Accurate and flexible estimation of effective population size history

Zhendong Huang[1], Yao-ban Chan[1], and David Balding[1]

[1]Melbourne Integrative Genomics, School of Mathematics & Statistics, University of Melbourne, Australia

December 4, 2025

Abstract

Current methods for inferring historical population sizes from DNA sequences often impose a heavy computational burden, or relieve that burden by imposing a fixed parametric form. In addition, they can be marred by sequencing errors or uncertainty about recombination rates, and the quality of inference is often poor in the recent past. We propose "InferNo" for flexible, nonparametric inference of effective population sizes. It requires modest computing resources and little prior knowledge of the recombination and mutation maps, and it is robust to sequencing error and gene conversion. We illustrate the statistical and computational advantages of InferNo over previous approaches using a range of simulation scenarios. In particular, we demonstrate the ability of InferNo to exploit biobank-scale datasets for accurate inference of rapid population size changes in the recent past. Applying InferNo to worldwide human data, we find remarkable similarities in inferences from different populations in the same region. The historical sizes of all non-African populations converge around

40 000 years ago, but they remain distinct from those of African populations for over 800 000 years, long predating estimates for the migration of modern humans out of Africa.

Keywords: effective population size, semi-analytical inference, human demographic history.

Introduction

We propose a new method, "InferNo", which takes DNA sequence data and infers effective population size history, providing clues about migrations, plagues and environmental events. Some features of InferNo:

- It is computationally and statistically efficient for the recent and distant past. In particular, it can exploit biobank-scale datasets to provide accurate inferences of recent population size changes.
- The inference is nonparametric, requiring no pre-specified population size model.
- Variation in mutation rate along the genome is accommodated, only an average needs to be pre-specified.
- An average recombination rate must be pre-specified, but InferNo is robust both to variation along the genome and misspecification of the average rate.
- It is robust to sequencing errors, requiring no information about the error rate.

Some existing approaches infer population size history from estimated times since the most recent common ancestor (TMRCA) along pairs of sequences, under the sequentially Markov coalescent (SMC) model. These include PSMC Li and Durbin (2011), MSMC Schiffels and Durbin (2014), diCal Druet et al. (2014), SMC++ Terhorst et al. (2017) and PHLASH Terhorst (2024). When the sample size n is small, these approaches can provide good inferences for ancient population sizes, but tend to perform poorly over recent history and

are computationally demanding for large n. Moreover, these approaches usually require good knowledge of the mutation and recombination rate maps across the genome to accurately specify the SMC, and they can be sensitive to sequencing errors.

A second class of approaches involves using the allele (or site) frequency spectrum (AFS, or SFS) of the genome sequences Gutenkunst et al. (2009); Excoffier et al. (2013); Bhaskar et al. (2015); Kamm et al. (2020). Variation in mutation rate along the genome presents a particular difficulty for these methods, and they can also be sensitive to sequencing errors. The AFS ignores the ordering of sites along the genome and so avoids any assumption about recombination, but the resulting loss of information can lead to an ill-posed inverse problem (a given AFS can arise from different demographic histories), especially for small n DeWitt et al. (2021).

A third class of approaches is based on inference of the ancestral recombination graph (ARG), which summarises the genealogical history of the sequences Palamara et al. (2012); Boitard et al. (2016); Speidel et al. (2019); Fournier et al. (2023). The true ARG provides all information available from the sample about population size history. However, ARG inference is imprecise, in part due to uncertainty about recombination and mutation rates, and computational costs can be high despite recent advances Rasmussen et al. (2014); Speidel et al. (2019); Kelleher et al. (2019); Mahmoudi et al. (2022); Deng et al. (2024b).

Population-size inference can also be based on inferred identity-by-descent (IBD) relationships or linkage disequilibrium (LD) coefficients Mezzavilla and Ghirotto (2015); Browning and Browning (2015); Browning et al. (2018); Ragsdale and Gravel (2019); Santiago et al. (2020). These methods are typically limited to recent history due to the challenging nature of inferring short IBD segments and the complex relationship of LD with demography.

InferNo uses the AFS in a novel way that improves on previous approaches, and provides further improvement by adding information from pairwise TMRCA estimates. It uses a simulation under the standard coalescent model to estimate the distribution of mutation events, then identifies a time scaling of the standard coalescent to best match the observed

AFS and TMRCA estimates. The inferred population size is proportional to the inverse of the time scaling.

In simulation studies, we find that InferNo is overall more accurate and computationally faster than the TMRCA-based PSMC Li and Durbin (2011) and CHIMP Upadhya and Steinrücken (2022), AFS-based Mushi DeWitt et al. (2021), ARG-based Relate Speidel et al. (2019), and IBD-based IBDNe Browning and Browning (2015). In application to human data, we demonstrate inferences of population size history that are remarkably similar for populations in the same region. We find evidence for two periods of population decline in the past 150 000 years for most non-African populations, and a single, more ancient population decline in ancestral African (AFR) populations. A substantial difference in the demographic histories of AFR and non-AFR populations dates back more than 800 000 years.

Methods

Overview of InferNo

The data are n homologous genome sequences of length ℓ, with alleles either 0 (ancestral) or 1 (derived); ancestral alleles can often be accurately inferred from related species. Required input also includes $\bar{\mu}$, a genome-average mutation rate (per generation per site). In deriving the method we assume that, at each site, at most one mutation has occurred since the MRCA, but we demonstrate robustness to this assumption by not incorporating it into the data generation models of the simulation study.

InferNo fits a coalescent model with effective population size that is a piece-wise constant function over time. By scaling time inversely proportional to population size, this size-varying model becomes equivalent to a standard coalescent model (the "null model") with population size constant over time. InferNo uses penalised least-squares regression to infer the time scaling and hence the population size function.

First, a simulation is performed under the null model, generating a dataset with the same

dimensions as the observed data. The null-model time axis is divided into intervals, and corresponding intervals are inferred in the size-varying model by relating observed statistics to sums over time intervals of simulation-generated statistics. One statistic is the AFS, which aggregates information across sequences and across sites. The AFS can suffice for large n, otherwise we also use TMRCA estimates from pairs of sequences within genome segments.

Below we describe the key features of each step. More details are given in Supporting Information (SI).

The null model and time scaling

The recombination rate r is assumed constant over sites and over time. The mutation rate is constant over time, but site-specific rates $\mu(s)$ can be pre-specified or estimated using a piecewise-constant smoothing of the fraction of sites that are polymorphic, scaled so that the average over all sites is $\bar{\mu}$ (see SI for details).

We initially assume that the observed sequences originate from a coalescent-with-recombination model with $N(g)$, the effective population size g generations ago, a piece-wise constant function. This variable-size model is equivalent to the constant-size null model (with population size $\tilde{N}$) with time scaled in proportion to $1/N(g)$ Griffiths and Tavaré (1994). We infer the time scaling by comparing the observed dataset to a new dataset with the same sample size and sequence length generated from the null model.

First, we partition the null-model time domain into intervals $\tilde{I}_t = [\tilde{g}_{t-1}, \tilde{g}_t)$, $t = 1, \ldots, T$, where $\tilde{g}_0 = 0$. By default, we choose the $\tilde{g}_t$ to approximately equalise the number of null-simulation mutations during each interval, but they can be pre-specified, for example, based on historical periods of interest or prior knowledge about $N(g)$.

We then infer time intervals $I_t = [g_{t-1}, g_t)$ in the variable-size model, with $g_0 = 0$. Within each I_t, the value of $N(g)$ is constant and such that the cumulative coalescence rate over I_t is the same as for $\tilde{I}_t$. It follows that the time interval lengths are inversely proportional to

population sizes, so that for $g \in I_t$,

$$\beta_t := \frac{N(g)}{\tilde{N}} = \frac{g_t - g_{t-1}}{\tilde{g}_t - \tilde{g}_{t-1}}, \qquad t = 1, \ldots, T, \tag{1}$$

where := indicates a definition.

The inference task is now, given our choices for $\tilde{N}$ and the $\tilde{g}_t$, to estimate the β_t and hence the g_t. The value of $\tilde{g}_T$ should be large enough that the inferred g_T predates the historic period of interest, but it should not be larger than the largest TMRCA arising in the null simulation as there is no information for inference prior to that time.

Using the AFS

The observed-data AFS is a length-n vector Y_1 with kth element equal to the number of sites at which the derived allele has frequency k. From the null simulation, we obtain the $n \times T$ matrix $\tilde{X}_1$ in which column t is an AFS restricted to sites at which the mutation arose during $\tilde{I}_t$ (oldest mutation if there is more than one). We can now infer β using the regression equation

$$Y_1 = \tilde{X}_1 \beta + \epsilon\,. \tag{2}$$

The resulting estimates are accurate when n is large, but for small n we seek additional estimating equations by relating TMRCA estimates between the observed data and null simulation.

Using pairwise TMRCA estimates

We partition the genome $[0, \ell)$ into segments $J_l = [h_{l-1}, h_l)$, $l = 1, \ldots, L$, containing roughly equal numbers of variable sites. Then, for each of n sequence pairs we count the sites in each J_l at which the sequences differ in order to estimate Y_2, a vector of quantiles of the

pairwise TMRCA distribution averaged over segments. Next, we estimate the corresponding quantiles from our null simulation, once again separated by the time interval of the mutation (the resulting matrix $\tilde{X}_2$ has a column for each time interval). Similar to (2), we can now infer β via

$$Y_2 = \tilde{X}_2\beta + \epsilon \ . \tag{3}$$

Combining the estimating equations

We combine [2] and [3] by defining $Y = ({Y_1}', \omega {Y_2}')'$ and $\tilde{X} = (\tilde{X}_1', \omega \tilde{X}_2')'$, where $'$ denotes matrix transpose and ω is a scalar weight. For large n, we set $\omega = 0$, thus ignoring [3] which greatly reduces computational effort.

We then estimate β by minimising a penalised least-squares based on

$$Y = \tilde{X}\beta + \epsilon, \tag{4}$$

with the smoothing penalty based on the second difference of β, parameterised by λ. See SI for selection of λ by minimising an adjusted Bayes Information Criterion (BIC).

An optional final step is to convert the piece-wise constant estimate $\tilde{N}\hat{\beta}$ into a continuous piece-wise linear function by connecting the interval midpoints.

Design of simulation studies

Table 1: Population sizes (in units of 1K sequences, K denotes 10^3) at growth-rate change points for eight simulation models. Growth rates are constant between change-points, see Table 5 for the rates. A blank entry indicates no change from the value to its left. $N(g) = N(30\text{K})$ for $g > 30\text{K}$.

Model	$N(0)$	N(25)	N(30)	N(300)	N(400)	N(1K)	N(1.8K)	N(1.9K)	N(3K)	N(9K)	N(18K)	N(18.1K)	N(30K)
1	20												
2	40	40	40	39	38	36	33	33	30	16	7	7	2
3	5	5	5	6	6	8	12	13	22	28	34		
4	40	40	40	38	37	33	28	27	22	12	16	16	23
5	40				10			33				20	
6	100	84	81	12	6	31				6	23		
7	3 000	20											
8	300	86	182	177	175	165	152	151	135	74	30	30	9

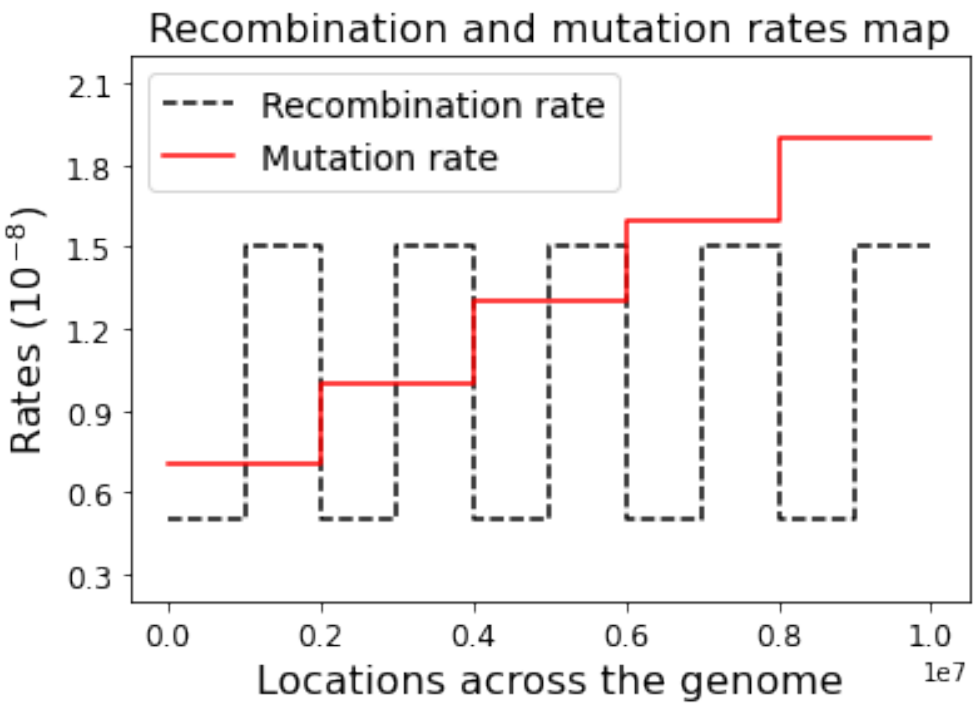

Figure 1: **Recombination and mutation maps used in the simulation study.** InferNo requires genome-wide average values as input, which for these maps are $r = 1$ and $\mu = 1.3$ per site per 10^8 generations.

To compare the performance of InferNo with Relate, PSMC, Mushi and CHIMP, we used msprime Baumdicker et al. (2021) to generate sequences of length $\ell = 10^7$, with sample sizes $n = 10$ and 200, under Models 1 to 6 (Table 1). We additionally compare with PSMC for $n = 2$, corresponding to a single diploid individual. We use 25 replicate datasets for each setting, and compare performance using RMISE (root mean integrated squared error).

Models 1 to 3 have constant, monotonic increasing and monotonic decreasing population sizes at varying rates. Models 4 to 6 have periods of both growth and decline, with population size changing smoothly in Model 4 and with jumps in Model 5, while Model 6 allows us to examine inference of an older bottleneck when its signal may be blurred by a more recent bottleneck.

See SI for implementation details of the comparison methods. For the InferNo analyses, genome-wide average recombination and mutation rates were set to the correct values for the mutation and recombination maps used in the simulations (Fig. 1), but the variation along

the genome is not input. To further challenge inferences, we include sequencing errors in the data simulations, with rate $\epsilon = 1$ per 10^5 sequence sites, and gene conversions with rate 2 per site per 10^8 generations and tract length 300 sites, which are close to recent estimates for humans Williams et al. (2015).

Using the same 6 models, we then assessed the performance of InferNo over sample sizes ($n = 10$, 50, 100, 500, and 1000 sequences) and under two variations to the data generating model that we now describe.

In the case $n = 100$, we checked the robustness of InferNo to misspecification of the recombination rate across the genome, by simulating datasets with (i) the recombination map given in Fig. 1, which has average rate $r = 10^{-8}$ per site and per generation, (ii) a constant rate $r = 10^{-8}$ (which matches the InferNo analysis model) and (iii) a constant rate $r = 1.5\times10^{-8}$.

Again with $n = 100$, we checked the robustness of InferNo to sequencing error. We ran InferNo on simulated datasets contaminated by sequencing error with rates $\epsilon = 0$, 1, 2, and 3 errors per 10^5 sites.

To focus on recent population size inference, we compared InferNo with the IBD-based approach IBDNe Browning and Browning (2015) using Model 7 (rapid, recent growth) with no sequencing error or gene conversion. For both methods, we used $n = 2\,000$, length $\ell = 10^8$, input the correct mutation and recombination rates $\mu = 1.3r = 1.3\times10^{-8}$ and set $\omega = 0$. Because of our focus on the recent past, we set time intervals for InferNo as $\tilde{g}_t = 0.08t$ for $t = 1, 2, 3$, and $\tilde{g}_t = 0.54\exp\{1.4(t-4)\}$ for $t = 4, \ldots, 12$. For IBDNe, we input IBD segments of length > 2 cM identified by hap-ibd Zhou et al. (2020) with the genetic map parameter set to true. Default values were used for other IBDNe and hap-ibd parameters.

Finally, we tested the ability of InferNo to infer very recent population size changes (Model 8) using a biobank-scale dataset ($n = 2\times10^5$) for which IBDNe is not tractable. We set $\ell = 5\times10^7$, similar to human chromosome 21, and kept $\mu = 1.3r = 1.3\times10^{-8}$. Sequencing error is included, at rate 1 per 10^5 sites, but not gene conversion. We set $\tilde{N} = 10^5$ in InferNo,

Table 2: Sample sizes (haploid) for human data analysis. The number of individuals is $n/2$.

Pop	Description	Sample size n	Region
LWK	Luhya in Webuye, Kenya	206	Africa (AFR)
MSL	Mende in Sierra Leone	180	
ESN	Esan in Nigeria	200	
YRI	Yoruba in Ibadan, Nigeria	214	
CHB	Han Chinese in Beijing, China	212	East Asia (EAS)
JPT	Japanese in Tokyo, Japan	210	
CHS	Southern Han Chinese	210	
CDX	Chinese Dai in Xishuangbanna, China	200	
FIN	Finnish in Finland	210	Europe (EUR)
GBR	British in England and Scotland	200	
IBS	Iberian Population in Spain	214	
TSI	Toscani in Italia	222	
BEB	Bengali from Bangladesh	172	South Asia (SAS)
ITU	Indian Telugu from the UK	204	
PJL	Punjabi from Lahore, Pakistan	192	
STU	Sri Lankan Tamil from the UK	204	
Total		3 250	

and to prioritize very recent inferences, $\tilde{g}_t = t/2$ for $t = 1, \ldots, 50$, and $\tilde{g}_t = 30 \exp\{1.6(t-51)\}$ for $t = 51, \ldots, 56$.

Human data analysis

We estimated $N(g)$ from all 22 autosomes for 16 contemporary human populations, including four populations from each of Africa, East Asia, Europe, and South Asia (Table 2). The samples came predominantly from the 1000 Genomes Project Fairley et al. (2020), supplemented by the Human Genome Diversity Project Bergström et al. (2020) and the Simons Genome Diversity Project Mallick et al. (2016). Rather than using the original sequence data, we extracted AFS and pairwise differences required for the InferNo analyses from a genealogy constructed using genome sequences from 3 601 modern and eight high-coverage ancient humans Wohns et al. (2022). Use of this unified genealogy helps ensure uniform

Table 3: RMISE over $g \in [200, 30\,000]$ ($g \in [1\,000, 30\,000]$) for five $N(g)$ inference methods. RMISE = root mean integrated squared error (in units of 10^5). M = 10^6, K = 10^3. Values are averages over the 25 curves of Fig. 2.

Model	InferNo		Relate		PSMC		Mushi		CHIMP	
1	**1.8**	(**1.7**)	7K	(9.9)	722	(4.3)	682	(2.0)	7.7	(7.6)
2	**2.0**	(**1.7**)	14K	(8.6)	0.9M	(10.1)	1K	(8.6)	0.2M	(0.2M)
3	**3.5**	(**3.5**)	4K	(9.5)	9.5	(6.1)	289	(8.1)	9.8	(9.8)
4	**5.5**	(**5.3**)	9K	(9.4)	3M	(7.3)	769	(7.0)	8.0	(7.8)
5	**7.6**	(**7.2**)	6K	(17.2)	4K	(10.8)	343	(9.8)	16.6	(14.9)
6	**7.7**	(**7.6**)	15K	(10.3)	5M	(8.4)	3K	(49.1)	9.1	(8.9)

data quality thanks to data consistency checks based on the full dataset. Our analysis used a subset of the genealogy corresponding to 1 625 modern humans.

For the InferNo analyses, to reduce computational effort we optimized the smoothing penalty λ using the chromosome 20 data only, and applied the resulting value to the analyses of all 22 autosomes. See SI for details.

Results

Simulation study results

For $n = 200$, Fig. 2 shows that all five methods perform relatively poorly at recent times ($g < 10^3$) and when $g > 3\times10^4$. In the former case, the small number of recent recombination events limits the number of distinct lineages that can contribute to inference, while in the latter case, the number of lineages has been reduced by coalescences. Rapid changes of $N(g)$ with g also challenge every method.

On the criterion of minimising RMISE over $g \in [200, 30\,000]$, InferNo is superior to all other methods, in most cases by orders of magnitude (Table 3). This primarily reflects its superior performance for $g \in [200, 1\,000)$. When we remove this time interval and re-compute RMISE for $g \in [1\,000, 30\,000]$, the methods become more comparable, but InferNo is still the best for all six models. In the special case of a single diploid individual ($n = 2$), InferNo

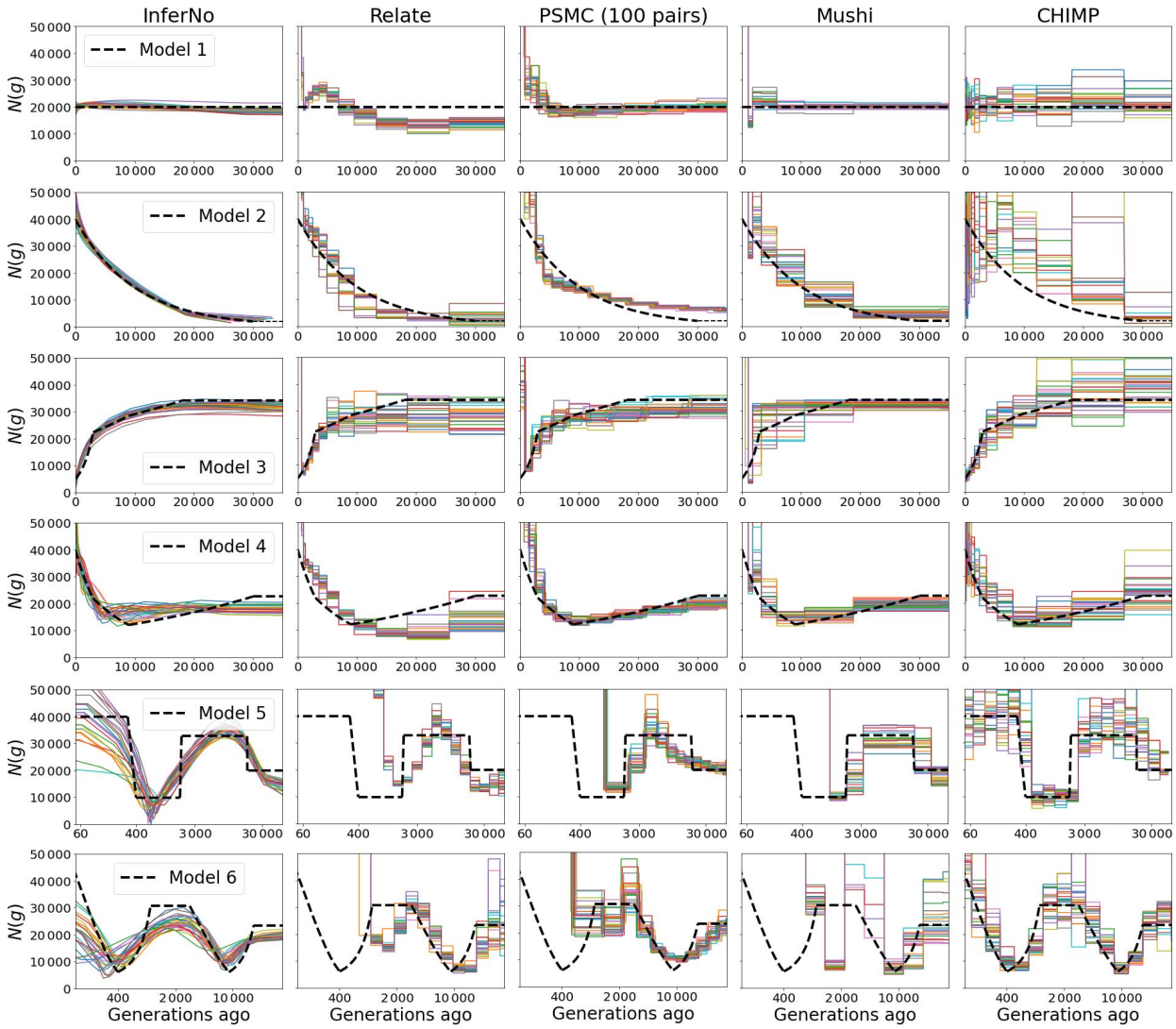

Figure 2: **Estimates of population size $N(g)$ in Models 1 to 6 using five inference methods.** Sample size $n = 200$ and sequence length $\ell = 10^7$ sites. Note logarithmic time scale for Models 5 and 6. See Table 3 for corresponding RMISE values, Fig. 8 for plots when $n = 10$, and Table 6 for computing times.

Table 4: RMISE (in units of 10^5) for InferNo and PSMC when $n = 2$. Values are averages over the 25 curves of Fig. 9.

Generations	Model:	1	2	3	4	5	6
$[200, 30\,000]$	InferNo	**8**	**6**	**13**	**16**	**21**	**14**
	PSMC	6M	9M	3M	9M	8M	6M
$[1\,000, 30\,000]$	InferNo	**7.3**	**6.1**	**11.7**	**15.2**	**18.2**	**12.2**
	PSMC	30K	3.0M	14.3	2.7M	51.0	900
$[5\,000, 30\,000]$	InferNo	**5.2**	**5.4**	**9.0**	7.8	**11.6**	8.6
	PSMC	5.9	6.0	12.0	**4.4**	11.8	**7.4**

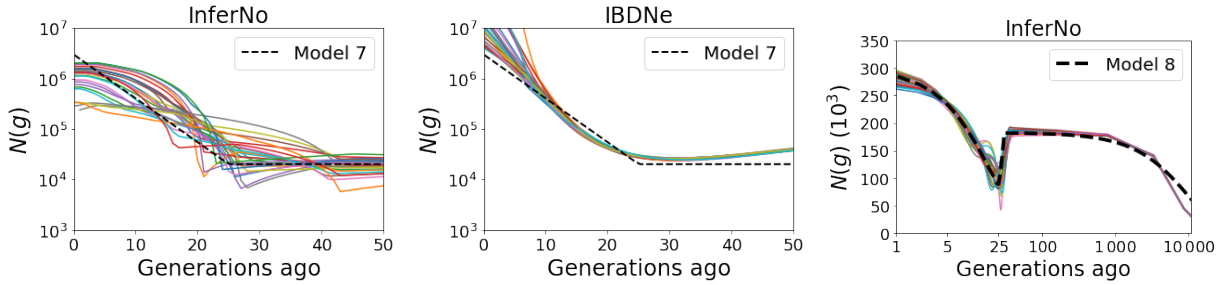

Figure 3: **InferNo and IBDNe estimates of $N(g)$ under Models 7 and 8.** The sample size and sequence length for Model 7 (top and middle panels) are $n = 2\,000$ and $\ell = 10^8$ sites. For Model 8 (bottom), they are $n = 2{\times}10^5$ and $\ell = 5{\times}10^7$ sites.

greatly outperforms PSMC for $g < 5\,000$ (Table 4), while for $g > 5\,000$ the two methods show similar accuracy.

InferNo is also the fastest among the five methods (Table 6), taking on average (over 25 replicate datasets for each of 6 models) 8.5 seconds per analysis when $n = 200$, just faster than 9.2 seconds for Mushi and much faster than 386 seconds for CHIMP, 406 seconds for Relate and 983 seconds for PSMC. Mushi has a computing cost similar to InferNo when $n = 200$, and is slower when $n = 10$.

Increasing n improves inference more for recent generations ($g < 1\,000$) than the distant past (Fig. 10), because many lineages coalesce in recent generations. The average RMISE for $g \in [200, 30\,000]$ decreases by 82% when n increases from 10 to 1 000 (Table 7), but when restricted to $g \in [1\,000, 30\,000]$ RMISE decreases by only 39%. Table 7 and Fig. 11 show little effect on inference of misspecifying recombination rates, with average RMISE increasing 15% or 18% when the recombination map of Fig. 1 or a constant rate 50% higher than that assumed by InferNo is used for the data simulation. InferNo is also robust to sequencing errors, with RMISE increasing by 9%, 20% and 29% when sequencing errors are introduced at rate 1, 2 and 3 errors per 10^5 sites (Table 7 and Fig. 12).

Because the computing time is approximately linear with n (Fig. 10), InferNo can exploit large sample sizes to obtain accurate estimates of $N(g)$ in the recent past (small g). Fig. 3 (top) demonstrates its capability for accurate estimation of recent population sizes, in comparison with IBDNe (middle). The RMISE for IBDNe is about 20-fold higher than for

InferNo (638 vs 32). With $n = 2{\times}10^5$ (Fig. 3, bottom), InferNo required only 326 seconds for each analysis to accurately infer rapid size changes over the past 30 generations. It is feasible to use InferNo on even the largest available human biobanks to infer changes in population size over recent centuries, which may be due to migration events, plagues or environment changes.

Results from human data analysis

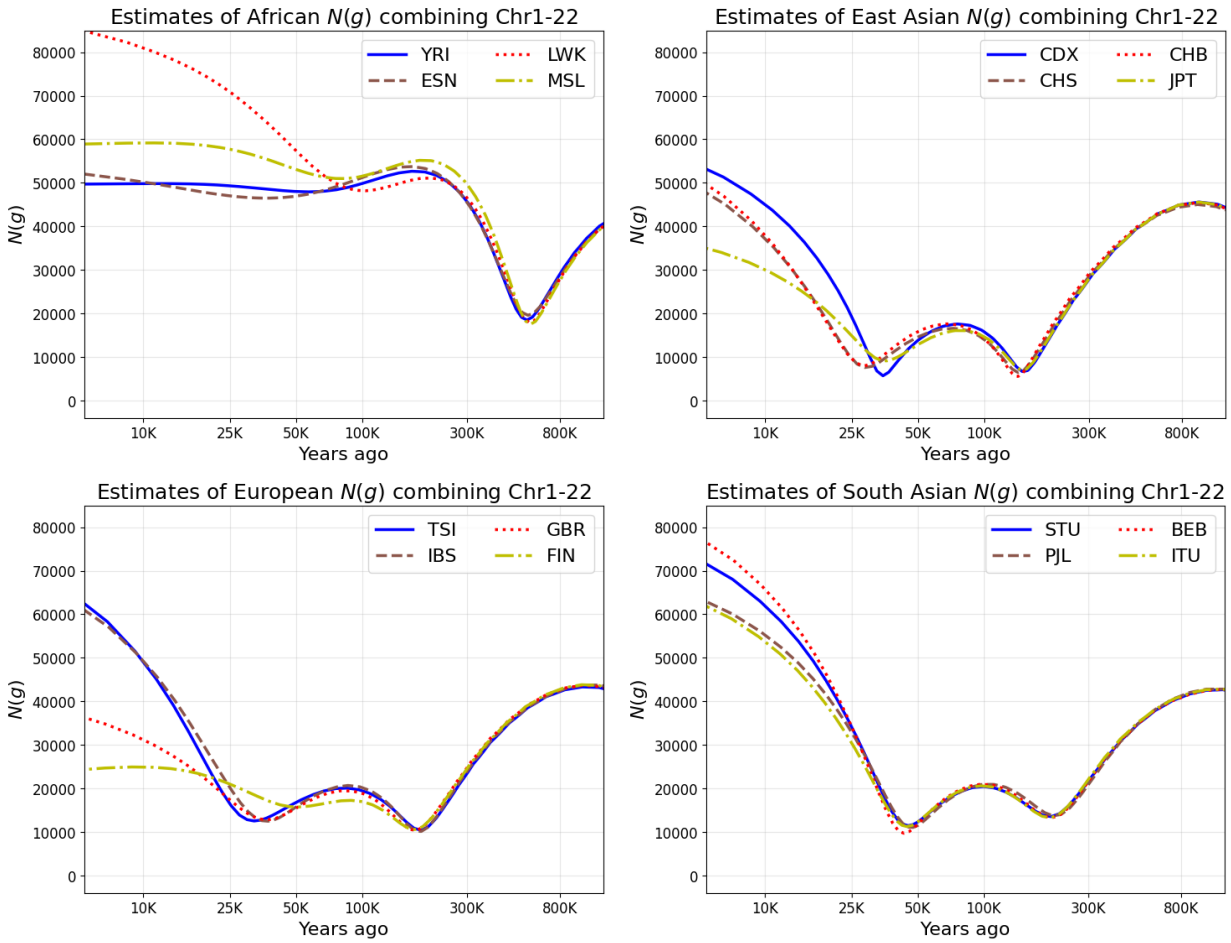

Figure 4: **InferNo estimates of population size N for four populations in each of four regions.** See Table 2 for the population codes. The time axes are on a logarithmic scale.

When presenting results for human demographic history, for consistency with other authors we will express time as years (t) rather than generations (g) before present. We will

also use K as shorthand for 10^3. Unfortunately there is no agreement among authors on generation time; we adopt a recent estimate of 27 years per generation (Wang et al., 2023). Since N counts haploid genomes, the number of humans is $N/2$.

For $t > 25$K there is a striking similarity of the inferred effective population size curves cross the four populations within each region (Fig. 4), despite the InferNo analysis being performed independently for each population. The four AFR populations have $N(t) > 40$K for $t < 350$K, while only for $550\text{K} < t < 600\text{K}$ are all four $N(t)$ values below 20K. The lowest $N(t)$ in any AFR population was 17.6K in MSL at $t = 590$K. The 12 non-AFR populations have $N(t) < 40$K for $30\text{K} < t < 500\text{K}$.

The population sizes of AFR and non-AFR populations remain distinct at $t = 800$K, estimated at 28K – 30K for all 4 AFR populations and at 41K – 45K for all 12 non-AFR populations. This time is well before the migration of modern humans out of Africa, which is typically estimated around $t \approx 50$K Ragsdale et al. (2023). Estimates of current size $N(0)$ are much more variable, ranging from 21K in FIN to 88K in LWK.

The non-AFR populations each experienced two local minima (bottlenecks). The most severe bottleneck arose in the ancestry of the EAS populations when $t \approx 135$K, with $N(t)$ estimates between 5.6K (CHB) and 7.1K (JPT). The corresponding bottleneck in EUR populations was inferred to be less severe (between 10.2K (IBS) and 10.7K (FIN)) and earlier ($t \approx 175$K). For the SAS populations, the more recent of the two bottlenecks ($t \approx 45$K) was slightly more severe than the earlier bottleneck ($t \approx 195$K), ranging from 9.7K (BEB) to 11.4K (STU). The more recent bottleneck was even more recent in EAS ($28\text{K} < t < 36\text{K}$), with $N(t)$ ranging from 5.7K (CDX) to 9.1K (JPT). In EUR, the more recent bottleneck was negligible in FIN but arose at a similar time to EAS in the other three populations ($32\text{K} < t < 37\text{K}$), with $N(t)$ ranging from 12.5K (IBS) to 12.8K (GBR).

Fig. 13 shows, for 4 of the 16 populations, that the major features of demographic history inferred by InferNo are detected using each type of statistic individually: AFS only ($\omega = 0$) and TMRCA estimates only ($\omega = \infty$), but with some differences in timing and severity of

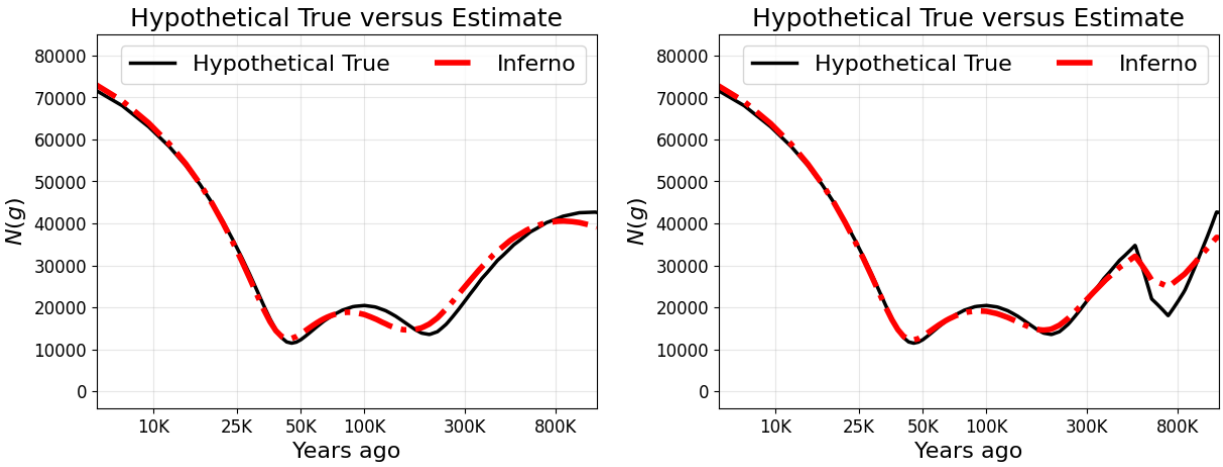

Figure 5: **Testing InferNo inferences from human data.** Left: data are simulated with $N(g)$ (black curve) close to those inferred for the non-AFR populations (Fig, 4), and then InferNo used for inference (red curve). See Fig. 14 for the corresponding plot for AFR populations. Right: we add an ancient bottleneck similar to that inferred for the AFR populations, and are able to recover it in the inference. The time axes are logarithmic. Both simulations generated data for all 22 autosomes with $n = 200$, mutation rate $\mu = 1.3 \times 10^{-8}$ and recombination rate $r = 10^{-8}$.

bottlenecks, particularly for the older bottleneck in the two non-AFR populations. These discrepancies reflect uncertainty in inferences arising from modelling assumptions that inevitably deviate from the complex reality of human history.

As a further check on our inferences about human demographic history, we performed two further simulation experiments. First, we simulated a dataset with $n = 200$ and an $N(g)$ curve similar to those inferred for the non-AFR populations. We then applied InferNo using the same λ value as for the real-data analysis. Fig. 5 (left) shows that InferNo accurately recovers the curve, which is also the case when the experiment is repeated for an AFR population (Fig. 14). We repeated the non-AFR simulation now including an additional, ancient bottleneck similar to that inferred for the AFR populations. Fig. 5 (right) shows that InferNo captures much of the signal from the ancient bottleneck, and that inference of the more recent bottlenecks remains accurate.

Discussion

InferNo achieves higher statistical and computational efficiency than existing approaches for estimating effective population sizes by combining analytical steps with simulation-based approximation. The constant-size simulation model is equivalent to the target variable-size model under a time scaling that is inferred from the observed AFS and (if required) from TMRCA estimates for sequence pairs in genome segments. This approach allows InferNo to achieve more precise population-size inferences, identifying features that are not detected by other methods.

We showed robustness of InferNo to sequencing errors and gene conversions, variation in the mutation rate along the genome, and misspecification of the recombination map. The better performance of InferNo over four alternative approaches is particularly marked for recent times ($g < 1\,000$ generations), and it also outperforms an IBD-based method that focusses on the recent past. InferNo is the most computationally efficient of the methods, with cost increasing linearly with n. We illustrated its ability to accurately infer very recent population sizes from biobank-scale datasets.

We used InferNo to estimate historical population sizes for 16 worldwide human populations. The four populations within each of four worldwide regions showed a striking similarity of the inferred $N(t)$ curves for $t > 25$K years before present. As expected, we found the greatest differences between African (AFR) and non-AFR populations, with these differences persisting more than 800K years into the past.

Our results show broad similarity with, but also important differences from, results of analyses of three of the populations studied here (Upadhya and Steinrücken, 2022, Fig 11), and different populations in similar regions (Bergström et al., 2020, Fig 4) using MSMC2 (Schiffels and Wang, 2020), which is related to PSMC and was found by Upadhya and Steinrücken (2022) to be inferior to their CHIMP software. Both these studies identified one long-duration bottleneck in the histories of EUR and EAS populations, with $N < 20$K for t between about 15K and 125K, with minima at $t \approx 38$K Bergström et al. (2020) and $t \approx 100$K

Upadhya and Steinrücken (2022). There is no ground truth available to conclude that any method is most accurate, but we point to the better performance of InferNo in simulations, including under Model 6 with two bottlenecks. InferNo also infers a single bottleneck when its smoothing parameter λ is increased from the minimum-BIC value; however, the consistency in the timing of the two bottlenecks across populations in each region is evidence against them being statistical artifacts.

An ancient, severe bottleneck reported in ancestral AFR populations was not evident in analyses of non-AFR populations Hu et al. (2023). Our inferred AFR bottleneck is much less severe ($N \approx 17$K versus $N \approx 1\,280$), of shorter duration and less ancient ($550\text{K} < t < 600\text{K}$ versus $813\text{K} < t < 930\text{K}$). The findings of Hu et al. (2023) have been challenged by Deng et al. (2024a), who suggest that the severe bottleneck may be an artifact of AFS-based inference. Although we also use the AFS, Fig. 13 shows that InferNo inferences restricted to pairwise TMRCA are similar to AFS-based inferences. Further criticism of Hu et al. (2023) has come from Cousins and Durvasula (2025) who found a better fit using Mushi, inferring a bottleneck that is comparable in severity with the one we infer, but its date agrees with that of Hu et al. (2023).

We agree with Hu et al. (2023), and previous authors cited therein, in finding markedly different demographic histories of AFR and non-AFR populations long before the migrations of modern humans out of Africa. This difference could be explained by persistent structure in the populations ancestral to modern humans, as has been reported Ragsdale et al. (2023); Cousins et al. (2025), with modern AFR and non-AFR populations having a different ancestry mix from ancient populations. Alternatively, Hu et al. (2023) propose that the ancient bottleneck is also present in the demographic histories of non-AFR populations but its signal has been erased by a more recent bottleneck corresponding to out-of-Africa migrations. This was tested and considered implausible by Cousins and Durvasula (2025), and we also show that such an erasure is implausible for the case of the bottlenecks that we infer: Fig. 5 illustrates that InferNo would have extracted a signal if the more ancient bottleneck we

inferred for the AFR populations also existed in the histories of the non-AFR populations.

InferNo does not report uncertainty in estimates, but users can gain an understanding of its magnitude from simulations (such as column 1 of Fig. 6) and from comparisons across similar populations (such as Fig. 4). In common with other methods for inferring $N(g)$, InferNo assumes a single, unstructured population, and hence it infers an effective population size that may differ from census size, for example in the presence of population structure. The method can be extended to two or more populations at additional computational cost. While InferNo avoids inferring recombination events, it requires enough recombination to generate replication along the genome. InferNo allows $\mu(s)$ to vary over sites s, but like other methods, it assumes that each $\mu(s)$ is constant over time. Subject to the above caveats, InferNo is applicable to many species, and can benefit inferences of other evolutionary parameters, such as species divergence times.

The InferNo software and the data and code used in this paper are available at: github.com/ZhendongHuang/Inference_for_demographic_history.

References

F. Baumdicker, G. Bisschop, D. Goldstein, G. Gower, A. P. Ragsdale, G. Tsambos, S. Zhu, B. Eldon, E. C. Ellerman, J. G. Galloway, A. L. Gladstein, G. Gorjanc, B. Guo, B. Jeffery, W. W. Kretzschumar, K. Lohse, M. Matschiner, D. Nelson, N. S. Pope, C. D. Quinto-Cortés, M. F. Rodrigues, K. Saunack, T. Sellinger, K. Thornton, H. van Kemenade, A. W. Wohns, Y. Wong, S. Gravel, A. D. Kern, J. Koskela, P. L. Ralph, and J. Kelleher. Efficient ancestry and mutation simulation with msprime 1.0. *Genetics*, 220(3):iyab229, 12 2021. ISSN 1943-2631. doi: 10.1093/genetics/iyab229. URL https://doi.org/10.1093/genetics/iyab229.

A. Bergström, S. McCarthy, R. Hui, M. Almarri, Q. Ayub, P. Danecek, Y. Chen, S. Felkel, P. Hallast, J. Kamm, H. Blanché, J. Deleuze, H. Cann, S. Mallick, D. Reich, M. Sandhu,

P. Skoglund, A. Scally, Y. Xue, R. Durbin, and C. Tyler-Smith. Insights into human genetic variation and population history from 929 diverse genomes. *Science*, 367:eaay5012, 2020.

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

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

S. Browning, B. Browning, M. Daviglus, R. Durazo-Arvizu, N. Schneiderman, R. Kaplan, and C. Laurie. Ancestry-specific recent effective population size in the Americas. *PLoS Genet*, 14(5):e1007385, 2018.

S. R. Browning and B. L. Browning. Accurate non-parametric estimation of recent effective population size from segments of identity by descent. *American Journal of Human Genetics*, 97(3):404–418, 2015.

T. Cousins and A. Durvasula. Insufficient evidence for a severe bottleneck in humans during the Early to Middle Pleistocene transition. *Molecular Biology and Evolution*, 42:msaf041, 2025. doi: 10.1093/molbev/msaf041.

T. Cousins, A. Scally, and R. Durbin. A structured coalescent model reveals deep ancestral structure shared by all modern humans. *Nature Genetics*, 57:856–864, 2025.

Y. Deng, R. Nielsen, and Y. Song. A previously reported bottleneck in human ancestry 900 kya is likely a statistical artifact. *Genetics*, 229:1–3, 2024a. doi: 10.1093/genetics/iyae192.

Y. Deng, R. Nielsen, and Y. S. Song. Robust and accurate Bayesian inference of genome-wide genealogies for large samples. *bioRxiv*, 2024.03.16.585351, 2024b.

W. S. DeWitt, K. D. Harris, A. P. Ragsdale, and K. Harris. Nonparametric coalescent inference of mutation spectrum history and demography. *Proceedings of the National Academy of Sciences*, 118(21):e2013798118, 2021.

T. Druet, I. Macleod, and B. Hayes. Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity*, 112(1):39–47, 2014.

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

S. Fairley, E. Lowy-Gallego, E. Perry, and P. Flicek. The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Research*, 48(D1):D941–D947, 2020.

R. Fournier, Z. Tsangalidou, D. Reich, and P. F. Palamara. Haplotype-based inference of recent effective population size in modern and ancient DNA samples. *Nature Communications*, 14(1):7945, 2023.

R. Griffiths and S. Tavaré. Sampling theory for neutral alleles in a varying environment. *Philosophical Transactions of the Royal Society of London*, A 344:403–10, 1994.

R. N. Gutenkunst, R. D. Hernandez, S. H. Williamson, and C. D. Bustamante. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, 5(10):e1000695, 2009.

W. Hu, Z. Hao, P. Du, F. Di Vincenzo, G. Manzi, J. Cui, Y. Fu, Y. Pan, and H. Li. Genomic inference of a severe human bottleneck during the Early to Middle Pleistocene transition. *Science*, 381:979—-984, 2023.

Z. Huang, J. Kelleher, Y. Chan, and D. Balding. Estimating evolutionary and demographic

parameters via ARG-derived IBD. *PLoS Genet*, page e1011537, 2025. doi: 10.1371/journal.pgen.1011537.

J. Kamm, J. Terhorst, R. Durbin, and Y. S. Song. Efficiently inferring the demographic history of many populations with allele count data. *Journal of the American Statistical Association*, 115(531):1472–1487, 2020.

J. Kelleher, Y. Wong, A. W. Wohns, C. Fadil, P. K. Albers, and G. McVean. Inferring whole-genome histories in large population datasets. *Nature Genetics*, 51(9):1330–1338, 2019.

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

A. Mahmoudi, J. Koskela, J. Kelleher, Y. Chan, and D. Balding. Bayesian inference of ancestral recombination graphs. *PLoS Computational Biology*, 18(3):e1009960, 2022.

S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, M. Zhao, N. Chennagiri, S. Nordenfelt, A. Tandon, P. Skoglund, I. Lazaridis, S. Sankararaman, Q. Fu, N. Rohland, G. Renaud, Y. Erlich, T. Willems, C. Gallo, J. Spence, Y. Song, G. Poletti, F. Balloux, G. van Driem, and D. Reich. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*, 538:201–206, 2016. doi: 10.1038/nature18964.

M. Mezzavilla and S. Ghirotto. Neon: An R package to estimate human effective population size and divergence time from patterns of linkage disequilibrium between SNPs. *Journal of Computer Science and Systems Biology*, 8(1), 2015.

P. F. Palamara, T. Lencz, A. Darvasi, and I. Pe'er. Length distributions of identity by descent reveal fine-scale demographic history. *American Journal of Human Genetics*, 91 (5):809–822, 2012.

A. P. Ragsdale and S. Gravel. Models of archaic admixture and recent history from two-locus statistics. *PLoS Genetics*, 15(6):e1008204, 2019.

A. P. Ragsdale, T. D. Weaver, E. G. Atkinson, E. G. Hoal, M. Möller, B. M. Henn, and S. Gravel. A weakly structured stem for human origins in Africa. *Nature*, 617(7962): 755–763, 2023.

M. D. Rasmussen, M. J. Hubisz, I. Gronau, and A. Siepel. Genome-wide inference of ancestral recombination graphs. *PLoS Genetics*, 10(5):e1004342, 2014.

E. Santiago, I. Novo, A. F. Pardiñas, M. Saura, J. Wang, and A. Caballero. Recent demographic history inferred by high-resolution analysis of linkage disequilibrium. *Molecular Biology and Evolution*, 37(12):3642–3653, 2020.

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

S. Schiffels and K. Wang. MSMC and MSMC2: The multiple sequentially Markovian coalescent. *Methods in Molecular Biology*, 2090:147–166, 2020.

G. Schwarz. The Bayesian information criterion. *Annals of Statistics*, 6:461–464, 1978.

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

J. Terhorst. Accelerated Bayesian inference of population size history from recombining sequence data. *bioRxiv*, 2024.03.25.586640, 2024.

J. Terhorst, J. A. Kamm, and Y. S. Song. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics*, 49(2):303–309, 2017.

G. Upadhya and M. Steinrücken. Robust inference of population size histories from genomic sequencing data. *PLoS Computational Biology*, 18(9):e1010419, 2022.

R. J. Wang, S. I. Al-Saffar, J. Rogers, and M. W. Hahn. Human generation times across the past 250,000 years. *Science Advances*, 9:eabm7047, 2023.

A. L. Williams, G. Genovese, T. Dyer, N. Altemose, K. Truax, G. Jun, N. Patterson, S. R. Myers, J. E. Curran, R. Duggirala, et al. Non-crossover gene conversions show strong GC bias and unexpected clustering in humans. *Elife*, 4:e04637, 2015.

A. W. Wohns, Y. Wong, B. Jeffery, A. Akbari, S. Mallick, R. Pinhasi, N. Patterson, D. Reich, J. Kelleher, and G. McVean. A unified genealogy of modern and ancient genomes. *Science*, 375(6583):eabi8264, 2022.

J. Ye. On measuring and correcting the effects of data mining and model selection. *Journal of the American Statistical Association*, 93(441):120–131, 1998.

Y. Zhou, S. Browning, and B. Browning. A fast and simple method for detecting identity-by-descent segments in large-scale data. *The American Journal of Human Genetics*, 106 (4):426–437, 2020.

Supplementary Methods

Inference of mutation rate

We estimate the mutation rate along the genome using the number of polymorphic sites close to any given site. More specifically, we partition the sites into L^* segments $J_1^*, \ldots, J_{L^*}^*$, each of equal length ℓ/L^*, and estimate $\mu(s)$ by a piece-wise constant map fitted to n_l^*, the number of polymorphic sites in J_l^*:

$$\hat{\mu}(s) = \frac{\sum_{l=1}^{L^*} n_l^* \mathbf{1}(s \in J_l^*)}{\sum_{l=1}^{L^*} n_l^*} \bar{\mu} L^* \ , \tag{5}$$

where $\mathbf{1}()$ is the indicator function.

The null model and time scaling

Although there is flexibility in setting the null model population size, it should be not too far from average values over the historic periods of interest. We use $\tilde{N} = 20\,000$. For large ℓ given the value of r, there is enough internal replication along the genome due to recombination that a single simulation is sufficient for precise estimation. We used one null-model simulation for all analyses reported here, but multiple simulations can be performed if desired for greater accuracy.

To relate the null model to the variable-size model, we consider the cumulative per-site coalescence rate over the interval I_t in the variable-size model, which is $(g_t - g_{t-1})/N(g)$. Eq. [1] is obtained by equating this value to $(\tilde{g}_t - \tilde{g}_{t-1})/\tilde{N}$, the corresponding quantity over $\tilde{I}_t$ in the null model.

Using the AFS

The $\{k, t\}$ entry of $\tilde{X}_1$ is the number of sites in the null simulation with a frequency-k derived allele for which the mutation arose during $\tilde{I}_t$. Because of the equivalence of null and size-varying models under the time scaling, the product of this entry with the time-scale ratio β_t equals in expectation the number of frequency-k sites in the size-varying model for which the mutation arose during I_t. Summing over t gives the expected value of the kth entry of the AFS. Eq. [2] specifies this relationship jointly for each k.

Using pairwise TMRCA estimates

From observed pairwise site differences, we estimate quantiles of the distribution of the TMRCA of pairs of observed sequences in short genome segments. We match these to the corresponding quantiles of the null model, obtained from simulated TMRCA values.

We choose the segments to obtain approximately c mutations per lineage per generation

in each J_l by minimizing

$$|c - \sum_{s \in J_l} \hat{\mu}(s)| \qquad \text{for} \quad l = 1, \ldots, L.$$

By default, we set $c = 2.6 \times 10^{-4}$, resulting in $|J_l| \approx 20$K sites for human populations, which is long enough to capture an informative number of site differences between sequence pairs, while limiting the within-segment variation in TMRCA. However, InferNo is robust to the choice of c because the same J_l are used for the null and variable-size models.

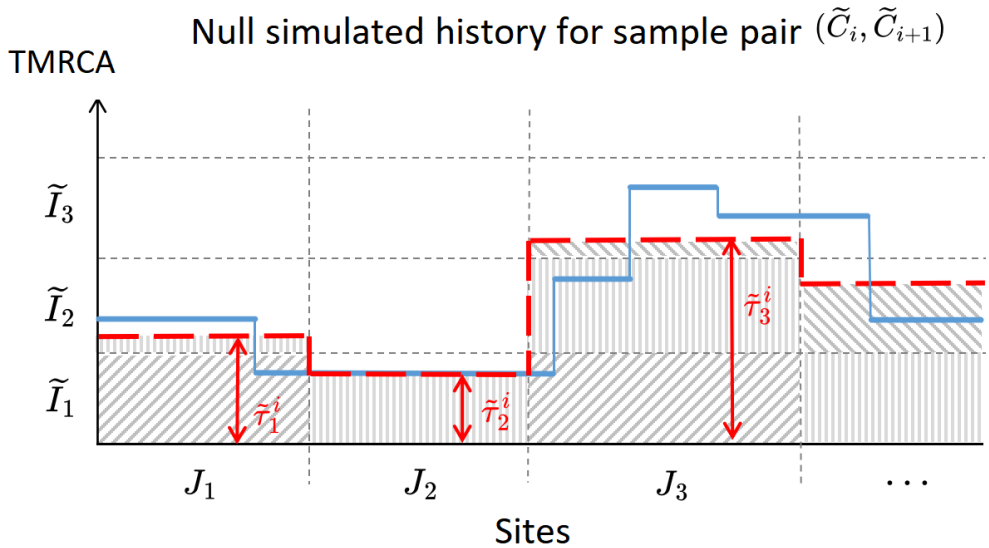

Figure 6: The TMRCA in a null simulation (solid blue) and its approximation $\tilde{\tau}$ (dashed red) which equals the average blue value in each interval J_l. The heights of the shaded rectangles correspond to the elements of $\tilde{W}^i$ defined in the text.

We write τ for the TMRCA of observed sequences C_i and C_{i+1} averaged over $s \in J_l$ (see Fig. 6 for the corresponding averaging in the null model). To reduce computational effort, we only use neighbouring pairs of sequences (plus (C_1, C_n) when $n > 2$), so the resulting inferences can depend on sequence order; however, this effect is weak because every mutation is represented by a site difference in at least two neighbour pairs. By symmetry, τ has the same distribution for each i. We also treat the distribution as invariant over l, which holds approximately, and we make the corresponding assumption for the null model.

We use the sequence data to approximate the quantiles of the distribution of τ. Let Z denote an $n \times L$ matrix with $\{i, l\}$ element equal to the number of site differences in J_l between the ith sequence pair. From coalescent theory we can model each element of Z as a Poisson($2c\tau$) random variable, conditional on τ. Let $\mathbb{P}_a(b)$ denote the probability mass function at b of the Poisson(a) distribution and $Z_{\max} = \max\{Z\}$. The probability mass function of each element z of Z is

$$p_z(m) = \int_0^\infty \mathbb{P}_{2ct}(m) f(t)\,\mathrm{d}t \ , \qquad m = 0, 1, 2, \ldots,$$

where f denotes the probability density function of τ. Discretising the integral at the points $\frac{t}{2c}, t = 0, \ldots, Z_{\max}$, gives

$$p_z(m) \approx \frac{1}{2c} \sum_{t=0}^{Z_{\max}} \mathbb{P}_t(m) f(t/2c) \ , \qquad m = 0, 1, 2, \ldots.$$

Restricting $p_z(m)$ to $m \leq Z_{\max}$, defining A as a matrix with $A_{i+1,j+1} = \mathbb{P}_j(i)$, for $i, j = 0, \ldots, Z_{\max}$, and defining a vector version of f evaluated at $t/2c$, for $t = 0, \ldots, Z_{\max}$, gives the matrix equation

$$p_z = \frac{1}{2c} A f \ . \tag{6}$$

We estimate p_z by empirical values

$$\hat{p}_z(m) = \frac{1}{nL} \sum_{i=1}^{n} \sum_{l=1}^{L} \mathbf{1}(Z_{i,l} = m) \ , \qquad m = 0, \ldots, Z_{\max},$$

and then invert (6) to estimate f by

$$\hat{f} = 2cA^{-1}\hat{p}_z \ .$$

Let the vector $\hat{F}$ be the estimated cumulative distribution function of τ with jth element $\hat{F}_j = \sum_{i=1}^{j} \hat{f}_i$ for $j = 1, \ldots, Z_{\max}+1$. Then the γ-quantile of τ, for $\gamma \in (0, 1)$, is estimated

by linear interpolation: it is 0 if $\gamma < \hat{F}_1$, and

$$\frac{1}{2c}\left[k + \frac{\gamma - \hat{F}_{k+1}}{\hat{F}_{k+2} - \hat{F}_{k+1}}\right]$$

otherwise, where $k \in \{0, \ldots, Z_{\max} - 1\}$ satisfies $\hat{F}_j \leq \gamma$ for $j = 1, \ldots, k+1$ and $\hat{F}_{k+2} > \gamma$. The mth element of the resulting length-nL vector Y_2 estimates the $m/(nL+1)$ quantile of τ.

We relate Y_2 to the corresponding estimated quantiles for the null model, based on TMRCA values extracted from the null simulation. For a null-simulation sequence pair, let $\tilde{\tau}$ denote the average TMRCA over J_l (Figure 6). Let $\tilde{W}^i$ be an $L \times T$ matrix with $\{l, t\}$ entry the contribution to $\tilde{\tau}$ from the time interval $\tilde{I}_t = [\tilde{g}_{t-1}, \tilde{g}_t)$,

$$\tilde{W}^i_{l,t} = \begin{cases} \tilde{g}_t - \tilde{g}_{t-1} & \text{if } \tilde{\tau} > \tilde{g}_t \ , \\ \tilde{\tau} - \tilde{g}_{t-1} & \text{if } \tilde{g}_{t-1} < \tilde{\tau} < \tilde{g}_t \ , \\ 0 & \text{if } \tilde{\tau} < \tilde{g}_{t-1} \ . \end{cases} \tag{7}$$

By construction, the sum of the lth row of $\tilde{W}_i$ is the average TMRCA $\tilde{\tau}$.

Now let $\tilde{X}_2$ be the $nL \times T$ matrix constructed by vertically stacking the $\tilde{W}^i$, $i = 1, \ldots, n$, then sorting the rows to have increasing row sums. Since each row of $\tilde{W}_i$ corresponds to a TMRCA, the rows of the sorted matrix correspond to the order statistics, which estimate the quantiles of $\tilde{\tau}$. We can now infer β from regression equation Eq. [3], similar to that for Y_1 and $\tilde{X}_1$.

Combining the estimating equations

The first row of $\tilde{X}_1$ and of Y_1 correspond to frequency-one alleles, which can be informative about recent $N(g)$ but distorted by sequencing errors. By default, InferNo removes these first rows to improve robustness, but they can be retained if desired, for example when the sequencing quality is known to be high.

The matrix $\tilde{X} = (\tilde{X}_1', \omega\tilde{X}_2')'$ has dimensions $K \times T$, where $K = (L+1)n-2$. For $n < 2\,000$, we use the InferNo default $\omega = \sqrt{400c/(nL)}$, where c is the constant used to define the genome partition. Intuitively, this balances the amount of information from the AFS (represented by the sample size n) and the pairwise TMRCA (represented by c/L, approximately the expected number of mutations at each site). The constant 400 was found to work well for human data. Other choices for ω can be specified in InferNo, with higher ω tending to improve precision in the distant past at a cost to recent inferences. For n large we can set $\omega = 0$, ignoring $\tilde{X}_2$ and Y_2 which greatly reduces computational effort.

To infer β from the combined estimating equations (Eq. [4]), we weight the kth equation by an approximation to the standard deviation of y_k, to give reduced weight to noisier observations. Let S be a diagonal matrix with kth diagonal element $s_k = (y_k+1)^{-1/2}$, $k = 1, \ldots, K$, so that $1/s_k$ approximates the (Poisson) standard deviation of y_k, adjusted to ensure finite values. We then estimate β by

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \|S(\tilde{X}\beta - Y)\|_2^2 + \lambda \sum_{t=2}^{T-1} (\beta_{t+1} - 2\beta_t + \beta_{t-1})^2 \right\}, \tag{8}$$

subject to $\beta_t > 0$, $t = 1, \ldots, T$. The second term in (8) controls the smoothness of the β_t by penalising their second differences.

To select λ, we minimise the adjusted Bayesian information criterion (BIC) Schwarz (1978); Ye (1998):

$$\text{BIC} = K \log \left(\frac{1}{K} \|S(\tilde{X}\hat{\beta} - Y)\|_2^2 \right) + d \log(K),$$

where d is the estimated degrees of freedom,

$$d = \text{trace}\{S\tilde{X}(\tilde{X}'S'S\tilde{X} + \lambda P'P)^{-1}\tilde{X}'S'\},$$

and P is a $(T-2) \times T$ matrix with entries $P_{i,i} = P_{i,i+2} = 1$ and $P_{i,i+1} = -2$ for $i = 1, \ldots, T-2$,

and $P_{i,j} = 0$ otherwise.

For simulation studies using Models 1 through 6, and for the real data analyses, we select λ to minimise BIC over $\log_{10}(\lambda) \in \{0,1,2,3,4,5,6,7\}$, although the extreme values 0 and 7 were never chosen in practice. We illustrate this approach for Models 4 and 5 with $n = 100$ (Figure 7). For Models 7 and 8, which focus on recent population size, BIC is computed using only the first ten rows of $\tilde{X}$ and Y, and the first ten rows and columns of S, which are informative about low-frequency alleles that predominantly arise from recent mutations. We select λ to minimise BIC over $\log_{10}(\lambda) \in \{-4, -3, -2, -1, 0\}$ and $\log_{10}(\lambda) \in \{2.5,3,3.5,4,4.5\}$ for Models 7 and 8, respectively.

The final (optional) step is to convert the piece-wise constant estimate $\hat{N}(g) = \tilde{N}\hat{\beta}$ into a continuous, piece-wise-linear function with each interval midpoint remaining fixed, by connecting the points $(0, \tilde{N}\hat{\beta}_1)$ and $((g_{t-1}+g_t)/2, \tilde{N}\hat{\beta}_t)$ for $t \geq 1$.

Implementation details for other software

Relate: we set $N(0) = 30\,000$ for both ARG inference and population size estimation.

PSMC: we set 20 free parameters, otherwise default settings were used. The original PSMC only uses a pair of sequences; for larger n, we implemented PSMC by partitioning the sequences into $n/2$ pairs and then averaging the resulting $n/2$ estimates.

Mushi: we set the maximum time investigated to 60 000 generations and the maximum number of iterations to 300. We obtain piece-wise constant estimates by fixing the trend order parameter to 0 and selecting the trend penalty parameter over $\{1, 3, 5, 10, 20, 40, 60, 100\}$ in each setting. In practice, the extreme values 1 and 100 were never chosen.

CHIMP: we set `base_n` $= 2$ and the right endpoints of the 20 time intervals as $\exp\{3 + 7.6(t-1)/19\}$ for $t = 1, \ldots, 20$. To accelerate the convergence, we further initialise the population sizes at each time interval to be $N(g) + \varepsilon$, where g is the

starting time of the interval and ε is a Gaussian error with mean zero and standard derivation $N(g)/7$. By doing so, we provided additional information of the demography, in return for faster computing time.

All other inputs are set by default. We use $T = 20$ time intervals for $N(g)$ in all methods compared using Model 1-6, except for Relate, which automatically determines the time partition.

Implementation details for human whole-genome data

Tree sequences data were downloaded from https://zenodo.org/records/5512994 and sequences extracted for both genomes of the 1 625 individuals whose population affiliation matched one of the 16 populations (Table 2).

We have previously shown Huang et al. (2025) that mutation rates can vary over chromosomes, and the pattern of variation may differ across populations. To allow for this variation, for each population we estimate the average mutation rate for chromosome i by $1.3 \times 10^{-8} \times (22M_i/\sum_{j=1}^{22} M_j)$ per site per generation, where M_i is the fraction of sites on Chromosome i that are polymorphic. Then we compute the matrices $\tilde{X}^{(i)}$ and $Y^{(i)}$ for chromosome i, and stacking them vertically to form $\tilde{X}$ and Y for the InferNo inference.

References

F. Baumdicker, G. Bisschop, D. Goldstein, G. Gower, A. P. Ragsdale, G. Tsambos, S. Zhu, B. Eldon, E. C. Ellerman, J. G. Galloway, A. L. Gladstein, G. Gorjanc, B. Guo, B. Jeffery, W. W. Kretzschumar, K. Lohse, M. Matschiner, D. Nelson, N. S. Pope, C. D. Quinto-Cortés, M. F. Rodrigues, K. Saunack, T. Sellinger, K. Thornton, H. van Kemenade, A. W. Wohns, Y. Wong, S. Gravel, A. D. Kern, J. Koskela, P. L. Ralph, and J. Kelleher. Efficient ancestry and mutation simulation with msprime 1.0. *Genet-*

ics, 220(3):iyab229, 12 2021. ISSN 1943-2631. doi: 10.1093/genetics/iyab229. URL https://doi.org/10.1093/genetics/iyab229.

A. Bergström, S. McCarthy, R. Hui, M. Almarri, Q. Ayub, P. Danecek, Y. Chen, S. Felkel, P. Hallast, J. Kamm, H. Blanché, J. Deleuze, H. Cann, S. Mallick, D. Reich, M. Sandhu, P. Skoglund, A. Scally, Y. Xue, R. Durbin, and C. Tyler-Smith. Insights into human genetic variation and population history from 929 diverse genomes. *Science*, 367:eaay5012, 2020.

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

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

S. Browning, B. Browning, M. Daviglus, R. Durazo-Arvizu, N. Schneiderman, R. Kaplan, and C. Laurie. Ancestry-specific recent effective population size in the Americas. *PLoS Genet*, 14(5):e1007385, 2018.

S. R. Browning and B. L. Browning. Accurate non-parametric estimation of recent effective population size from segments of identity by descent. *American Journal of Human Genetics*, 97(3):404–418, 2015.

T. Cousins and A. Durvasula. Insufficient evidence for a severe bottleneck in humans during the Early to Middle Pleistocene transition. *Molecular Biology and Evolution*, 42:msaf041, 2025. doi: 10.1093/molbev/msaf041.

T. Cousins, A. Scally, and R. Durbin. A structured coalescent model reveals deep ancestral structure shared by all modern humans. *Nature Genetics*, 57:856–864, 2025.

Y. Deng, R. Nielsen, and Y. Song. A previously reported bottleneck in human ancestry 900 kya is likely a statistical artifact. *Genetics*, 229:1–3, 2024a. doi: 10.1093/genetics/iyae192.

Y. Deng, R. Nielsen, and Y. S. Song. Robust and accurate Bayesian inference of genome-wide genealogies for large samples. *bioRxiv*, 2024.03.16.585351, 2024b.

W. S. DeWitt, K. D. Harris, A. P. Ragsdale, and K. Harris. Nonparametric coalescent inference of mutation spectrum history and demography. *Proceedings of the National Academy of Sciences*, 118(21):e2013798118, 2021.

T. Druet, I. Macleod, and B. Hayes. Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity*, 112(1):39–47, 2014.

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

S. Fairley, E. Lowy-Gallego, E. Perry, and P. Flicek. The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Research*, 48(D1):D941–D947, 2020.

R. Fournier, Z. Tsangalidou, D. Reich, and P. F. Palamara. Haplotype-based inference of recent effective population size in modern and ancient DNA samples. *Nature Communications*, 14(1):7945, 2023.

R. Griffiths and S. Tavaré. Sampling theory for neutral alleles in a varying environment. *Philosophical Transactions of the Royal Society of London*, A 344:403–10, 1994.

R. N. Gutenkunst, R. D. Hernandez, S. H. Williamson, and C. D. Bustamante. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, 5(10):e1000695, 2009.

W. Hu, Z. Hao, P. Du, F. Di Vincenzo, G. Manzi, J. Cui, Y. Fu, Y. Pan, and H. Li. Genomic inference of a severe human bottleneck during the Early to Middle Pleistocene transition. *Science*, 381:979—-984, 2023.

Z. Huang, J. Kelleher, Y. Chan, and D. Balding. Estimating evolutionary and demographic parameters via ARG-derived IBD. *PLoS Genet*, page e1011537, 2025. doi: 10.1371/journal.pgen.1011537.

J. Kamm, J. Terhorst, R. Durbin, and Y. S. Song. Efficiently inferring the demographic history of many populations with allele count data. *Journal of the American Statistical Association*, 115(531):1472–1487, 2020.

J. Kelleher, Y. Wong, A. W. Wohns, C. Fadil, P. K. Albers, and G. McVean. Inferring whole-genome histories in large population datasets. *Nature Genetics*, 51(9):1330–1338, 2019.

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

A. Mahmoudi, J. Koskela, J. Kelleher, Y. Chan, and D. Balding. Bayesian inference of ancestral recombination graphs. *PLoS Computational Biology*, 18(3):e1009960, 2022.

S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, M. Zhao, N. Chennagiri, S. Nordenfelt, A. Tandon, P. Skoglund, I. Lazaridis, S. Sankararaman, Q. Fu, N. Rohland, G. Renaud, Y. Erlich, T. Willems, C. Gallo, J. Spence, Y. Song, G. Poletti, F. Balloux, G. van Driem, and D. Reich. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*, 538:201–206, 2016. doi: 10.1038/nature18964.

M. Mezzavilla and S. Ghirotto. Neon: An R package to estimate human effective population size and divergence time from patterns of linkage disequilibrium between SNPs. *Journal of Computer Science and Systems Biology*, 8(1), 2015.

P. F. Palamara, T. Lencz, A. Darvasi, and I. Pe'er. Length distributions of identity by descent reveal fine-scale demographic history. *American Journal of Human Genetics*, 91 (5):809–822, 2012.

A. P. Ragsdale and S. Gravel. Models of archaic admixture and recent history from two-locus statistics. *PLoS Genetics*, 15(6):e1008204, 2019.

A. P. Ragsdale, T. D. Weaver, E. G. Atkinson, E. G. Hoal, M. Möller, B. M. Henn, and S. Gravel. A weakly structured stem for human origins in Africa. *Nature*, 617(7962): 755–763, 2023.

M. D. Rasmussen, M. J. Hubisz, I. Gronau, and A. Siepel. Genome-wide inference of ancestral recombination graphs. *PLoS Genetics*, 10(5):e1004342, 2014.

E. Santiago, I. Novo, A. F. Pardiñas, M. Saura, J. Wang, and A. Caballero. Recent demographic history inferred by high-resolution analysis of linkage disequilibrium. *Molecular Biology and Evolution*, 37(12):3642–3653, 2020.

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

S. Schiffels and K. Wang. MSMC and MSMC2: The multiple sequentially Markovian coalescent. *Methods in Molecular Biology*, 2090:147–166, 2020.

G. Schwarz. The Bayesian information criterion. *Annals of Statistics*, 6:461–464, 1978.

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

J. Terhorst. Accelerated Bayesian inference of population size history from recombining sequence data. *bioRxiv*, 2024.03.25.586640, 2024.

J. Terhorst, J. A. Kamm, and Y. S. Song. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics*, 49(2):303–309, 2017.

G. Upadhya and M. Steinrücken. Robust inference of population size histories from genomic sequencing data. *PLoS Computational Biology*, 18(9):e1010419, 2022.

R. J. Wang, S. I. Al-Saffar, J. Rogers, and M. W. Hahn. Human generation times across the past 250,000 years. *Science Advances*, 9:eabm7047, 2023.

A. L. Williams, G. Genovese, T. Dyer, N. Altemose, K. Truax, G. Jun, N. Patterson, S. R. Myers, J. E. Curran, R. Duggirala, et al. Non-crossover gene conversions show strong GC bias and unexpected clustering in humans. *Elife*, 4:e04637, 2015.

A. W. Wohns, Y. Wong, B. Jeffery, A. Akbari, S. Mallick, R. Pinhasi, N. Patterson, D. Reich, J. Kelleher, and G. McVean. A unified genealogy of modern and ancient genomes. *Science*, 375(6583):eabi8264, 2022.

J. Ye. On measuring and correcting the effects of data mining and model selection. *Journal of the American Statistical Association*, 93(441):120–131, 1998.

Y. Zhou, S. Browning, and B. Browning. A fast and simple method for detecting identity-by-descent segments in large-scale data. *The American Journal of Human Genetics*, 106 (4):426–437, 2020.

Model	Time before present in units of 10^3 generations												
	0	0.025	0.03	0.3	0.4	1	1.8	1.9	3	9	18	18.1	30
1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	10	10	10	10	10	10	10	10	10	10	10	10	0
3	−50	−50	−50	−50	−50	−50	−50	−50	−4	−2	0	0	0
4	20	20	20	20	20	20	20	20	10	−3	−3	−3	0
5	0	0	0	1 400	0	0	−1 200	0	0	0	500	0	0
6	700	700	700	700	−270	0	0	0	27	−15	0	0	0
7	20 000	0	0	0	0	0	0	0	0	0	0	0	0
8	5 000	−15 000	10	10	10	10	10	10	10	10	10	10	10

Table 5: **Growth rates (per 10^5 generations) for eight simulation models.** Although increasing g corresponds to backwards in time, growth is measured forward in time so a positive growth rate means that $N(g)$ decreases as g increases. The rates in each column apply from the generation shown at the top of that column until that shown in the next column, or for all $g > 30K$ in the case of the final column. See Table 1 for population sizes at the same time points.

	$n = 10$					$n = 200$				
Model	InferNo	Relate	PSMC	Mushi	CHIMP	InferNo	Relate	PSMC	Mushi	CHIMP
1	0.64	20.2	52.9	6.7	190	9.2	464	985	9.5	271
2	0.65	17.8	53.3	6.8	850	8.7	380	979	9.2	947
3	0.65	17.0	52.9	7.2	113	7.8	345	982	9.6	243
4	0.64	22.2	52.8	7.3	144	8.5	479	982	9.4	252
5	0.68	25.7	52.2	7.3	192	9.9	483	985	8.8	270
6	0.56	15.7	52.7	3.0	130	6.7	287	984	8.8	334
Mean	0.64	19.8	52.8	6.4	270	8.5	406	983	9.2	386

Table 6: **CPU time (seconds).** Average computing time (over the 25 replicates in each setting) for the results in Figure 8 ($n = 10$) and Figure 3 ($n = 200$).

	Sample size n						Recombination rate			Error rate ϵ			
Model	10		100		1 000		map	$r = 1$	$r = 1.5$	0	1	2	3
1	1.6	(3.8)	0.5	(2.8)	0.3	(3.6)	2.8	1.9	3.5	1.7	1.9	2.4	2.4
2	0.9	(3.2)	1.0	(2.1)	0.5	(2.2)	2.6	2.3	2.0	2.0	2.2	2.3	3.3
3	1.1	(6.2)	0.7	(6.7)	0.6	(5.5)	6.3	3.2	3.7	3.3	3.4	3.4	3.1
4	8.3	(18.2)	2.4	(5.1)	0.8	(4.9)	5.5	5.4	7.1	4.6	5.3	6.7	7.1
5	6.1	(10.8)	2.6	(8.2)	1.6	(7.4)	7.8	7.9	8.3	7.3	8.3	8.9	10.0
6	9.0	(9.2)	1.9	(8.0)	1.2	(8.4)	8.0	8.1	9.5	7.5	7.5	7.9	8.1
Mean	ref	(ref)	0.34	(0.63)	0.18	(0.61)	1.15	ref	1.18	ref	1.09	1.20	1.29

Table 7: **RMISE of InferNo (in units of 10^5) under perturbations of the data generating model.** Entries in the bottom row are expressed relative to a baseline indicated by "ref". r is stated in units of 10^{-8} and ϵ is per 10^5 sites. Entries under "Sample size" correspond to the curves in Figure 10, with integration over $g \in [200, 1\,000]$ ($g \in [1\,000, 30\,000]$). Entries under "Recombination rate" correspond to the curves in Figure 11 and those under "Error rate" correspond to the curves in Figure 12; in both cases the integration is over $g \in [200, 30\,000]$.

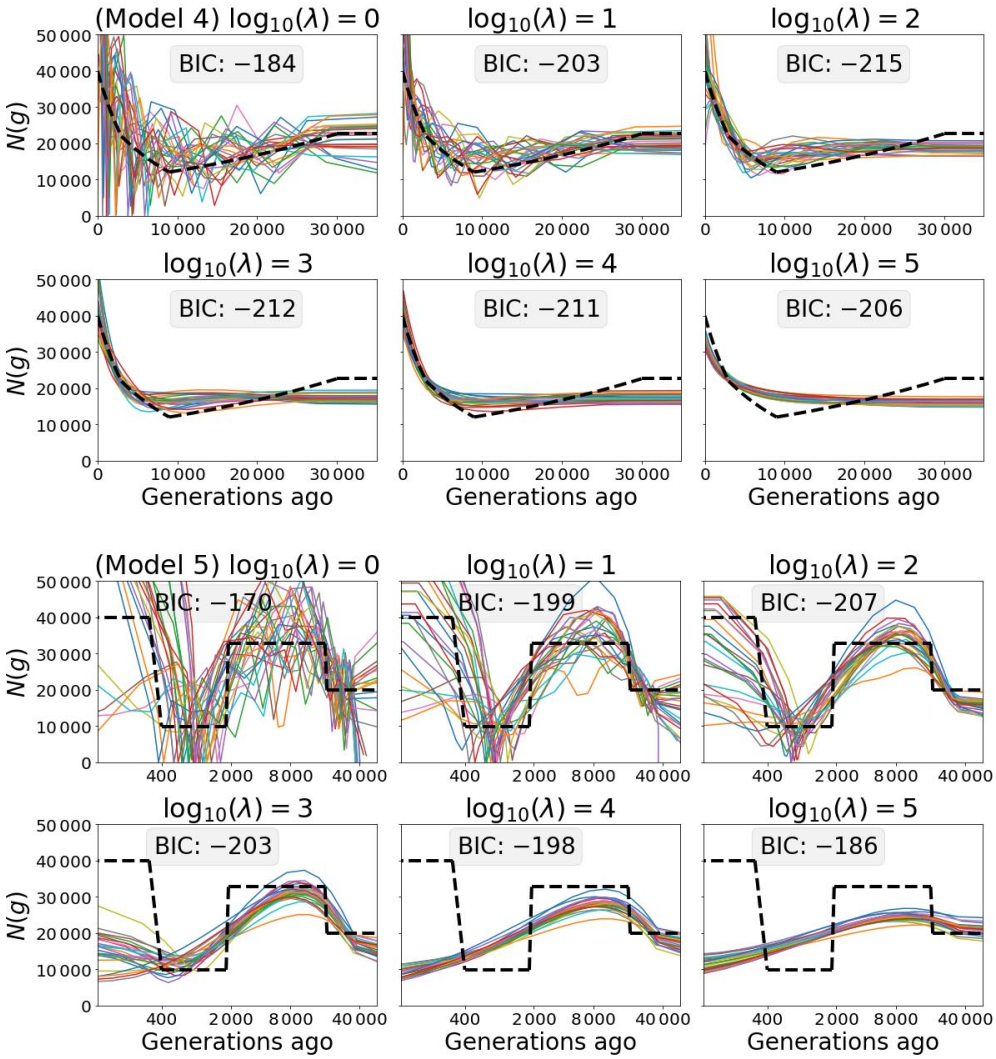

Figure 7: **Selecting smoothing parameter λ using BIC (in units of 10^3).** Estimates of population size $N(g)$, with $n = 100$, for Model 4 (top two rows) and Model 5 (bottom two rows, logarithmic time scale). The λ value is indicated above each panel. The minimum BIC is obtained for $\lambda = 100$ in both cases.

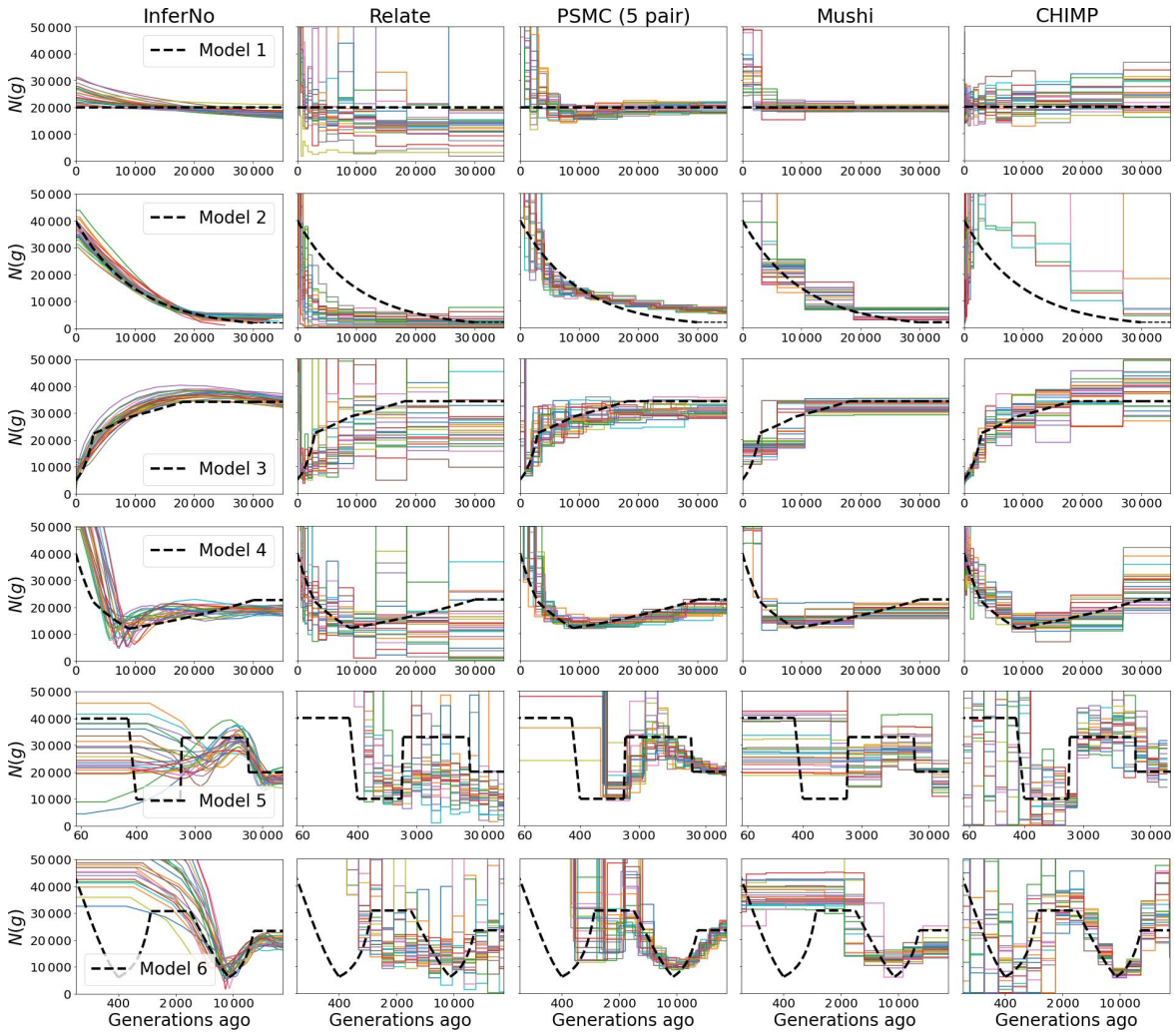

Figure 8: **Comparison with other methods** ($n = 10$). Other details are the same as for Figure 2.

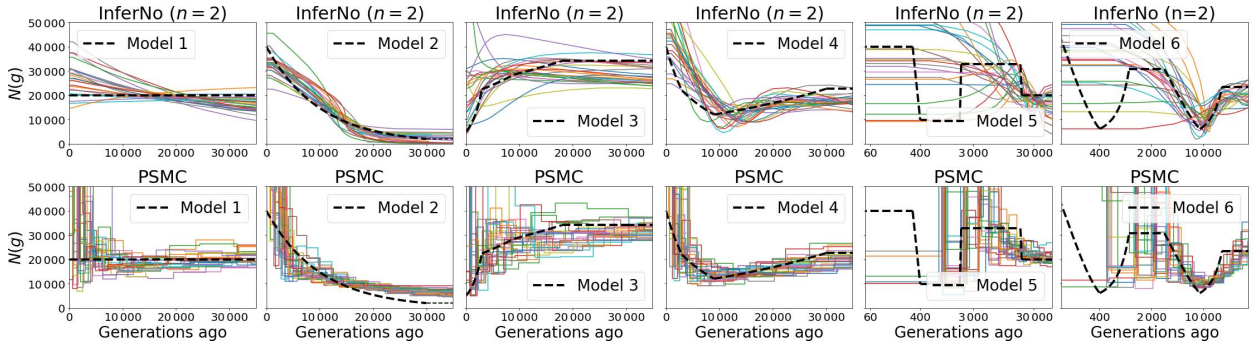

Figure 9: **Comparison of InferNo with PSMC when** $n = 2$. Sequence length $\ell = 10^7$ sites. Time scale is logarithmic for Models 5 and 6. See Table 4 for corresponding RMISE values.

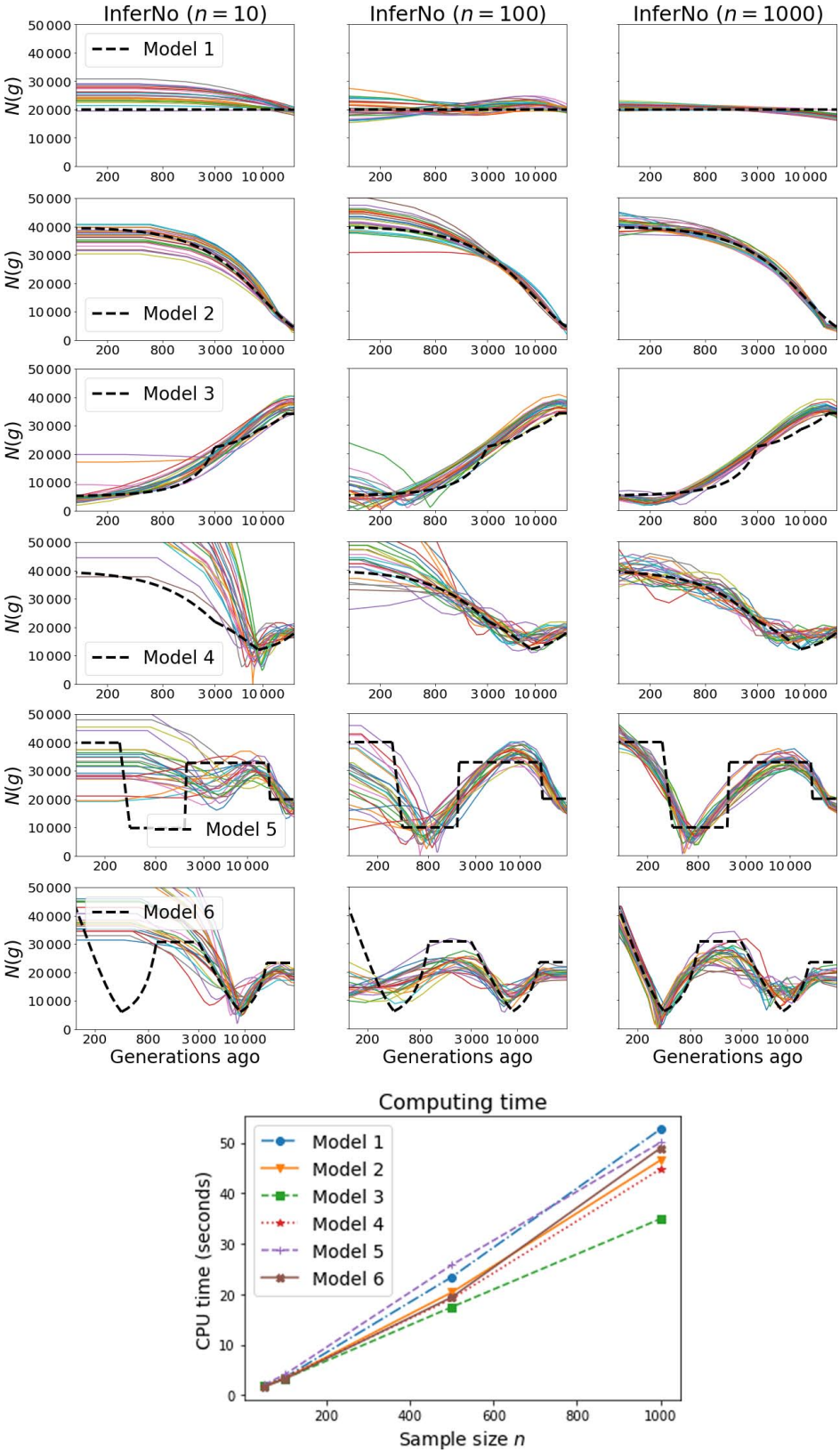

Figure 10: **Estimation and computational performance with varying sample sizes**. Rows 1-6: InferNo estimates of population size $N(g)$ for three values of n in Models 1-6, with g on a logarithmic scale. Bottom: run time with $n = 10$, 50, 100, 500 and 1 000.

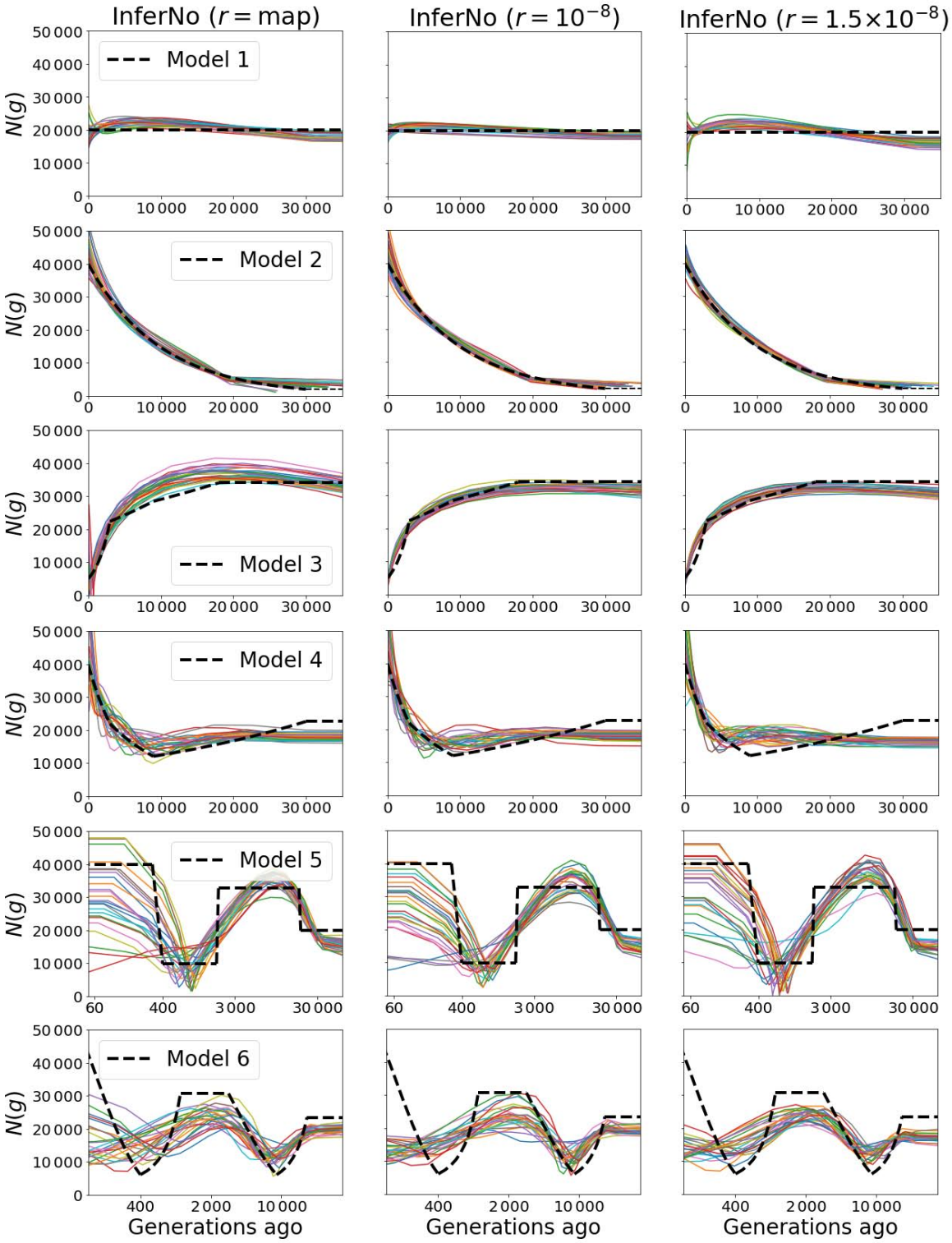

Figure 11: **Performance of InferNo with misspecified recombination rates**. Estimates of population size $N(g)$ with $n = 100$ when the observed data is generated with the recombination map in Figure 2 (column 1), and with constant rates $r = 1\times10^{-8}$ (column 2) and $r = 1.5\times10^{-8}$ (column 3). The middle column corresponds to the recombination model assumed by InferNo. Rows correspond to simulation Models 1 through 6.

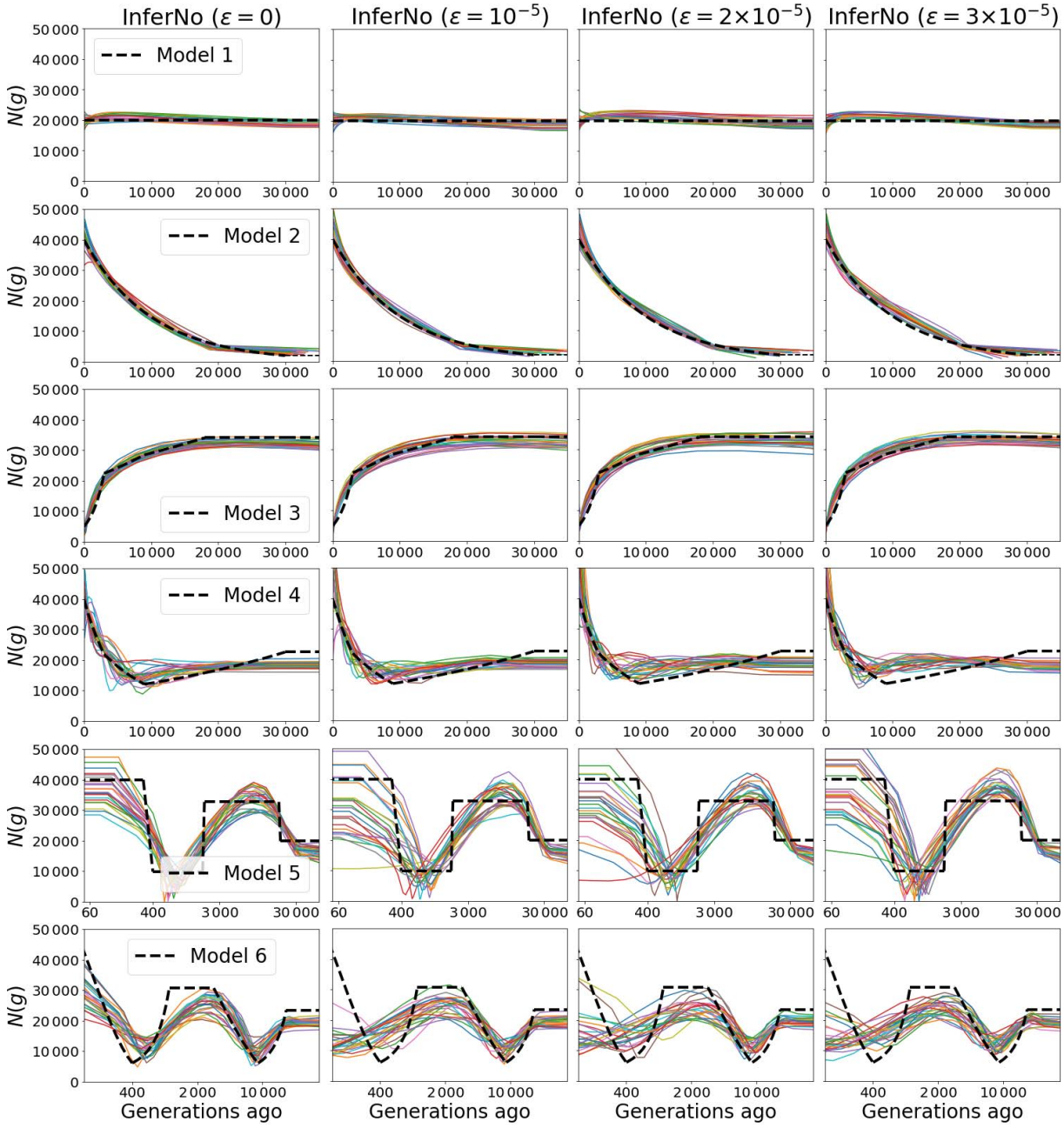

Figure 12: **Performance of InferNo with different levels of sequencing error**. Estimates of population size $N(g)$ with $n = 100$. Columns correspond to sequencing error rates $\epsilon = 0$, 1, 2, and 3 (per 10^5 generations). Rows correspond to simulation Models 1 through 6.

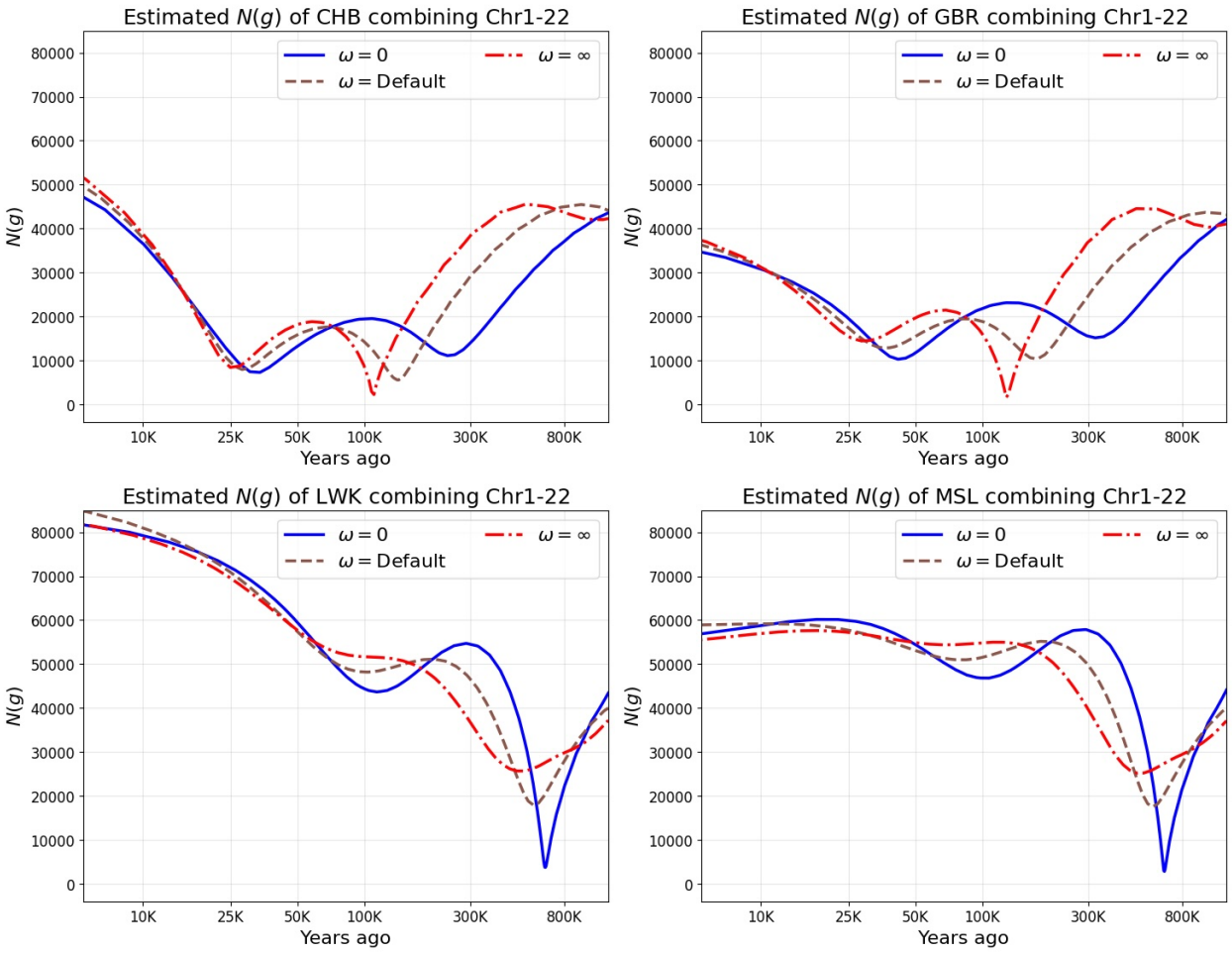

Figure 13: **Components of InferNo inference for four human populations**. Estimates of population size $N(g)$ using only AFS ($\omega = 0$), only TMRCA estimates ($\omega = \infty$) and the default ω setting (used for all other plots).

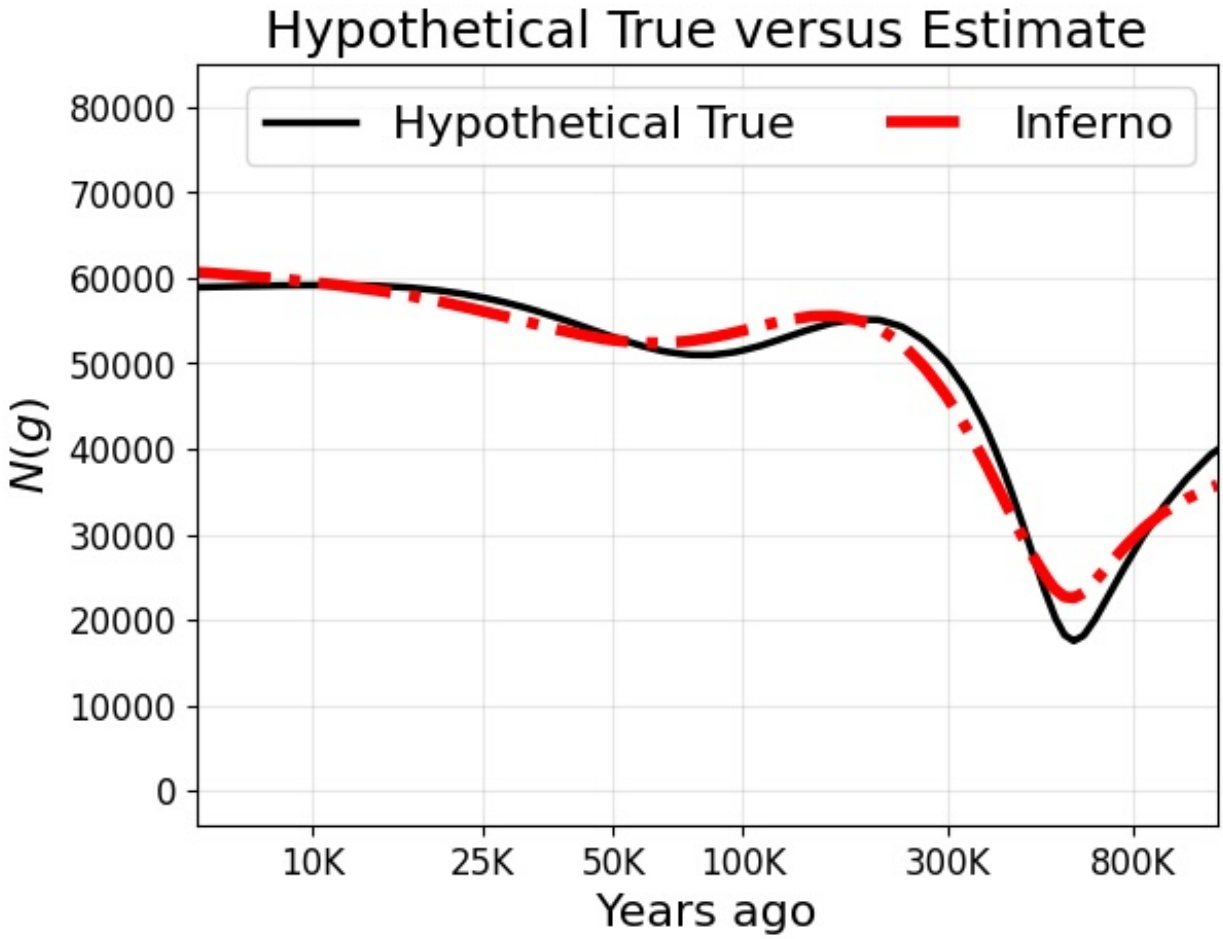

Figure 14: **Testing InferNo inference for an AFR population**. Similar to Fig. 5 (left) which shows inference for a non-AFR population. The red curve shows InferNo inference from $n = 200$ human sequences simulated with $N(g)$ as shown in the black curve, similar to the inferred $N(g)$ curve for AFR populations (Fig. 4, top left).