Effects of linked selective sweeps on demographic inference and model selection

Daniel R. Schrider[1,2,*,†], Alexander G. Shanku[1,3,*], and Andrew D. Kern[1,2]

[1]Department of Genetics, Rutgers University, Piscataway, NJ, 08854, USA

[2]Human Genetics Institute of New Jersey, Rutgers University, Piscataway, NJ, 08554, USA

[3]Institute for Quantitative Biomedicine, Rutgers University, Piscataway, NJ, 08554, USA

[†]Corresponding author: Email: dan.schrider@rutgers.edu

[*]These authors contributed equally to this work.

Keywords: population genetics, positive selection, demographic inference

ABSTRACT

The availability of large-scale population genomic sequence data has resulted in an explosion in efforts to infer the demographic histories of natural populations across a broad range of organisms. As demographic events alter coalescent genealogies they leave detectable signatures in patterns of genetic variation within and between populations. Accordingly, a variety of approaches have been designed to leverage population genetic data to uncover the footprints of demographic change in the genome, thereby elucidating population histories. The vast majority of these methods make the simplifying assumption that the measures of genetic variation used as their input are unaffected by natural selection. However, natural selection can dramatically skew patterns of variation not only at selected sites, but at linked, neutral loci as well. Here we assess the impact of recent positive selection on demographic inference by characterizing the performance of three popular methods through extensive simulation of datasets with varying numbers of linked selective sweeps. In particular, we examined three different demographic models relevant to a number of species, finding that positive selection can bias parameter estimates of each of these models—often severely. Moreover, we find that selection can lead to incorrect inferences of population size changes when none have occurred. We argue that the amount of recent positive selection required to skew inferences may often be acting in natural populations. These results suggest that demographic studies conducted in many species to date may have exaggerated the extent and frequency of population size changes.

INTRODUCTION

The widespread availability of population genomic data has spurred a new generation of studies aimed at understanding the histories of natural populations from a host of model and non-model organisms alike. In particular, genome-scale variation data allows for inference of demographic factors such as population size changes, the timing and ordering of population splits, migration rates between populations, and the founding of admixed populations (Pool et al. 2010; Pickrell and Pritchard 2012; Sousa and Hey 2013). Such efforts can refine our picture of demographic events inferred from the archaeological record (e.g. Fagundes et al. 2007; Goebel et al. 2008), or reveal such events in species where no archaeological data are available, and can aid conservation efforts by complementing census data (e.g. Hájková et al. 2007; Garrick et al. 2015).

Population genomic datasets are well suited for this task simply because demographic changes leave their mark on patterns of genetic variation. Recent population growth for example will result in an excess of rare variation compared to equilibrium expectations (Fu 1997), while population contraction will result in an excess of intermediate frequency alleles (Maruyama and Fuerst 1985). In recent years researchers have devised a variety of methods that seek to detect the population genetic signatures of these demographic events. These include approximate Bayesian computation (ABC) methods, where simulation is used to approximate the posterior probability distributions of a demographic model's parameters through the use of a collection of population genetic summary statistics without specification of an explicit likelihood function (Tavaré et al. 1997; Pritchard et al. 1999; Beaumont et al. 2002; Marjoram et al. 2003; Excoffier et al. 2005; Wegmann et al. 2010). Other approaches, such as ∂a∂i (Diffusion Approximations

for Demographic Inference; Gutenkunst et al. 2009), use the probability density of the site frequency spectrum (SFS) under a given demographic model and parameterization to calculate the likelihood of the observed SFS (Marth et al. 2004; Gutenkunst et al. 2009), thereby allowing for optimization of model parameters. More recently, methods based on the sequentially Markovian coalescent (SMC; McVean and Cardin 2005; Marjoram and Wall 2006) have been devised (Li and Durbin 2011; Sheehan et al. 2013; Schiffels and Durbin 2014), to infer how a population's size has changed over time through the description of patterns of genetic variation along a recombining chromosome.

Applications of these inference methods have revealed much about the demographic histories of various species. For instance, early studies of human genomic variation found that non-African populations experienced a considerable population bottleneck (Marth et al. 2003), most likely associated with migration out of Africa (Reich et al. 2001; Adams and Hudson 2004; Voight et al. 2005), followed by more recent recovery. Later studies refined the estimated timing of this bottleneck to ~50 kya (Fagundes et al. 2007; Gravel et al. 2011; Lukić and Hey 2012), uncovered a second bottleneck associated with the divergence of European and Asian populations (Keinan et al. 2007; Gutenkunst et al. 2009; Gravel et al. 2011), and inferred that the recovery from this bottleneck proceeded through continuous exponential growth (Fagundes et al. 2007; Gutenkunst et al. 2009). Numerous studies have found strong genetic signals of admixture in many different human subpopulations (Meinilä et al. 2001; Parra et al. 2001; Martínez-Cortés et al. 2012; Patterson et al. 2012; Moorjani et al. 2013; Auton et al. 2015). Finally, recent studies of large population samples capable of observing very rare alleles found evidence that this growth has accelerated dramatically within the last several thousand years (Coventry et al. 2010; Tennessen et al. 2012; Gao and Keinan 2016). Similarly, recent studies in *Drosophila*

melanogaster show evidence of a severe out-of-Africa bottleneck (Begun and Aquadro 1993), but occurring within the last 20,000 years (Li and Stephan 2006; Thornton and Andolfatto 2006). There is also growing support from population genetic studies for African-European admixture in the North American population of *D. melanogaster* (Duchen et al. 2013; Bergland et al. 2015; Kao et al. 2015).

While demographic inference from population genomic data in its various forms has proven to be successful technique, a unifying assumption of these various inference methods (ABC, SFS-based, and SMC-based approaches.) is that the genetic data in question are strictly neutral and free from the effects of linked selection in the genome. While this is an important simplifying assumption, it may be the case that in many populations a sizeable fraction of the genome is influenced by natural selection (Hahn 2008; Sella et al. 2009; Corbett-Detig et al. 2015). Indeed natural selection can produce skews in patterns of genetic variation that are quite similar to those generated by certain non-equilibrium demographic histories. For example, positive selection driving a mutation to fixation (i.e. a selective sweep; Maynard Smith and Haigh 1974) may resemble a population bottleneck (Simonsen et al. 1995). Conversely, many demographic perturbations are well known to cause unacceptably high rates of false positives for many classical tests for selection (Simonsen et al. 1995; Przeworski 2002; Akey et al. 2004; Jensen et al. 2005; Nielsen et al. 2005). Thus, if natural selection has a substantial impact on genome-wide patterns of variation, then many demographic parameter estimates could be biased (Hahn 2008; Gazave et al. 2014). Indeed this has been shown to be the case for at least some scenarios of background selection (Ewing and Jensen 2016), where purifying selection reduces levels of neutral polymorphism at linked sites.

Here, we examine the potential impact of linked positive selection on three of the most widely used methods for demographic inference: ABC (Pritchard et al. 1999; Beaumont et al. 2002), ∂a∂i (Gutenkunst et al. 2009), and PSMC (pairwise sequentially Markovian Coalescent; Li and Durbin 2011). We demonstrate that selection can substantially bias parameter estimates, often leading to overestimates of the severity of population bottlenecks and/or the rate of population growth. Moreover, we show that selective sweeps can result in the selection of the incorrect demographic model: if a reasonably small fraction of loci used for inference are linked to a selective sweep, one may incorrectly infer that a constant-size population experienced a bottleneck. Finally, we discuss the implications of our results for inferences made in humans and *Drosophila*, and recommend steps that could partially mitigate the bias caused by selection.

RESULTS

The impact of positive selection on variation under four demographic scenarios

In order to test the potential impact of positive selection on demographic inference, we simulated four different population size histories: a constant population size model, a bottleneck model, an exponential growth model, and a population contraction followed by later exponential growth (fig. 1; Methods). We begin by demonstrating the impact of linked selective sweeps on genetic diversity, as measured by π (Nei and Li 1979) and Tajima's D (Tajima 1989), under each of these models. For each model, the selection coefficient, s, for each sweep was set to 0.05, and selection was modeled as a "hard" sweep involving the single origin of an advantageous mutation going to fixation (Methods). In fig. 1, we show the mean values of these two statistics

at increasing distances from a selective sweep (where distance is measured by the total crossover rate, *c*, over *s*) under each of the four demographic models examined.

Under our constant population size model, a recent selective sweep with s=0.05 has a marked effect on genetic polymorphism at the site of the sweep (fig. 1A): on average π is reduced approximately 8-fold (6.03, versus a neutral expectation of 47.99), and Tajima's *D* is well below zero (mean *D*: -2.16). At increasing distances, both of these statistics recover toward their expectation under neutrality, which they have nearly reached at c/s=1. At intermediate distances Tajima's *D* passes through a range where its value is above the expectation of approximately zero, as has been observed previously (Teshima et al. 2006; Schrider et al. 2015). Under our bottleneck scenario (fig. 1B), which is Marth et al.'s (2004) model of European human population size history, a selective sweep causes a very similar reduction in diversity and skew away from intermediate allele frequencies (mean π: 6.49, and mean *D*: -2.27, versus neutral expectations of 47.33 and -0.67, respectively). With increasing genetic distance from a sweep under the bottleneck model, π and Tajima's *D* recover in a similar manner as under constant population size.

Exponential growth is often used to model recent population expansions in lieu of instantaneous population size change, and indeed such growth appears to be a key feature of human population history (Fagundes et al. 2007; Gravel et al. 2011; Tennessen et al. 2012). We therefore examined a model based on the same parameterization of strong growth from Gravel et al.'s (2011) estimated European model but omitting the population contractions and most ancient expansion (fig. 1C). Under this model we find that the impact of positive selection is less severe than under either the equilibrium or bottleneck models: π is reduced only ~2-fold from a neutral expectation of 6.57 to 3.17 when a sweep occurs immediately adjacent to the sampled locus,

while Tajima's *D* only drops from -2.06 to -2.40. This demonstrates in part how demography can mimic the action of selection. At increasing distances from the sweep these summary statistics approach their neutral expectations very quickly relative to the equilibrium and bottleneck models. For instance, at c/s of 0.2, π and Tajima's *D* have recovered to 6.34 and -2.00, respectively. Thus, while positive selection still has a fairly far-reaching effect on linked neutral polymorphism under exponential growth, it is subtler than under constant population size because even in unselected regions summaries of variation deviate for equilibrium expectations in a manner that resembles selection.

Finally, we examined a three-epoch model with a population contraction and then subsequent exponential growth. Currently, this contraction-then-growth model is used to represent non-African human population size histories (e.g. Fagundes et al. 2007; Gutenkunst et al. 2009; Gravel et al. 2011; Tennessen et al. 2012). Our parameterization shown in fig. 1D is a simplified version of the European model from Gravel et al. (2011). Under this contraction-then-growth model, we find that π experiences a ~3-fold reduction (from 40.08 to 13.91 on average) when a recent sweep is adjacent to the sampled locus—more modest than the equilibrium and bottleneck models, but slightly more severe than the growth model. Tajima's *D* experiences a dramatic reduction when $c/s=0$, more similar to that under equilibrium and bottleneck models (from a mean of 0.22 to -1.55). As c/s increases, both π and *D* quickly recover toward their neutral expectations as in the growth scenario: when $c/s=0.2$, π is 38.40 on average, while *D* is 0.26. We note that the shallower and narrower valleys of diversity experienced around selective sweeps under both this model and the growth model could impair the sensitivity of efforts to identify selective sweeps in humans and other species where such population size histories appear common (Schrider and Kern 2016).

Demographic parameter estimates are biased by positive selection

We sought to quantify the impact of positive selection on demographic parameter estimation under our bottleneck, growth, and contraction-then-growth models. First, we simulated population samples experiencing no selection and asked how well we could recover the true parameters of the model using diffusion approximations to the SFS via the ∂a∂i software package (Gutenkunst et al. 2009), or with a set of commonly used summary statistics via ABC (Thornton 2009; Csilléry et al. 2012); we address the effects of selection on PSMC later, as this method requires a different sampling scheme. Briefly, we used both of these methods to fit the focal demographic model to data sampled from 500 unlinked simulated loci, and repeated this process on 100 replicate simulated "genomes" (Methods). We then gradually increased the value of *f*, the fraction of these sampled loci linked to hard selective sweeps (within a distance of $c/s \leq 1.0$; Methods). At values of *f* ranging from 0 to 1, we repeated parameter estimation to assess the extent to which a given amount of selection biases our inference.

Population bottleneck

When using ∂a∂i to infer the optimal set of parameters of a bottlenecked population (fig. 1B) experiencing no positive selection, our estimates were quite accurate (fig. 2): our average parameter estimate for the ancestral effective population size, Ne_A, was 10,060 individuals (a 0.6% deviation from the true parameter value); our mean estimate for the time of recovery from the bottleneck, T_R, was 3,120 generations ago (4.0% deviation); our estimated effective population size during the bottleneck, Ne_B was 1,999 individuals (0.05% deviation) on average; and our mean estimate of the present-day effective population size, Ne_0, was 20,465 (2.3%

deviation). Moreover, our inferences were fairly consistent, with most parameter estimates being fairly close to the true value (fig. 2). However, while repeating this analysis with increasing numbers of loci linked to a selective sweep, our parameter estimates became increasingly biased. Even a small value of f produces significant underestimates of the population sizes Ne_0 and Ne_B (fig. 2). For example, the mean inferred value of Ne_B decreases to 1,764 when f=0.2 (an 11.8% underestimate), to 1,402 when f=0.5 (29.9% underestimate), and to 717 when f=1.0 (64.2% underestimate). A more subtle but consistent downward bias of Ne_0 also appears with increasing f: Ne_0 is estimated at 19,650 at f=0.2 (1.8% underestimate), 18,371 when f=0.5 (8.1% underestimate), and 15,561 when f=1.0 (22.2% underestimate). By contrast, estimates of the ancestral effective population size (Ne_A) and the time since the recovery (T_R) are largely unaffected unless f is fairly high (≥0.8), in which case the values of these two parameters are somewhat overestimated (fig. 2).

Like inference using the SFS (i.e. ∂a∂i), our ABC procedure was able to infer the true parameters with minimal bias when run on simulated population samples experiencing no positive selection: the mean estimates were 10,104 for Ne_A (1.4% difference from true value), 2,917 for T_R (2.7% difference), 2,311 for Ne_B (15.6% difference), and 20,078 for Ne_0 (0.4% difference). However, we note that the Ne_B estimate was fairly inconsistent (with the middle 50% of estimates ranging from 1,447 to 3,373), while other parameter estimates exhibited much lower variance. When positive selection is introduced, we obtain significantly biased estimates of all parameters when f≥0.2, and all but Ne_A are significantly biased when f is only 0.1. These biases are in the same direction as observed using ∂a∂i (underestimates for Ne_0 and Ne_B, and overestimates for Ne_A and T_R), but almost always substantially larger. Indeed, for f≥0.2, our estimates of Ne_B and T_R are at the boundaries of our prior parameter ranges, respectively (the

upper bound of 3,500 for T_R and the lower bound of 100 for Ne_B). Also note that when f increases to >0.2, estimates of T_R are also at the upper bound of our prior: we are inferring a very short but extreme bottleneck. Thus, for our bottleneck model, ABC based on our set of summary statistics appears to be more sensitive to selection than ∂a∂i. Overall, the presence of positive selection seems to cause both methods to overestimate the extent of population contraction, and to underestimate the degree of recovery from the bottleneck. For the simulated datasets used in these analyses, whenever a locus was linked to a selective sweep, the distance from the sweep, c/s was drawn uniformly between 0 and 1. We also repeated these analyses when fixing the value of c/s, and in supplementary fig. S1 we show our distribution of parameter estimates obtained using both ∂a∂i and ABC on 111 different combinations of f and c/s. This figure demonstrates that, for a given fraction of neutral loci linked to a selective sweep, increasing the proximity to the sweep increases bias, as expected.

Note that for our ABC inference we examined only the means of several population genetic summary statistics (Methods). Including the variances caused estimates to behave non-monotonically, because whenever f is not equal to one or zero the distribution of summary statistic values is a mixture of two models, and therefore has inflated variance, resulting in less accurate parameter estimation. We also show our parameter estimates when including variances in supplementary fig. S1.

Population growth

Next, we examined the impact of positive selection on parameter estimates for our model of population growth (fig 1C). When our simulated genomes experienced no recent selective sweeps, we again achieved good accuracy when using ∂a∂i (fig. 3): our mean estimates of Ne_A,

T_G, and Ne_0 were 1,040 (0.8% difference from true value), 955 (3.8% difference), and 36,610 (2.0% difference), respectively. Increasing f again biases our estimates, but the effect is subtler than for the bottleneck case. This is probably a consequence of the reduced scale of the impact of positive selection on flanking variation under this model relative to the bottleneck model (fig 1). The most notable pattern that we observe for this model is that T_G decreases with increasing f, while the population size estimates are largely unaffected: when f=0.5 our average estimate is 905 (1.6% difference from true value), versus 872 when f=0.8 (5.3% difference), and 855 when f=1.0 (7.1% difference). In other words, widespread selective sweeps will cause one to infer slightly more recent but more pronounced exponential growth. When c/s is relatively small, our error rates are substantially higher (supplementary fig. S2). Thus, stronger positive selection could still seriously impair ∂a∂i's demographic inferences under this population size history.

We then used ABC to perform parameter estimation under the growth model. In the neutral case, our estimated parameters were largely concordant with the true values, with the exception of some bias observed for T_G (mean estimate of 801, which is 13% below the true value). Our estimates of Ne_A were also far more dispersed than those obtained from ∂a∂i (fig. 3). Further, unlike our estimates with ∂a∂i, increasing the value of f substantially biases our ABC estimates. For example, Ne_0 is 39,497 when f=0 (10% greater than the true value), but increases to 42,851 when f=0.5 (an overestimate of 19%), 49,030 when f=0.8 (an overestimate of 37%), and 62,889 when f=1.0 (an overestimate of 75% plus a dramatic increase in variance). The degree to which T_G is underestimated also increases with f: the average estimate is 769 at f=0.2 (16% below the true value), 717 at f=0.5 (22% bias), 667 at f=0.8 (27.5% bias), and 623 at f=1.0 (32.2% bias). Again, we demonstrate the effect of varying the distance c/s of sampled loci from the selective sweep, as well as the effect of performing ABC on the variances of summary

statistics in addition to their means, in supplementary fig. S2. Overall, we observe that under our growth model positive selection will cause inferences of more recent, faster population growth, with this effect being far subtler when using ∂a∂i than ABC with our set of summary statistics.

Population contraction followed by growth

Finally, we assessed our ability to recover the parameters of our contraction-then-growth model (fig. 1D) with increasing amounts of positive selection. Without selection, ∂a∂i estimates Ne_A, T_G, and Ne_0 with reasonably high accuracy (fig. 4): 14,773 on average for Ne_A (2.1% over the true value), 862 for T_G (6.3% under the true value), and 37,999 for Ne_0 (5.8% over the true value). However, T_C and Ne_C are substantially overestimated at 2,530 (24.0% over the true value) and 1,350 on average (30.8% over the true value). As we increase *f*, our estimates of Ne_A, T_C, and Ne_C are inflated, T_G is increasingly underestimated, and Ne_0 is largely unaffected. The effect on T_C is the largest resulting in a seemingly linear increase with *f*: our estimate is 2,886 when f=0.2 (an overestimate of 41.47%), 3,712 when f=0.5 (overestimate of 82.0%), 5,329 when f=0.8 (overestimate of 161.2%), and 7,082 when f=1.0 (overestimate of 247.2%). Ne_A and Ne_C increase more slowly: to 29,069 (an overestimate of 100.8%) and 2183 (an overestimate of 111.6%) when f=1.0, respectively, while T_G on the other hand decreases to 642 when f=1.0 (underestimate of 30.2%). Thus, positive selection typically results in our ∂a∂i-estimated demographic model to have more protracted population contraction, with larger initial and contracted population sizes. Results for varying values of *c*/*s* are shown in supplementary fig. S3.

When repeating these analyses using ABC given our set of summary statistics (fig. 4), we find that under neutrality T_G is grossly underestimated, Ne_C is slightly overestimated, and Ne_A and Ne_0 are estimated with greater accuracy (14,435 or 0.27% under the true value, and 34,373,

or 4.3% under the true value, respectively). Thus, we infer a more protracted but slightly less severe population contraction than the true population size history. Even so we proceed to characterize what the effect of linked selection on parameter estimates using ABC as before. Indeed our estimates become more biased as we add increasing amounts of positive selection. Most notably, Ne_0 exhibits a substantial downward bias as we increase f, and is estimated at 31,970 when f=0.2 (11% underestimate), 28,105 when f=0.5 (21.7% underestimate), 25,829 when f=0.8 (28% underestimate), and 24,735 when f=1.0 (31.1% underestimate). Also, as f becomes large Ne_C shifts from being slightly overestimated to significantly underestimated, and estimates of Ne_A become slightly upwardly biased. Thus we find that under this model positive selection again biases parameter estimates, though not in the same manner for ∂a∂i and ABC: while ∂a∂i infers a longer phase of reduced population size along with an inflated ancestral size and less severe contraction, our ABC procedure infers a more severe contraction followed by a less complete recovery. We show our inference results on the full grid of c/s and f values, as well as when including variances of summary statistics, in supplementary fig. S3.

Effect of positive selection on population size history inference using PSMC

The pairwise sequentially Markovian coalescent (PSMC) is a widely used method that infers a discretized history of population size changes from a single recombining diploid genome (Li and Durbin 2011). Such inference is possible because coalescence times between the two allelic copies in a diploid, which are governed by the effective population size, will change at the breakpoints of historical recombination events, and the resulting distribution of coalescence times across the genome thus contains information about population size history. However this method necessitates sampling a large stretch of a recombining chromosome. In order to test the

impact of positive selection on inferences from PSMC, we simulated constant-size populations of 10,000 individuals, sampling a 15 Mb chromosomal region from two haploid individuals. We performed 100 replicates of this simulation for each of four scenarios (Methods): the standard neutral model; a population experiencing one fairly recent sweep (reaching fixation 0.2×*Ne* generations ago) somewhere in this region; a population experiencing three recurrent sweeps (fixed 0, 0.2, and 0.4×*Ne* generations ago); and a population experiencing five sweeps (0, 0.1, 0.2, 0.3, and 0.4×*Ne* generations ago). We find that under neutrality, very little population size change is inferred on average (though there is a fair bit of variance; fig. 5A). However, when there has been only a single selective sweep, a population bottleneck near the time of the sweep is inferred, in which the population contracts to approximately two-thirds of its original size before recovering (fig. 5B). When there have been three or five recurrent selective sweeps the inferred population contraction becomes increasingly severe (fig. 5C-D). In the five-sweep case, we typically infer a contraction down to roughly one-fourth of the original size, often with no subsequent recovery in (perhaps because these scenarios involve a very recent sweep; Methods). Thus, we find that positive selection can dramatically skew population size histories deduced by PSMC.

Positive selection produces spurious support for non-equilibrium demographic histories

Demographic inference methods are often used not only to infer parameters of a model, but increasingly to select the best fitting among several competing models (e.g. Adams and Hudson 2004; Fagundes et al. 2007; Duchen et al. 2013). To ask whether positive selection might affect the outcome of demographic model selection, we simulated genomes with constant population size, again sampling loci for which some fraction, *f*, is located within $c/s \leq 1$ of a selective

sweep. We then performed model selection among our four demographic histories (fig. 1) using both ∂a∂i and ABC (Methods).

Prior to performing model selection with ∂a∂i, we first examined the degree of support for each model when fit to each dataset using the Akaike information criterion (AIC). Examining the differences in AIC between models, we found that even a moderate number of selective sweeps will cause non-equilibrium demographic scenarios to have far stronger support than the true equilibrium history (supplementary fig. S4). This is especially so for the bottleneck and contraction-then-growth models, which achieve better support than the equilibrium model even at small values of f. For example, when f=0.2 the bottleneck model receives an AIC lower than the equilibrium model in 90% of cases, and the contraction-then-growth model has a lower AIC 72% of the time (supplementary fig. S4). By contrast, the pure growth model is supported to a lesser extent (a lower AIC in 54% of cases), and occasionally failed to optimize properly, settling on a very low-likelihood parameterization—an indication of a poorly fitting model. The better fit of the bottleneck and contraction-then-growth models is likely because they better model the genealogy of a region experiencing a selective sweep: much of the ancestral variation flanking the selected site is removed during the sweep (analogous to contraction), while being replaced by the subset of alleles within the rapidly expanding class of individuals containing the selected mutation (analogous to expansion).

We conducted formal model selection as described in the Methods, conservatively selecting the equilibrium model unless one of the other models had an AIC at least 50 units higher. We note that it would be preferable to perform parametric bootstraps from competing models to compare the distributions of AIC values, but in the interest of computational efficiency we instead choose this heuristic. Even with this conservative cutoff, we selected a non-

equilibrium model for 15% of simulated data sets with *f*=0.2, for 47% of datasets with *f*=0.3, and for 91% of datasets when *f*=0.6 (Table 1). Thus even if a minority of loci are linked to a recent selective sweep then SFS-likelihood based approaches may prefer the wrong demographic model. Interestingly, in every case where a non-equilibrium model was the unambiguous best fit to the data this model choice was the bottleneck scenario.

Next, we performed model selection on our constant-size population samples using ABC (Methods). For each of these datasets, we estimated Bayes factors for each pairwise comparison of demographic models. Again, we find that non-equilibrium demographic models may begin to receive stronger support than the constant-size model once a sizable fraction of loci are linked to selective sweeps. For example, when *f*=0.4, the bottleneck model has nominally stronger support (Bayes factor >1) than the equilibrium model for 55% of datasets, the growth model has stronger support than equilibrium in 9% of datasets, and the contraction-then-growth model has stronger support in 4% of datasets (supplementary fig. S5). When *f* is increased to 0.8, we observe even stronger support for non-equilibrium models, with 100% of the bottlenecks datasets, 79% of the contraction-the-growth datasets, and 26% of the growth model datasets having a Bayes factor > 1 when compared to the constant-size model. We used these Bayes factors to perform model selection in a manner similar to our analysis with ∂a∂i, conservatively selecting the equilibrium model if there was no alternative model that was a significantly better fit to the data (i.e. having a Bayes factor relative to the equilibrium model of ≥20). Again, we find that even if a minority of loci are linked to a sweep, then there is a substantial probability that the constant-size model will not be selected: for 6% of datasets we select a non-equilibrium model when *f*=0.2, for 34% of datasets with *f*=0.3, and for 99% of datasets when *f*=0.6 (Table 2). As with ∂a∂i-optimized models, we found that in every instance where we were able to unambiguously select a single

non-equilibrium demographic history as the best fit we chose the bottleneck model. When we include the variances of our set of summary statistics in our ABC procedure, we find that non-equilibrium models are strongly supported in an even higher proportion of simulated data sets, though in this case we typically select the contraction-then-growth model rather than the bottleneck model (supplementary table S1).

DISCUSSION

It is well known that natural selection profoundly affects genealogies and therefore patterns of genetic polymorphism (Kaplan et al. 1989; Hudson and Kaplan 1994), thus it is reasonable to expect that linked selection will bias demographic inference that assumes strict neutrality of population genomic data. Indeed, background selection has recently been shown to skew demographic inferences using the site frequency spectrum (Ewing and Jensen 2016). Here, we show through extensive simulation that positive selection can severely impair demographic model selection and parameter estimation based on the SFS, summary statistics of variation, and reduced approximations of the ancestral recombination graph (i.e. PSMC). The extent to which this is so depends on the fraction of genetic loci examined during inference that are affected by a recent sweep, and the ratio of the genetic distance between the locus and the target of selection to the selection coefficient (c/s).

When the fraction of loci affected by linked selection is low, we have shown that point estimates of population parameters estimated under the correct demographic scenario are reasonably accurate using both SFS-based inference and ABC with summary statistics (figs. 2-4)—the exact fraction, however, depends on the model in question. However, unless f is quite

low our results indicate that when model selection is applied using either SFS-based or ABC inference, linked selection can bias model choice (Tables 1 and 2). In many of our simulated datasets we have assumed that loci linked to sweeps are on average a *c*/*s* distance of 0.5 away from the sweep (i.e. drawn uniformly from between 0 and 1). In real genomes this may correspond to quite a large physical distance: for instance if we assume a selection coefficient of 0.05 (i.e. selection as strong as in our simulations) and a crossover rate of 2 cM/Mb (similar to estimates in Drosophila; e.g. Comeron et al. 2012) this corresponds to a physical distance of 1.25 Mb. If instead we assume a crossover rate of 1 cM/Mb (similar to estimates from humans; e.g. Kong et al. 2010), this corresponds to a physical distance of 2.5 Mb. While we have assumed a fairly high value of *s* that may not be representative of all selective sweeps, known sweeps in humans may often have selection coefficients fairly close to 0.05 (Peter et al. 2012).

Thus, even if there are a small number of recent selective sweeps, the majority of the genome may nonetheless be sufficiently impacted by linked selection to produce biased demographic inferences. For example, if a human population experienced 1,200 recent sweeps fairly evenly spaced across the genome (the equivalent of one recent sweep in ~5% of genes), every site in the genome would be within a *c*/*s* distance of 0.5 from the nearest sweep (i.e. f = 1.0). In Drosophila, this would be the case if there were 120 evenly spaced recent sweeps. This is a very small number indeed, equivalent to a single recent sweep affecting <1% of all genes. Indeed in Drosophila the fraction of loci affected by recent positive selection may be quite large (Begun et al. 2007; Langley et al. 2012). Numerous studies have estimated that the fraction of adaptive amino acid substitutions in *D. melanogaster* is considerable, with estimates ranging from 10-50% (Smith and Eyre-Walker 2002; Bierne and Eyre-Walker 2004; Langley et al. 2012; Mackay et al. 2012). Positive selection may therefore be particularly troublesome for

demographic inference in Drosophila and other organisms were adaptive natural selection is similarly pervasive. In humans, where positive selection is perhaps less common (Hernandez et al. 2011), this may be less of a thorny issue. However, some have argued that selection may be pervasive in the human genome as well (Boyko et al. 2008; Enard et al. 2014) , and certainly humans show many adaptations to local environments (e.g. Li et al. 2007; Perry et al. 2007; Tishkoff et al. 2007; Barreiro et al. 2008; Bryk et al. 2008). Nonetheless, the 10-fold larger number of sweeps (under our parameterization) required to produce the same level of bias suggests that the confounding effect of demography in humans may be less problematic than in Drosophila. However, given uncertainty in the number, location, strength, and type (see below) of selective sweeps we are unable to quantify the extent to which demographic inferences in either species are skewed by adaptation.

A new and promising class of methods for inferring demographic histories rely on estimating approximations to the ancestral recombination graph (ARG) using a sequentially Markovian coalescent (Li and Durbin 2011; Schiffels and Durbin 2014). In particular, because PSMC is applied to a single diploid genome, it has been used to infer population size histories in numerous species for which one or more genome sequences are available, in many cases finding support for large changes in population size (Groenen et al. 2012; Albert et al. 2013; Prado-Martinez et al. 2013; Zhan et al. 2013; Freedman et al. 2014; Green et al. 2014; Wallberg et al. 2014; Auton et al. 2015; Lamichhaney et al. 2015). Our findings suggest that natural selection may alter the shape of, and inflate the degree of change in, these inferred histories. Indeed because of the specific way in which a sweep perturbs the ARG locally during the coalescent history of a chromosome, PSMC inference on regions that have experienced one or a few sweeps in the past may lead to erroneous estimation of a population bottleneck. If sweeps continue until

the present day, PSMC inference might appear to support a population contraction rather than a bottleneck, though this may very well be a result of PSMC having lower power for very recent population dynamics (Li and Durbin 2011). Under a truly recurrent sweep model (Stephan et al. 1992) it is unclear what the behavior of inference using PSMC might be. Note that with PSMC we infer population size changes from a large simulated chromosomal segment—corresponding to ~15 Mb in the human genome—experiencing only a single selective sweep $0.2\times Ne$ generations ago. This is equivalent a total of ~200 fairly recent sweeps across the human genome.

Our results are broadly concordant with a recent examination of the ability of the McDonald–Kreitman (or MK) test (McDonald and Kreitman 1991) to infer the fraction of substitutions that were adaptive (α) under a simulated recurrent hitchhiking scenario with constant population size (Messer and Petrov 2013). Their study found that Eyre-Walker and Keightley's DFE-alpha method (Eyre-Walker and Keightley 2009), which simultaneously estimates α, the distribution of fitness effects, and a two-epoch population size history, incorrectly inferred the presence of population size changes (Messer and Petrov 2013). It is therefore reasonable to assume that positive selection could have a substantial confounding effect on a variety of population genomic methods for demographic inference in practice, beyond those considered here. In the empirical literature, numerous recent studies of demographic history have found support for contractions and recent expansions of natural populations (Thornton and Andolfatto 2006; Fagundes et al. 2007; Gravel et al. 2011; Tennessen et al. 2012; Duchen et al. 2013). While such population size changes are probably common, and our results do not call the major findings of these studies into question, they do suggest that natural selection exaggerates the inferred intensity of these changes.

Our study has examined only a single model of adaptive natural selection, and therefore has several limitations. Throughout we have assumed that positive selection occurs only through completed hard selective sweeps. Indeed soft sweeps (Innan and Kim 2004; Hermisson and Pennings 2005; Pennings and Hermisson 2006; Garud et al. 2015) and partial sweeps (Hudson et al. 1994; Sabeti et al. 2002; Voight et al. 2006), may be widespread, and differ in their effects on linked polymorphism (Orr and Betancourt 2001; Meiklejohn et al. 2004; Przeworski et al. 2005; Teshima et al. 2006; Schrider et al. 2015; Vy and Kim 2015). Polygenic selection, in which alleles at several different loci underlying a trait under selection will experience a change in frequency, is also thought to be widespread (Pritchard et al. 2010; Berg and Coop 2014). Such polygenic adaptation is known to leave its own unique signature on patterns of population genetic variation (Berg and Coop 2014). These alternative modes of positive selection could skew demographic inferences in a different manner than what we have observed in this study. Positive selection may also affect estimation of multi-population demographic scenarios: though we did not examine this here, Mathew and Jensen recently showed that selective sweeps will impair parameter estimates for a two-population isolation-with-migration model (Mathew and Jensen 2015). Thus our results, combined with those of Mathew and Jenson (2015), Ewing and Jensen (2016), and Messer and Petrov (2013), strongly suggest that the problem of natural selection skewing demographic inference is a general one.

The observations we have made here also suggest some steps that can be taken to mitigate the impact of positive selection. First, we note that in general ∂a∂i (i.e. SFS-based inference) appears to be somewhat more robust to selection than our ABC approach based on summary statistics. Perhaps this is because ∂a∂i uses an SFS summed across loci, such that more polymorphic regions will have a greater weight on the shape of the SFS (simply because they

contribute more observations). Thus the extent to which regions most affected by sweeps contribute to the SFS is diminished implicitly, as these regions will exhibit less variation. Relying on the SFS rather than summaries of variation that to a greater extent depend on the number of segregating sites may therefore reduce selection's confounding effect on inferred relative population size changes, though estimates of $4N\mu$ and therefore the absolute population size may be biased. We also found that including variances of summary statistics when performing ABC can dramatically inflate error, especially when an intermediate number of loci are linked to sweeps, perhaps because this mixture of two evolutionary models (neutrality and positive selection) inflates the variance. Omitting variances may therefore reduce the confounding effect of selection in some cases.

Finally, we have shown convincingly that the proximity of selective sweeps to genomic regions used for inference (as measured by c/s) has a large effect on the magnitude of bias (supplementary figs. 1-3). Thus, it is of paramount importance to select regions located as far away in genetic distance as possible from genes and other functional DNA elements (Gazave et al. 2014). While this is so, it may not be possible to move far enough away from potential targets of selection to completely eliminate any bias (as discussed above). Moreover, it is essential to omit regions with lowered recombination rates, where the impact of linked selection will be strongest (Begun and Aquadro 1992). Our results also motivate the challenging task of simultaneous estimation of parameters related to natural selection and demographic history (Eyre-Walker and Keightley 2009; Mathew and Jensen 2015; Sheehan and Song 2016). Until an approach to obtain accurate estimates of demographic parameters in the face of natural selection is devised, population size histories inferred from population genetic datasets will remain significantly biased.

METHODS

Simulating demographic and selective histories to test inference methods

To test the robustness of ∂a∂i and ABC to positive selection, we generated coalescent simulations from four different demographic scenarios: 1) a constant population size model; 2) a three-epoch population bottleneck (the European model from Marth et al. 2004); 3) a model of recent exponential population growth; 4) and a three-epoch model with a population contraction followed by a period of stasis and then recent exponential population growth (a simplified version of the European model from Gravel et al. 2011). We refer to this scenario as the contraction-then-growth model. These models and their parameters are shown in fig. 1.

For each demographic model, we simulated 100 observed genomes experiencing no natural selection, each of which was summarized by a collection of 500 unlinked loci sampled from 200 haploid individuals. We then repeated these simulations while stipulating that a specified fraction of loci (f) were linked to a recent selective sweep where the selected mutation reached fixation immediately prior to sampling. The selection coefficient, s, for this mutation was always set to 0.05, with a completely additive fitness effect (h=0.5). For each simulation including a selective sweep, we specified the genetic distance of the sweep from the sampled locus by the ratio c/s, where c is the crossover rate per base pair multiplied by the physical distance to the sweep, and s is again the selection coefficient. We examined values of f that were multiples of 0.1 between 0.1 (10% of loci linked to a sweep) and 1.0 (100% of loci linked to a sweep). Values of c/s examined were multiples of 0.1, raging from 0.0 (the sweep occurred immediately adjacent to the locus being used for inference) to 1.0 (~4.17 Mb given our value of s

and our recombination rate; see below). We generated sets of simulations with a given value of *f* by combining the appropriate numbers of neutral simulations and simulated loci linked to a sweep. For each combination of *f* and *c*/*s* (110 combinations in total), we generated 100 sets of 500 unlinked loci.

We also simulated large chromosomal regions to which we applied PSMC. These simulations were of 15 Mb regions with a constant-size population (*Ne*=10,000), from which two individuals were sampled. These 15 Mb regions either experienced no selective sweeps, one selective sweep fixing 0.4×*Ne* generations ago, three selective sweeps (fixing 0.4×*Ne* generations ago, 0.2×*Ne* generations ago, and immediately prior to sampling) or five selective sweeps (0.4×*Ne*, 0.3×*Ne*, 0.2×*Ne*, 0.1×*Ne*, or 0 generations prior to sampling). The location of each sweep was thrown down randomly along the chromosome. For each scenario, 100 replicate simulations were generated.

For all simulations we used parameters relevant to human populations: a recombination rate of 1.0×10^{-8}, (approximately equal to the sex-averaged rate from Kong et al. 2010) and a mutation rate of 1.2×10^{-8} (from Kong et al. 2012). Simulations were performed with our coalescent simulator discoal_multipop (https://github.com/kern-lab/discoal_multipop), and example command lines with the appropriate population mutation/recombination rates and population size changes for each demographic scenario are shown in supplementary table S2.

Parameter estimation and model selection with ∂a∂i

We downloaded version 1.6.3 of ∂a∂i (Gutenkunst et al. 2009), which we programmed to optimize the parameters of the bottleneck, growth, and contraction-then-growth models. For each model we used a two-step constrained optimization procedure to find the combination of

demographic parameters that have the highest likelihood given the site frequency spectrum measured across all 500 unlinked loci in the simulated genome. First we performed a coarse optimization using the Augmented Lagrangian Particle Swarm Optimizer (Jansen and Perez 2011), and then refined this solution using Sequential Least Squares Quadratic Programming (Kraft 1988). Both of these techniques are implemented in the pyOpt package (version 1.2.0) for optimization in python (Perez et al. 2012).

To asses the accuracy of point estimation of parameters in the face of varying amounts of and genetic distances to selective sweeps, we optimized the parameters of each demographic model against each data set simulated under that model, comparing estimated values to the true values. As shown in the Results, this approach was quite successful recovering the true parameter values of each demographic model when applied to data simulated under neutrality. However, one exception was the bottleneck model, for which the optimal solution was typically a shorter but more severe bottleneck than the one we had simulated. We therefore fixed the bottleneck duration to the true value (500 generations), after which ∂a∂i was able to estimate the remaining parameter values with acceptable accuracy.

To assess the support for a given demographic model, we obtained for a simulated data set the likelihood of each demographic model under the optimal parameters estimated by ∂a∂i, and then from this likelihood and the number of parameters of the model calculated the AIC (Akaike 1974). For the constant population size model, there are no optimized parameters, so the AIC is simply minus 2 times the log-likelihood of the model. Model selection was performed for each data set simulated with constant population size, with or without selection. For each model with variable population size, the python script we used to perform parameter optimization and obtain the likelihood of the optimal parameterization has been deposited on GitHub

(https://github.com/kern-lab/demogPosSelDadiScripts), as has the script used to obtain the likelihood under the constant population size model.

To perform formal model selection, we asked for a given simulated data set whether any non-equilibrium model had an AIC at least 50 units greater than that of the equilibrium model. If so, we asked whether any of our three non-equilibrium models had an AIC at least 50 units than the other two, in which case we selected that model; otherwise we classified the simulated data set as "ambiguous but non-equilibrium." If no non-equilibrium model had an AIC at least 50 units greater than the equilibrium model, then we conservatively classified the simulated data set as "equilibrium."

Parameter estimation and model selection using ABC

For each of our three non-equilibrium demographic models, we used ABC to estimate the model parameters in each "observed genome" simulated under the model. To this end, we summarized patterns of variation within each genome by calculating the means and variances of π, the number of segregating sites, Tajima's D, $\hat{\theta}_H$ (Fay and Wu 2000), and the number of distinct haplotypes for each of the 500 sampled unlinked loci. We then created a sampling dataset for each demographic model consisting of 5.0×10^5 simulated genomes evolving in the absence of selection, again with each genome represented by 500 unlinked loci, each of sample size 200. For these sampling datasets, the demographic parameters were drawn uniformly from prior distributions (shown for each model in supplementary table S3), and were summarized by the same set of statistics used for the "observed genomes." For the bottleneck and contraction-then-growth models, our initial efforts to estimate parameters under neutrality failed to approximate the true parameterizations. We therefore fixed the values of the times of population contraction

parameters of these models (referred to as T_B and T_C respectively) during both parameter estimation and model selection: these parameters were always set to the true values during sampling simulations, and their values were not estimated. After this change, we were able to estimate the parameters of each model under neutrality with reasonable accuracy.

We utilized the ABCreg software package (https://github.com/molpopgen/ABCreg) to perform parameter estimation for each of our four demographic models (Thornton 2009). We applied the conventional tangent transformation procedure to the parameters sampled from our prior distributions via passing the **–T** flag. We also set the tolerance parameter, **-t = 0.001**, thus retaining 0.1% of our sample data for use in estimating the posterior parameter distributions. ABCreg uses a weighted linear regression approach to the retained sampling simulations in order to estimate the posterior probability densities of the parameters (Beaumont et al. 2002). From each parameter's estimated posterior distribution we used the maximum *a posteriori* estimator (posterior mode) as our point estimate. We also repeated parameter estimation on each observed dataset using the means of summary statistics only.

We performed model selection on our simulated observed genomes with constant population size and varying degrees of positive selection. Our model selection procedure considered each of our four demographic models. For the constant-size model we constructed a new sampling set of 5.0×10^5 simulated datasets, and for the three variable-size models we used the same sampling sets generated for parameter estimation. We used the R package *abc* to conduct model choice, performing logistic regression-based estimation of the posterior probabilities of a model (Csilléry et al. 2012). We separately examined all pair-wise model selection scenarios: constant size vs. growth; constant size vs. contraction-then-growth; constant size vs. bottleneck; bottleneck vs. contraction-then-growth; bottleneck vs. growth demography;

and contraction-then-growth vs. growth. We summarized model support with the Bayes factor as calculated by the *abc* package; in cases where this calculation resulted in division by zero (i.e. where the posterior probability of a model was estimated to be zero), we set the Bayes factor to be equal to the largest one observed among datasets with the same combination of demographic model, *f*, and *c*/*s*.

We performed model selection in a manner similar to our approach with ∂a∂i, asking for a given simulated data set whether any non-equilibrium model had a Bayes factor of at least 20 when compared to the equilibrium model. If so, we asked whether any of our three non-equilibrium models had a Bayes factor ≥20 when compared to each of the two, in which we selected that model; otherwise we classified the simulated data set as "ambiguous but non-equilibrium." Analogous to our model selection with ∂a∂i, if no non-equilibrium model had a Bayes factor ≥20 when compared to the equilibrium model, then we conservatively classified the simulated data set as "equilibrium."

Inferring population size histories with PSMC

We ran PSMC in order to infer the history of population size changes of our 15 Mb simulations from which two individuals were sampled (see above). Briefly, we converted our simulation output to the same format generated by running msHOT-lite (https://github.com/lh3/foreign/tree/master/msHOT-lite) with the –l flag. We then ran PSMC's ms2psmcfa.pl script with default parameters to generate input for PSMC, which we ran with default parameters. Finally, we ran PSMC's psmc2history.pl script with default parameters to output the inferred population size history. We then rescaled the output from units of *Ne* to years (after rescaling to generations and assuming a 30 year generation time) and numbers of

individuals using the estimated value of θ. For each selective scenario, we ran PSMC separately on all 100 simulated population samples. Finally, for the purposes of visualization we obtained a median estimate of population size across time by examining a large number of time points (one every 100 years) across the entire period examined, and at each time point taking the median population size estimate from all 100 simulations.

ACKNOWLEDGMENTS

We thank Jody Hey and Matthew Hahn for feedback on the manuscript. D.R.S. was partially supported by the National Institutes of Health under Ruth L. Kirschstein National Research Service Award F32 GM105231. A.D.K. was supported in part by NIH award no. R01GM078204.

REFERENCES

Adams AM and Hudson RR. 2004. Maximum-likelihood estimation of demographic parameters using the frequency spectrum of unlinked single-nucleotide polymorphisms. *Genetics* 168: 1699-1712.

Akaike H. 1974. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on* 19: 716-723.

Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA and Kruglyak L. 2004. Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biol* 2: e286.

Albert VA, Barbazuk WB, Der JP, et al. 2013. The Amborella genome and the evolution of flowering plants. *Science* 342: 1241089.

Auton A, Brooks LD, Durbin RM, et al. 2015. A global reference for human genetic variation. *Nature* 526: 68-74.

Barreiro LB, Laval G, Quach H, Patin E and Quintana-Murci L. 2008. Natural selection has driven population differentiation in modern humans. *Nat Genet* 40: 340-345.

Beaumont MA, Zhang W and Balding DJ. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162: 2025-2035.

Begun DJ and Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* 356: 519-520.

Begun DJ and Aquadro CF. 1993. African and North American populations of Drosophila melanogaster are very different at the DNA level.

Begun DJ, Holloway AK, Stevens K, et al. 2007. Population genomics: whole-genome analysis of polymorphism and divergence in Drosophila simulans. *PLoS Biol* 5: e310.

Berg JJ and Coop G. 2014. A population genetic signal of polygenic adaptation. *PLoS Genet* 10: e1004412.

Bergland AO, Tobler R, González J, Schmidt P and Petrov D. 2015. Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in Drosophila melanogaster. *Mol Ecol*: doi: 10.1111/mec.13455.

Bierne N and Eyre-Walker A. 2004. The genomic rate of adaptive amino acid substitution in Drosophila. *Mol Biol Evol* 21: 1350-1360.

Boyko AR, Williamson SH, Indap AR, et al. 2008. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet* 4: e1000083.

Bryk J, Hardouin E, Pugach I, Hughes D, Strotmann R, Stoneking M and Myles S. 2008. Positive selection in East Asians for an EDAR allele that enhances NF-κB activation. *PLoS ONE* 3: e2209.

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

Corbett-Detig RB, Hartl DL and Sackton TB. 2015. Natural selection constrains neutral diversity across a wide range of species. *PLoS Biol* 13: e1002112.

Coventry A, Bull-Otterson LM, Liu X, et al. 2010. Deep resequencing reveals excess rare recent variants consistent with explosive population growth. *Nature Communications* 1: 131.

Csilléry K, François O and Blum MG. 2012. abc: an R package for approximate Bayesian computation (ABC). *Methods in ecology and evolution* 3: 475-479.

Duchen P, Živković D, Hutter S, Stephan W and Laurent S. 2013. Demographic inference reveals African and European admixture in the North American Drosophila melanogaster population. *Genetics* 193: 291-301.

Enard D, Messer PW and Petrov DA. 2014. Genome-wide signals of positive selection in human evolution. *Genome Res*.

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

Excoffier L, Estoup A and Cornuet J-M. 2005. Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics* 169: 1727-1738.

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

Fagundes NJ, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL and Excoffier L. 2007. Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences* 104: 17614-17619.

Fay JC and Wu C-I. 2000. Hitchhiking under positive Darwinian selection. *Genetics* 155: 1405-1413.

Freedman AH, Gronau I, Schweizer RM, et al. 2014. Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet* 10: e1004016.

Fu Y-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147: 915-925.

Gao F and Keinan A. 2016. Inference of super-exponential human population growth via efficient computation of the site frequency spectrum for generalized models. *Genetics* 202: 235-245.

Garrick RC, Kajdacsi B, Russello MA, Benavides E, Hyseni C, Gibbs JP, Tapia W and Caccone A. 2015. Naturally rare versus newly rare: demographic inferences on two timescales inform conservation of Galápagos giant tortoises. *Ecology and evolution* 5: 676-694.

Garud NR, Messer PW, Buzbas EO and Petrov DA. 2015. Recent selective sweeps in North American Drosophila melanogaster show signatures of soft sweeps. *PLoS Genet* 11: e1005004.

Gazave E, Ma L, Chang D, et al. 2014. Neutral genomic regions refine models of recent rapid human population growth. *Proceedings of the National Academy of Sciences* 111: 757-762.

Goebel T, Waters MR and O'Rourke DH. 2008. The late Pleistocene dispersal of modern humans in the Americas. *Science* 319: 1497-1502.

Gravel S, Henn BM, Gutenkunst RN, et al. 2011. Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences* 108: 11983-11988.

Green RE, Braun EL, Armstrong J, et al. 2014. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science* 346: 1254449.

Groenen MA, Archibald AL, Uenishi H, et al. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491: 393-398.

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

Hahn MW. 2008. Toward a selection theory of molecular evolution. *Evolution* 62: 255-265.

Hájková P, Pertoldi C, Zemanová B, Roche K, Hájek B, Bryja J and Zima J. 2007. Genetic structure and evidence for recent population decline in Eurasian otter populations in the Czech and Slovak Republics: implications for conservation. *J Zool* 272: 1-9.

Hermisson J and Pennings PS. 2005. Soft sweeps molecular population genetics of adaptation from standing genetic variation. *Genetics* 169: 2335-2352.

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

Hudson RR, Bailey K, Skarecky D, Kwiatowski J and Ayala FJ. 1994. Evidence for positive selection in the superoxide dismutase (Sod) region of *Drosophila melanogaster*. *Genetics* 136: 1329-1340.

Hudson RR and Kaplan NL. 1994. Gene trees with background selection. In. Non-Neutral Evolution: Springer. p. 140-153.

Innan H and Kim Y. 2004. Pattern of polymorphism after strong artificial selection in a domestication event. *Proc Natl Acad Sci U S A* 101: 10667-10672.

Jansen PW and Perez RE. 2011. Constrained structural design optimization via a parallel augmented Lagrangian particle swarm optimization approach. *Computers & Structures* 89: 1352-1366.

Jensen JD, Kim Y, DuMont VB, Aquadro CF and Bustamante CD. 2005. Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics* 170: 1401-1410.

Kao JY, Zubair A, Salomon MP, Nuzhdin SV and Campo D. 2015. Population genomic analysis uncovers African and European admixture in *Drosophila melanogaster* populations from the south-eastern United States and Caribbean Islands. *Mol Ecol* 24: 1499-1509.

Kaplan NL, Hudson R and Langley C. 1989. The" hitchhiking effect" revisited. *Genetics* 123: 887-899.

Keinan A, Mullikin JC, Patterson N and Reich D. 2007. Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nat Genet* 39: 1251-1255.

Kong A, Frigge ML, Masson G, et al. 2012. Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature* 488: 471-475.

Kong A, Thorleifsson G, Gudbjartsson DF, et al. 2010. Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* 467: 1099-1103.

Kraft D. 1988. A software package for sequential quadratic programming: DFVLR Obersfaffeuhofen, Germany.

Lamichhaney S, Berglund J, Almén MS, et al. 2015. Evolution of Darwin/'s finches and their beaks revealed by genome sequencing. *Nature* 518: 371-375.

Langley CH, Stevens K, Cardeno C, et al. 2012. Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics* 192: 533-598.

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

Li H, Mukherjee N, Soundararajan U, et al. 2007. Geographically separate increases in the frequency of the derived ADH1B* 47His allele in eastern and western Asia. *The American Journal of Human Genetics* 81: 842-846.

Li H and Stephan W. 2006. Inferring the demographic history and rate of adaptive substitution in Drosophila. *PLoS Genet* 2: e166.

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

Mackay TF, Richards S, Stone EA, et al. 2012. The Drosophila melanogaster genetic reference panel. *Nature* 482: 173-178.

Marjoram P, Molitor J, Plagnol V and Tavaré S. 2003. Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences* 100: 15324-15328.

Marjoram P and Wall JD. 2006. Fast" coalescent" simulation. *BMC genetics* 7: 1.

Marth G, Schuler G, Yeh R, et al. 2003. Sequence variations in the public human genome data reflect a bottlenecked population history. *Proceedings of the National Academy of Sciences* 100: 376-381.

Marth GT, Czabarka E, Murvai J and Sherry ST. 2004. The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics* 166: 351-372.

Martínez-Cortés G, Salazar-Flores J, Fernández-Rodríguez LG, Rubi-Castellanos R, Rodríguez-Loya C, Velarde-Félix JS, Muñoz-Valle JF, Parra-Rojas I and Rangel-Villalobos H. 2012. Admixture and population structure in Mexican-Mestizos based on paternal lineages. *J Hum Genet* 57: 568-574.

Maruyama T and Fuerst PA. 1985. Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics* 111: 675-689.

Mathew LA and Jensen JD. 2015. Evaluating the ability of the pairwise joint site frequency spectrum to co-estimate selection and demography. *Frontiers in genetics* 6.

Maynard Smith J and Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res* 23: 23-35.

McDonald JH and Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* 351: 652-654.

McVean GA and Cardin NJ. 2005. Approximating the coalescent with recombination. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 360: 1387-1393.

Meiklejohn CD, Kim Y, Hartl DL and Parsch J. 2004. Identification of a locus under complex positive selection in Drosophila simulans by haplotype mapping and composite-likelihood estimation. *Genetics* 168: 265-279.

Meinilä M, Finnilä S and Majamaa K. 2001. Evidence for mtDNA admixture between the Finns and the Saami. *Hum Hered* 52: 160-170.

Messer PW and Petrov DA. 2013. Frequent adaptation and the McDonald–Kreitman test. *Proceedings of the National Academy of Sciences* 110: 8615-8620.

Moorjani P, Thangaraj K, Patterson N, Lipson M, Loh P-R, Govindaraj P, Berger B, Reich D and Singh L. 2013. Genetic evidence for recent population mixture in India. *The American Journal of Human Genetics* 93: 422-438.

Nei M and Li W-H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences* 76: 5269-5273.

Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG and Bustamante C. 2005. Genomic scans for selective sweeps using SNP data. *Genome Res* 15: 1566-1575.

Orr HA and Betancourt AJ. 2001. Haldane's sieve and adaptation from the standing genetic variation. *Genetics* 157: 875-884.

Parra EJ, Kittles RA, Argyropoulos G, et al. 2001. Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. *Am J Phys Anthropol* 114: 18-29.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T and Reich D. 2012. Ancient admixture in human history. *Genetics* 192: 1065-1093.

Pennings PS and Hermisson J. 2006. Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. *Mol Biol Evol* 23: 1076-1084.

Perez RE, Jansen PW and Martins JR. 2012. pyOpt: a Python-based object-oriented framework for nonlinear constrained optimization. *Structural and Multidisciplinary Optimization* 45: 101-118.

Perry GH, Dominy NJ, Claw KG, et al. 2007. Diet and the evolution of human amylase gene copy number variation. *Nat Genet* 39: 1256-1260.

Peter BM, Huerta-Sanchez E and Nielsen R. 2012. Distinguishing between selective sweeps from standing variation and from a *de novo* mutation. *PLoS Genet* 8: e1003011.

Pickrell JK and Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* 8: e1002967.

Pool JE, Hellmann I, Jensen JD and Nielsen R. 2010. Population genetic inference from genomic sequence variation. *Genome Res* 20: 291-300.

Prado-Martinez J, Sudmant PH, Kidd JM, et al. 2013. Great ape genetic diversity and population history. *Nature* 499: 471-475.

Pritchard JK, Pickrell JK and Coop G. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol* 20: R208-R215.

Pritchard JK, Seielstad MT, Perez-Lezaun A and Feldman MW. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol Biol Evol* 16: 1791-1798.

Przeworski M. 2002. The signature of positive selection at randomly chosen loci. *Genetics* 160: 1179-1189.

Przeworski M, Coop G and Wall JD. 2005. The signature of positive selection on standing genetic variation. *Evolution* 59: 2312-2323.

Reich DE, Cargill M, Bolk S, et al. 2001. Linkage disequilibrium in the human genome. *Nature* 411: 199-204.

Sabeti PC, Reich DE, Higgins JM, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419: 832-837.

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

Schrider DR and Kern AD. 2016. S/HIC: Robust Identification of Soft and Hard Sweeps Using Machine Learning. *PLoS Genet* 12: e1005928.

Schrider DR, Mendes FK, Hahn MW and Kern AD. 2015. Soft shoulders ahead: spurious signatures of soft and partial selective sweeps result from linked hard sweeps. *Genetics* 200: 267-284.

Sella G, Petrov DA, Przeworski M and Andolfatto P. 2009. Pervasive natural selection in the Drosophila genome? *PLoS Genet* 5: e1000495.

Sheehan S, Harris K and Song YS. 2013. Estimating variable effective population sizes from multiple genomes: a sequentially Markov conditional sampling distribution approach. *Genetics* 194: 647-662.

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

Simonsen KL, Churchill GA and Aquadro CF. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* 141: 413-429.

Smith NG and Eyre-Walker A. 2002. Adaptive protein evolution in Drosophila. *Nature* 415: 1022-1024.

Sousa V and Hey J. 2013. Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics* 14: 404-414.

Stephan W, Wiehe TH and Lenz MW. 1992. The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. *Theor Popul Biol* 41: 237-254.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.

Tavaré S, Balding DJ, Griffiths RC and Donnelly P. 1997. Inferring coalescence times from DNA sequence data. *Genetics* 145: 505-518.

Tennessen JA, Bigham AW, O'Connor TD, et al. 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337: 64-69.

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

Thornton K and Andolfatto P. 2006. Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of Drosophila melanogaster. *Genetics* 172: 1607-1619.

Thornton KR. 2009. Automating approximate Bayesian computation by local linear regression. *BMC genetics* 10: 1.

Tishkoff SA, Reed FA, Ranciaro A, et al. 2007. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet* 39: 31-40.

Voight BF, Adams AM, Frisse LA, Qian Y, Hudson RR and Di Rienzo A. 2005. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc Natl Acad Sci U S A* 102: 18508-18513.

Voight BF, Kudaravalli S, Wen X and Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol* 4: e72.

Vy HMT and Kim Y. 2015. A Composite-Likelihood Method for Detecting Incomplete Selective Sweep from Population Genomic Data. *Genetics* 200: 633-649.

Wallberg A, Han F, Wellhagen G, et al. 2014. A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee Apis mellifera. *Nat Genet* 46: 1081-1088.

Wegmann D, Leuenberger C, Neuenschwander S and Excoffier L. 2010. ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics* 11: 116.

Zhan X, Pan S, Wang J, et al. 2013. Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle. *Nat Genet* 45: 563-566.

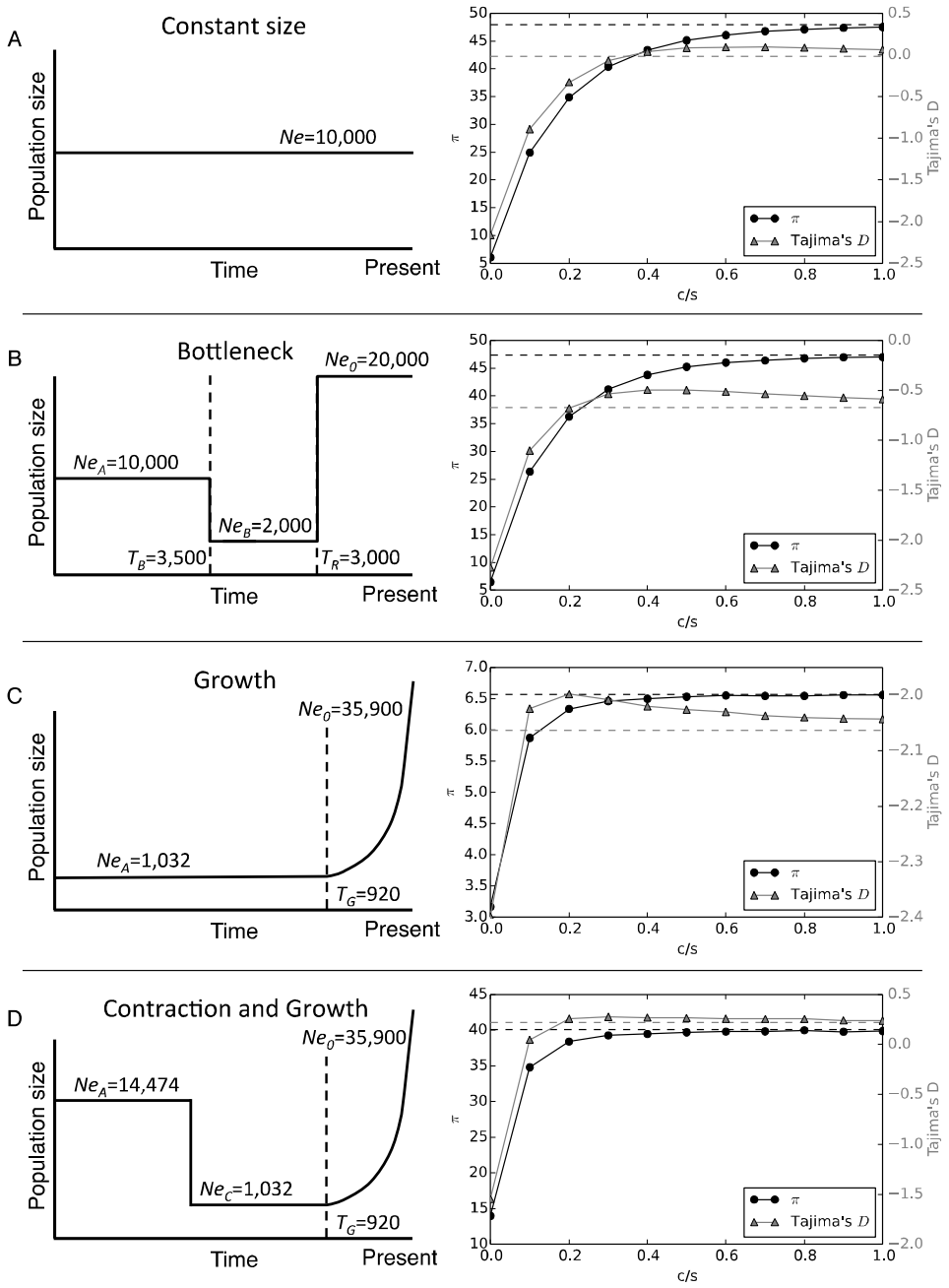

Fig. 1. Demographic models used in this study. For each model, a diagram of the population size history is shown on the left (not to scale) along with the values of each parameter. On the right, the values of π and Tajima's D are shown for windows sampled at varying distances (measured by the total crossover rate over the selection coefficient, c/s) from a hard selective sweep. (*A*) A model with constant population size. (*B*) A population bottleneck (parameterization from Marth et al. 2004). (*C*) Recent exponential population growth. (*D*) A three-epoch model with a population contraction and recent exponential growth (a simplified version of the European model from Gravel et al. 2011).

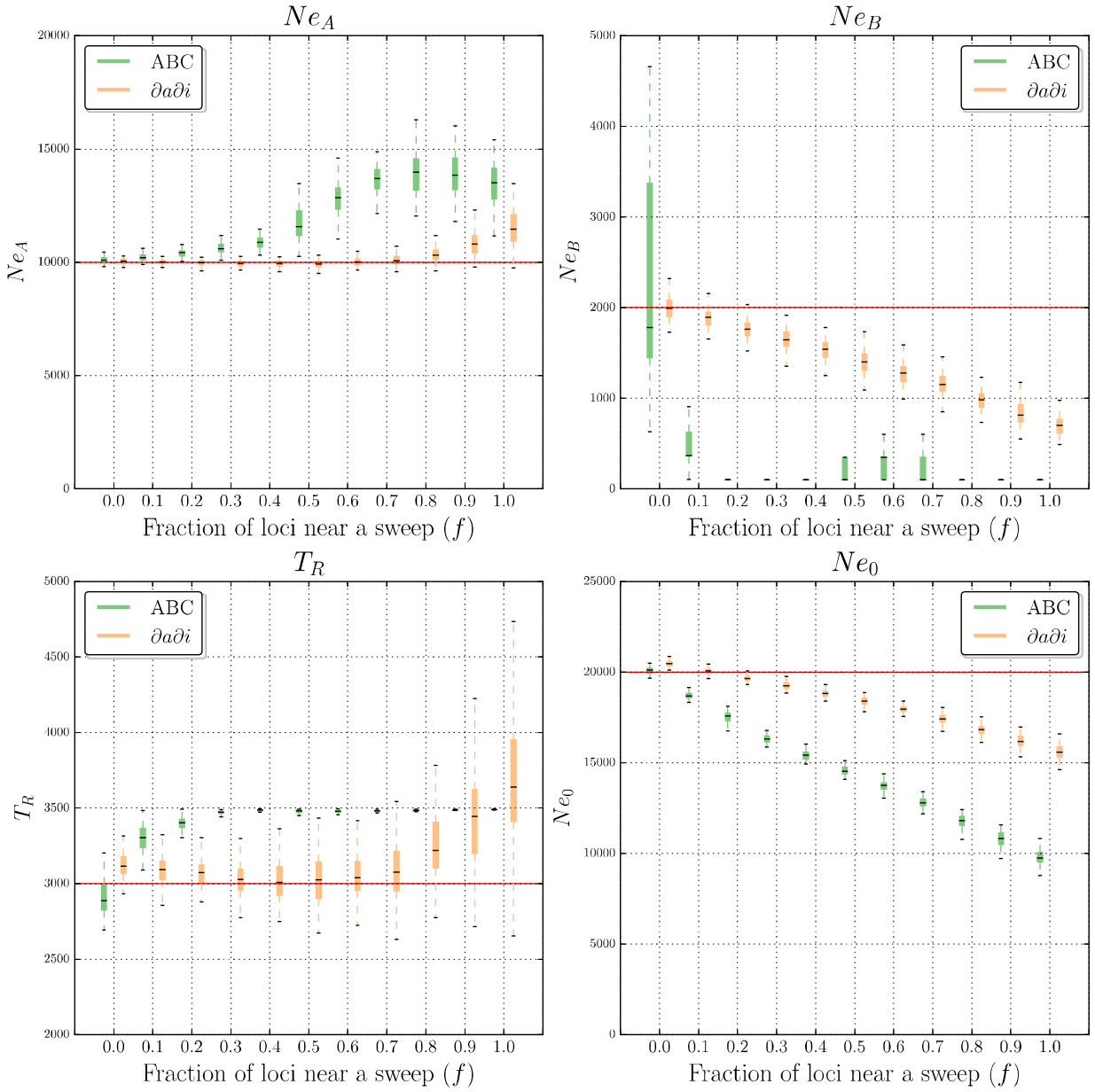

Fig. 2. Bottleneck model parameter estimates from ∂a∂i and ABC. Parameter estimation was performed on simulated data sets either evolving neutrally, or with some fraction of loci (*f*) used for inference linked to a selective sweep. Each box plot summarizes estimates from 100 replicates for each scenario. Note that T_B, the bottleneck onset time, is absent from this figure because it was fixed it to the true value (Methods).

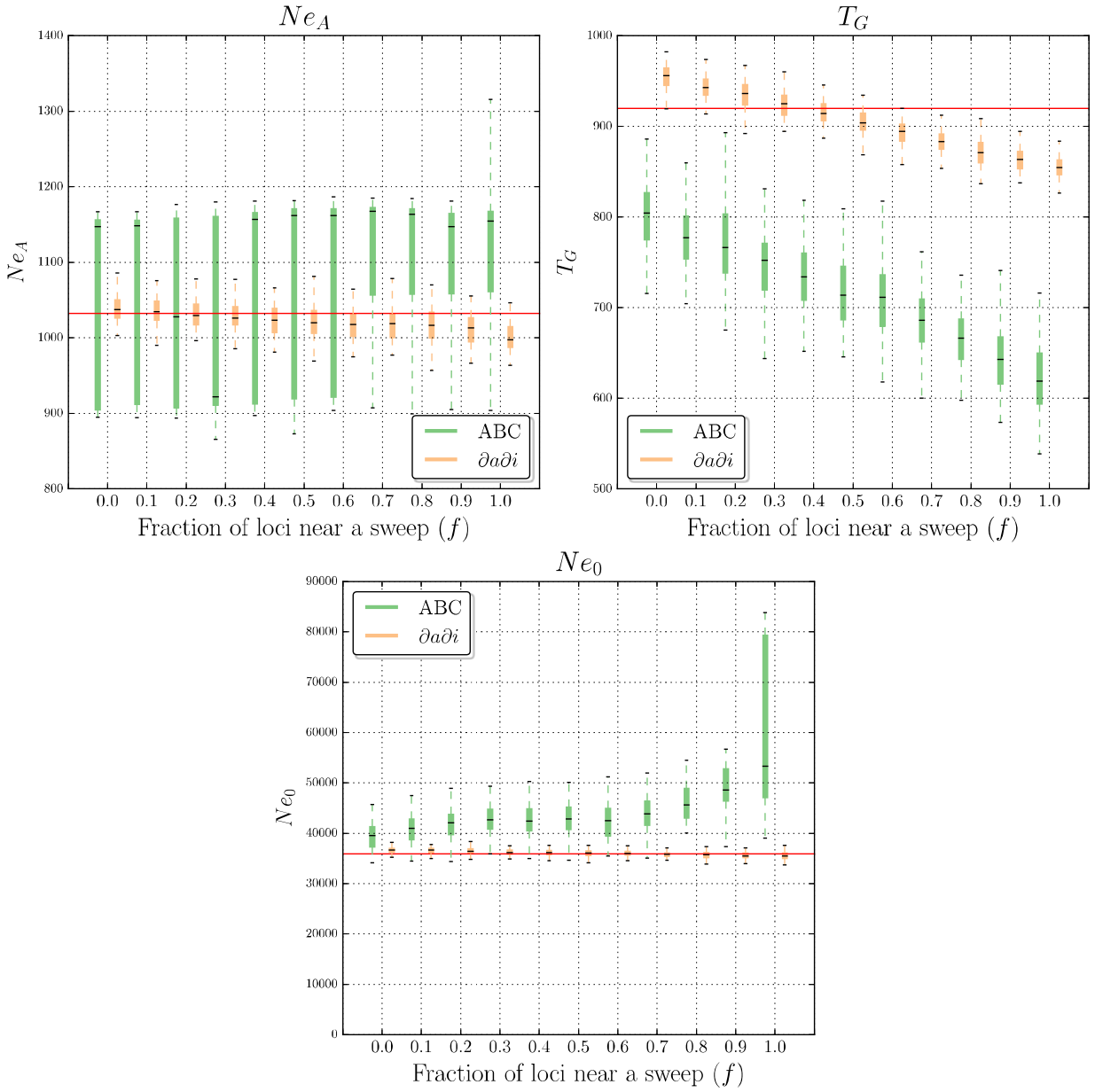

Fig. 3. Exponential growth model parameter estimates from ∂a∂i and ABC. Parameter estimation was performed on simulated data sets either evolving neutrally, or with some fraction (*f*) of loci used for inference linked to a selective sweep. Each box plot summarizes estimates from 100 replicates for each scenario.

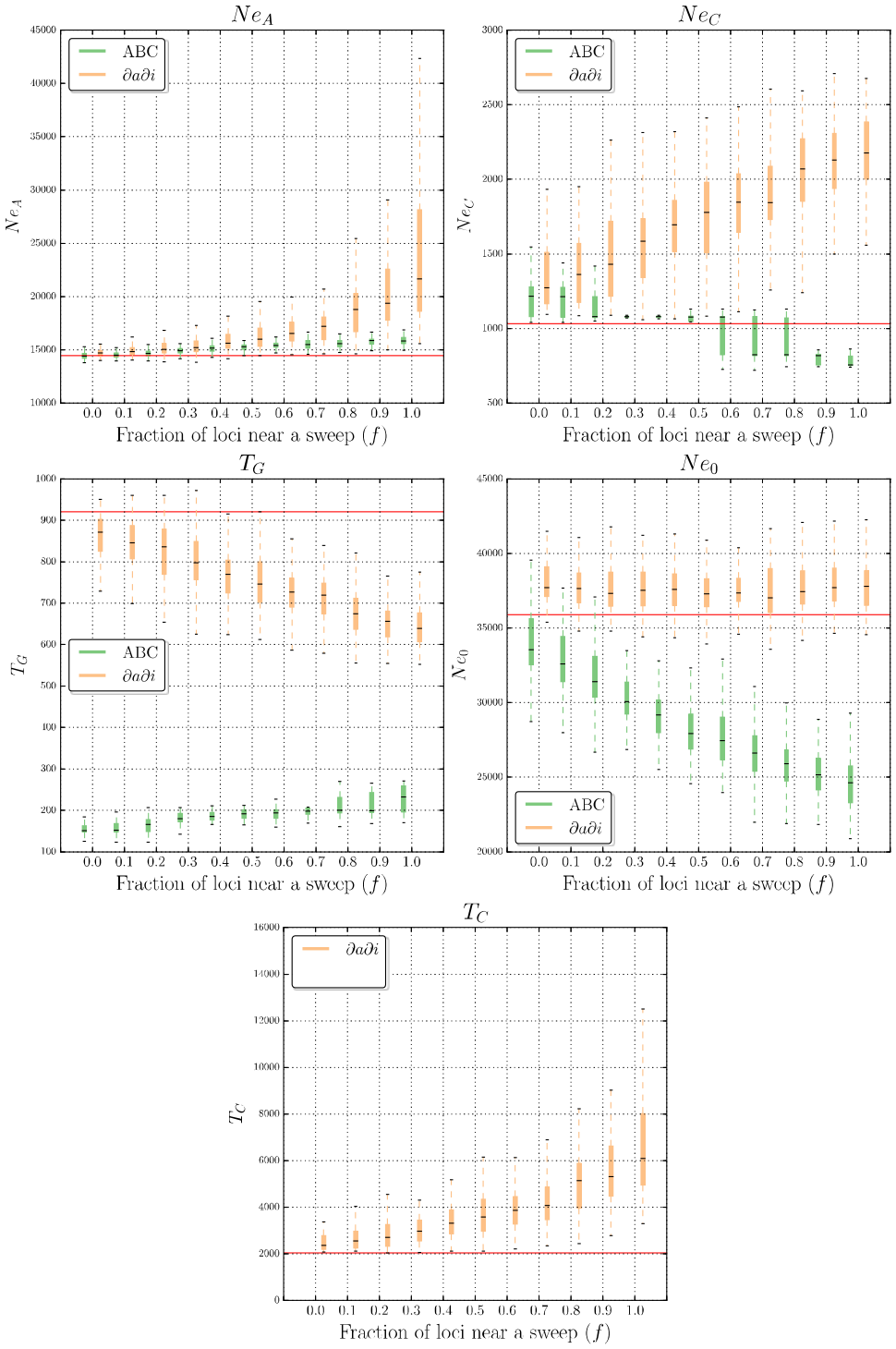

Fig. 4. Contraction-then-growth model parameter estimates from ∂a∂i and ABC. Parameter estimation was performed on simulated data sets either evolving neutrally, or with some fraction of loci (*f*) used for inference linked to a selective sweep. Each box plot summarizes estimates from 100 replicates for each scenario. Note that when performing ABC, time of population contraction (T_C), was fixed to the true value and therefore this parameter is only shown for ∂a∂i.

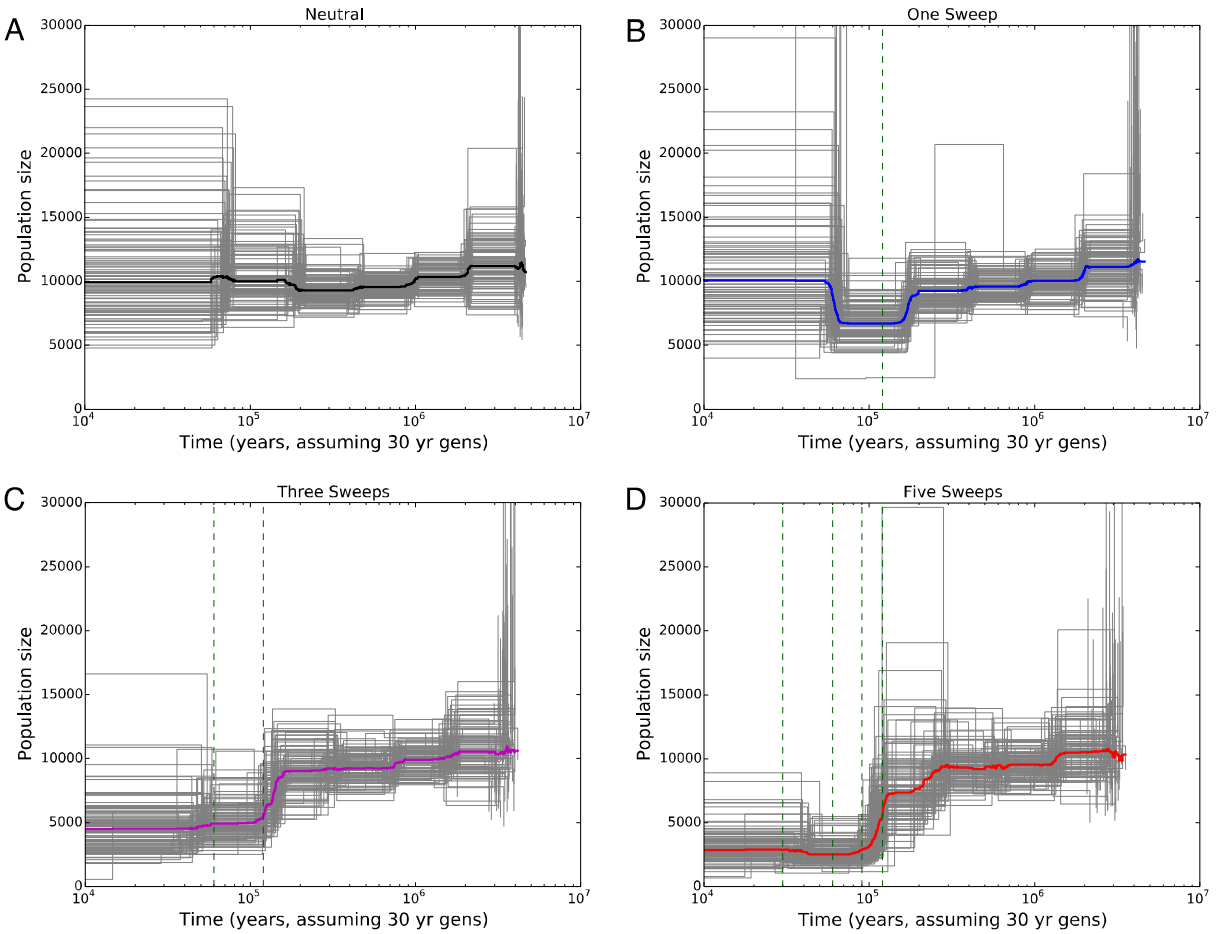

Fig. 5. Population size histories inferred by PSMC. Inferred size histories for each of 100 replicate simulations are shown as thin gray lines, and the median across all replicates is shown as the thicker line. In all cases the simulated population's size was constant throughout. (*A*) Population size histories inferred from neutral simulations. (*B*) Inferences from simulations with one selective sweep, for which the fixation time is shown as a dashed green vertical line. (*C*) Inferences from simulations with three recurrent selective sweeps. Fixation times for the two older sweeps are shown as dashed green vertical lines, while the most recent sweep fixed immediately prior to sampling. (*D*) Five recurrent selective sweeps, with fixation times for the four oldest shown as dashed vertical lines; again, the most recent sweep fixed immediately prior to sampling.

Table 1: The fraction of simulated data sets for which each demographic model was selected by ∂a∂i.

f	Bottleneck	Growth	Contraction-then-growth	Non-equilibrium (ambiguous)	Equilibrium
0	0.0%	0.0%	0.0%	0.0%	100.0%
0.1	1.0%	0.0%	0.0%	1.0%	98.0%
0.2	5.0%	0.0%	0.0%	10.0%	85.0%
0.3	14.0%	0.0%	0.0%	33.0%	53.0%
0.4	41.0%	0.0%	0.0%	55.0%	4.0%
0.5	77.0%	0.0%	0.0%	23.0%	0.0%
0.6	91.0%	0.0%	0.0%	9.0%	0.0%
0.7	100.0%	0.0%	0.0%	0.0%	0.0%
0.8	99.0%	0.0%	0.0%	1.0%	0.0%
0.9	100.0%	0.0%	0.0%	0.0%	0.0%
1	100.0%	0.0%	0.0%	0.0%	0.0%

Table 2: The fraction of simulated data sets for which each demographic model was selected by ABC.

f	Bottleneck	Growth	Contraction-then-growth	Non-equilibrium (ambiguous)	Equilibrium
0	0.0%	0.0%	0.0%	0.0%	100.0%
0.1	0.0%	0.0%	0.0%	1.0%	99.0%
0.2	0.0%	0.0%	0.0%	6.0%	94.0%
0.3	2.0%	0.0%	0.0%	32.0%	66.0%
0.4	25.0%	0.0%	0.0%	45.0%	30.0%
0.5	88.0%	0.0%	0.0%	2.0%	10.0%
0.6	99.0%	0.0%	0.0%	0.0%	1.0%
0.7	100.0%	0.0%	0.0%	0.0%	0.0%
0.8	100.0%	0.0%	0.0%	0.0%	0.0%
0.9	100.0%	0.0%	0.0%	0.0%	0.0%
1	100.0%	0.0%	0.0%	0.0%	0.0%

SUPPLEMENTARY FIGURE AND TABLE LEGENDS

Supplementary fig. S1. Bottleneck model parameter estimates from ∂a∂i, ABC using summary statistic means, and ABC using both means and variances. Parameter estimation was performed on simulated data sets either evolving neutrally, or with some fraction (*f*) of loci used for inference linked to a selective sweep at some distance (measured by *c*/*s*). Each box plot summarizes estimates from 100 replicates for each scenario. Note that T_B, the bottleneck onset time, is absent from this figure because it was fixed it to the true value (Methods).

Supplementary fig. S2. Growth model parameter estimates from ∂a∂i, ABC using summary statistic means, and ABC using both means and variances. Parameter estimation was performed on simulated data sets either evolving neutrally, or with some fraction (*f*) of loci used for inference linked to a selective sweep at some distance (measured by *c*/*s*). Each box plot summarizes estimates from 100 replicates for each scenario. Note that T_B, the bottleneck onset time, is absent from this figure because it was fixed to the true value (Methods).

Supplementary fig. S3. Contraction-then-growth model parameter estimates from ∂a∂i, ABC using summary statistic means, and ABC using both means and variances. Parameter estimation was performed on simulated data sets either evolving neutrally, or with some fraction (*f*) of loci used for inference linked to a selective sweep at some distance (measured by *c*/*s*). Each box plot summarizes estimates from 100 replicates for each scenario. Note that T_C, the time of population contraction, is present only for ∂a∂i because for ABC this parameter was fixed to the true value (Methods).

Supplementary fig. S4. Differences in AIC between equilibrium and non-equilibrium models when fitted by ∂a∂i to simulated constant-size populations with varying degrees of positive selection. For the growth model, a small number of simulated optimized very poorly, leading to large AICs, and therefore large differences between the growth and equilibrium AIC. The upper limit of the *y*-axis of this plot was truncated to allow visualization of AIC differences for the bulk of the data for which optimization was more successful (though box and whisker lengths still reflect the presence of these outliers in the set).

Supplementary fig. S5. Bayes factors from ABC's model selection comparing equilibrium and non-equilibrium demographic models.

Supplementary table S1. The fraction of simulated data sets for which each demographic model was selected by ABC when including variances of summary statistics.

Supplementary table S2. Example command lines to simulate each demographic model using discoal_multipop.

Supplementary table S3. Priors on parameter values of each demographic model for ABC sampling simulations.