Comparison of Single Genome and Allele Frequency Data Reveals Discordant Demographic Histories

Annabel C. Beichman,* Tanya N. Phung,† and Kirk E. Lohmueller*,†,‡,1

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

ORCID IDs: 0000-0002-6991-587X (A.C.B.); 0000-0003-2685-6977 (T.N.P.); 0000-0002-3874-369X (K.E.L.)

ABSTRACT Inference of demographic history from genetic data is a primary goal of population genetics of model and nonmodel organisms. Whole genome-based approaches such as the pairwise/multiple sequentially Markovian coalescent methods use genomic data from one to four individuals to infer the demographic history of an entire population, while site frequency spectrum (SFS)-based methods use the distribution of allele frequencies in a sample to reconstruct the same historical events. Although both methods are extensively used in empirical studies and perform well on data simulated under simple models, there have been only limited comparisons of them in more complex and realistic settings. Here we use published demographic models based on data from three human populations (Yoruba, descendants of northwest-Europeans, and Han Chinese) as an empirical test case to study the behavior of both inference procedures. We find that several of the demographic histories inferred by the whole genome-based methods do not predict the genome-wide distribution of heterozygosity, nor do they predict the empirical SFS. However, using simulated data, we also find that the whole genome methods can reconstruct the complex demographic models inferred by SFS-based methods, suggesting that the discordant patterns of genetic variation are not attributable to a lack of statistical power, but may reflect unmodeled complexities in the underlying demography. More generally, our findings indicate that demographic inference from a small number of genomes, routine in genomic studies of nonmodel organisms, should be interpreted cautiously, as these models cannot recapitulate other summaries of the data.

The pairwise sequentially Markovian coalescent (PSMC) and related methods have become a popular tool to estimate the history of a population from genetic variation data (McVean and Cardin 2005; Li and Durbin 2011; Schiffels and Durbin 2014). These methods use whole genome sequences from one to four individuals to infer the demographic history of an entire population. Specifically, they estimate the local time to the most recent common ancestor (TMRCA) for small regions in the genome and then use the distribution of these coalescent times to infer an overarching demographic history. For instance, if many regions of the genome coalesce at a specific time, it may be evidence for a population contraction, which would reduce the number of genetic lineages. The great appeal of these methods is that they do not rely on deep sequencing of multiple individuals in a population; instead, a single genome can be used to infer the demographic history of an entire population. PSMC and its successors have been used to infer the demographic histories and split times of many human populations (Li and Durbin 2011; Kidd et al. 2012; Schiffels and Durbin 2014; 1000 Genomes Project Consortium 2015; Henn et al. 2016), and were recently featured in three prominent articles that reconstructed human history using whole genome sequencing data from over 20 populations (Malaspinas et al. 2016; Mallick et al. 2016; Pagani et al. 2016).

PSMC plots have also become a cornerstone of many studies of nonmodel organisms lacking resources for the sequencing of numerous individuals, including archaic hominins (Meyer et al. 2012; Prufer et al. 2014), great apes (Prado-Martinez et al. 2013), wild boars and domestic pigs (Groenen et al. 2012; Bosse et al. 2014), canids (Freedman et al.

2014; Wang *et al.* 2016), horses (Orlando *et al.* 2013), over 38 bird species (Nadachowska-Brzyska *et al.* 2013, 2015, 2016; Hung *et al.* 2014; Murray *et al.* 2017), pandas (Zhao *et al.* 2012), dromedaries (Fitak *et al.* 2016), flowering plants (Albert *et al.* 2013; Ibarra-Laclette *et al.* 2013; Holliday *et al.* 2016), and even woolly mammoths (Palkopoulou *et al.* 2015).

Despite their widespread prominence, there is concern over the validity of demographic models obtained from this set of whole genome-based methods. Particularly, Mazet *et al.* (2015b) found that PSMC captures the inverse instantaneous coalescent rate (IICR) rather than an absolute measure of population size. The IICR corresponds to the effective population size if the population is panmictic, but it can differ from the population size due to gene flow and population structure, which affect the time to coalescence between subgroups. Thus, population structure can give a false signal of population growth or contraction, a notorious problem in demographic inference (Ptak and Przeworski 2002; Chikhi *et al.* 2010; Peter *et al.* 2010; Gattepaille *et al.* 2013; Heller *et al.* 2013; Mazet *et al.* 2015a,b; Orozco-terWengel 2016). Given these possible confounders, the degree to which whole genome-based plots derived from PSMC and its successors correspond to actual population size changes, rather than other demographic phenomena, remains unclear.

An alternative approach to infer population demography from genetic data uses the site frequency spectrum (SFS). The SFS represents the distribution of alleles at different frequencies in a sample of individuals from a population (Nielsen 2000; Wakeley 2009). The distribution of single nucleotide polymorphisms (SNPs), ranging from rare "singletons" which appear only once in the sample, to high-frequency variants that may appear in the majority of individuals, is directly affected by the demographic history of the population (Nielsen 2000; Wakeley 2009). Population contractions ("bottlenecks") can lead to a dearth of rare variants (Nei *et al.* 1975), whereas a rapid population expansion can lead to an overabundance (Tajima 1989; Slatkin and Hudson 1991; Keinan and Clark 2012). The SFS is a sufficient statistic for unlinked SNPs and has been used extensively in population genetic inference of demography (Nielsen 2000; Polanski and Kimmel 2003; Adams and Hudson 2004; Marth *et al.* 2004; Keinan *et al.* 2007; Gutenkunst *et al.* 2009; Gravel *et al.* 2011; Excoffier *et al.* 2013). SFS-based demographic inference has been implemented in programs such as $\partial a \partial i$ (Gutenkunst *et al.* 2009), moments (Jouganous *et al.* 2017), fastsimcoal2 (Excoffier *et al.* 2013), stairway plot (Liu and Fu 2015), fastNeutrino (Bhaskar *et al.* 2015), and others (Schraiber and Akey 2015). The SFS requires less sequence data per individual than the whole genome methods, but requires a greater number of individuals to be studied, with a minimum of 10 per population typically used (Gutenkunst *et al.* 2009; Excoffier *et al.* 2013). While the SFS is impractical if one can only sequence one or two individuals per population, population genomic studies based on many short loci scattered throughout the genome are beginning to be carried out on nonmodel organisms. RAD-seq data or gene transcript data from RNA-seq can readily be used for SFS-based demographic inference (McCoy *et al.* 2014; Trucchi *et al.* 2014; Sovic *et al.* 2016).

SFS-based and whole genome-based methods may have different strengths and weaknesses for demographic inference (Schraiber and Akey 2015). Theoretical and empirical data show that SFS-based approaches using large numbers of individuals can accurately estimate recent population growth (Nelson *et al.* 2012; Tennessen *et al.* 2012; Gazave *et al.* 2014; Bhaskar *et al.* 2015; Gao and Keinan 2016). In contrast, whole genome-based methods are less able to do so (Li and Durbin 2011). Recently, however, Schiffels and Durbin (2014) developed the multiple sequentially Markovian coalescent (MSMC),

an extension to PSMC that uses the SMC′ algorithm (Marjoram and Wall 2006) and can infer demography from two, four, or eight haplotypes (also known as PSMC′ when inferring from two haplotypes). The incorporation of multiple genomes in MSMC is specifically meant to improve estimates of recent growth (Schiffels and Durbin 2014).

The SFS may be limited in the degree to which it can detect ancient bottlenecks $>2N_e$ (effective population size) generations ago and in its ability to detect population declines (Bunnefeld *et al.* 2015; Terhorst and Song 2015; Boitard *et al.* 2016). Whole genome-based approaches are not constrained *a priori* by the number of population size changes as is common in the SFS-based approaches (but see the "stairway plot" approach of Liu and Fu 2015). They therefore often give information about events occurring millions of years ago, but the reliability of those results remains uncertain (Li and Durbin 2011). Further, demographic models inferred from human populations using the SFS were unable to recapitulate the empirical distribution of identity-by-state (IBS) tracts across the genome, while PSMC-derived models and an IBS-derived model were better able to match the IBS tract distribution (Harris and Nielsen 2013). However, the IBS-derived model did not predict the empirical SFS.

Due to these different strengths and weaknesses of approaches using single types of data, new methods have been developed which attempt to combine linkage disequilibrium (LD) information and the SFS (Bunnefeld *et al.* 2015; Boitard *et al.* 2016; Terhorst *et al.* 2017; Weissman and Hallatschek 2017). One of the most recent methods is Terhorst *et al.*'s (2017) SMC++, which combines a PSMC-like approach with the SFS to condition an SFS calculated from many individuals on the distribution of TMRCA from a single unphased genome. This approach is fast and potentially very powerful, but has the same barrier to entry for those studying nonmodel organisms as the other SFS methods, as it requires sequence data from many individuals.

Due to anthropological and biomedical interest, humans have been extensively studied using numerous demographic inference methods and provide a means to quantitatively compare these demographic inference approaches using the same empirical populations. Gutenkunst *et al.* (2009) and Gravel *et al.* (2011) carried out SFS-based inference of human demography using the diffusion approximation in $\partial a \partial i$, while Li and Durbin (2011) and Schiffels and Durbin (2014) estimated human demography from the same populations using PSMC and MSMC, respectively. Although the results are in some ways generally similar, the demographic models inferred for three human populations using MSMC (Schiffels and Durbin 2014) differ from demographic models for the same populations derived from SFS-based methods (Gutenkunst *et al.* 2009). MSMC infers ancient ancestral sizes and periods of growth and decline (the characteristic "humps" in MSMC trajectories) that were not detected in the SFS-derived models as well as inferring greater recent growth (Figure 1). The models inferred using MSMC also vary depending on the number of genomes used for the inference (Figure 1).

Terhorst *et al.* (2017) analyzed the same populations with the combined whole genome and SFS method, SMC++, finding an ancestral size more in line with Gutenkunst *et al.*'s (2009) model, but with greater recent growth and ancestral bottlenecks more resembling the MSMC models (Figure 1). The reasons why these approaches to demographic inference yield different estimates remain poorly understood.

Here we leverage humans as a model system to perform an empirical comparison of the performance of whole genome, SFS, and combined methods of demographic inference. Specifically, we determine which published models of human demography described above (Figure 1) best fit the empirical distributions of genome-wide heterozygosity, LD decay, and the observed SFS.

We find that the models inferred using the SFS or the combined method, SMC++, accurately recapitulate heterozygosity and the observed SFS. Among the MSMC models inferred by Schiffels and Durbin (2014), only the MSMC models based on a single genome were able to accurately recapitulate heterozygosity, and none of the MSMC models predicted an SFS that matched the empirical SFS. None of the demographic histories accurately predicted LD decay, but the histories derived from MSMC using four genomes (eight haplotypes), the SFS, and SMC++-based models fit better than the MSMC models based on one or two genomes. Our results provide a cautionary tale against the literal interpretation of demographic models inferred using one type of data, instead arguing for considering multiple summaries of the data when making detailed demographic inferences in nonmodel species.

METHODS

Published demographic models used in this study

We determined which, if any, of the published models of human demography (Figure 1) described below could accurately predict multiple summaries of the genetic variation data. Demographic models that fit the data well should produce patterns of genetic variation that match the empirical patterns in the data. We focused on three human populations: Utah residents with Northern and Western European ancestry from the Centre d'Etude du Polymorphism Humain collection (CEU); Han Chinese in Beijing, China (CHB); and Yoruba in Ibadan, Nigeria (YRI).

The first set of demographic models was jointly inferred for the three populations in ∂a∂i by Gutenkunst et al. (2009) using a three-population joint SFS based on data from intronic regions. Their model parameters were made available both in ∂a∂i and Hudson's ms (Hudson 2002) format, and include gene flow between the three populations (here referred to as the "Gutenkunst" model).

The next nine models were inferred by Schiffels and Durbin (2014) using whole genome Complete Genomics (Drmanac et al. 2010) sequence data of two, four, and eight statistically phased genomic haplotypes (one, two, and three individual genomes) per population to infer demographic histories using MSMC (here referred to as the "MSMC 2-haplotype," "MSMC 4-haplotype," and "MSMC 8-haplotype" models; Supplemental Material, Supplementary Note 1 in File S1).

To analyze their models with ∂a∂i, we converted these nine demographic models (CEU, CHB, and YRI populations, each based on two, four, and eight haplotypes) into stepwise models of population size changes over small time intervals (Figure S1 and Supplementary Note 2 in File S1).

The final set of models was inferred by Terhorst et al. (2017) in SMC++, a combined SFS plus whole genome approach. For the whole genome portion of the analysis, they used high coverage sequence data from Complete Genomics, and generated an SFS based on a combination of 1000 Genomes and Complete Genomics whole genome data for each population (Drmanac et al. 2010; 1000 Genomes Project Consortium 2015; Terhorst et al. 2017). We converted these SMC++ models to ∂a∂i and ms format in the same manner as the MSMC models (here referred to as the "SMC++" models; Supplementary Note 2 in File S1).

Heterozygosity predicted by demographic models

We compared the distribution of expected heterozygosity from data simulated under each demographic model to empirical 1000 Genomes data from the same populations to determine which models most accurately predict this broad summary of the data (Figure 2 and Table S1 in File S1). While heterozygosity is a summary of the SFS, we considered it valuable to examine both statistics since information regarding the spatial correlation among SNPs along the genome is lost in

the genome-wide SFS. The distribution of heterozygosity across windows of the genome retains some spatial information and is more similar to what is used by the MSMC inference approach.

Empirical heterozygosity: 1000 Genomes data from the CEU, CHB, and YRI populations were downloaded. 10 unrelated individuals per population (see Supplementary Note 3 in File S1 for sequence IDs) were randomly chosen so that comparisons could be made with Gutenkunst et al.'s (2009) empirical SFS, described below. For all our empirical analyses, only sites that passed the 1000 Genomes "Strict Mask" filter were considered (1000 Genomes Project Consortium 2015).

Expected heterozygosity per site (π) was calculated in nonoverlapping 100-kb windows from the whole genome data (Supplementary Note 3 in File S1) as:

$$\pi = \frac{n}{n-1} \frac{\sum_{i=1}^{L} 2p_i(1-p_i)}{L} \; ,$$

where p is the frequency of one allele, L is the total number of callable sites in the window, and n is number of sampled chromosomes ($n = 20$ for 10 diploid individuals).

Because genetic variation can be affected by linked natural selection (Gazave et al. 2014; Schrider et al. 2016), we also calculated expected heterozygosity for a set of 6333 × 10-kb neutral windows that were selected using the Neutral Region Explorer (NRE) (Arbiza et al. 2012) (Figure S2 and Supplementary Note 3 in File S1). The NRE is a useful tool that allows for the quick identification of putatively neutral regions that have high recombination rates and high B-values (indicating less linked selection). For the full set of parameters used in selection of putatively neutral regions, see Supplementary Note 3 in File S1.

Simulated heterozygosity: For each demographic model, whole genome data for 10 individuals were simulated in MaCS (Chen et al. 2009) over 20,000 × 100-kb independent blocks, each with a different recombination rate drawn from the distribution of recombination rates calculated by Phung et al. (2016) from the pedigree-based genetic map assembled by the deCODE project (Kong et al. 2010). Additionally, 6300 × 10-kb independent blocks per 10 individuals were simulated for comparison to the neutral regions from the 1000 Genomes data set (1000 Genomes Project Consortium 2015). Each 10-kb block was simulated using a recombination rate matched to that of one of the empirically neutral 10-kb windows, linearly interpolated from the deCODE project (Kong et al. 2010). For both sets of simulations, the expected heterozygosity across the 10 individuals was calculated using the equation above in msstats (Hudson 2002).

LD decay predicted by demographic models

We calculated LD between pairs of SNPs using genotype data from 10 individuals from each of the four populations in the 1000 Genomes Project data. We removed singletons and sites where all 10 individuals were homozygous for the reference allele and then calculated genotype r^2 using vcftools (Danecek et al. 2011). All pairs of SNPs were then placed into bins based on their physical distance (bp) between each other, from 0–1000 bp (bin 1) to 50,000–51,000 bp (bin 51). Within each bin, the average r^2 was calculated by dividing the sum of r^2 values of each pair of SNPs in the bin by the total number of SNP pairs in that bin.

The same procedure was carried out for the data simulated in MaCS (Chen et al. 2009) that were used for the calculations of heterozygosity above. The MaCS output was converted to vcf format using a custom bash script. Genotype r^2 was calculated in vcftools (Danecek et al. 2011) for each 100-kb simulated window, the SNP pairs were binned by distance, and
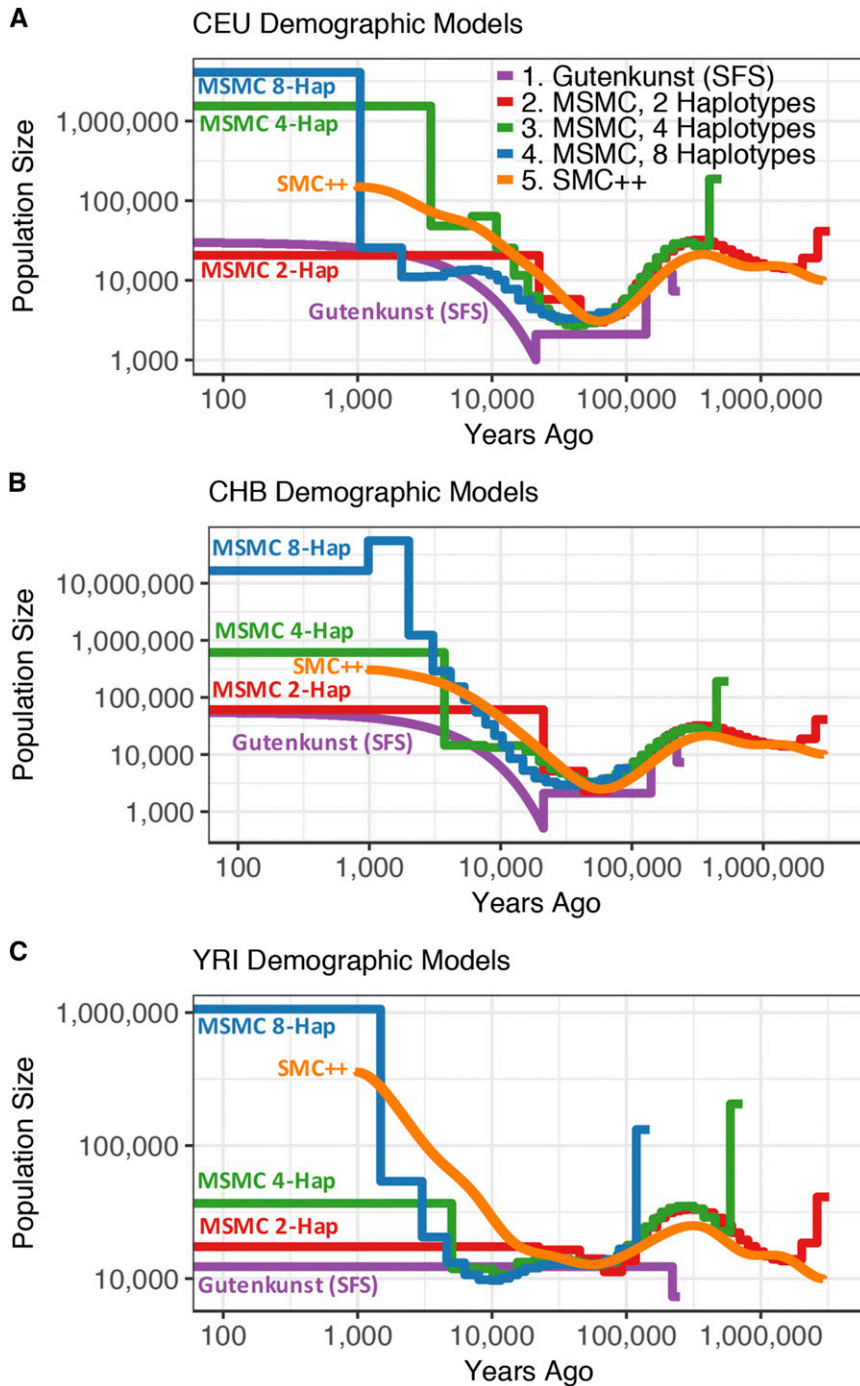
Figure 1 Demographic histories for the (A) CEU, (B) CHB, and (C) YRI populations. Trajectories are log scaled and in terms of physical units (diploid individuals and years). Models were either inferred using SFS-based methods (Gutenkunst) by Gutenkunst *et al.* (2009); from a sequentially Markovian coalescent-based approach (MSMC) from two, four, and eight haplotypes by Schiffels and Durbin (2014); or using a combined SFS and whole genome approach (SMC++) by Terhorst *et al.* (2017). The Gutenkunst models also include migration between all three populations, not depicted here. Models are scaled by the generation times used in each study [Gutenkunst *et al.* (2009): 25 yr/generation; Schiffels and Durbin (2014): 30 yr/generation; Terhorst *et al.* (2017): 29 yr/generation].

average r^2 was calculated as described above. The MSMC 8-haplotype YRI and MSMC 4-haplotype CEU, CHB, and YRI models have extremely large ancestral sizes, and so their simulations involve so many SNPs that the LD calculations become highly computationally intensive. Therefore, for these models only 5000 × 100-kb blocks were used for LD decay calculations, with 20,000 × 100-kb blocks used for the other models. We experimented with down-sampling the results and found no change in the LD decay curve due to the smaller amount of data.

To demonstrate that the use of the SMC′ approximation in the MaCS (Chen *et al.* 2009) simulator was not biasing our estimates of

LD, we simulated data in the manner described above under a simple model of extreme population decline (from 100,000 ancestral individuals to 1000) using both MaCS and MSMS (Ewing and Hermisson 2010) (which does not use the SMC′ approximation) and ran it through the same LD decay pipeline used for our other simulated data (Figure S3 in File S1).

SFS predicted by demographic models

We used the diffusion approximation in ∂a∂i (Gutenkunst *et al.* 2009) to calculate the expected SFSs under the Gutenkunst, MSMC
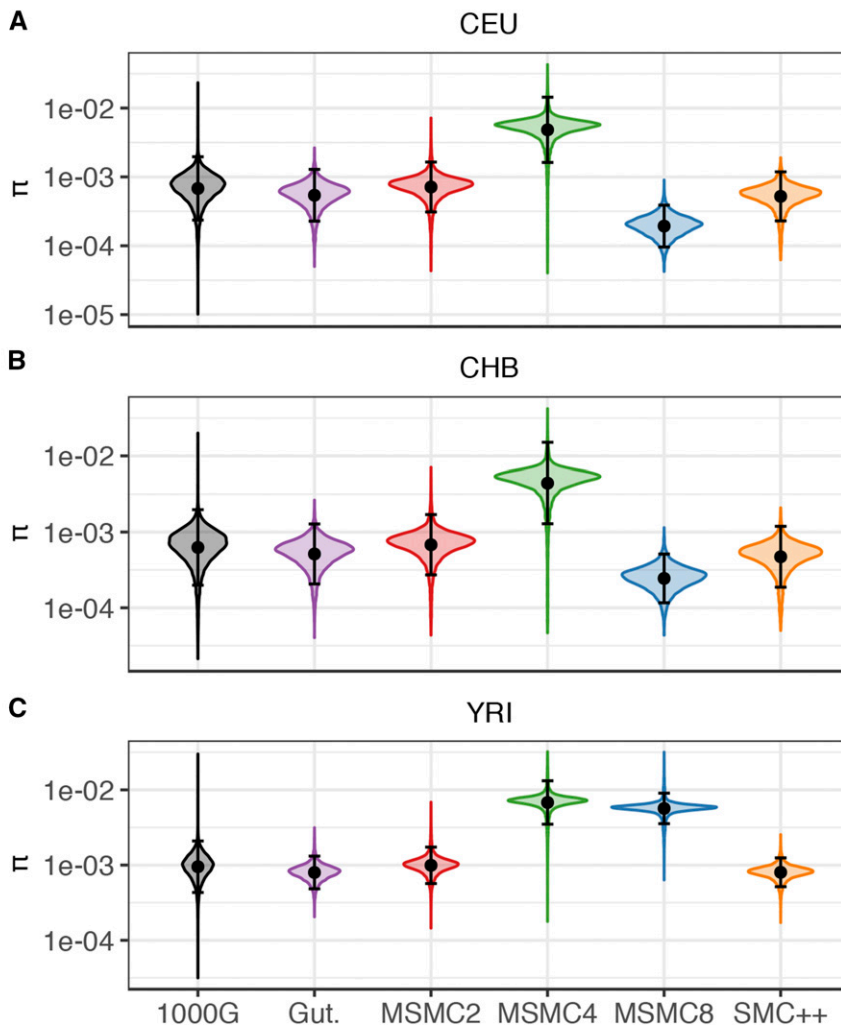
Figure 2 Kernel density distribution of expected heterozygosity (π per site). Heterozygosity was calculated across 100-kb windows from whole genome 1000 Genomes Project data for (A) CEU, (B) CHB, and (C) YRI, and from 20,000 × 100-kb blocks for data simulated under each demographic model. The black dot and bars indicate the mean ± 2 SD for each distribution. Note the log-10 scaling on the y-axis.

2-haplotype, MSMC 4-haplotype, MSMC 8-haplotype, and SMC++ models for the CEU, CHB, and YRI populations. We compared the SFSs expected under each of these models both to the empirical SFS used by Gutenkunst *et al.* (2009) to infer the demographic histories of these three populations ["observed (Gutenkunst)," Figure 4 and Figure 5], as well as to the SFSs based on low-coverage 1000 Genomes whole genome sequencing data ["1000 Genomes (whole genome)," Figure 6] and SFSs based on putatively neutral regions in the 1000 Genomes data set ["1000 Genomes (neutral)", Figure 6]. We assessed the fit of different models to the observed SFS by comparing their log-likelihoods (see below, Table 1, Supplementary Note 4 in File S1, and Tables S2–S4 in File S1).

Empirical SFSs: The primary empirical SFSs used in our comparisons were produced by Gutenkunst *et al.* (2009) and used to infer the joint demographic histories of CEU, CHB, and YRI populations in their study [observed (Gutenkunst)]. As described in their supplementary material, the joint SFS represents 4.04 Mb of Sanger sequencing data from 12 YRI, 12 CHB, and 22 CEU individuals projected down to

20 chromosomes for a total of 17,446 segregating SNPs polarized against chimp, with a correction for ancestral misidentification applied. We marginalized the SFS using $\partial a \partial i$ (Gutenkunst *et al.* 2009) to have one SFS per population (Figure 4 and Figure 5).

To make sure our results were consistent with SFSs derived from other sequencing methodologies and different genomic regions, we also generated folded proportional genome-wide and neutral SFSs from the 1000 Genomes data described above [1000 Genomes (whole genome) and 1000 Genomes (neutral)] (1000 Genomes Project Consortium 2015; Figure 6, Figure S7 in File S1, and Supplementary Note 3 in File S1).

Expected SFSs under published demographic models: Expected SFSs for a sample size of 10 diploid individuals were calculated in $\partial a \partial i$ (Gutenkunst *et al.* 2009) for each of the published demographic models extrapolating calculations across three grid points (40, 50, and 60) (Figure 4 and Figure 5). To test whether the effect of differences in mutation rate between the studies may be responsible for discrepancies, we also considered an alternative scaling of the MSMC models using a higher mutation rate (Supplementary Note 5 in File S1).

■ **Table 1 Multinomial log-likelihoods comparing the fit of various models to the observed SFS derived from Sanger sequencing data and used by Gutenkunst *et al.* (2009) for their inference (SFSs in Figure 4)**

Model	Multinomial LL	ΔLL (Model − Data)
CEU		
Data to data	−21,546	0
Gutenkunst	−21,555	−9
SMC++	−21,599	−53
MSMC 2-haplotype	−21,698	−152
MSMC 8-haplotype	−21,816	−270
MSMC 4-haplotype	−22,760	−1214
CHB		
Data to data	−20,154	0
Gutenkunst	−20,202	−48
SMC++	−20,277	−123
MSMC 8-haplotype	−20,343	−188
MSMC 2-haplotype	−20,370	−216
MSMC 4-haplotype	−21,411	−1257
YRI		
Data to data	−29,630	0
Gutenkunst	−29,647	−17
SMC++	−29,779	−150
MSMC 2-haplotype	−30,003	−373
MSMC 8-haplotype	−31,282	−1652
MSMC 4-haplotype	−32,976	−3346

Data to data denotes the best log-likelihood possible when replacing the proportions predicted by the model with the observed proportions from the SFS used in Gutenkunst *et al.*'s (2009) study (see Supplementary Note 4 in File S1). Gutenkunst denotes the model inferred by Gutenkunst *et al.* (2009) fit to the observed SFS. SMC++ denotes the model inferred by Terhorst *et al.* (2017) using a combined whole genome and SFS approach. MSMC 2-, 4-, and 8-haplotype denote the demographic models inferred by Schiffels and Durbin (2014) using MSMC on two, four, and eight haplotypes, respectively.

We generated both the proportional (Figure 4 and Figure S5 in File S1) and absolute (*i.e.*, SFS based on SNP counts) SFSs (Figure 5 and Figure S6 in File). The proportional SFS was calculated by dividing each bin of the SFS output by ∂a∂i by the sum of the bins. The absolute SFS was calculated by scaling the SFS output by ∂a∂i (which is relative to $\theta = 1$) by

$$\theta = 4N_{Ai}\mu L,$$

where N_{Ai} is the oldest ancestral size inferred in each model and L is the sequence length (4.04 Mbp) in Gutenkunst *et al.* (2009). θ for the Gutenkunst model used the authors' preferred mutation rate of $\mu = 2.35 \times 10^{-8}$ mutations per base per generation, and θ for the MSMC and SMC++ models used the authors' preferred mutation rate of $\mu = 1.25 \times 10^{-8}$ mutations per base per generation (see Supplementary Note 5 in File S1 for scaling using alternate mutation rates).

Assessing SFS fit: Log-likelihoods were calculated for each proportional SFS relative to each of the three observed SFSs [observed (Gutenkunst), 1000 Genomes (whole genome), and 1000 Genomes (neutral)] using a multinomial log-likelihood (Table 1, Supplementary Note 4 in File S1, and Tables S2 and S4 in File S1). The fit of different models was compared by examining their decrease in log-likelihood compared to that of each of the observed SFSs to itself (Table 1, Supplementary Note 4 in File S1, and Tables S2 and S4 in File S1). Due to the uncertainty of singleton SNP calls using high-throughput sequencing data, log-likelihoods were calculated both with singletons and with the SFS renormalized without the singletons category when comparing to the 1000 Genomes SFSs (Figure S7 and Table S4 in File S1).

Log-likelihoods were calculated for each absolute SFS (in terms of SNP counts) using a Poisson likelihood relative to the observed (Gutenkunst) SFS (Supplementary Note 4 and Table S3 in File S1).

Effect of uncertainty in ancestral population size

To investigate whether changing the ancestral population size (N_A) in the MSMC trajectories would result in SFSs that better fit the observed SFS, we adjusted the CEU MSMC 2-haplotype model to have a variety of N_A values. We also trimmed the model to remove ancient events (older than 225.5 KYA) to better match the time period (in years) encompassed by the Gutenkunst *et al.*'s (2009) model. These adjusted stepwise models were then used to calculate the expected SFS in ∂a∂i, as above. Supplementary Note 7 in File S1 describes the values of N_A used when testing the trimmed and untrimmed models (Figures S10–S13 in File S1).

MSMC population size trajectories for demographic models inferred from the SFS

To determine whether MSMC is capable of inferring a demography as complex as the one inferred in the Gutenkunst model, we used coalescent simulations to generate long chromosomal sequence data for each population under the Gutenkunst *et al.* (2009) inferred demographic model [see Gutenkunst *et al.*'s (2009) figure 2B and table 1 for full model], and then ran MSMC on these simulated data sets to assess whether the program is capable of recovering the underlying demographic model.

Simulations were carried out using MaCS (Chen *et al.* 2009). For each population, we simulated 50 replicate "genomes," made up of 80 independent 30-Mb "chromosomes," each made up of 300 linked 100-kb recombination blocks, with per-block recombination rates calculated by Phung *et al.* (2016) from the pedigree-based genetic map assembled by the deCODE project (Kong *et al.* 2010).

Each simulated genome was then used for a separate MSMC inference, using the default parameters (Schiffels and Durbin 2014) (Figure 7A). To determine whether these inferred MSMC trajectories would lead to SFSs matching those predicted by Gutenkunst *et al.*'s (2009) model, the MSMC trajectories were averaged and the average was converted into a stepwise ∂a∂i model. This model was then used to calculate the expected SFS under the averaged model based on simulated data (Figure 7, B and C). The multinomial and Poisson log-likelihoods for the proportional and SNP count SFSs were calculated as described in Supplementary Note 4 in File S1 (Tables S2 and S3 in File S1).

Extreme recent growth and Neanderthal admixture: We simulated data under more complex demographic histories, first to explore the relative abilities of the MSMC 2-haplotype and 8-haplotype models to infer extreme recent growth, and then to determine whether the addition of Neanderthal admixture may lead to MSMC trajectories resembling those inferred from real data by Schiffels and Durbin (2014) (Supplementary Note 6 and Figures S8 and S9 in File S1).

Data availability

All code to simulate data under each demographic model, calculate heterozygosity, and generate the SFS from simulated and empirical data are available on GitHub: github.com/LohmuellerLab/Compare_Demographic_Models.

RESULTS

We compared published models of demography for three human populations (CEU, CHB, and YRI) inferred using different methods
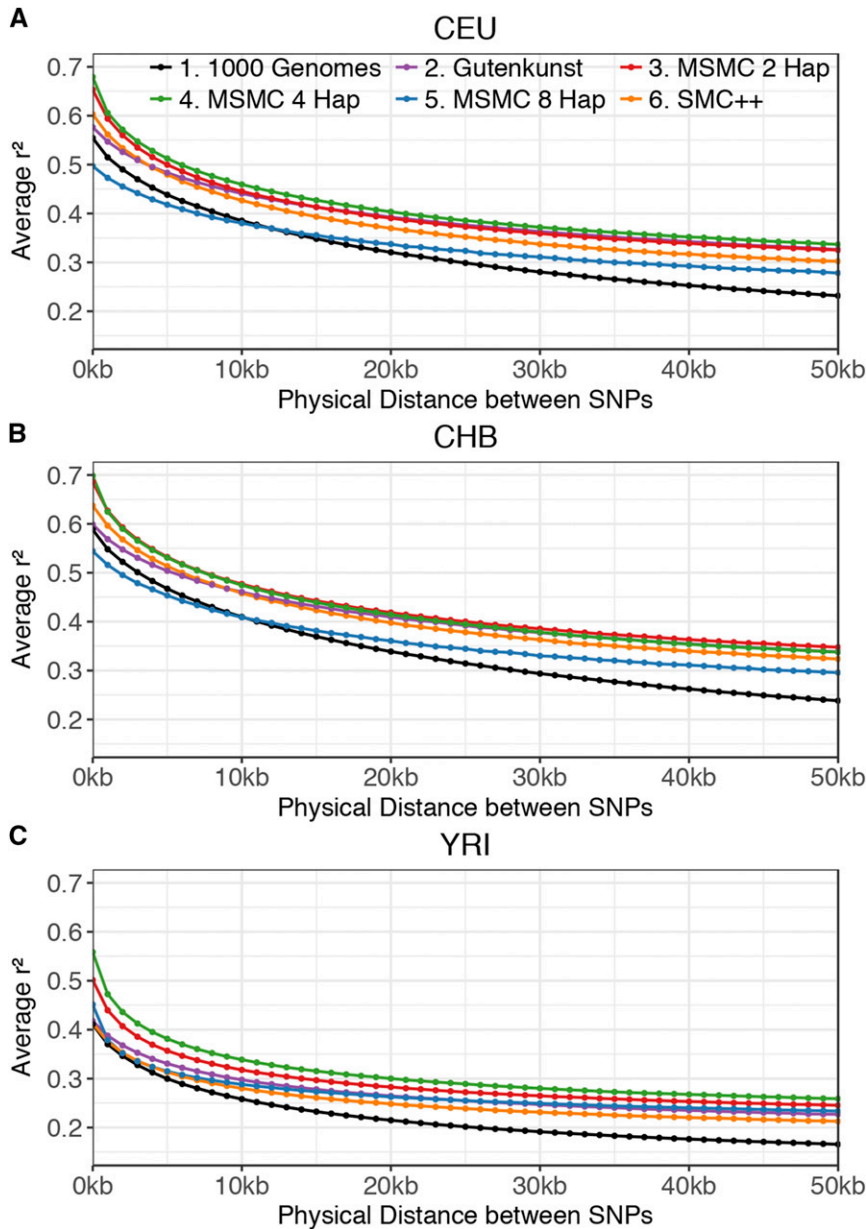
Figure 3 LD decay patterns. LD decay was calculated across 100-kb windows from 1000 Genomes data and simulated data under each demographic model for (A) CEU, (B) CHB, and (C) YRI. Pairs of SNPs are binned based on physical distance (bp) between them, up to 51 kb. Average genotype r^2 is calculated within each distance bin.

for demographic inference: (1) using the SFS in ∂a∂i (Gutenkunst) (Gutenkunst *et al.* 2009), (2) using whole genomes in MSMC (MSMC 2-, 4-, and 8-haplotype) (Schiffels and Durbin 2014), and (3) using a combined SFS plus whole genome approach in SMC++ (SMC++) (Terhorst *et al.* 2017). The evaluation of the MSMC models involves three models per population because Schiffels and Durbin's (2014) inference was carried out using two, three, or eight chromosomal haplotypes (from one, two, and four individuals), sometimes resulting in fundamentally different demographic parameter estimates. We evaluated whether the method's performance was improved using certain numbers of haplotypes.

Heterozygosity predicted by demographic models

The distribution of expected heterozygosity across 100- and 10-kb blocks was calculated from data simulated under each published demographic model for each of the three populations and compared to empirical

distributions of heterozygosity based on whole genome and putatively neutral sequence data from the 1000 Genomes Project.

We find that the Gutenkunst demographic model inferred from the SFS, the MSMC 2-haplotype model, and the SMC++ model all yielded distributions of heterozygosity that resemble the empirical whole genome distribution of heterozygosity, with MSMC 2-haplotype fitting the mean most closely (Figure 2). However, we found that the higher haplotype MSMC models (MSMC 4-haplotype and 8-haplotype) yielded distributions of heterozygosity that were highly divergent from the empirical distribution (Figure 2 and Table S1 in File S1).

The MSMC 4-haplotype models fit worst due to their extremely high inferred ancestral size across all three populations (Figure 1 and Table S2 in File S1) (CEU, 187,514; CHB, 191,238; YRI, 205,845 individuals; compared to 4000–40,000 individuals in the other models), with mean whole genome heterozygosity distributions nearly seven times larger than that of the empirical whole genome distribution (Figure 2 and
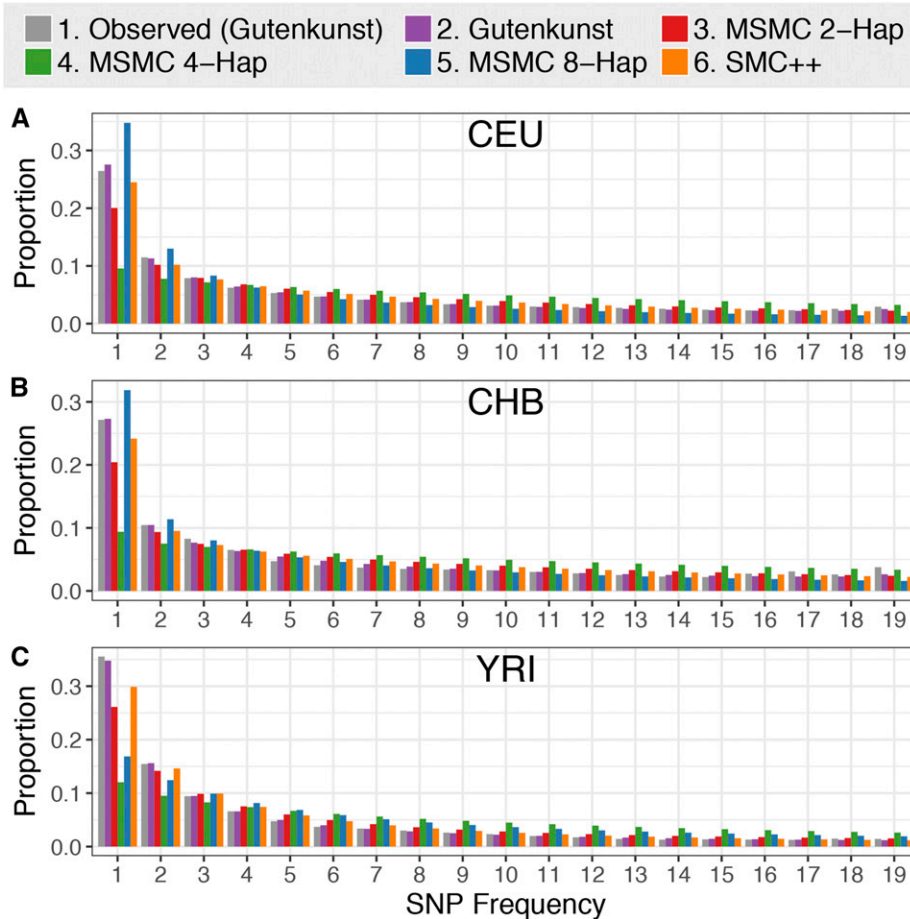
Figure 4 Unfolded proportional site frequency spectra for (A) CEU, (B) CHB (B), and (C) YRI populations. The "observed" SFS is from noncoding sequence used by Gutenkunst *et al.* (2009) to infer demographic histories for these three populations. See Figure S5 in File S1 for scaling using alternative mutation rates.

Table S1 in File S1). The MSMC 8-haplotye model for YRI infers a similarly large ancestral size and has a similarly high mean heterozygosity as the 4-haplotype YRI model. The MSMC 8-haplotype models for CEU and CHB, however, infer much lower ancestral sizes (CEU, 2147; CHB, 5666) (Figure 1). Due to the low ancestral size, these models also do not fit the empirical distribution well, yielding distributions of heterozygosity with means that are two to four times lower than the empirical distributions.

When examining the 1000 Genomes data, we found that heterozygosity in the neutral regions was higher than that seen for the genome-wide distribution of heterozygosity calculated in 10-kb windows (Table S1 in File S1; *e.g.*, CEU mean heterozygosity per site for whole genome: 7.8×10^{-4} *vs.* neutral: 9.4×10^{-4}), suggesting that natural selection has directly and/or indirectly affected genome-wide patterns of heterozygosity. When the published demographic models were compared to the neutral heterozygosity distributions, we found similar trends to those seen for the whole genome data (Figure S2 in File S1).

LD predicted by demographic models
None of the published demographic models could perfectly recapitulate the empirical LD decay curve (Figure 3). For SNP pairs <10 kb apart, the MSMC-8 haplotype model comes closest to the empirical curve for the CEU and CHB populations (Figure 3, A and B), but underestimates the amount of LD, while all other models predict too much LD. The Gutenkunst and SMC++ models predict similar LD curves and are closer to the empirical curve than the MSMC 2-haplotype and

4-haplotype models. For YRI SNP pairs <10 kb apart, SMC++ and MSMC 8-haplotype predict similar LD decay curves and are close to the empirical distribution, with Gutenkunst still fitting better than MSMC 2-haplotype and 4-haplotype (Figure 3C). At distances >10 kb apart, all demographic models predict there to be more LD than seen in the empirical data (Figure 3).

We found that the lack of fit is not due to the use of the SMC′ approximation in the simulator MaCS (Chen *et al.* 2009), as both MaCS and MSMS (Ewing and Hermisson 2010), a coalescent simulator which does not use the SMC′ approximation, yielded highly similar LD decay curves when simulating data under the same simple population contraction model (Figure S3 in File S1).

SFS predicted by demographic models
Lastly, we examined which of the demographic models could match the SFS of the empirical data. To account for the possibility of overfitting the SFS-based Gutenkunst model to the SFS it was inferred from, we also compared all models to empirical SFSs based on low-coverage, high-throughput 1000 Genomes sequence data from the same three populations.

Comparing to the observed Gutenkunst SFS: For each population, the SFSs predicted by the three MSMC models do not match the empirical proportional SFS from Gutenkunst *et al.* (2009), regardless of the mutation rate or number of genomes used (Figure 4, Table 1, and Figure S5 and Table S2 in File S1). The expected SFS based on the Gutenkunst
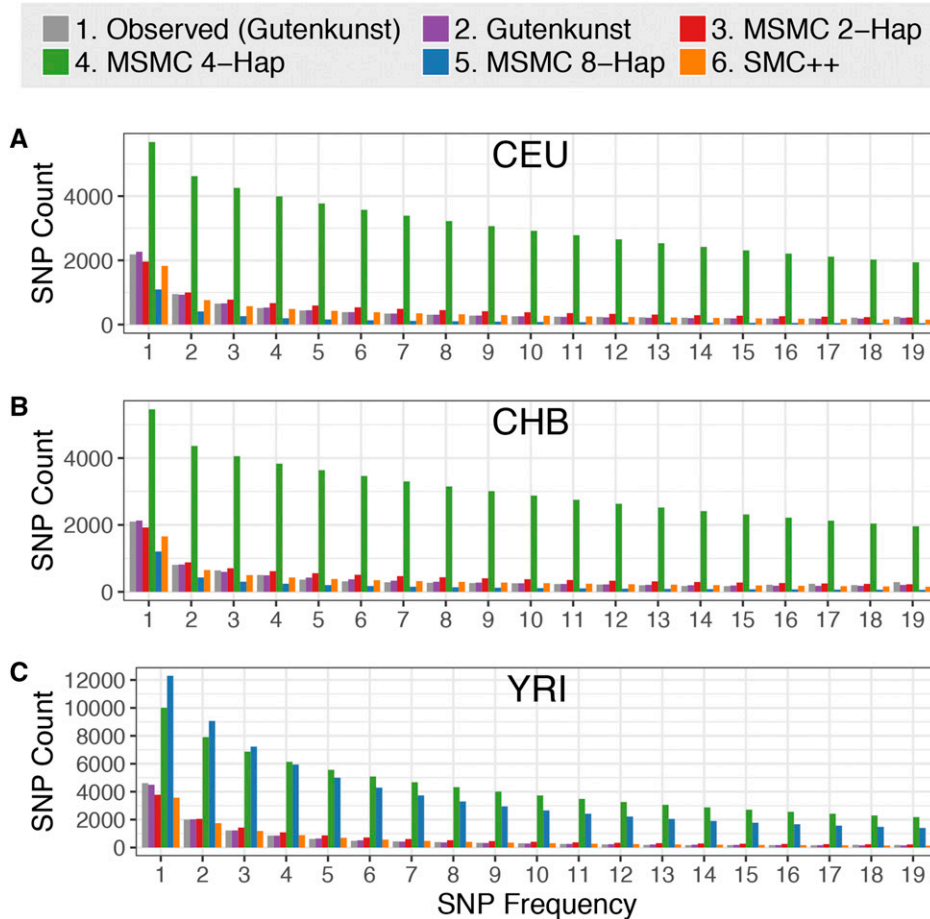
Figure 5 SNP count SFSs using the counts of SNPs for the (A) CEU, (B) CHB, and (C) YRI populations. The "observed" SFS is from noncoding sequence used by Gutenkunst *et al.* (2009) to infer demographic histories for these three populations. SFSs are scaled using the ancestral population size given by each model, the mutation rate used to scale each model by the authors, and the sequence length of the empirical data set (4.04 Mb). See Figure S6 in File S1 for scaling using alternative mutation rates.

et al. (2009) demographic history matches the observed SFS closely, being only 9 log-likelihood units worse than the best possible fit (comparing the empirical SFS to itself) for CEU, 48 units worse for CHB, and 17 units worse for YRI (Table 1). In comparison, the best-fitting MSMC models for each population are 152, 188, and 373 log-likelihood units below the best possible fit (Table 1). The combined whole genome plus SFS method SMC++ has an intermediate fit, with a log-likelihood well below the Gutenkunst model, but consistently better than any of the MSMC models (Table 1).

Interestingly, there is not a consistent improvement in fit to the observed SFS when increasing the number of individuals used for the MSMC inference. For each population, the 4-haplotype model has the worst fit (Figure 4 and Table 1). For CEU and YRI, the MSMC 2-haplotype models fit best out of the MSMC models, but both are >100 log-likelihood units worse than the Gutenkunst model. For CHB, the 8-haplotype model fits best, but is still 140 units worse than the Gutenkunst model (Table 1).

The above comparisons considered the proportions of SNPs at specific frequencies in the sample. We also performed a comparison of the number of SNPs in each bin of the SFS, the absolute SFS, to the observed absolute SFS used in Gutenkunst *et al.*'s (2009) inference using a Poisson likelihood. The absolute SFS expected under the Gutenkunst *et al.* (2009) model fits the observed SFS best (Figure 5 and Table S3 in File S1), and is only 9, 49, and 17 log-likelihood units below the best possible fits for CEU, CHB, and YRI models, respectively. The SMC++ models have the next-best fit to the absolute SFS, but come 86 (CEU), 176 (CHB), and 193 (YRI) log-likelihood units below the best possible

fit; followed by MSMC 2-haplotype which fell 278 (CEU), 378 (CHB), and 455 (YRI) log-likelihood units below the optimal fit (Table S3 in File S1). In all three populations, the MSMC 4-haplotype and 8-haplotype models are thousands of log-likelihood units worse than the best possible fit, showing no improvement based on using a larger number of individuals in the inference (Table S3 in File S1). The overestimation of SNPs in the 4-haplotype model is due to the model's extremely high predicted ancestral size (~200,000 individuals for each population) (Table S3 in File S1).

For both the proportional and absolute SFSs, we found that rescaling the models using a higher mutation rate did not produce large qualitative differences in how the MSMC models fit the observed (Gutenkunst) SFS (Supplementary Note 5 and Figures S4–S6 in File S1).

Comparing to the folded low-coverage 1000 Genomes SFS: To avoid giving the Gutenkunst model an unfair advantage by fitting all models to the SFS used to infer that particular model, we also compared all models to proportional folded SFSs based on whole genome and neutral data from the 1000 Genomes Project (Figure 6 and Figure S7 in File S1). The fit to the empirical singletons bin was poor for all models, except for SMC++, which was, in part, fit to an SFS based on 1000 Genomes data. Calling singletons is notoriously difficult in low-coverage data, making that bin the least reliable in the 1000 Genomes data (Kim *et al.* 2011; Nielsen *et al.* 2011; Han *et al.* 2014, 2015). We therefore calculated likelihoods for all models relative to the data both with singletons included and again with the SFSs renormalized without the singletons category (Figure S7 and Table S4 in File S1).
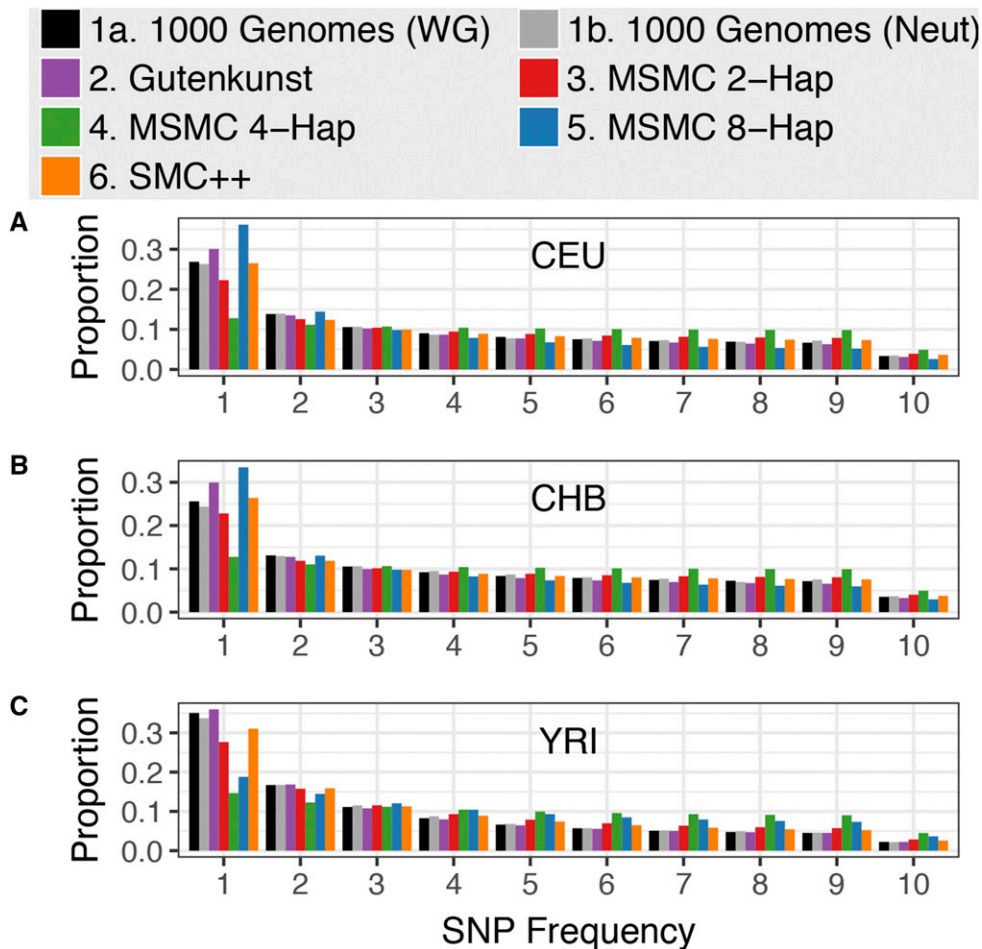
Figure 6 Folded proportional SFSs for (A) CEU, (B) CHB, and (C) YRI populations. The "1000 Genomes (WG)" SFS is from low-coverage whole genome 1000 Genomes data, and the "1000 Genomes (Neut)" SFS is from 6333 × 10-kb putatively neutral regions in the 1000 Genomes data.

For YRI, the Gutenkunst model is the best-fitting model for the whole genome and neutral 1000 Genomes SFSs, both with and without singletons, with all other models having a much worse fit (the next best model, SMC++, is hundreds to thousands of log-likelihood units below the fit of the Gutenkunst model) (Figure 6C and Table S4 in File S1). For CEU and CHB, if singletons are included, SMC++ fits the whole genome and neutral 1000 Genomes SFSs best. For CEU, the Gutenkunst model then fits second best, with the MSMC models far behind (Figure 6A and Table S4 in File S1). For CHB, the MSMC 2-haplotype fits second best after SMC++, with the Gutenkunst model coming third, but both are >10,000 log-likelihood units below SMC++ (Figure 6B and Table S4 in File S1). If singletons are excluded for CEU and CHB, then the Gutenkunst model fits best, with SMC++ coming in second, and the MSMC models all ranking far below (Table S4 in File S1).

Effect of uncertain ancestral population size

The accuracy of ancient ancestral population sizes, particularly >3 MYA (>100,000 generations), using the whole genome-based methods remains unclear (Li and Durbin 2011). As discussed above, the MSMC 2-haplotype and 4-haplotype models infer large ancestral sizes for each population that are not supported by previous inferences of human demographic history (Adams and Hudson 2004; Keinan *et al.* 2007; Boyko *et al.* 2008; Gutenkunst *et al.* 2009; Nielsen *et al.* 2009; Gravel *et al.* 2011). We hypothesized that these extreme ancestral sizes, as well as ancient bottlenecks and population growth (the signature humps of MSMC trajectories), which do not appear in demographic models

inferred using other methods, could be artifacts that are causing the SFS predicted by these models to deviate from the true SFS.

To test this hypothesis, we took the best fitting of the MSMC models, the CEU 2-haplotype model, and carried out a series of adjustment experiments to determine whether changes to the model could provide a better fit to the observed SFS. Without adjusting the time period encompassed by the model, we altered the ancestral population size to a variety of values including those inferred by Gutenkunst *et al.* (2009) (Supplementary Note 7 and Figures S10 and S11 in File S1). We also truncated the MSMC trajectory to remove ancient events and better match the time period (in years) encompassed by the Gutenkunst *et al.* (2009) model. We again adjusted the ancestral population size to a variety of plausible values (Supplementary Note 7 and Figures S12 and S13 in File S1).

We found trimming away the ancient (older than ~225 KYA) part of the demographic trajectory and lowering the ancestral population size to 10,000–12,300 (compared to 41,261 inferred initially) dramatically improved the fit of the proportional SFSs predicted under these adjusted models to the observed (Gutenkunst) SFS (Figure S12 and Table S5 in File S1). The best-fit model with ancestral size (N_A) equal to 12,300 was brought to within 38 log-likelihood units of the best possible likelihood (Figure S12D and Table S5 in File S1), only 29 units below the Gutenkunst model. When repeating this procedure using the SFS based on counts, the SFSs under these adjusted models showed a different pattern of improvement. Here the untrimmed models that did *not* have ancient events >225 KYA trimmed away, but had a lowered
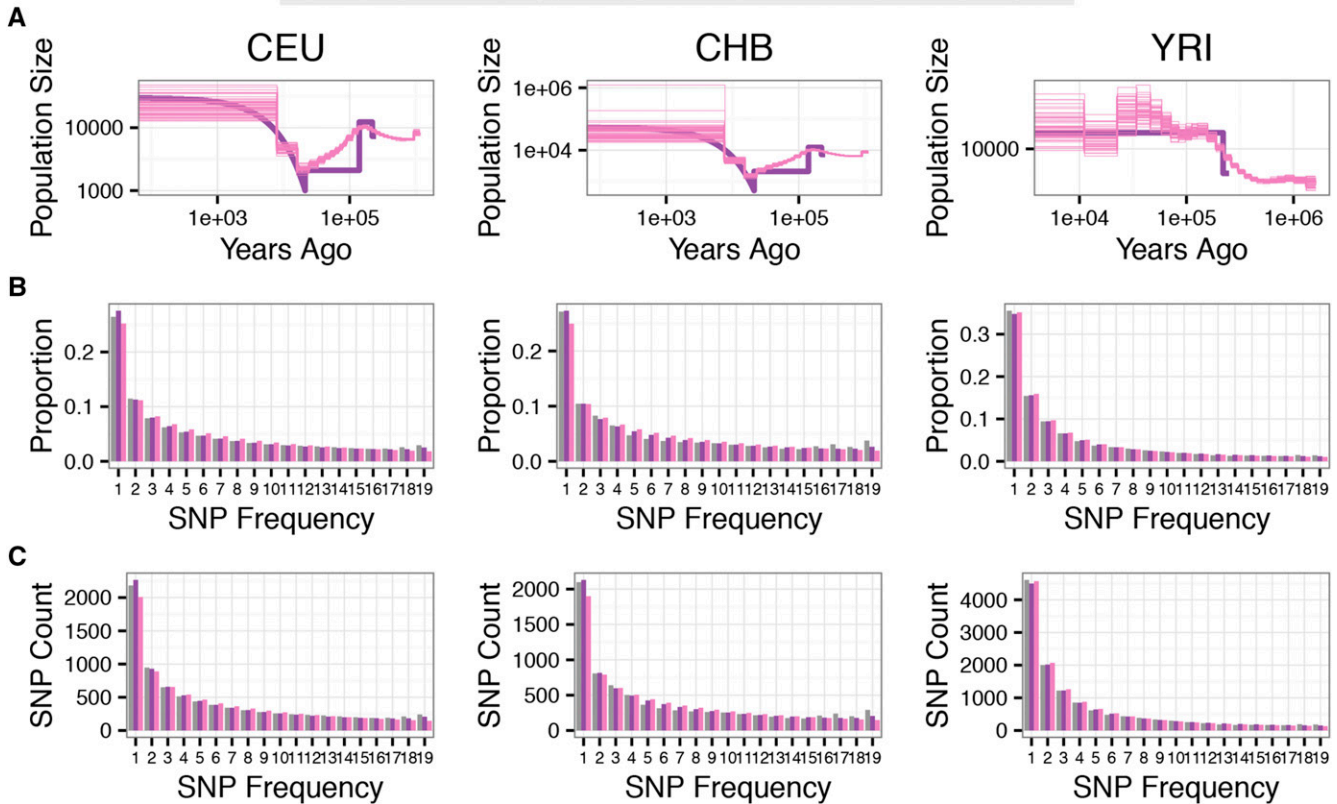
Figure 7 MSMC 2-haplotype can accurately infer the demographic model predicted by Gutenkunst *et al.* (2009). (A) The results of running MSMC 2-haplotype on 50 independent two-haplotype data sets simulated under the Gutenkunst *et al.* (2009) model of human demographic history (Gutenkunst, heavy purple line). The resulting MSMC 2-haplotype trajectories ("MSMC Sim. Gut. Data," fine pink lines) show the MSMC trajectories inferred from these 50 data sets. Note that these trajectories accurately track the demographic model used to simulate the data. (B) and (C) show proportional and SNP count SFSs for each population, respectively. The gray bars (observed) denote the empirical SFS used by Gutenkunst *et al.* (2009). The purple bars denote the expected SFS under the inferred Gutenkunst demographic models. The pink bars denote the expected SFS under the average of the 50 MSMC 2-haplotype demographic model trajectories for each population. Note that these three SFSs agree.

ancestral population size of 7300–12,300, showed the most improvement (Figures S11 and S12 in File S1). However, their fit was still >100 log-likelihood units worse than the Gutenkunst model (Figure S12 and Table S6 in File S1).

MSMC population size trajectories for demographic models inferred from the SFS

Given that the SFSs predicted by the demographic models inferred using MSMC do not fit the observed SFS, we examined whether MSMC is capable of recovering a complex demography such as the one inferred by Gutenkunst *et al.* (2009) from a single simulated genome. We find that MSMC performs relatively well at inferring the underlying demography from the simulated data. Figure 7A shows the underlying Gutenkunst demographic model for each population (purple) (as in the other Gutenkunst model simulations, migration is included in the model, but is not depicted in our diagrams), with the results of 50 independent MSMC inferences on each 2-haplotype simulated data set coming close to the underlying demography. However, sharp bottlenecks are inferred as long population declines (as noted by Li and Durbin 2011 and Schiffels and Durbin 2014). Additionally, we found evidence of MSMC detecting a false spurt of growth in the YRI population 1350 generations ago

(Figure 7A). Both of these phenomena were also noted by Bunnefeld *et al.* (2015).

The SFSs predicted by the demographic models inferred using MSMC on the simulated data fit the SFS expected under the Gutenkunst model and the observed Gutenkunst SFSs better than the MSMC demographic models inferred by Schiffels and Durbin (2014) (Figure 7, B and C). The proportional MSMC simulated data SFSs were only 40, 74, and 10 log-likelihood units below the Gutenkunst model SFS (Table S2 in File S1), with the SFSs based on SNP counts showing a similar pattern (Table S3 in File S1). Therefore, if the Gutenkunst model is the true demographic model for human history, MSMC accurately captures the population size changes and produces an appropriate SFS.

It is well established that two-haplotype, whole genome-based inference (PSMC and MSMC 2-haplotype, also known as PSMC') is not able to detect recent demographic events (Li and Durbin 2011; Schiffels and Durbin 2014). However, the ability to detect recent growth by using more than two haplotypes in the inference is cited as a feature of MSMC (Schiffels and Durbin 2014). We ran MSMC 2-haplotype and 8-haplotype on data sets simulated under the Gutenkunst model and a Gutenkunst model plus extreme recent growth (Figure S8 and Supplementary Note 6 in File S1). Unsurprisingly, MSMC 2-haplotype was not able to detect extreme recent growth. Its estimates of current population size were fairly accurate for the original Gutenkunst model

(Figure 7A), but the method dramatically underestimated the growth for data simulated under the Gutenkunst plus growth model (Figure S8 in File S1). The results from 8-haplotype MSMC inference were most surprising. We found that for both models, MSMC 8-haplotype inferred extreme recent growth as many as four orders of magnitude beyond that in the underlying model, with a high degree of variance between replicates (Figure S8 in File S1). Despite the high degree of variance, the average of the MSMC trajectories all showed a strong upward bias in estimates of the recent past (Figure S8 in File S1). While the ability to detect recent growth is meant to be a feature of MSMC, our findings indicate that the magnitude of growth may not be estimated well.

We had hypothesized that Neanderthal admixture could cause deviation between the MSMC and Gutenkunst demographic models, but found that the addition of Neanderthal admixture to our Gutenkunst model simulations did not substantively change the MSMC trajectories or expected SFSs (Figure S9, Supplementary Note 6, and Tables S2 and S3 in File S1).

DISCUSSION

We tested which published models of human demographic history, inferred using either whole genome sequence data, the SFS, or a combined approach, can recapitulate multiple summaries of human genetic variation data. We found that no model was able to recapitulate all summaries of the data, but some models still performed better than others. In particular, none of the models was able recapitulate LD decay, but the Gutenkunst SFS-based models and the combined whole genome and SFS-based SMC++ models were able to recapitulate empirical heterozygosity and the SFS. MSMC 2-haplotype was able to recapitulate heterozygosity, but not the SFS, and MSMC 4-haplotype and 8-haplotype could fit neither heterozygosity nor the SFS, though MSMC 8-haplotype did fit LD decay slightly better than the other models. These results highlight the uncertainties of demographic inference from one, or even two, types of data and the need to assess the fit of demographic models using multiple summaries of the data.

We found that the models based on MSMC inference from four or eight haplotypes did not improve the fit of the expected SFS compared to that based on two haplotypes; in fact, in most cases the 4- and 8-haplotype models fit much worse than the 2-haplotype models. The 4-haplotype models for CEU, CHB, and YRI and the 8-haplotype model for YRI appear to fit poorly due to their extremely high ancestral sizes and ancient humps of growth and decline (Figure 1). The expected SFSs under the 8-haplotype models for CEU and CHB show a skew toward low-frequency variants that may be due to their low ancestral size followed by extreme recent growth (Figure 1). We find that MSMC 8-haplotype vastly overestimates recent growth in simulated data, which may be contributing to the lack of fit to the SFS (Figure S8 in File S1). This result is at odds with the findings of Schiffels and Durbin (2014), who suggested that using eight haplotypes instead of two should increase accuracy of population size inference in the recent past, though they also noted a bias toward smaller ancient population sizes when using an increased number of haplotypes. Changing the scaling of the mutation rate did not generally help the MSMC models to fit the expected SFS better (Figures S4–S6 in File S1). It is worth noting that the model inferred in SMC++ used the same mutation rate as MSMC, yet fit the empirical SFSs much better (Figure 4, Figure 5, Figure 6, Table 1, and Tables S2–S4 in File S1), indicating that mutation rate differences between the whole genome- and SFS-based studies is not the source of the discrepancies.

We found that in addition to not fitting the empirical SFS, the MSMC 4-haplotype and 8-haplotype models did not predict the genome-wide

distribution of heterozygosity (Figure 2), which may be surprising as the genome-wide distribution of heterozygosity is a major feature of the data used by MSMC. As with the SFS, the reason for the lack of fit for these models appears to be the extremely high ancestral size inferred in the 4-haplotype models for all three populations and in the 8-haplotype YRI model, and the low ancestral size inferred in the 8-haplotype models for CEU and CHB (Figure 1).

Since the most ancient size in the MSMC trajectory will have a large influence on heterozygosity and the SFS and the most ancient bin of the MSMC trajectory may be unreliable (Li and Durbin 2011; Schiffels and Durbin 2014), we explored the effect of altering this ancient size and removing ancient growth events in the CEU MSMC 2-haplotype model. We found that selective trimming could improve the fit to the SFS (Figures S10–S13 in File S1). However, the final bin of the model cannot explain all of the lack of fit of the MSMC models to the data as the CEU and CHB MSMC 8-haplotype trajectories do not show the extreme ancestral sizes in the last bin, yet these models also dramatically deviate from empirical heterozygosity and the SFS. In other words, simple exclusion of the final high ancestral size is not sufficient to improve model fit to other summaries of the data. Our trimming experiments were only made possible by the abundance of human sequence data and demographic models previously fit to the data. Since many MSMC trajectories are calculated for species for which there is no prior information about ancient demographic history, the "informed trimming" we carried out is not a practicable solution to improve the reliability of MSMC inference.

While our results indicate that features of MSMC trajectories, particularly ancient events, should be regarded with caution, we also found that MSMC 2-haplotype is able to accurately recapitulate a complex demography (with the exception of steep drops in population size, extreme recent growth, and some false periods of growth) from simulated data, supporting the validity of the method, at least for use on simulated data (Figure 7). Migration between populations did not appear to cause deviations in MSMC trajectories from the underlying model (Figure 7), nor did a small degree of Neanderthal admixture (Figure S9 in File S1), indicating that MSMC is robust to small amounts of gene flow. The fact that the 2-haplotype model based on real data did not fit the observed SFS very well (Figure 4, Figure 5, Figure 6, Table 1, and Tables S2–S4 in File S1) suggests that the true underlying pattern of human demography is more complex than either type of inference ($\partial a \partial i$ or MSMC) is capturing, potentially revealing weaknesses in both methods.

Alternatively, if the Gutenkunst *et al.* (2009) demographic model is largely accurate, biases or other factors that exist in real data but not in simulated data may be affecting MSMC inference, resulting in the method failing to recover an underlying demography that matches Gutenkunst *et al.*'s (2009) model. For example, Song *et al.* (2016) found that statistical phasing could affect MSMC estimates of population split times; and Nadachowska-Brzyska *et al.* (2016) found that per-site sequencing depth, mean genome coverage, and the amount of missing data led to differences in PSMC curve amplitudes, expansions and contractions, and the timing and values of N_e. They therefore recommended only using samples with a mean genome coverage of $\geq 18 \times$ and $<25\%$ missing data, and employing a per-site sequencing depth filter of ≥ 10 (Nadachowska-Brzyska *et al.* 2016). The Complete Genomics genomes used by Li and Durbin (2011) were $>40 \times$ coverage (Drmanac *et al.* 2010), indicating that lack of coverage is not responsible for their divergence from estimates based on the SFS. However, the standards suggested by Nadachowska-Brzyska *et al.* (2016) may not always be attainable in *de novo* genome projects and, thus, data quality issues may affect nonmodel organism PSMC and MSMC inferences

more acutely. Future work should also examine the impact of artifacts of genome assembly errors and structural variants on PSMC inference. For example, collapsing duplicate regions of the genome on top of each other could result in regions of the genome having excess heterozygosity, which could in turn affect inference of demography.

We found that no model was able to accurately recapitulate the empirical distribution of LD decay. The lack of fit of the SFS-based models is perhaps unsurprising, as Harris and Nielsen (2013) found that the Gutenkunst model cannot recapitulate empirical IBS distributions (a finer-scale summary of the data related to LD), and Garud et al. (2015) found that they could not recover empirical LD patterns in Drosophila, despite matching the SFS, number of segregating sites (S), and number of pairwise differences (π). Garud et al. (2015) suggested the lack of fit could either be due to linked positive selection or to an incompleteness of the demographic model, demonstrating how models that fit some summaries of the data may not recapitulate others. It is more surprising that the MSMC 2-haplotype and 4-haplotype models do not fit the data well, as the method uses LD information in its inference, though different summaries of LD may be affected by demography in distinct ways (Plagnol and Wall 2006). Other possible factors that could lead to the lack of fit of all models to empirical LD decay patterns include the absence of natural selection, gene conversion, and fine-scale recombination hotspots in our simulations (Ardlie et al. 2001; Frisse et al. 2001; Wall and Pritchard 2003). Further, if the true mutation rate is actually smaller than the relatively high value used by Gutenkunst et al. ($\mu = 2.35 \times 10^{-8}$ mutations/bp/generation), then the population sizes would have to be larger than those estimated by Gutenkunst et al. (2009). Larger population sizes would yield larger values of the population scaled recombination rate (ρ) than what was used in our simulations under the Gutenkunst model. Larger values of ρ would then lead to a decrease in LD in the simulations, which might better match the empirical LD decay curves.

Natural selection may affect both SFS- and whole genome-based methods of demographic inference. Li and Durbin (2011) found that masking exonic sequence did not alter PSMC trajectories. However, Schrider et al. (2016) examined the impact of selective sweeps on demographic inference using the SFS in $\partial a \partial i$, approximate Bayesian computation (ABC), and PSMC and found that all three methods were influenced to varying degrees and in slightly different directions by the presence of selective sweeps, with $\partial a \partial i$ the most robust to these effects. This is a concern for published human demographic models as Gutenkunst et al. (2009) used noncoding sequence from autosomal genes in their study, which may be subject to linked selection (Gazave et al. 2014; Schrider et al. 2016). Schiffels and Durbin (2014) used whole genome sequences that included genic and nongenic regions, some of which are certainly under selection. Thus, the sensitivity of these methods to selection may partially explain why both perform well on simulated data without selection, yet have such divergent results when run on empirical data.

Our results have implications for understanding human demographic history. First, there has been controversy concerning the presence of ancient bottlenecks (>100 KYA) in human populations (Takahata et al. 1995; Harpending et al. 1998; Takahata and Satta 1998; Hawks et al. 2000; Garrigan and Hammer 2006; Fagundes et al. 2007; Scholz et al. 2007; Blum and Jakobsson 2011; Sjödin et al. 2012). The inferred humps in the ancient portions of MSMC plots (Figure 1) tended to lend support to these ancient population size changes that appeared to be absent from SFS demographic estimates. Our results suggest that if these ancient population size changes did indeed occur, the resulting SFS would appear very different from the SFSs seen in human populations (Figure 4, Figure 5, Figure 6, and

Figures S10–S13 in File S1). The fact that they are not seen in the observed SFS suggests that either the size changes did not occur, and the inferred size changes are artifacts, or instead, the true demography is more complex than currently modeled using either approach. Our conclusion of finding little evidence for the ancient population size changes is supported by the study of Sjödin et al. (2012). They employed an ABC approach to directly test models with ancient population size changes in Africa and found little support for such ancient bottlenecks.

Deep ancestral structure has been put forward as explanation for the humps detected by the whole genome-based methods by the developers of PSMC and others (Li and Durbin 2011; Henn et al. 2012; Mazet et al. 2015a,b; Orozco-terWengel 2016). While Blum and Jakobsson (2011) used the TMRCA to postulate an ancient bottleneck 150 KYA, they also were not able to reject a model of ancestral structure. Strikingly, Mazet et al. (2015b) were able to perfectly recapitulate the human PSMC humps without invoking a single size change in the population by simulating data from a highly structured ancestral population (10 subpopulations) and modulating the amount of gene flow between these populations. Therefore, the large "population size changes" inferred in MSMC, which cause the models not to match the empirical SFS, may in fact be due to complex structure and large-scale changes in gene flow. This ancient structure may have a large effect on MSMC trajectories and LD patterns, but may not strongly influence the SFS (see figure 7 in Lohmueller et al. 2009), potentially resolving the discrepancy between the methods (Henn et al. 2012).

Our work provides a cautionary tale for understanding population history in nonmodel organisms. Our results argue against a literal interpretation of humps and other jumps in MSMC plots as reflecting population size changes. This problem is exacerbated for putative ancient size changes. Given the ever-increasing generation of genomic data from nonmodel taxa and the application of whole genome-based approaches to such data (Groenen et al. 2012; Meyer et al. 2012; Zhao et al. 2012; Albert et al. 2013; Ibarra-Laclette et al. 2013; Nadachowska-Brzyska et al. 2013, 2015, 2016; Orlando et al. 2013; Prado-Martinez et al. 2013; Bosse et al. 2014; Freedman et al. 2014; Hung et al. 2014; Prufer et al. 2014; Palkopoulou et al. 2015; Holliday et al. 2016; Wang et al. 2016), our findings are especially concerning. We recommend employing other model-based types of demographic inference leveraging either SFS-based or other summary statistics in an ABC framework to test whether important demographic features suggested by PSMC or MSMC plots can be recapitulated using other features in the data. We also recommend, as done in Thornton and Andolfatto (2006), Freedman et al. (2014), Cahill et al. (2016), Hsieh et al. (2016), and Song et al. (2016), that the PSMC or MSMC plots and TMRCA estimates be used themselves as summary statistics for model comparison, rather than the actual population size estimates. In other words, more complex demographic models can be simulated and tested to see whether they recapitulate the observed whole genome-based trajectories. Of course, this approach will not be successful if the trajectories are strongly influenced by bioinformatic artifacts or other features not captured within the simulations, such as natural selection. For both PSMC/MSMC and SFS-based inference methods, we also recommend testing whether the estimated models can predict multiple features of the data. Specifically, researchers should check whether their inferred model can recapitulate the genome-wide distribution of heterozygosity. The genome-wide distribution of heterozygosity may be the most practical and useful statistic for studies of nonmodel organisms that only have a handful of genomes available to them. SMC++ and other new approaches that leverage multiple types of data (Bunnefeld et al. 2015; Boitard et al. 2016; Weissman and Hallatschek 2017) are promising

alternatives, though our results indicate that SMC++ still cannot recapitulate all summaries of the data.

Testing more complex demographic scenarios using multiple summaries of the data may help to resolve uncertainties about our own species' history and will improve demographic inference for nonmodel organisms. Incorporating the potential complexity of possible demographic histories to produce models that better recapitulate the data may in fact present the greatest challenge.

Author contributions: K.E.L. and A.C.B. conceived the study. A.C.B. carried out all analyses based on the demographic models, and T.N.P. carried out all empirical analyses based on the 1000 Genomes data. A.C.B. generated all figures. A.C.B., T.N.P., and K.E.L. all participated in manuscript preparation.

LITERATURE CITED

1000 Genomes Project Consortium, 2015   A global reference for human genetic variation. Nature 526: 68–74.

Adams, A. M., and R. R. Hudson, 2004   Maximum-likelihood estimation of demographic parameters using the frequency spectrum of unlinked single-nucleotide polymorphisms. Genetics 168: 1699–1712.

Albert, V. A., W. B. Barbazuk, J. P. Der, J. Leebens-Mack, H. Ma et al., 2013   The Amborella genome and the evolution of flowering plants. Science 342: 1241089.

Arbiza, L., E. Zhong, and A. Keinan, 2012   NRE: a tool for exploring neutral loci in the human genome. BMC Bioinformatics 13: 301.

Ardlie, K., S. N. Liu-Cordero, M. A. Eberle, M. Daly, J. Barrett et al., 2001   Lower-than-expected linkage disequilibrium between tightly linked markers in humans suggests a role for gene conversion. Am. J. Hum. Genet. 69: 582–589.

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

Blum, M. G., and M. Jakobsson, 2011   Deep divergences of human gene trees and models of human origins. Mol. Biol. Evol. 28: 889–898.

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

Bosse, M., H.-J. Megens, O. Madsen, L. A. Frantz, Y. Paudel et al., 2014   Untangling the hybrid nature of modern pig genomes: a mosaic derived from biogeographically distinct and highly divergent Sus scrofa populations. Mol. Ecol. 23: 4089–4102.

Boyko, A. R., S. H. Williamson, A. R. Indap, J. D. Degenhardt, R. D. Hernandez et al., 2008   Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet. 4: e1000083.

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

Cahill, J. A., A. E. Soares, R. E. Green, and B. Shapiro, 2016   Inferring species divergence times using pairwise sequential Markovian coalescent

modelling and low-coverage genomic data. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 371: 20150138.

Chen, G. K., P. Marjoram, and J. D. Wall, 2009   Fast and flexible simulation of DNA sequence data. Genome Res. 19: 136–142.

Chikhi, L., V. C. Sousa, P. Luisi, B. Goossens, and M. A. Beaumont, 2010   The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. Genetics 186: 983–995.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks et al., 2011   The variant call format and VCFtools. Bioinformatics 27: 2156–2158.

Drmanac, R., A. B. Sparks, M. J. Callow, A. L. Halpern, N. L. Burns et al., 2010   Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science 327: 78–81.

Ewing, G., and J. Hermisson, 2010   MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. Bioinformatics 26: 2064–2065.

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

Fagundes, N. J., N. Ray, M. Beaumont, S. Neuenschwander, F. M. Salzano et al., 2007   Statistical evaluation of alternative models of human evolution. Proc. Natl. Acad. Sci. USA 104: 17614–17619.

Fitak, R. R., E. Mohandesan, J. Corander, and P. A. Burger, 2016   The de novo genome assembly and annotation of a female domestic dromedary of North African origin. Mol. Ecol. Resour. 16: 314–324.

Freedman, A. H., I. Gronau, R. M. Schweizer, D. Ortega-Del Vecchyo, E. Han et al., 2014   Genome sequencing highlights the dynamic early history of dogs. PLoS Genet. 10: e1004016.

Frisse, L., R. R. Hudson, A. Bartoszewicz, J. D. Wall, J. Donfack et al., 2001   Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. Am. J. Hum. Genet. 69: 831–843.

Gao, F., and A. Keinan, 2016   Explosive genetic evidence for explosive human population growth. Curr. Opin. Genet. Dev. 41: 130–139.

Garrigan, D., and M. F. Hammer, 2006   Reconstructing human origins in the genomic era. Nat. Rev. Genet. 7: 669–680.

Garud, N. R., P. W. Messer, E. O. Buzbas, and D. A. Petrov, 2015   Recent selective sweeps in North American Drosophila melanogaster show signatures of soft sweeps. PLoS Genet. 11: e1005004.

Gattepaille, L. M., M. Jakobsson, and M. G. Blum, 2013   Inferring population size changes with sequence and SNP data: lessons from human bottlenecks. Heredity 110: 409–419.

Gazave, E., L. Ma, D. Chang, A. Coventry, F. Gao et al., 2014   Neutral genomic regions refine models of recent rapid human population growth. Proc. Natl. Acad. Sci. USA 111: 757–762.

Gravel, S., B. M. Henn, R. N. Gutenkunst, A. R. Indap, G. T. Marth et al., 2011   Demographic history and rare allele sharing among human populations. Proc. Natl. Acad. Sci. USA 108: 11983–11988.

Groenen, M. A. M., A. L. Archibald, H. Uenishi, C. K. Tuggle, Y. Takeuchi et al., 2012   Analyses of pig genomes provide insight into porcine demography and evolution. Nature 491: 393–398.

Gutenkunst, R. N., R. D. Hernandez, S. H. Williamson, and C. D. Bustamante, 2009   Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5: e1000695.

Han, E., J. S. Sinsheimer, and J. Novembre, 2014   Characterizing bias in population genetic inferences from low-coverage sequencing data. Mol. Biol. Evol. 31: 723–735.

Han, E., J. S. Sinsheimer, and J. Novembre, 2015   Fast and accurate site frequency spectrum estimation from low coverage sequence data. Bioinformatics 31: 720–725.

Harpending, H. C., M. A. Batzer, M. Gurven, L. B. Jorde, A. R. Rogers et al., 1998   Genetic traces of ancient demography. Proc. Natl. Acad. Sci. USA 95: 1961–1967.

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

Hawks, J., K. Hunley, S.-H. Lee, and M. Wolpoff, 2000   Population bottlenecks and Pleistocene human evolution. Mol. Biol. Evol. 17: 2–22.

Heller, R., L. Chikhi, and H. R. Siegismund, 2013   The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. PLoS One 8: 1–10.

Henn, B. M., L. L. Cavalli-Sforza, and M. W. Feldman, 2012   The great human expansion. Proc. Natl. Acad. Sci. USA 109: 17758–17764.

Henn, B. M., L. R. Botigué, S. Peischl, I. Dupanloup, M. Lipatov *et al.*, 2016   Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. Proc. Natl. Acad. Sci. USA 113: E440–E449.

Holliday, J. A., L. Zhou, R. Bawa, M. Zhang, and R. W. Oubida, 2016   Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in *Populus trichocarpa*. New Phytol. 209: 1240–1251.

Hsieh, P., K. R. Veeramah, J. Lachance, S. A. Tishkoff, J. D. Wall *et al.*, 2016   Whole-genome sequence analyses of Western Central African Pygmy hunter-gatherers reveal a complex demographic history and identify candidate genes under positive natural selection. Genome Res. 26: 279–290.

Hudson, R. R., 2002   Generating samples under a Wright–Fisher neutral model of genetic variation. Bioinformatics 18: 337–338.

Hung, C.-M., P.-J. L. Shaner, R. M. Zink, W.-C. Liu, T.-C. Chu *et al.*, 2014   Drastic population fluctuations explain the rapid extinction of the passenger pigeon. Proc. Natl. Acad. Sci. USA 111: 10636–10641.

Ibarra-Laclette, E., E. Lyons, G. Hernández-Guzmán, C. A. Pérez-Torres, L. Carretero-Paulet *et al.*, 2013   Architecture and evolution of a minute plant genome. Nature 498: 94–98.

Jouganous, J., W. Long, A. P. Ragsdale, and S. Gravel, 2017   Inferring the joint demographic history of multiple populations: beyond the diffusion approximation. Genetics 206: 1549–1567.

Keinan, A., and A. G. Clark, 2012   Recent explosive human population growth has resulted in an excess of rare genetic variants. Science 336: 740–743.

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

Kidd, J. M., S. Gravel, J. Byrnes, A. Moreno-Estrada, S. Musharoff *et al.*, 2012   Population genetic inference from personal genome data: impact of ancestry and admixture on human genomic variation. Am. J. Hum. Genet. 91: 660–671.

Kim, S. Y., K. E. Lohmueller, A. Albrechtsen, Y. Li, T. Korneliussen *et al.*, 2011   Estimation of allele frequency and association mapping using next-generation sequencing data. BMC Bioinformatics 12: 231.

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

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

Liu, X., and Y.-X. Fu, 2015   Exploring population size changes using SNP frequency spectra. Nat. Genet. 47: 555–559.

Lohmueller, K. E., C. D. Bustamante, and A. G. Clark, 2009   Methods for human demographic inference using haplotype patterns from genome-wide single-nucleotide polymorphism data. Genetics 182: 217–231.

Malaspinas, A. S., M. C. Westaway, C. Muller, V. C. Sousa, O. Lao *et al.*, 2016   A genomic history of Aboriginal Australia. Nature 538: 207–214.

Mallick, S., H. Li, M. Lipson, I. Mathieson, M. Gymrek *et al.*, 2016   The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature 538: 201–206.

Marjoram, P., and J. D. Wall, 2006   Fast "coalescent" simulation. BMC Genet. 7: 1–9.

Marth, G. T., E. Czabarka, J. Murvai, and S. T. Sherry, 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.

Mazet, O., W. Rodriguez, and L. Chikhi, 2015a   Demographic inference using genetic data from a single individual: separating population size variation from population structure. Theor. Popul. Biol. 104: 46–58.

Mazet, O., W. Rodriguez, S. Grusea, S. Boitard, and L. Chikhi, 2015b   On the importance of being structured: instantaneous coalescence rates and

human evolution—lessons for ancestral population size inference? Heredity 116: 362–371.

McCoy, R. C., N. R. Garud, J. L. Kelley, C. L. Boggs, and D. A. Petrov, 2014   Genomic inference accurately predicts the timing and severity of a recent bottleneck in a nonmodel insect population. Mol. Ecol. 23: 136–150.

McVean, G. A., and N. J. Cardin, 2005   Approximating the coalescent with recombination. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360: 1387–1393.

Meyer, M., M. Kircher, M.-T. Gansauge, H. Li, F. Racimo *et al.*, 2012   A high-coverage genome sequence from an archaic Denisovan individual. Science 338: 222–226.

Murray, G. G. R., A. E. R. Soares, B. J. Novak, N. K. Schaefer, J. A. Cahill *et al.*, 2017   Natural selection shaped the rise and fall of passenger pigeon genomic diversity. bioRxiv . Available at:: https://doi.org/10.1101/154294.

Nadachowska-Brzyska, K., R. Burri, P. I. Olason, T. Kawakami, L. A. Smeds *et al.*, 2013   Demographic divergence history of pied flycatcher and collared flycatcher inferred from whole-genome re-sequencing data. PLoS Genet. 9: e1003942.

Nadachowska-Brzyska, K., C. Li, L. Smeds, G. Zhang, and H. Ellegren, 2015   Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. Curr. Biol. 25: 1375–1380.

Nadachowska-Brzyska, K., R. Burri, L. Smeds, and H. Ellegren, 2016   PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. Mol. Ecol. 25: 1058–1072.

Nei, M., T. Maruyama, and R. Chakraborty, 1975   The bottleneck effect and genetic variability in populations. Evolution 29: 1–10.

Nelson, M. R., D. Wegmann, M. G. Ehm, D. Kessner, P. S. Jean *et al.*, 2012   An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science 337: 100–104.

Nielsen, R., 2000   Estimation of population parameters and recombination rates from single nucleotide polymorphisms. Genetics 154: 931–942.

Nielsen, R., M. J. Hubisz, I. Hellmann, D. Torgerson, A. M. Andres *et al.*, 2009   Darwinian and demographic forces affecting human protein coding genes. Genome Res. 19: 838–849.

Nielsen, R., J. S. Paul, A. Albrechtsen, and Y. S. Song, 2011   Genotype and SNP calling from next-generation sequencing data. Nat. Rev. Genet. 12: 443–451.

Orlando, L., A. Ginolhac, G. Zhang, D. Froese, A. Albrechtsen *et al.*, 2013   Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. Nature 499: 74–78.

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

Pagani, L., D. J. Lawson, E. Jagoda, A. Mörseburg, A. Eriksson *et al.*, 2016   Genomic analyses inform on migration events during the peopling of Eurasia. Nature 538: 238–242.

Palkopoulou, E., S. Mallick, P. Skoglund, J. Enk, N. Rohland *et al.*, 2015   Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. Curr. Biol. 25: 1395–1400.

Peter, B. M., D. Wegmann, and L. Excoffier, 2010   Distinguishing between population bottleneck and population subdivision by a Bayesian model choice procedure. Mol. Ecol. 19: 4648–4660.

Phung, T. N., C. D. Huber, and K. E. Lohmueller, 2016   Determining the effect of natural selection on linked neutral divergence across species. PLoS Genet. 12: e1006199.

Plagnol, V., and J. D. Wall, 2006   Possible ancestral structure in human populations. PLoS Genet. 2: e105.

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

Prado-Martinez, J., P. H. Sudmant, J. M. Kidd, H. Li, J. L. Kelley *et al.*, 2013   Great ape genetic diversity and population history. Nature 499: 471–475.

Prufer, K., F. Racimo, N. Patterson, F. Jay, S. Sankararaman *et al.*, 2014   The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505: 43–49.

Ptak, S. E., and M. Przeworski, 2002 Evidence for population growth in humans is confounded by fine-scale population structure. Trends Genet. 18: 559–563.

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

Scholz, C. A., T. C. Johnson, A. S. Cohen, J. W. King, J. A. Peck *et al.*, 2007 East African megadroughts between 135 and 75 thousand years ago and bearing on early-modern human origins. Proc. Natl. Acad. Sci. USA 104: 16416–16421.

Schraiber, J. G., and J. M. Akey, 2015 Methods and models for unravelling human evolutionary history. Nat. Rev. Genet. 16: 727–740.

Schrider, D. R., A. G. Shanku, and A. D. Kern, 2016 Effects of linked selective sweeps on demographic inference and model selection. Genetics 204: 1207–1223.

Sjödin, P., A. E. Sjöstrand, M. Jakobsson, and M. G. Blum, 2012 Resequencing data provide no evidence for a human bottleneck in Africa during the penultimate glacial period. Mol. Biol. Evol. 29: 1851–1860.

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

Song, S., E. Sliwerska, S. Emery, and J. M. Kidd, 2016 Modeling human population separation history using physically phased genomes. Genetics 205: 385–395.

Sovic, M. G., B. C. Carstens, and H. L. Gibbs, 2016 Genetic diversity in migratory bats: results from RADseq data for three tree bat species at an Ohio windfarm. PeerJ 4: e1647.

Tajima, F., 1989 The effect of change in population size on DNA polymorphism. Genetics 123: 597–601.

Takahata, N., and Y. Satta, 1998 Footprints of intragenic recombination at HLA loci. Immunogenetics 47: 430–441.

Takahata, N., Y. Satta, and J. Klein, 1995 Divergence time and population size in the lineage leading to modern humans. Theor. Popul. Biol. 48: 198–221.

Tennessen, J. A., A. W. Bigham, T. D. O'Connor, W. Fu, E. E. Kenny *et al.*, 2012 Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337: 64–69.

Terhorst, J., and Y. S. Song, 2015 Fundamental limits on the accuracy of demographic inference based on the sample frequency spectrum. Proc. Natl. Acad. Sci. USA 112: 7677–7682.

Terhorst, J., J. A. Kamm, and Y. S. Song, 2017 Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat. Genet. 49: 303–309.

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

Trucchi, E., P. Gratton, J. D. Whittington, R. Cristofari, Y. Le Maho *et al.*, 2014 King penguin demography since the last glaciation inferred from genome-wide data. Proc. Biol. Sci. 281: 20140528.

Wakeley, J., 2009 *Coalescent Theory: An Introduction*. Roberts & Co. Publishers, Greenwood Village, CO.

Wall, J. D., and J. K. Pritchard, 2003 Haplotype blocks and linkage disequilibrium in the human genome. Nat. Rev. Genet. 4: 587–597.

Wang, G.-D., W. Zhai, H.-C. Yang, L. Wang, L. Zhong *et al.*, 2016 Out of southern East Asia: the natural history of domestic dogs across the world. Cell Res. 26: 21–33.

Weissman, D. B., and O. Hallatschek, 2017 Minimal-assumption inference from population-genomic data. Elife 6: e24836.

Zhao, S., P. Zheng, S. Dong, X. Zhan, Q. Wu *et al.*, 2012 Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. Nat. Genet. 45: 67–71.

Communicating editor: B. Andrews