Accurate and flexible estimation of effective population size history

Zhendong Huang[a], **David Balding**[a,b,1], **and Yao-ban Chan**[a,1,2]

This manuscript was compiled on September 29, 2025

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.

effective population size | semi-analytical inference | human demographic history

Significance Statement

InferNo estimates population-size history from a sample of genome sequences, faster and more accurately than previous methods. The superiority of InferNo is most marked in the recent past. It requires no pre-specified population-size model, and is robust to variation in mutation and recombination rates along the genome, and to sequencing error and gene conversion. It can exploit very large sample sizes to infer human population sizes in recent centuries. Applied to worldwide human data, InferNo identifies novel features of demographic history: two local minima ("bottlenecks") in the histories of most non-African populations, and one ancient bottleneck for African populations. We infer distinct demographic histories for African and non-African populations, long before the usual dates for out-of-Africa migrations of modern humans.

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 (1), MSMC (2), diCal (3), SMC++ (4) and PHLASH (5). 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 (6–9). 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 (10).

A third class of approaches is based on inference of the ancestral recombination graph (ARG), which summarises the genealogical history of the sequences (11–14). 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

Author affiliations: [a]Melbourne Integrative Genomics, School of Mathematics & Statistics, University of Melbourne, VIC 3010, Australia; [b]UCL Genetics Institute, London WC1E 6BT, England

ZH had the original idea for the method, wrote the computer code, performed the simulation studies and write the first manuscript draft. YBC and DB supervised the project, suggested improvements and edited the manuscript.

The authors declare no competing interests.

[1] These authors contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: yaoban@unimelb.edu.au

recombination and mutation rates, and computational costs can be high despite recent advances (13, 15–18).

Population-size inference can also be based on inferred identity-by-descent (IBD) relationships or linkage disequilibrium (LD) coefficients (19–23). 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 (1) and CHIMP (24), AFS-based Mushi (10), ARG-based Relate (13), and IBD-based IBDNe (20). 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 $1/N(g)$ (25). 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 $g \in I_t$,

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

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

$$Y_1 = \tilde{X}_1 \beta + \epsilon\,. \qquad [2]$$

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 Eq. (2), we can now infer β via

$$Y_2 = \tilde{X}_2 \beta + \epsilon\,. \qquad [3]$$

Table 1. Population sizes (in units of 1K sequences, K denotes 10^3) at growth-rate change points for eight simulation models.

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

Growth rates are constant between change-points, see Table S1 for the rates. A blank entry indicates no change from the value to its left. $N(g) = N(30\text{K})$ for $g > 30\text{K}$.

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

$$Y = \tilde{X}\beta + \epsilon, \qquad [4]$$

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. To compare the performance of InferNo with Relate, PSMC, Mushi and CHIMP, we used msprime (26) 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 (27).

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 $r = 1.5\times10^{-8}$.

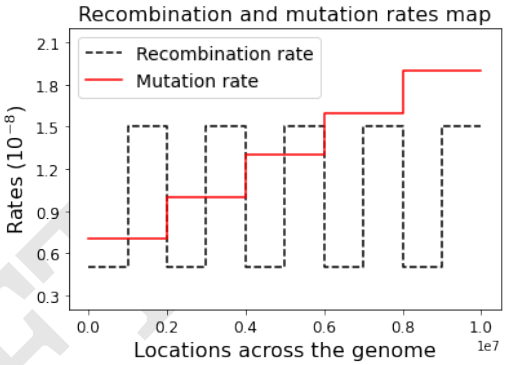

Fig. 1. Recombination and mutation maps used in the simulation study. InferNo requires genome-wide average values as input, which for these maps are $r = 1$ and $\mu = 1.3$ per site per 10^8 generations.

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 (20) 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 (28) 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 $\ell = 5 \times 10^7$, 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, 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 $t = 51, \ldots, 56$.

Table 2. Sample sizes (haploid) for human data analysis.

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

The number of individuals is $n/2$.

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 (29), supplemented by the Human Genome Diversity Project (30) and the Simons Genome Diversity Project (31). 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 (32). Use of this unified genealogy helps ensure uniform 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 $g > 3{\times}10^4$. 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.

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 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 S2), 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. S5), 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 S3), but when restricted to $g \in [1\,000, 30\,000]$ RMISE decreases by only 39%. Table S3 and Fig. S6 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 S3 and Fig. S7).

Because the computing time is approximately linear with n (Fig. S5), 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. 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 (33). 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 (34). Estimates of current size $N(0)$ are much more variable, ranging from 21K in FIN to 88K in LWK.

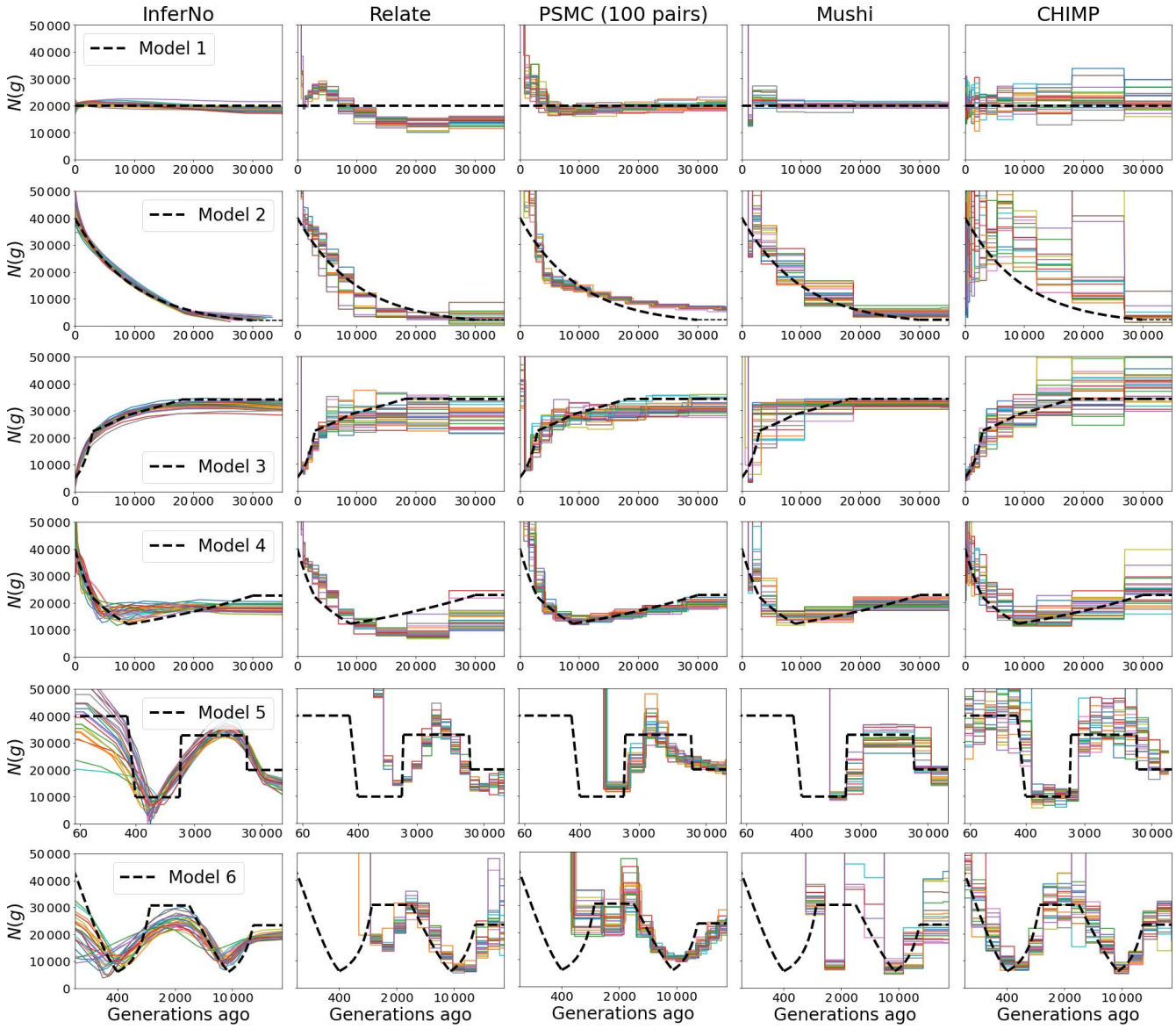

Fig. 2. Estimates of population size $N(g)$ in Models 1 to 6 using five inference methods. Sample size $n = 200$ and sequence length $\ell = 10^7$ sites. Note logarithmic time scale for Models 5 and 6. See Table 3 for corresponding RMISE values, Fig. S3 for plots when $n = 10$, and Table S2 for computing times.

Table 3. RMISE over $g \in [200, 30\,000]$ ($g \in [1\,000, 30\,000]$) for five $N(g)$ inference methods.

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)

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.

Table 4. RMISE (in units of 10^5) for InferNo and PSMC when $n = 2$.

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**

Values are averages over the 25 curves of Fig. S4.

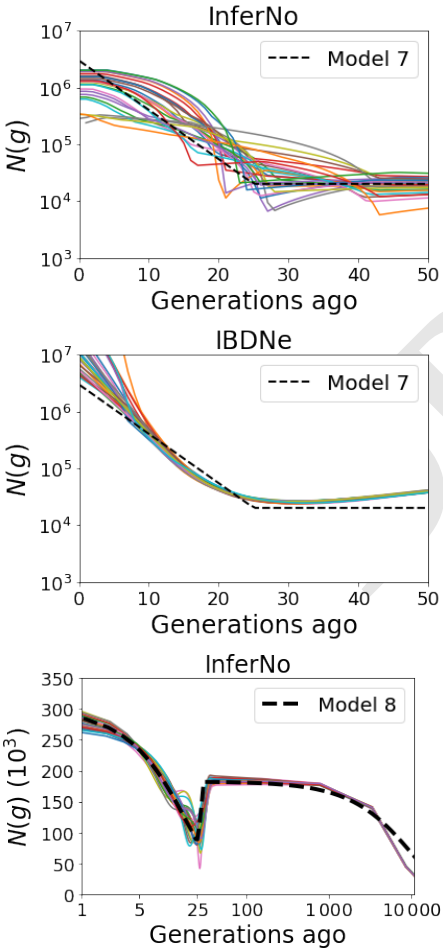

Fig. 3. InferNo and IBDNe estimates of $N(g)$ under Models 7 and 8. The sample size and sequence length for Model 7 (top and middle panels) are $n = 2\,000$ and $\ell = 10^8$ sites. For Model 8 (bottom), they are $n = 2{\times}10^5$ and $\ell = 5{\times}10^7$ sites.

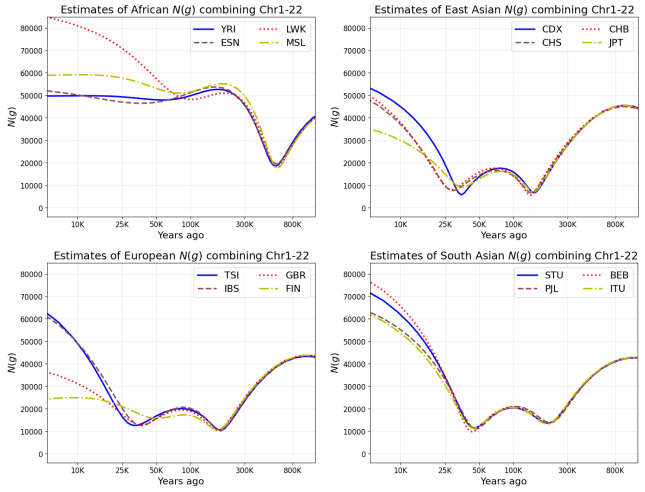

Fig. 4. InferNo estimates of population size N for four populations in each of four regions. See Table 2 for the population codes. The time axes are on a logarithmic scale.

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. S8 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 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. S9). 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.

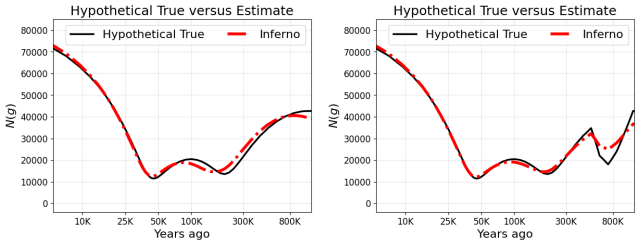

Fig. 5. Testing InferNo inferences from human data. Left: data are simulated with $N(g)$ (black curve) close to those inferred for the non-AFR populations (Fig, 4), and then InferNo used for inference (red curve). See Fig. S9 for the corresponding plot for AFR populations. Right: we add an ancient bottleneck 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 $r = 10^{-8}$.

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 (24, Fig 11), and different populations in similar regions (30, Fig 4) using MSMC2 (35), which is related to PSMC and was found by (24) 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 (30) and $t \approx 100$K (24). 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 (36). 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 (36) have been challenged by (37), who suggest that the severe bottleneck may be an artifact of AFS-based inference. Although we also use the AFS, Fig. S8 shows that InferNo inferences restricted to pairwise TMRCA are similar to AFS-based inferences. Further criticism of (36) has come from (38) 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 (36).

We agree with (36), 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 (34, 39), with modern AFR and non-AFR populations having a different ancestry mix from ancient populations. Alternatively, (36) 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 (38), 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. S1) 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.

ACKNOWLEDGMENTS. ZH was funded in part by Australian Research Council grant DP210102168 awarded to YBC and DB. Thanks to Garrett Hellenthal and Leo Speidel for valuable comments on drafts.

1. H Li, R Durbin, Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
2. S Schiffels, R Durbin, Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919–925 (2014).
3. T Druet, I Macleod, B Hayes, Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity* **112**, 39–47 (2014).
4. J Terhorst, JA Kamm, YS Song, Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017).
5. J Terhorst, Accelerated Bayesian inference of population size history from recombining sequence data. *bioRxiv* **2024.03.25.586640** (2024).
6. RN Gutenkunst, RD Hernandez, SH Williamson, CD Bustamante, Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695 (2009).
7. L Excoffier, I Dupanloup, E Huerta-Sánchez, VC Sousa, M Foll, Robust demographic inference from genomic and SNP data. *PLoS Genet.* **9**, e1003905 (2013).
8. A Bhaskar, YR Wang, YS Song, Efficient inference of population size histories and locus-specific mutation rates from large-sample genomic variation data. *Genome Res.* **25**, 268–279 (2015).
9. J Kamm, J Terhorst, R Durbin, YS Song, Efficiently inferring the demographic history of many populations with allele count data. *J. Am. Stat. Assoc.* **115**, 1472–1487 (2020).
10. WS DeWitt, KD Harris, AP Ragsdale, K Harris, Nonparametric coalescent inference of mutation spectrum history and demography. *Proc. Natl. Acad. Sci.* **118**, e2013798118 (2021).
11. PF Palamara, T Lencz, A Darvasi, I Pe'er, Length distributions of identity by descent reveal fine-scale demographic history. *Am. J. Hum. Genet.* **91**, 809–822 (2012).
12. S Boitard, W Rodríguez, F Jay, S Mona, F Austerlitz, Inferring population size history from large samples of genome-wide molecular data – an approximate Bayesian computation approach. *PLoS Genet.* **12**, e1005877 (2016).
13. L Speidel, M Forest, S Shi, SR Myers, A method for genome-wide genealogy estimation for thousands of samples. *Nat. Genet.* **51**, 1321–1329 (2019).
14. R Fournier, Z Tsangalidou, D Reich, PF Palamara, Haplotype-based inference of recent effective population size in modern and ancient DNA samples. *Nat. Commun.* **14**, 7945 (2023).
15. MD Rasmussen, MJ Hubisz, I Gronau, A Siepel, Genome-wide inference of ancestral recombination graphs. *PLoS Genet.* **10**, e1004342 (2014).
16. J Kelleher, et al., Inferring whole-genome histories in large population datasets. *Nat. Genet.* **51**, 1330–1338 (2019).
17. A Mahmoudi, J Koskela, J Kelleher, Y Chan, D Balding, Bayesian inference of ancestral recombination graphs. *PLoS Comput. Biol.* **18**, e1009960 (2022).
18. Y Deng, R Nielsen, YS Song, Robust and accurate Bayesian inference of genome-wide genealogies for large samples. *bioRxiv* **2024.03.16.585351** (2024).
19. M Mezzavilla, S Ghirotto, Neon: An R package to estimate human effective population size and divergence time from patterns of linkage disequilibrium between SNPs. *J. Comput. Sci. Syst. Biol.* **8** (2015).
20. SR Browning, BL Browning, Accurate non-parametric estimation of recent effective population size from segments of identity by descent. *Am. J. Hum. Genet.* **97**, 404–418 (2015).
21. S Browning, et al., Ancestry-specific recent effective population size in the Americas. *PLoS Genet.* **14**, e1007385 (2018).
22. AP Ragsdale, S Gravel, Models of archaic admixture and recent history from two-locus statistics. *PLoS Genet.* **15**, e1008204 (2019).
23. E Santiago, et al., Recent demographic history inferred by high-resolution analysis of linkage disequilibrium. *Mol. Biol. Evol.* **37**, 3642–3653 (2020).
24. G Upadhya, M Steinrücken, Robust inference of population size histories from genomic sequencing data. *PLoS Comput. Biol.* **18**, e1010419 (2022).
25. R Griffiths, S Tavaré, Sampling theory for neutral alleles in a varying environment. *Philos. Transactions Royal Soc. Lond.* **A 344**, 403–10 (1994).
26. F Baumdicker, et al., Efficient ancestry and mutation simulation with msprime 1.0. *Genetics* **220**, iyab229 (2021).
27. AL Williams, et al., Non-crossover gene conversions show strong GC bias and unexpected clustering in humans. *Elife* **4**, e04637 (2015).
28. Y Zhou, S Browning, B Browning, A fast and simple method for detecting identity-by-descent segments in large-scale data. *The Am. J. Hum. Genet.* **106**, 426–437 (2020).
29. S Fairley, E Lowy-Gallego, E Perry, P Flicek, The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Res.* **48**, D941–D947 (2020).
30. A Bergström, et al., Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, eaay5012 (2020).
31. S Mallick, et al., The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
32. AW Wohns, et al., A unified genealogy of modern and ancient genomes. *Science* **375**, eabi8264 (2022).
33. RJ Wang, SI Al-Saffar, J Rogers, MW Hahn, Human generation times across the past 250,000 years. *Sci. Adv.* **9**, eabm7047 (2023).
34. AP Ragsdale, et al., A weakly structured stem for human origins in Africa. *Nature* **617**, 755–763 (2023).
35. S Schiffels, K Wang, MSMC and MSMC2: The multiple sequentially Markovian coalescent. *Methods Mol. Biol.* **2090**, 147–166 (2020).
36. W Hu, et al., Genomic inference of a severe human bottleneck during the Early to Middle Pleistocene transition. *Science* **381**, 979—-984 (2023).
37. Y Deng, R Nielsen, Y Song, A previously reported bottleneck in human ancestry 900 kya is likely a statistical artifact. *Genetics* **229**, 1–3 (2024).
38. T Cousins, A Durvasula, Insufficient evidence for a severe bottleneck in humans during the Early to Middle Pleistocene transition. *Mol. Biol. Evol.* **42**, msaf041 (2025).
39. T Cousins, A Scally, R Durbin, A structured coalescent model reveals deep ancestral structure shared by all modern humans. *Nat. Genet.* **57**, 856–864 (2025).