# 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

October 16, 2024

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 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. We also apply InferNo to worldwide human data, finding remarkable similarities in inferences from different populations in the same region. Unlike previous studies, we show two historic bottlenecks for most of the non-African populations.

**Keywords**: effective population size, semi-analytical inference.

# 1 Introduction

We propose a new method, "InferNo", to infer effective population size history from a sample of sequences, which can provide 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 are based on the sequential Markov coalescent (SMC) model, and infer population size history from the estimated time since the most recent common ancestor (MRCA) along the genome. 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 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 sample sizes. 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 leads to an ill-posed inverse problem (a given AFS can arise from different demographic histories), especially for small sample sizes (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., 2024).

Population-size inference can also be based on inferred identity-by-descent (IBD) relationships or linkage disequilibrium (LD) coefficients (Mezzavilla et al., 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 combines features of the AFS and SMC approaches to improve on both of them. It uses an ARG simulated under the standard coalescent model to estimate the distribution of mutation events in time and genome intervals. InferNo then identifies a time scaling of the standard coalescent to best match the observed AFS and pairwise site differences. This time scaling is inversely proportional to population size.

In simulation studies, we find that InferNo is overall more accurate and computationally faster than the SMC-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 real human data, we demonstrate inferences of population size history that are remarkably similar for populations on the same continent, but different among continents. In particular, we find evidence for two historic bottlenecks in most non-African populations and a substantial difference in the demographic histories of African and non-African populations dating back more than 10 thousand generations.

# 2 Methods

Our goal is to estimate $N(g)$, the effective haploid population size $g$ generations ago, based on a sample $C_i$, $i = 1, \ldots, n$, of $n$ homologous genome sequences each of length $\ell$. For $s = 1, \ldots, \ell$, the state at position $s$, denoted $C_i(s)$, is either 0 (ancestral) or 1 (derived); ancestral allele states can often be accurately inferred from related species.

## 2.1 Inference of mutation rate

In the derivations below we assume that at most one mutation has occurred at any site since the MRCA, but we demonstrate robustness to this assumption by not incorporating it into the data generation models of the simulation study.

We require an average mutation rate (per generation per site) across the genome as input, denoted $\bar{\mu}$. Site-specific mutation rates $\mu(s)$, $s = 1, \ldots, \ell$, can then be pre-specified if desired, or are otherwise estimated from the number of polymorphic sites in the neighbourhood of $s$, scaled so that the average rate across the genome is $\bar{\mu}$. Specifically, we partition the sites into $L^*$ intervals $J_1^*, \ldots, J_{L^*}^*$, each of length $\ell/L^*$, and estimate $\mu(s)$ by a piece-wise constant map fitted to $n_l^*$, the number of polymorphic sites in $J_l^*$:

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

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

## 2.2 Inferring the time scaling using AFS

We assume a variable-size coalescent with recombination model, with coalescence rate $1/N(g)$ per generation for each pair of lineages, where $N(g)$ is the population size $g$ generations ago. A key idea underlying InferNo is that this variable-size model is equivalent to a constant-size model with time scaled in proportion to $1/N(g)$ (Griffiths and Tavaré, 1994). The mutation rate is assumed constant over time in the variable-size model, but the mapping onto the constant-size model makes it proportional to $N(g)$.

To exploit this relationship, we simulate under a "null model" with constant population size $\tilde{N}$, sample size $n$, sequence length $\ell$, and recombination rate $r = 10^{-8}$ per site per generation. Here we use $\tilde{N} = 20\,000$. Although there is much flexibility in setting this scaling factor, it should be not too far from average values over the historic periods of interest. For large $\ell$ given the value of $r$, there can be enough internal replication along the length of the genome due to recombination that a single simulation is used in all analyses reported here, but multiple null-model simulations can be performed if desired for greater accuracy. The mutation rate in the null model, $\mu(s), s = 1, \ldots, \ell$, can be pre-specified or estimated using (1). The mutation rate is constant over time in the null model, but proportional to $N(g)$ in the variable-size model. Thus we can infer $N(g)$ by mapping mutation rates in the null model to the observed allele frequency data.

To this end, we partition the time domain for the null model into intervals $\tilde{I}_t = [\tilde{g}_{t-1}, \tilde{g}_t)$, $t = 1, \ldots, T$, where $\tilde{g}_0 = 0$ and $\tilde{g}_T = \infty$. The $\tilde{g}_t$ can be pre-specified, for example, based on historical periods of interest or prior knowledge about $N(g)$. By default, we set them to approximately equalise the number of null-simulation mutations during each $\tilde{I}_t$, except that when we focus on recent inferences we choose the $\tilde{g}_t$ so that the resulting $g_t$ are concentrated in the time period of interest.

We then approximate $N(g)$ by a piece-wise constant function $N(g) \approx \tilde{N}\beta_t$ for $g \in I_t =$

$[g_{t-1}, g_t)$, where $g_0 = 0$, $g_T = \infty$, and $I_t$ is defined so that the cumulative per-site coalescence rate over $I_t$ in the variable-size model, which is $(g_t - g_{t-1})/(\tilde{N}\beta_t)$, equals that over $\tilde{I}_t$ in the null model, or $(\tilde{g}_t - \tilde{g}_{t-1})/\tilde{N}$. From this equality, we obtain

$$\beta_t = \frac{g_t - g_{t-1}}{\tilde{g}_t - \tilde{g}_{t-1}}, \qquad t = 1, \ldots, T-1. \tag{2}$$

The inference task is now, given our choices for $\tilde{N}$ and the $\tilde{g}_t$, to estimate $\beta = (\beta_1, \ldots, \beta_T)$ and the $g_t$ from the sequence data.

We first attack this inference task by relating the mutations observed in a null simulation to the sample AFS, which is a length-$(n-1)$ vector $Y_1$ with $k$th element equal to the number of sites at which the derived allele has frequency $k$. From the null simulation we construct $\tilde{X}_1$, an $(n-1) \times T$ matrix with $\{k, t\}$ element equal to the number of sites with a frequency-$k$ derived allele where the mutation arose during the time interval $\tilde{I}_t$ (oldest mutation if there is more than one). Mutations arise during $I_t$ in proportion to $\beta_t$, but are uniform over time in the null model, so we estimate $\beta$ via the regression equation

$$Y_1 = \tilde{X}_1 \beta + \epsilon \ . \tag{3}$$

The resulting inferences are accurate when $n$ is large, but for small $n$ we next seek additional estimating equations by refining the AFS into sequence pair mismatches in genome intervals.

## 2.3 Localising mutations to sequence pairs and genome intervals

We partition the genome $[0, \ell)$ into intervals $J_l = [h_{l-1}, h_l)$, $l = 1, \ldots, L$, such that approximately $c$ mutations are expected per lineage per generation in each $J_l$. This is obtained by minimizing $|c - \sum_{s \in J_l} \hat{\mu}(s)|$ for $l = 1, \ldots, L$. Here we fix $c = 2.6 \times 10^{-4}$, but InferNo is robust to the choice of $c$, which controls the number of intervals $L$.

Let $Z$ denote a length-$nL$ vector with element $\{L(i-1) + l\}$ the number of site differences in $J_l$ between the sequences $C_i$ and $C_{i+1}$, where $C_{n+1} = C_1$. We only use neighbouring pairs of

sequences to reduce computational effort, 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 of the neighbour pairs. We assume that the $\{L(i-1)+l\}$th element $z$ of $Z$ is a Poisson$(2c\tau_l^i)$ random variable, conditional on $\tau_l^i$, the average TMRCA (time since the MRCA) of $C_i$ and $C_{i+1}$ in the interval $J_l$. See Supplementary Methods for estimates of $f_\tau$ and $p_z$, the probability density and mass functions of $\tau_l^i$ and $z$, and their use to derive $Y_2$, a length-$nL$ vector whose $m$th element estimates the $m/(nL+1)$ quantile of $f_\tau$.

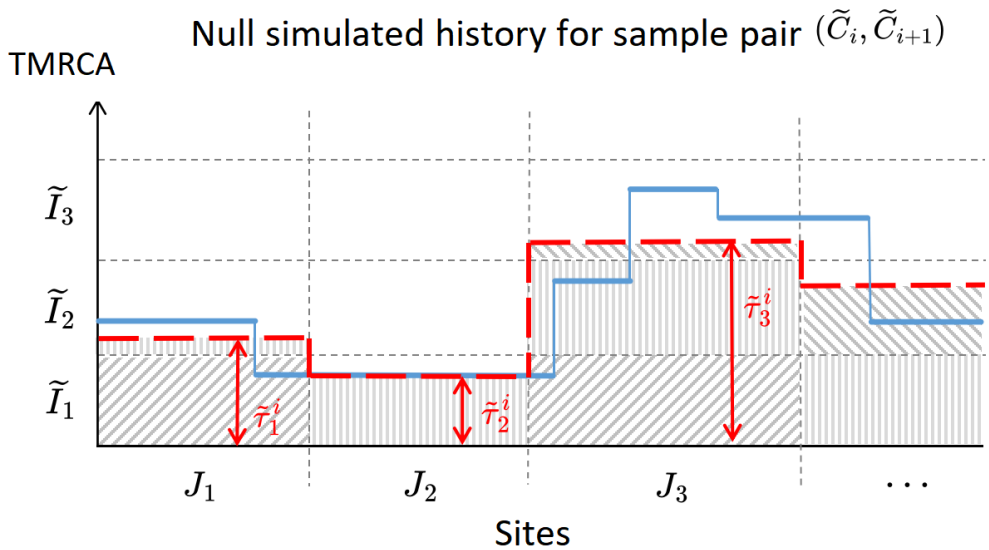


Figure 1: The TMRCA in a null simulation (solid blue) and its approximation $\tilde{\tau}_l^i$ (dashed red) which equals the average blue value in each interval $J_l$. The heights of the shaded rectangles correspond to the elements of $\tilde{W}^i$ defined in the text.

Let $\{\tilde{C}_1, \ldots, \tilde{C}_n\}$ be the sequences generated in the null simulation. As we did for the variable-size model, for each sequence pair $(\tilde{C}_i, \tilde{C}_{i+1})$ we use the average TMRCA over $J_l$, denoted $\tilde{\tau}_l^i$ (see Figure 1). Let $\tilde{W}^i$ be an $L \times T$ matrix with entry $\{l, t\}$ the contribution to $\tilde{\tau}_l^i$ from the time interval $\tilde{I}_t = [\tilde{g}_{t-1}, \tilde{g}_t)$,

$$\tilde{W}_{l,t}^i = \begin{cases} \tilde{g}_t - \tilde{g}_{t-1} & \text{if } \tilde{\tau}_l^i > \tilde{g}_t \ , \\ \tilde{\tau}_l^i - \tilde{g}_{t-1} & \text{if } \tilde{g}_{t-1} < \tilde{\tau}_l^i < \tilde{g}_t \ , \\ 0 & \text{if } \tilde{\tau}_l^i < \tilde{g}_{t-1} \ . \end{cases} \tag{4}$$

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 (= TMRCA, $\tilde{\tau}_l^i$). Because $\beta > 0$ and the columns of $\tilde{X}_2$ are non-decreasing, $\tilde{X}_2\beta$ is in increasing order and we can infer $\beta$ from the regression equation

$$Y_2 = \tilde{X}_2\beta + \epsilon \ . \tag{5}$$

## 2.4 Combining the two estimating equations for $\beta$

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 the sequencing quality is known to be high. Let $Y = ({Y_1}', \omega {Y_2}')'$, where $'$ denotes matrix transpose, and $\tilde{X} = (\tilde{X}_1', \omega \tilde{X}_2')'$ with dimensions $K \times T$, where $K = (L+1)n-2$. Then (3) and (5) give $\tilde{X}\beta = Y$. To balance the information from the AFS, which is more useful for larger sample sizes, with that from sequence pairs, for $n < 2\,000$ we use the InferNo default $\omega = \sqrt{400c/(nL)}$, where $c$ is the constant used to define the genome partition. For larger $n$, we set $\omega = 0$, thus ignoring $\tilde{X}_2$ and $Y_2$ which greatly reduces computational effort.

Let $S$ be a diagonal matrix with $k$th diagonal element $s_k = (y_k+1)^{-1/2}$, $k = 1, \ldots, K$, where $y_k$ is the $k$th element of $Y$. Thus $1/s_k$ approximates the (Poisson) standard deviation of $y_k$, adjusted to ensure finite values. We estimate $\beta$ by

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \|S(\tilde{X}\beta - Y)\|_2^2 + \lambda \sum_{t=2}^{T-1} (\beta_{t+1} - 2\beta_t + \beta_{t-1})^2 \right\}, \tag{6}$$

subject to $\beta_t > 0$, $t = 1, \ldots, T$. The summation in (6) controls the smoothness of the $\beta_t$ by penalizing their second differences. See Supplementary Methods for details of selecting $\lambda$ using the adjusted Bayes information criterion (BIC, Schwarz, 1978; Ye, 1998).

The final step is to convert the piece-wise constant estimate $N(g) = \tilde{N}\hat{\beta}$ into a continuous piece-wise linear function by connecting the points of $\{g, N(g)\}$ at $\{0, \tilde{N}\hat{\beta}_1\}$ and

$\{(g_{t-1}+g_t)/2, \tilde{N}\hat{\beta}_t\}$ for $t \geq 1$, such that there is no change at each interval midpoint.

## 2.5 Design of simulation studies

| | | $g$ (in units of $10^3$ generations) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $N(0)$ | 0 | 0.025 | 0.03 | 0.3 | 0.4 | 1 | 1.8 | 1.9 | 3 | 9 | 18 | 18.1 | 30 |
| 1 | 20 000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 40 000 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 0 |
| 3 | 5 000 | -50 | -50 | -50 | -50 | -50 | -50 | -50 | -50 | -4 | -2 | 0 | 0 | 0 |
| 4 | 40 000 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 10 | −3 | −3 | −3 | 0 |
| 5 | 40 000 | 0 | 0 | 0 | 1 400 | 0 | 0 | −1 200 | 0 | 0 | 0 | 500 | 0 | 0 |
| 6 | 100 000 | 700 | 700 | 700 | 700 | −270 | 0 | 0 | 0 | 27 | −15 | 0 | 0 | 0 |
| 7 | 3 000 000 | 20 000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 300 000 | 5 000 | −15 000 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |

Table 1: **Growth rates (per $10^5$ generations) for the simulation models.** The rates in each column apply from the generation shown in the top row of that column until that shown in the next column, or indefinitely in the case of the final column. $N(0)$ is the present-day (haploid) population size.

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 show 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, with jumps in Model 5, while the population of Model 6 experienced two bottlenecks, allowing us to examine estimation of an older bottleneck when its signal may be blurred by a more recent bottleneck.

See Supplementary Methods 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 (Figure 2), 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

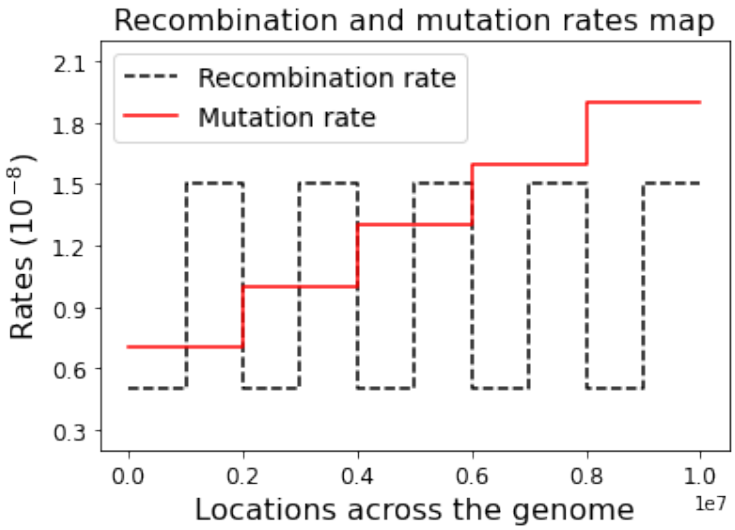


Figure 2: Recombination and mutation maps used in the simulation study. InferNo requires genome-wide average values as input, which for these maps are $r = 1$ and $\mu = 1.3$ per site per $10^8$ generations.

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 (1) the recombination map given in Figure 2, which has average rate $r = 10^{-8}$ per site and per generation, (2) a constant rate $r = 10^{-8}$ (which matches the InferNo analysis model) and (3) a constant rate $r = 1.5\times10^{-8}$.

Again with $n = 100$, we next 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 first compared InferNo with the IBD-based approach IBDNe (Browning and Browning, 2015) using Model 7 (fast recent exponential growth) with no sequencing error or gene conversion. For both methods, we used

sample size $n = 2\,000$, sequence length $\ell = 10^8$, input the correct mutation and recombination rates $\mu = 1.3r = 1.3{\times}10^{-8}$ 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.

Next, we tested the ability of InferNo to infer a very recent bottleneck ($g = 25$, Model 8) using a biobank-scale dataset ($n = 2{\times}10^5$). IBDNe is not tractable for $n$ at this scale. We set the sequence length to $\ell = 5 \times 10^7$, similar to human Chromosome 21, and kept $\mu = 1.3r = 1.3{\times}10^{-8}$. Sequencing error is included, at rate 1 per $10^5$ sites, but not gene conversion. We set $\tilde{N} = 10^5$ in InferNo, and to prioritize very recent inferences, we set $\tilde{g}_t = t/2$ for $t = 1, \ldots, 50$, and $\tilde{g}_t = 30\exp\{1.6(t-51)\}$ for $t = 51, \ldots, 56$.

## 2.6 Real data analysis

| Population | Description | Sample size | Region |
|---|---|---|---|
| LWK | Luhya in Webuye, Kenya | 198 | Africa (AFR) |
| MSL | Mende in Sierra Leone | 170 | |
| ESN | Esan in Nigeria | 198 | |
| YRI | Yoruba in Ibadan, Nigeria | 216 | |
| CHB | Han Chinese in Beijing, China | 206 | East Asia (EAS) |
| JPT | Japanese in Tokyo, Japan | 208 | |
| CHS | Southern Han Chinese | 210 | |
| CDX | Chinese Dai in Xishuangbanna, China | 186 | |
| FIN | Finnish in Finland | 198 | Europe (EUR) |
| GBR | British in England and Scotland | 182 | |
| IBS | Iberian Population in Spain | 214 | |
| TSI | Toscani in Italia | 214 | |
| 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 | |

Table 2: Populations from the 1 000 Genomes Project (Fairley et al., 2020) analysed here.

We estimated the population sizes $N(g)$ using the Chromosomes 20 ($\ell = 63\,025\,522$ sites)

and 21 ($\ell = 48\,129\,897$ sites) data from 16 populations of the 1000 Genomes Project (1KGP, Fairley et al., 2020), including four populations from each of Africa, East Asia, Europe, and South Asia (Table 2). See Supplementary Methods for data access details. On average over the 16 populations, 296 762 and 185 178 sites are polymorphic on Chromosome 20 and 21, respectively.

# 3 Results

## 3.1 Simulation study results

For a sample size of $n = 200$, Figure 3 shows that all five methods perform relatively poorly at recent times ($g < 10^3$) and when $g$ is very large ($g > 3 \times 10^4$). 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.

| 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) |

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

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 over $g \in [1\,000, 30\,000]$, 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

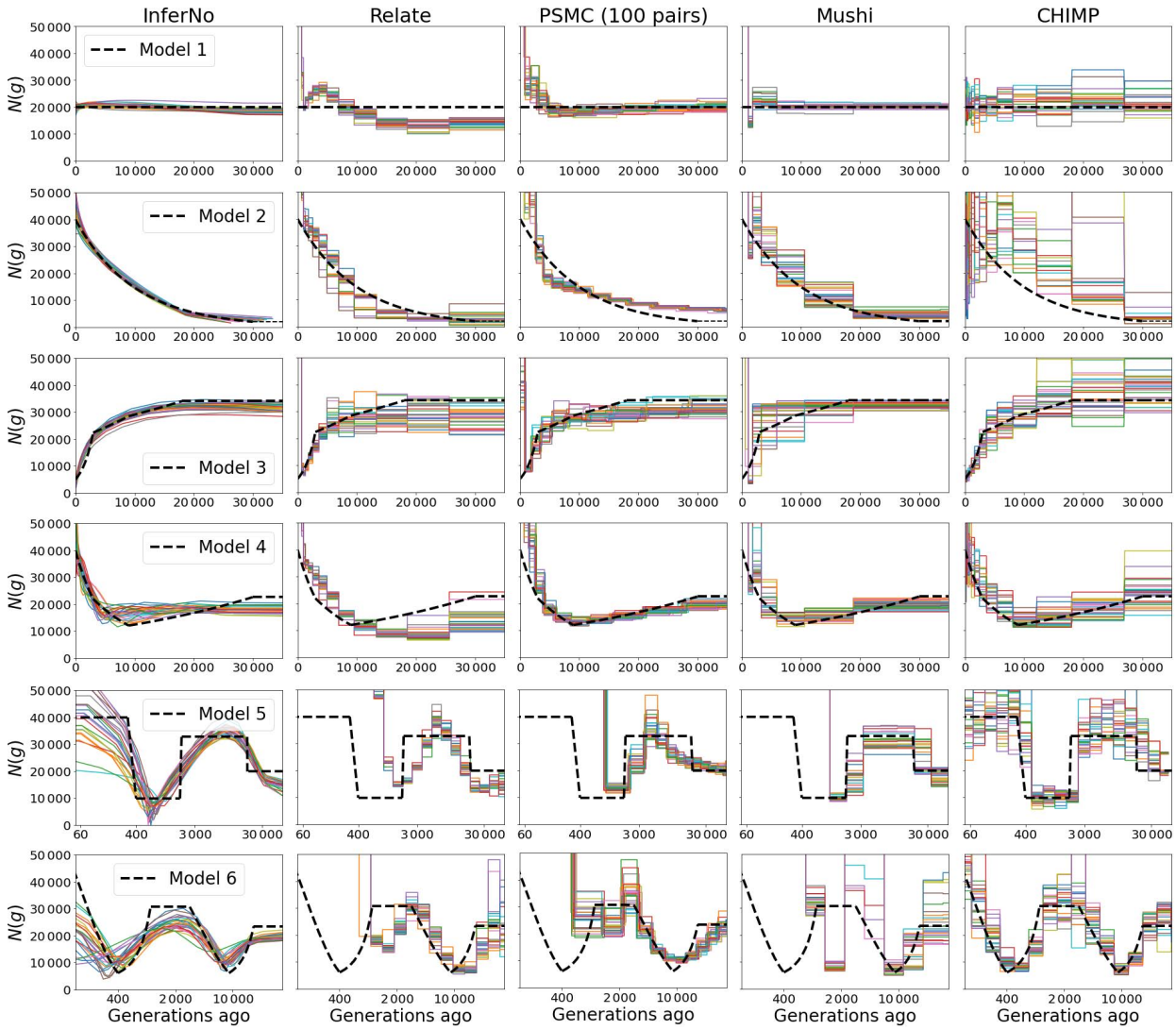


Figure 3: **Comparison of InferNo with four alternative methods for estimating population size** $N(g)$. Sample size $n = 200$ and sequence length $\ell = 10^7$ sites. Note logarithmic time scale for Models 5 and 6. See Table 3 for corresponding RMISE values, Figure S2 for plots when $n = 10$, and Table S1 for computing times.

| $n$=2 | | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
|---|---|---|---|---|---|---|---|
| $g \in [200, 30\,000]$ | InferNo | **8** | **6** | **13** | **16** | **21** | **14** |
| | PSMC | 6M | 9M | 3M | 9M | 8M | 6M |
| $g \in [1\,000, 30\,000]$ | InferNo | **7.3** | **6.1** | **11.7** | **15.2** | **18.2** | **12.2** |
| | PSMC | 30K | 3.0M | 14.3 | 2.7M | 51.0 | 0.9K |
| $g \in [5\,000, 30\,000]$ | 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** |

Table 4: **RMISE (in units of $10^5$) of three time intervals for InferNo and PSMC when** $n = 2$. Values are averages over the 25 curves of Figure S3.

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

InferNo is also the fastest among the five methods, taking on average (over 6 models and 25 replicate datasets for each model) 8.5 seconds per analysis for $n = 200$, 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 (Table S1). Mushi has a computing cost similar to InferNo when $n = 200$, and is slower when $n = 10$.

| | Sample Size $n$ | | | | | | Recombination | | | Error rate $\epsilon$ (per $10^5$ sites) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 5: **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". The sample size values correspond to the curves in Figure S4, with integration over $g \in [200, 1\,000]$ ($g \in [1\,000, 30\,000]$). The recombination values correspond to the curves in Figure S5 (see its caption for details of the recombination models) and the sequencing error values correspond to the curves in Figure S6; in both cases the integration is over $g \in [200, 30\,000]$.

Larger $n$ improves inference more for recent generations ($g < 1\,000$) than the distant past (Figure S4), because many lineages coalesce in recent generations. The average RMISE for $g \in [1\,000, 30\,000]$ decreases by only 39% when $n$ increases 100-fold from 10 to 1 000 (Table 5), but for $g \in [200, 30\,000]$, RMISE decreases by 82%. Table 5 (see also Figure S5) also shows little effect on inference of misspecifying recombination rates, with average RMISE increasing 15% or 18%, when the recombination map of Figure 2 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.

Because the computing time is approximately linear with $n$ for each model (Figure S4),

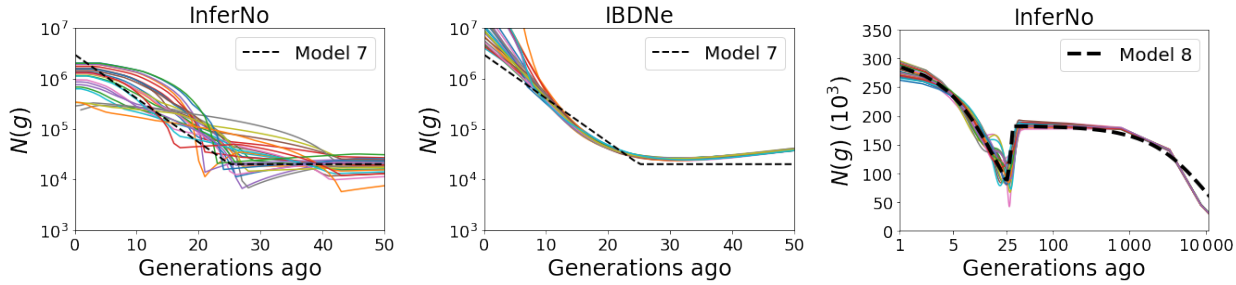


Figure 4: **InferNo and IBDNe estimates of population size $N(g)$ with large sample size under Model 7 and Model 8**. The sample size and sequence length for Model 7 (left and middle panels) are $n = 2\,000$ and $\ell = 10^8$ sites. For Model 8 (right panel), they are $n = 2{\times}10^5$ and $\ell = 5{\times}10^7$ sites.

InferNo can exploit large sample sizes to obtain accurate estimates of $N(g)$ in the recent past (small $g$). Figure 4 (left) 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 sample size $n = 2{\times}10^5$ (Figure 4, right), InferNo required only 326 seconds for each analysis to accurately infer a bottleneck 25 generations ago. 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.

## 3.2 Results from 1KGP data analysis

The $N(g)$ curves inferred from Chromosomes 20 and 21 are similar (Figure S7), and they are combined in Figure 5 which shows two groups of populations: two African and six non-African. The African populations have $N(g) > 40\,000$ ($= 20\,000$ individuals) for $g < 12\,000$ (up to 324 000 years ago if we assume 27 years per generation (Wang et al., 2023)) and $N(g)$ only briefly dips below 20 000, at about $g = 18\,000$. In contrast, the non-African populations mostly have $N(g) < 40\,000$ for $g < 12\,000$ and they each experience bottlenecks. The most severe bottleneck arose in the ancestry of the two EUR populations, with $N(g) \approx 5\,000$ at $g \approx 5\,200$ (about 140 000 years ago). The African and non-African populations remain distinct at $g = 20\,000$, well before the date usually given for the migrations of modern

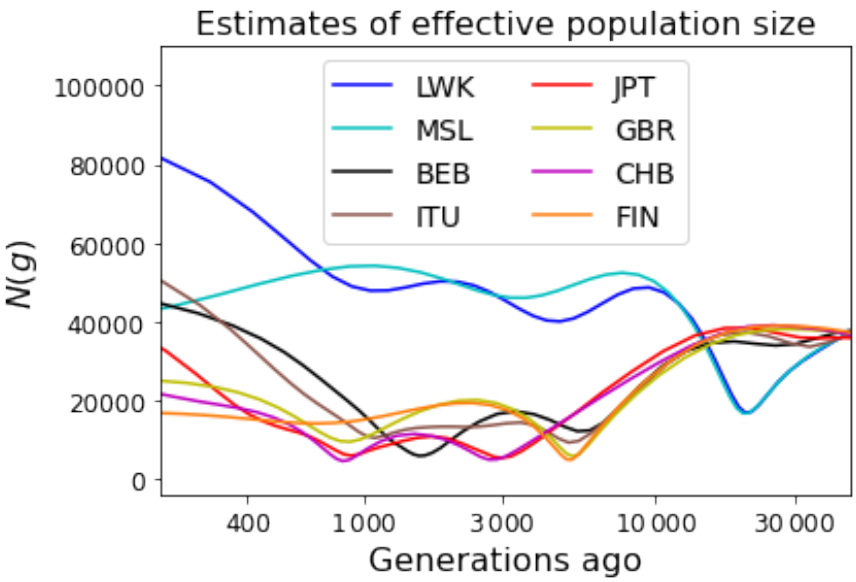


Figure 5: **InferNo estimates of population size $N(g)$ for eight 1KGP populations**. See Table 2 for the population codes. The results are weighted averages of the estimates from Chromosomes 20 and 21 (Figure S7), with weights proportional to the numbers of sites that are polymorphic in the eight populations combined. The time axis is on a logarithmic scale.

humans out of Africa, which is typically around $g = 3\,000$.

Figure 6 shows a striking similarity of the inferred $N(g)$ curves for $g > 1\,000$ across the four populations within each of four worldwide regions, despite the InferNo analysis being performed independently for each population. Most of the 12 non-African populations show two bottlenecks, with the most severe bottleneck in EAS, SAS and EUR populations occurring when $g \approx 900$, $g \approx 1\,600$ and $g \approx 5\,200$, respectively.

Our results show broad similarity with, but also important differences from, the results of Bergström et al. (2020, Fig 4), who analysed samples from different populations in the same regions, and Upadhya and Steinrücken (2022) who analysed three of the 1KGP populations studied here. Bergström et al. (2020) used 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 only one bottleneck in the histories of EUR and EAS populations, dated around $g = 1\,400$ and $g = 3\,700$, respectively. There is no ground truth available to conclude with certainty that any method is most accurate, but we

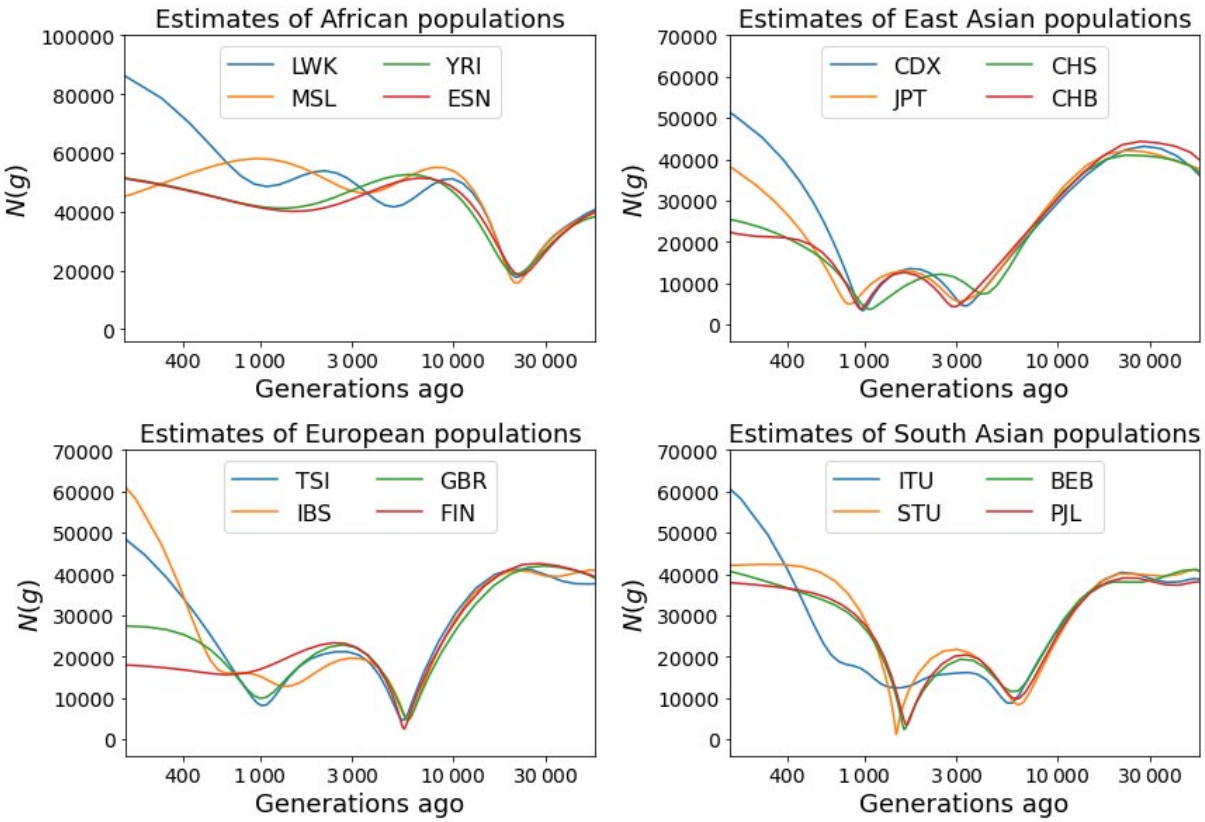


Figure 6: **InferNo estimates of population size $N(g)$ for four populations in each of four worldwide regions.** Based on data from Chromosome 20 only.

point to the better performance of InferNo in simulations, including under Model 6 with two bottlenecks. We suggest that signals from the two bottlenecks shown in the InferNo results may have been misinterpreted by the other methods as single bottleneck at a compromise time. InferNo also infers a single bottleneck when its smoothing parameter $\lambda$ is increased from the minimum-BIC value.

# 4 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 rescaling, which is inferred from the observed AFS and site differences

for sequence pairs in different genome regions. The accuracy and flexibility of this approach allows InferNo to achieve higher resolution population-size inferences, potentially identifying events that are undetectable by other methods, back in time until the MRCA is attained at most sites.

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 also the most computationally efficient of the methods, with cost increasing linearly with $n$, and we illustrated its ability to accurately infer very recent population sizes from biobank-scale datasets.

We used InferNo to estimate historical population size for 16 worldwide human populations from the 1000 Genomes Project. We found a striking similarity of the inferred $N(g)$ curves over the four populations within each of the four worldwide regions for $g > 1\,000$. As expected, we found the greatest differences between African and non-African populations, with the former showing larger $N(g)$ for $g < 12\,000$ and the latter showing bottlenecks with $N(g) < 10\,000$.

While InferNo avoids the challenging task of inferring recombination events, it requires sequences long enough that recombination generates replication along the genome, which is feasible for humans and many other species. InferNo allows $\mu(s)$ to vary over sites $s$, but like all other methods for inferring $N(g)$, it assumes that $\mu(s)$ is constant over time, which is unlikely to be strictly accurate.

Here we have inferred $N(g)$ for a single population, but the approach can be extended to multiple populations. Our real data example used human data, but InferNo can be applied to any populations that are approximately panmictic. More precise estimates of $N(g)$ can benefit inference for other evolutionary parameters, such as species divergence times.

# Data availability

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

# Acknowledgment

ZH is funded by Australian Research Council grant DP210102168 awarded to YBC and DJB.

# Supplementary methods

## Derivation of quantile approximation $Y_2$

Let $\mathbb{P}_a(b)$ denote the probability mass function at $b$ of the Poisson($a$) distribution and $Z_{\max} = \max\{Z\}$. Then

$$p_z(m) = \int_0^\infty \mathbb{P}_{2ct}(m) f_\tau(t)\,\mathrm{d}t \approx \frac{1}{2c}\sum_{t=0}^{Z_{\max}} \mathbb{P}_t(m) f_\tau(t/2c), \tag{7}$$

for $m = 0, 1, 2, \ldots$. Restricting $p_z(m)$ to $m \leq Z_{\max}$, (7) can be expressed as $p_z = Af_\tau/2c$, where $A$ is a matrix with $A_{i+1,j+1} = \mathbb{P}_j(i), i, j = 0, \ldots, Z_{\max}$, and the vector version of $f_\tau$ is evaluated at $t/2c$, for $t = 0, \ldots, Z_{\max}$. We estimate $p_z$ by $\hat{p}_z$, with $k$th element

$$\hat{p}_z^k = \frac{1}{nL}\sum_{m=1}^{nL} \mathbf{1}(Z_m = k), \qquad k = 0, \ldots, Z_{\max},$$

and then estimate $f_\tau$ by $\hat{f}_\tau = 2cA^{-1}\hat{p}_z$.

Let $\hat{F}_\tau$ be the vector estimating the cumulative distribution function of $\tau$, with its $j$th element being the summation of the first $j$ elements of $\hat{f}_\tau/2c$ with final entry one. Then the $\gamma$th-quantile ($\gamma \in (0, 1)$) is estimated by 0 if $\gamma < \hat{F}_{\tau,1}$, and $[k+(\gamma-\hat{F}_{\tau,k+1})/(\hat{F}_{\tau,k+2}-\hat{F}_{\tau,k+1})]/2c$ otherwise, where $\hat{F}_{\tau,j}$ refers to the $j$th element of $\hat{F}_\tau$ and $k \in \{0, \ldots, Z_{\max}\}$ is the index such that $\hat{F}_{\tau,j} \leq \gamma$ for all $j = 1, \ldots, k+1$ and $\hat{F}_{\tau,k+2} > \gamma$.

## Selecting the smoothing parameter $\lambda$ using BIC

To select $\lambda$, InferNo computes:

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

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 $\lambda$ 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 S1). 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 $\lambda$ 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.

## Software implementation details

For Relate, we set $N(0) = 30\,000$ for both ARG inference and population size estimation. For 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. For 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. For 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 of CHIMP, 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 $\varepsilon$ is a normal error with mean zero and standard derivation $N(g)/7$. By doing so, we provided additional information of the demography for CHIMP, in return for faster computing time. The 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 1KGP data

The sequence data were downloaded as .vcf files from ftp.1000genomes.ebi.ac.uk. Then, we converted them to the .samples format for the Python implementation of InferNo. Specifically, we first cloned the Github repository from the website github.com/awohns/unified_genealogy_paper (Wohns et al., 2022) and installed the software, packages and modules listed in the requirements.txt file and the tools sub-folder. Separately for Chromosome 20 and 21, in the all-data sub-folder we used Makefile to build the tree sequence from the variant data, and produce a .samples file from the .vcf file.

## Supplementary Tables and Figures

| | $n = 10$ | | | | | $n = 200$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 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 S1: **CPU time (seconds)**. Average computing time (over the 25 replicates in each setting) for the results in Figure 3 ($n = 200$) and Figure S2 ($n = 10$).

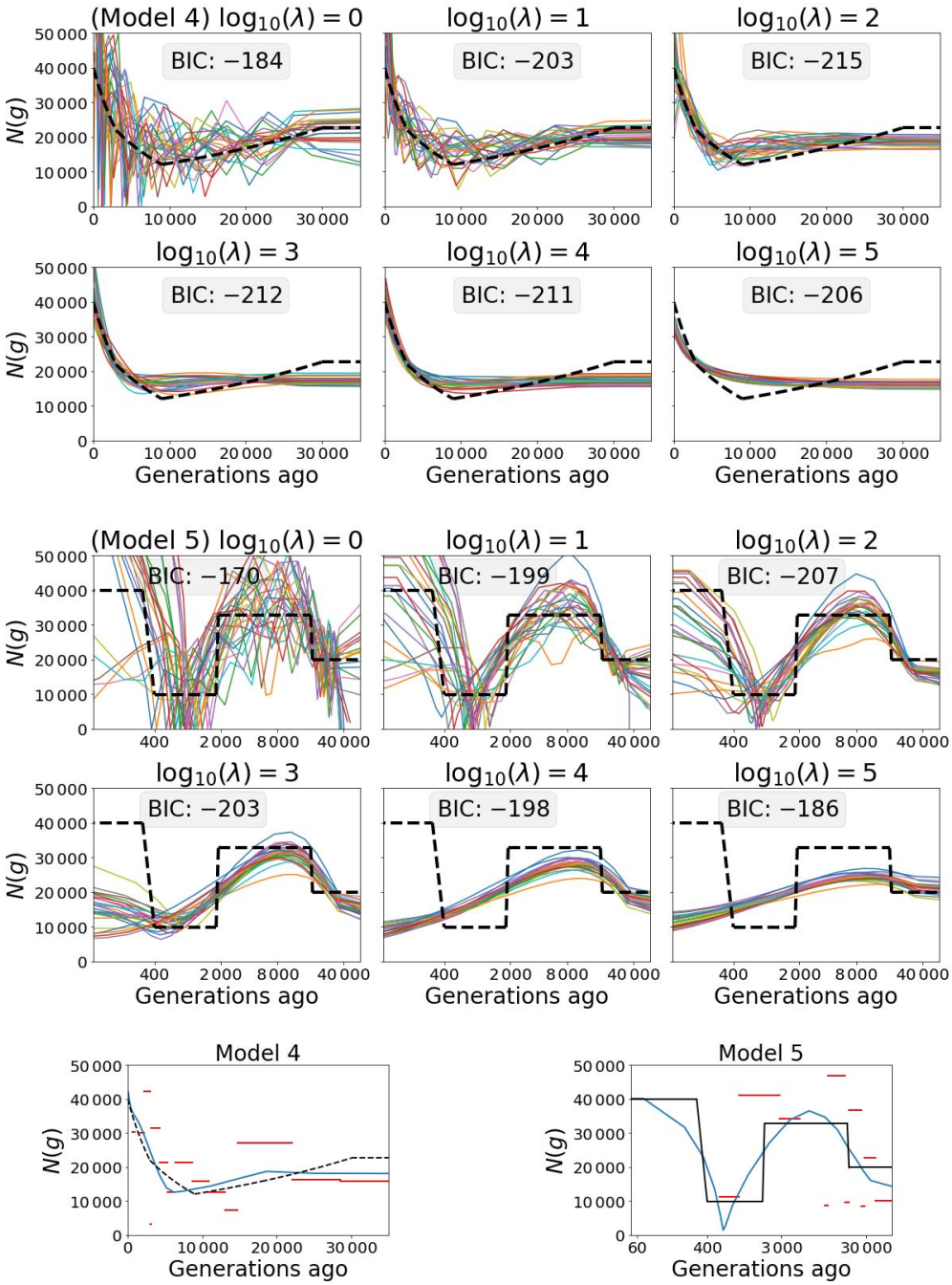


Figure S1: **Effects of smoothing parameter $\lambda$ and the corresponding BIC (in units of $10^3$).** Estimates of population size $N(g)$, with $n = 100$, for Model 4 (top two rows) and Model 5 (rows 3 and 4, logarithmic time scale). For each model, the $\lambda$ value is indicated above each of the six panels and the best fit (minimum BIC) is obtained for $\lambda = 10^2$ in both cases. The bottom row compares, for one example from each of Models 4 and 5, the final estimates (blue curve) with the piece-wise constant estimates obtained when $\lambda = 0$.

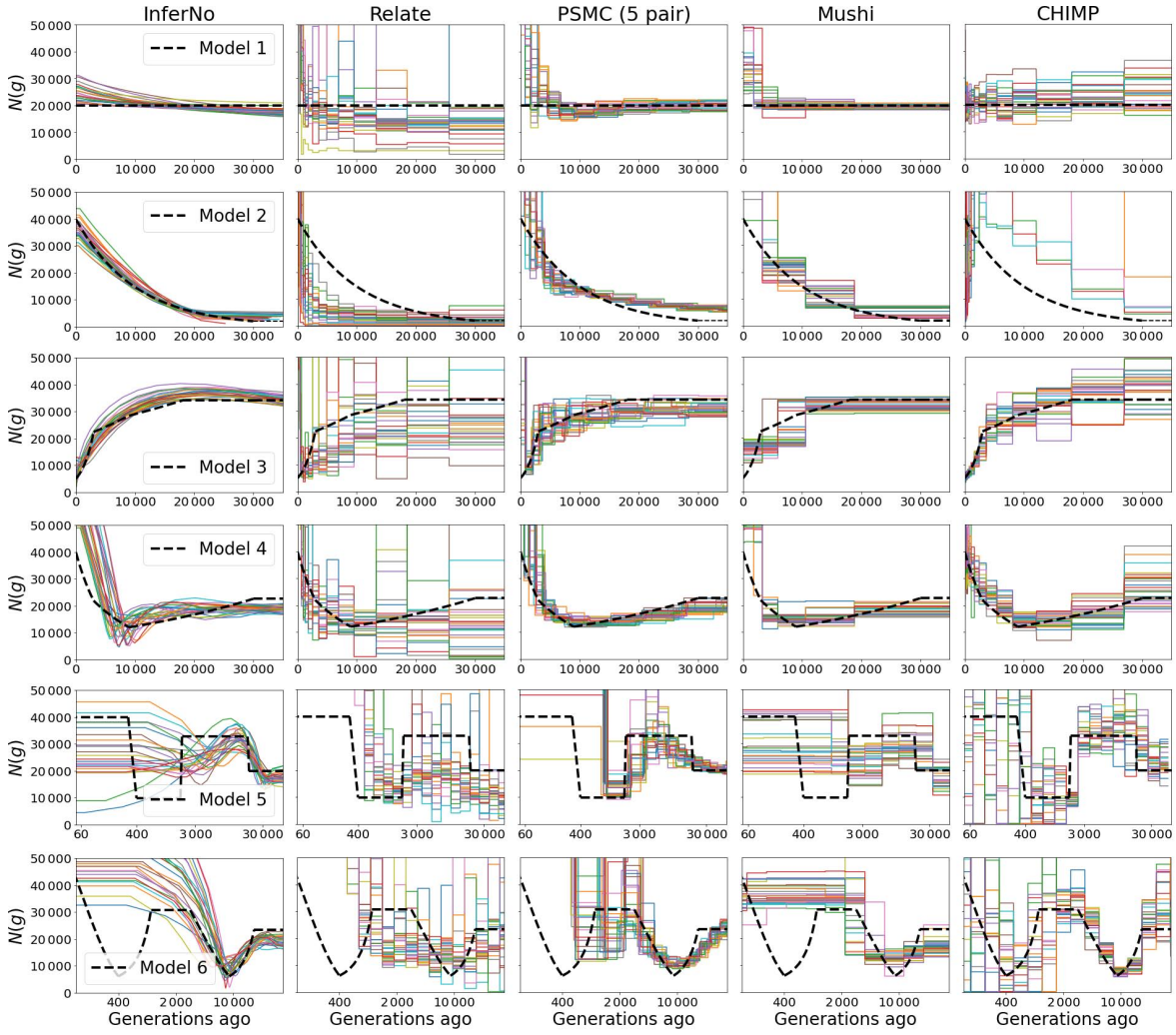


Figure S2: **Comparison with other methods** ($n = 10$). Other details are the same as for Figure 3.

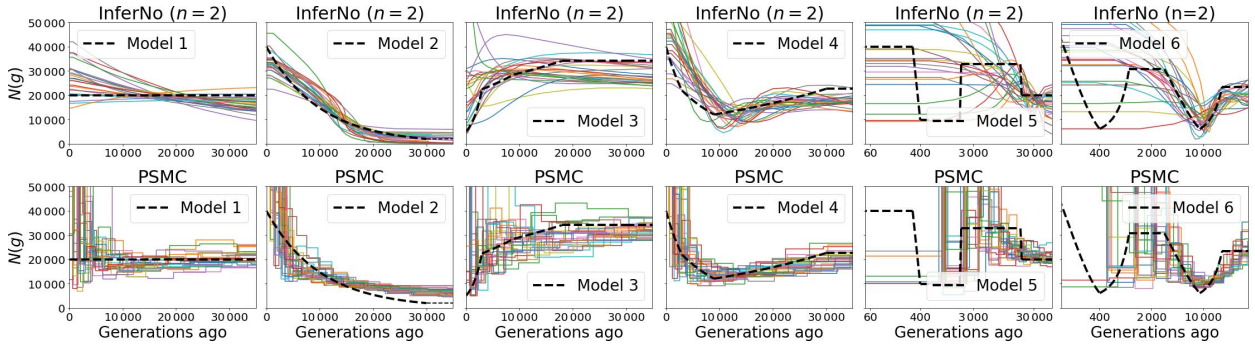


Figure S3: **Comparison of InferNo with PSMC when** $n = 2$. Sequence length $\ell = 10^7$ sites. Note logarithmic time scale for Models 5 and 6. See Table 4 for corresponding RMISE values.

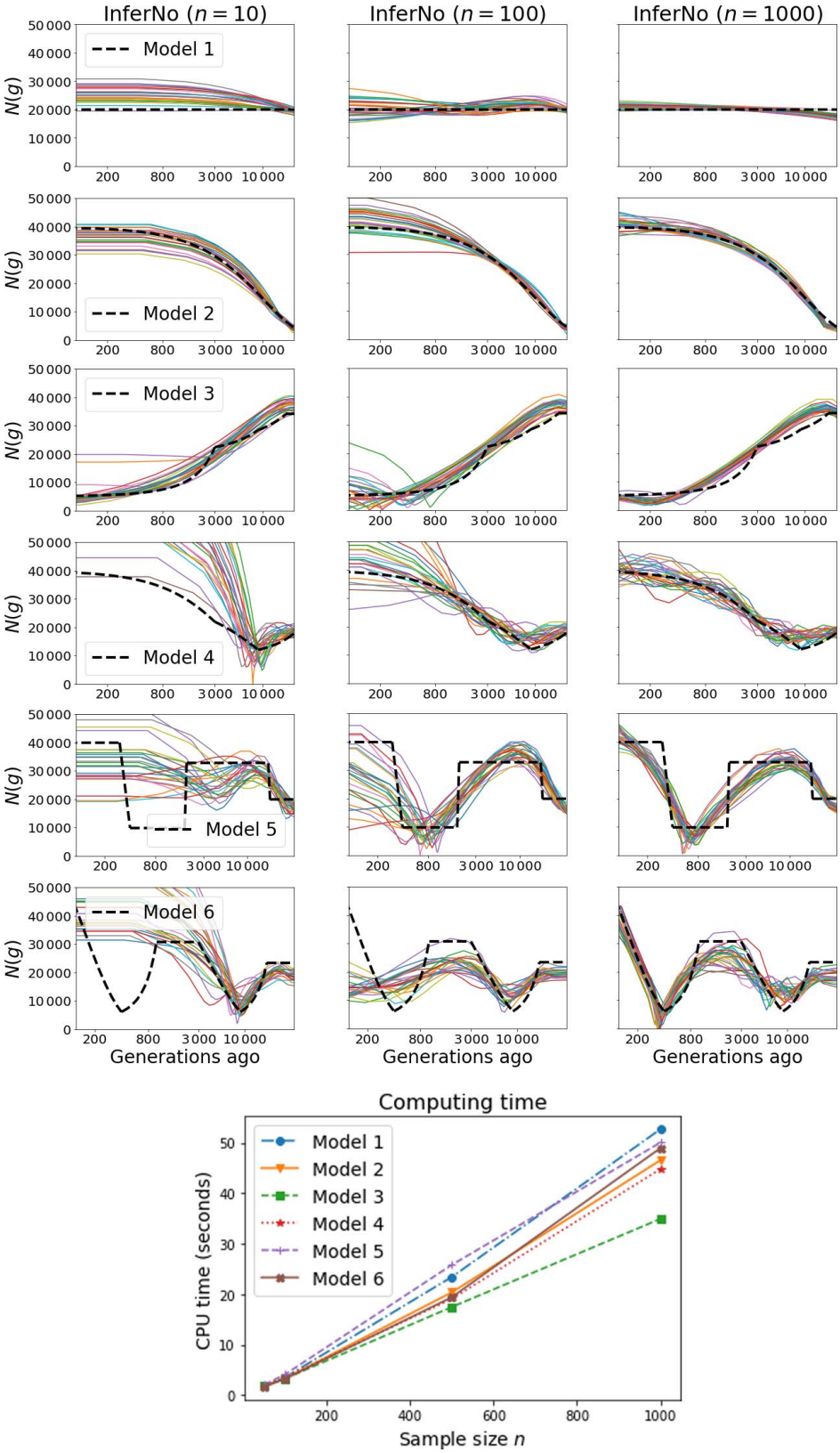


Figure S4: **Performance and computing time with different sample sizes**. InferNo estimates of population size $N(g)$ and run time with $n$ = 10, 50, 100, 500 and 1 000, for Models 1-6. Each time axis is on a logarithmic scale.

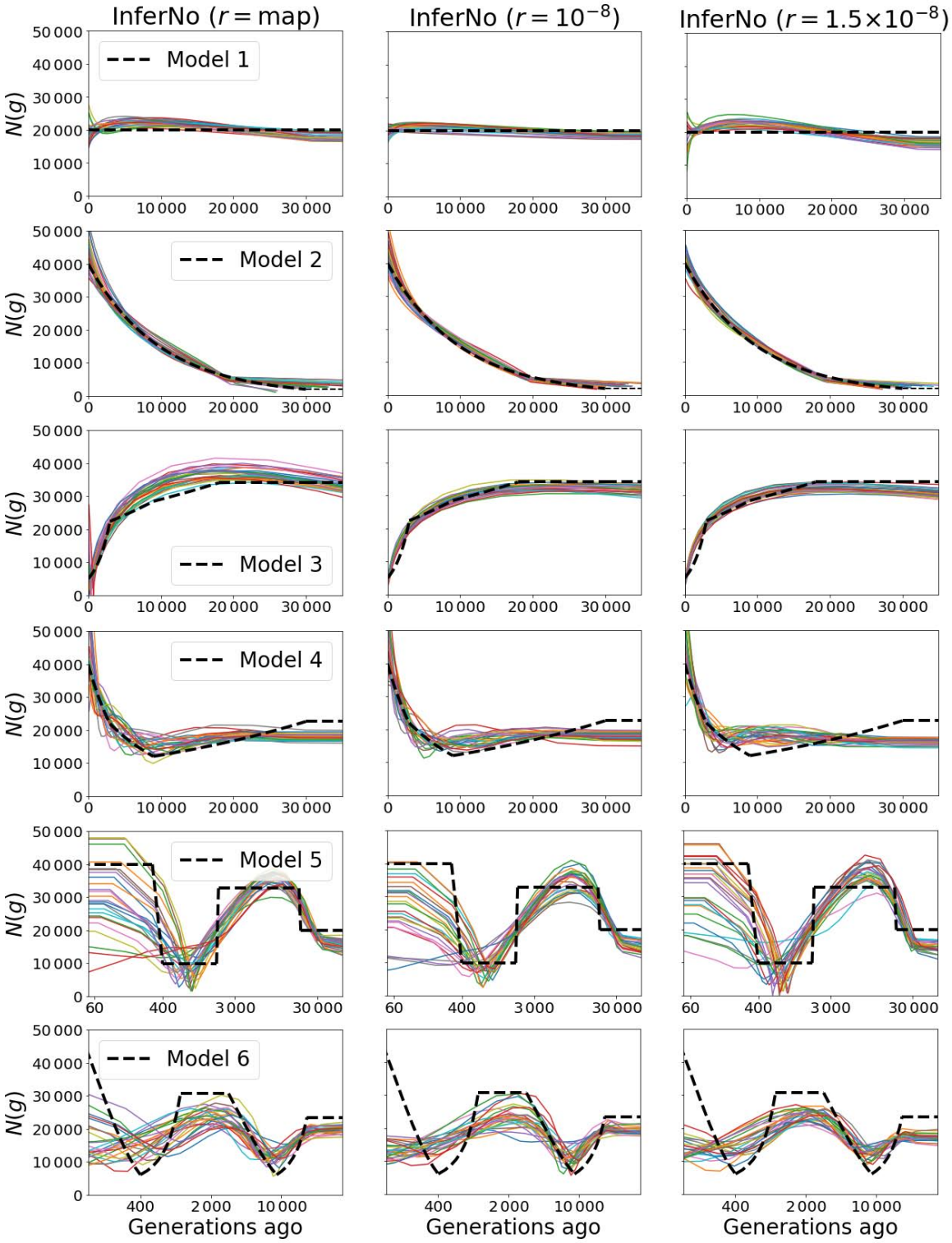


Figure S5: **Performance of InferNo with misspecified recombination rates**. Estimates of population size $N(g)$ with $n = 100$ when the observed data is generated with the recombination map in Figure 2 (column 1), and with constant rates $r = 1\times10^{-8}$ (column 2) and $r = 1.5\times10^{-8}$ (column 3). The InferNo analysis assumes a constant $r = 1\times10^{-8}$ and so the first and last columns each correspond to a misspecified recombination model. Rows correspond to simulation models 1 through 6.

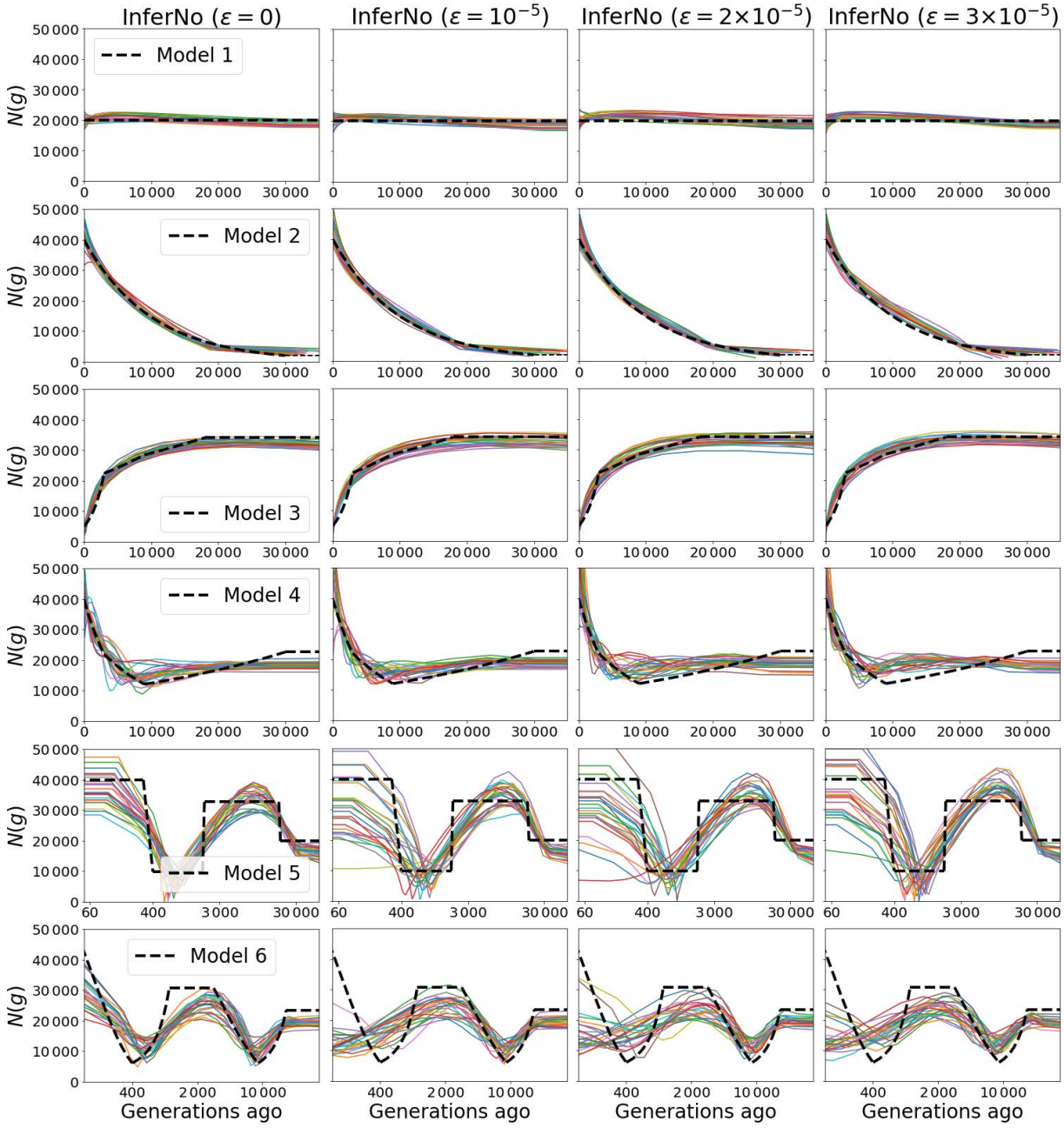


Figure S6: **Performance of InferNo with different levels of sequencing error**. Estimates of population size $N(g)$ with $n = 100$. Columns correspond to sequencing error rates $\epsilon = 0$, 1, 2, and 3 (per $10^5$ generations). Rows correspond to simulation models 1 through 6.

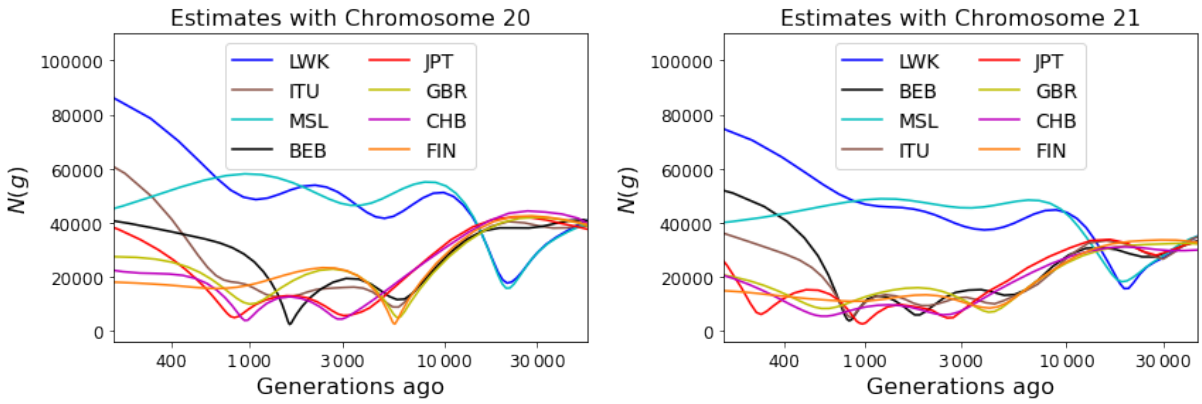


Figure S7: **InferNo estimates of population size $N(g)$ for eight 1KGP populations**. The data are from Chromosome 20 (left) and Chromosome 21 (right). The time axis is on a logarithmic scale.

# 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.

Y. Deng, R. Nielsen, and Y. S. Song. Robust and accurate Bayesian inference of genome-wide genealogies for large samples. *bioRxiv*, pages 2024–03, 2024.

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.

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.

M. Mezzavilla et al. 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.

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.