeric region is masked, (c) there is variation in recombination and mutation rates, and short regions resembling repeats (comprising 10% of each chromosome) are randomly masked across the genome, and (d) there is variation in recombination and mutation rates, and the centromere as well as small-sized repeats are randomly masked across the genome. Exons comprise of 20% of the genome, experience purifying selection given by DFE4 ($f_0 = f_1 = f_2 = f_3 = 0.25$), and are masked when performing inference. Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure15-20.txt.



Supp Figure 21: Performance of the ABC method when recombination rate is mis-specified. (a) The true recombination rate is 2-fold higher than that assumed, and (b) the true recombination rate is 2-fold lower than that assumed. Boxplots in green show the posterior estimates when the recombination rate is higher or lower than assumed. For comparison, boxplots in red show the posterior inferred when the corresponding recombination rate is correctly specified. The black line displays the true ancestral population size (N_{anc}) and the gray line represents the true current population size (N_{cur}). Detailed methods including command lines can be found here: https://github.com/paruljohri/demographic_inference_with_selection/blob/main/CommandLines/SuppFigure21.txt.



Supp Figure 22: 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 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 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)).