

<sup>1</sup> Supporting Information: Accurate and flexible  
<sup>2</sup> estimation of effective population size history

<sup>3</sup> Zhendong Huang, David Balding, and Yao-ban Chan

<sup>4</sup> Corresponding author: Yao-ban Chan, yaoban@unimelb.edu.au

<sup>5</sup> September 29, 2025

<sup>6</sup> This PDF file includes Supplementary Methods SM1 to SM7, Tables S1 to S3 and Figures

<sup>7</sup> S1 to S9.

<sup>8</sup> **Supplementary Methods**

<sup>9</sup> **SM1 Inference of mutation rate**

<sup>10</sup> We estimate the mutation rate along the genome using the number of polymorphic sites  
<sup>11</sup> close to any given site. More specifically, we partition the sites into  $L^*$  segments  $J_1^*, \dots, J_{L^*}^*$ ,  
<sup>12</sup> each of equal length  $\ell/L^*$ , and estimate  $\mu(s)$  by a piece-wise constant map fitted to  $n_l^*$ , the  
<sup>13</sup> number of polymorphic sites in  $J_l^*$ :

$$\hat{\mu}(s) = \frac{\sum_{l=1}^{L^*} n_l^* \mathbf{1}(s \in J_l^*)}{\sum_{l=1}^{L^*} n_l^*} \bar{\mu} L^*, \quad (\text{S1})$$

<sup>14</sup> where  $\mathbf{1}()$  is the indicator function.

<sub>15</sub> **SM2 The null model and time scaling**

<sub>16</sub> Although there is flexibility in setting the null model population size, it should be not too  
<sub>17</sub> far from average values over the historic periods of interest. We use  $\tilde{N} = 20\,000$ . For  
<sub>18</sub> large  $\ell$  given the value of  $r$ , there is enough internal replication along the genome due to  
<sub>19</sub> recombination that a single simulation is sufficient for precise estimation. We used one null-  
<sub>20</sub> model simulation for all analyses reported here, but multiple simulations can be performed  
<sub>21</sub> if desired for greater accuracy.

<sub>22</sub> To relate the null model to the variable-size model, we consider the cumulative per-site  
<sub>23</sub> coalescence rate over the interval  $I_t$  in the variable-size model, which is  $(g_t - g_{t-1})/N(g)$ .  
<sub>24</sub> Eq. [1] is obtained by equating this value to  $(\tilde{g}_t - \tilde{g}_{t-1})/\tilde{N}$ , the corresponding quantity over  
<sub>25</sub>  $\tilde{I}_t$  in the null model.

<sub>26</sub> **SM3 Using the AFS**

<sub>27</sub> The  $\{k, t\}$  entry of  $\tilde{X}_1$  is the number of sites in the null simulation with a frequency- $k$  derived  
<sub>28</sub> allele for which the mutation arose during  $\tilde{I}_t$ . Because of the equivalence of null and size-  
<sub>29</sub> varying models under the time scaling, the product of this entry with the time-scale ratio  
<sub>30</sub>  $\beta_t$  equals in expectation the number of frequency- $k$  sites in the size-varying model for which  
<sub>31</sub> the mutation arose during  $I_t$ . Summing over  $t$  gives the expected value of the  $k$ th entry of  
<sub>32</sub> the AFS. Eq. [2] specifies this relationship jointly for each  $k$ .

<sub>33</sub> **SM4 Using pairwise TMRCA estimates**

<sub>34</sub> From observed pairwise site differences, we estimate quantiles of the distribution of the  
<sub>35</sub> TMRCA of pairs of observed sequences in short genome segments. We match these to the  
<sub>36</sub> corresponding quantiles of the null model, obtained from simulated TMRCA values.

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

in each  $J_l$  by minimizing

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

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

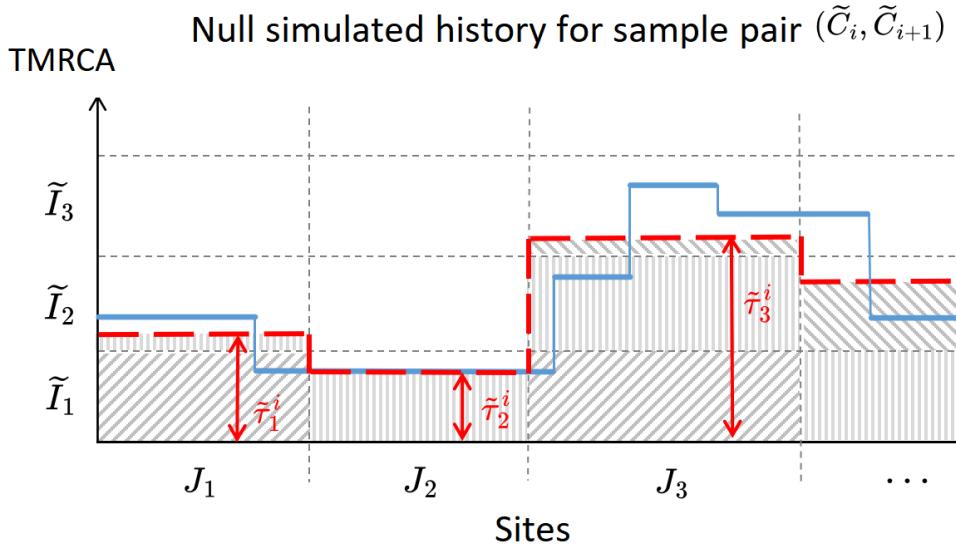


Figure S1: The TMRCA in a null simulation (solid blue) and its approximation  $\tilde{\tau}$  (dashed red) which equals the average blue value in each interval  $J_l$ . The heights of the shaded rectangles correspond to the elements of  $\tilde{W}^i$  defined in the text.

41 We write  $\tau$  for the TMRCA of observed sequences  $C_i$  and  $C_{i+1}$  averaged over  $s \in J_l$  (see  
 42 Fig. S1 for the corresponding averaging in the null model). To reduce computational effort,  
 43 we only use neighbouring pairs of sequences (plus  $(C_1, C_n)$  when  $n > 2$ ), so the resulting  
 44 inferences can depend on sequence order; however, this effect is weak because every mutation  
 45 is represented by a site difference in at least two neighbour pairs. By symmetry,  $\tau$  has the  
 46 same distribution for each  $i$ . We also treat the distribution as invariant over  $l$ , which holds  
 47 approximately, and we make the corresponding assumption for the null model.

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

$$p_z(m) = \int_0^\infty \mathbb{P}_{2ct}(m) f(t) dt , \quad m = 0, 1, 2, \dots,$$

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

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

56 Restricting  $p_z(m)$  to  $m \leq Z_{\max}$ , defining  $A$  as a matrix with  $A_{i+1,j+1} = \mathbb{P}_j(i)$ , for  $i, j =$   
 57  $0, \dots, Z_{\max}$ , and defining a vector version of  $f$  evaluated at  $t/2c$ , for  $t = 0, \dots, Z_{\max}$ , gives  
 58 the matrix equation

$$p_z = \frac{1}{2c} A f . \tag{S2}$$

59 We estimate  $p_z$  by empirical values

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

60 and then invert (S2) to estimate  $f$  by

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

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

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

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

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

63 We relate  $Y_2$  to the corresponding estimated quantiles for the null model, based on  
64 TMRCA values extracted from the null simulation. For a null-simulation sequence pair, let  
65  $\tilde{\tau}$  denote the average TMRCA over  $J_l$  (Figure S1). Let  $\tilde{W}^i$  be an  $L \times T$  matrix with  $\{l, t\}$   
66 entry the contribution to  $\tilde{\tau}$  from the time interval  $\tilde{I}_t = [\tilde{g}_{t-1}, \tilde{g}_t]$ ,

$$\tilde{W}_{l,t}^i = \begin{cases} \tilde{g}_t - \tilde{g}_{t-1} & \text{if } \tilde{\tau} > \tilde{g}_t , \\ \tilde{\tau} - \tilde{g}_{t-1} & \text{if } \tilde{g}_{t-1} < \tilde{\tau} < \tilde{g}_t , \\ 0 & \text{if } \tilde{\tau} < \tilde{g}_{t-1} . \end{cases} \quad (\text{S3})$$

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

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

## 73 SM5 Combining the estimating equations

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

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

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

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \|S(\tilde{X}\beta - Y)\|_2^2 + \lambda \sum_{t=2}^{T-1} (\beta_{t+1} - 2\beta_t + \beta_{t-1})^2 \right\}, \quad (\text{S4})$$

91        subject to  $\beta_t > 0$ ,  $t = 1, \dots, T$ . The second term in (S4) controls the smoothness of the  $\beta_t$   
 92        by penalising their second differences.

93        To select  $\lambda$ , we minimise the adjusted Bayesian information criterion (BIC) [Schwarz, 1978,  
 94        Ye, 1998]:

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

95        where  $d$  is the estimated degrees of freedom,

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

96        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  $i = 1, \dots, T-2$ ,

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

98 For simulation studies using Models 1 through 6, and for the real data analyses, we select  
99  $\lambda$  to minimise BIC over  $\log_{10}(\lambda) \in \{0,1,2,3,4,5,6,7\}$ , although the extreme values 0 and 7  
100 were never chosen in practice. We illustrate this approach for Models 4 and 5 with  $n = 100$   
101 (Figure S2). For Models 7 and 8, which focus on recent population size, BIC is computed  
102 using only the first ten rows of  $\tilde{X}$  and  $Y$ , and the first ten rows and columns of  $S$ , which are  
103 informative about low-frequency alleles that predominantly arise from recent mutations. We  
104 select  $\lambda$  to minimise BIC over  $\log_{10}(\lambda) \in \{-4, -3, -2, -1, 0\}$  and  $\log_{10}(\lambda) \in \{2.5, 3, 3.5, 4, 4.5\}$   
105 for Models 7 and 8, respectively.

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

## 109 SM6 Implementation details for other software

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

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

114 **Mushi:** we set the maximum time investigated to 60 000 generations and the maximum  
115 number of iterations to 300. We obtain piece-wise constant estimates by fixing  
116 the trend order parameter to 0 and selecting the trend penalty parameter over  
117  $\{1, 3, 5, 10, 20, 40, 60, 100\}$  in each setting. In practice, the extreme values 1 and 100  
118 were never chosen.

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

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

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

## 128 **SM7 Implementation details for human whole-genome data**

129 Tree sequences data were downloaded from <https://zenodo.org/records/5512994> and  
130 sequences extracted for both genomes of the 1625 individuals whose population affiliation  
131 matched one of the 16 populations (Table 2).

132 We have previously shown [Huang et al., 2025] that mutation rates can vary over chro-  
133 mosomes, and the pattern of variation may differ across populations. To allow for this  
134 variation, for each population we estimate the average mutation rate for chromosome  $i$  by  
135  $1.3 \times 10^{-8} \times (22M_i / \sum_{j=1}^{22} M_j)$  per site per generation, where  $M_i$  is the fraction of sites  
136 on Chromosome  $i$  that are polymorphic. Then we compute the matrices  $\tilde{X}^{(i)}$  and  $Y^{(i)}$  for  
137 chromosome  $i$ , and stacking them vertically to form  $\tilde{X}$  and  $Y$  for the InferNo inference.

## 138 **References**

- 139 [Huang et al., 2025] Huang, Z., Kelleher, J., Chan, Y., and Balding, D. (2025). Estimat-  
140 ing evolutionary and demographic parameters via ARG-derived IBD. *PLoS Genet*, page  
141 e1011537.
- 142 [Schwarz, 1978] Schwarz, G. (1978). The Bayesian information criterion. *Annals of Statistics*,  
143 6:461–464.

- <sup>144</sup> [Ye, 1998] Ye, J. (1998). On measuring and correcting the effects of data mining and model  
<sup>145</sup> selection. *Journal of the American Statistical Association*, 93(441):120–131.

Model	Time before present in units of $10^3$ generations												
	0	0.025	0.03	0.3	0.4	1	1.8	1.9	3	9	18	18.1	30
1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	10	10	10	10	10	10	10	10	10	10	10	10	0
3	-50	-50	-50	-50	-50	-50	-50	-50	-4	-2	0	0	0
4	20	20	20	20	20	20	20	20	10	-3	-3	-3	0
5	0	0	0	1400	0	0	-1200	0	0	0	500	0	0
6	700	700	700	700	-270	0	0	0	27	-15	0	0	0
7	20 000	0	0	0	0	0	0	0	0	0	0	0	0
8	5 000	-15 000	10	10	10	10	10	10	10	10	10	10	10

Table S1: **Growth rates (per  $10^5$  generations) for eight simulation models.** Although increasing  $g$  corresponds to backwards in time, growth is measured forward in time so a positive growth rate means that  $N(g)$  decreases as  $g$  increases. The rates in each column apply from the generation shown at the top of that column until that shown in the next column, or for all  $g > 30K$  in the case of the final column. See Table 1 for population sizes at the same time points.

Model	$n = 10$					$n = 200$				
	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 S2: **CPU time (seconds).** Average computing time (over the 25 replicates in each setting) for the results in Figure S3 ( $n = 10$ ) and Figure 3 ( $n = 200$ ).

Model	Sample size $n$			Recombination rate			Error rate $\epsilon$			
	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 S3: **RMISE of InferNo (in units of  $10^5$ ) under perturbations of the data generating model.** Entries in the bottom row are expressed relative to a baseline indicated by “ref”.  $r$  is stated in units of  $10^{-8}$  and  $\epsilon$  is per  $10^5$  sites. Entries under “Sample size” correspond to the curves in Figure S5, with integration over  $g \in [200, 1 000]$  ( $g \in [1 000, 30 000]$ ). Entries under “Recombination rate” correspond to the curves in Figure S6 and those under “Error rate” correspond to the curves in Figure S7; in both cases the integration is over  $g \in [200, 30 000]$ .

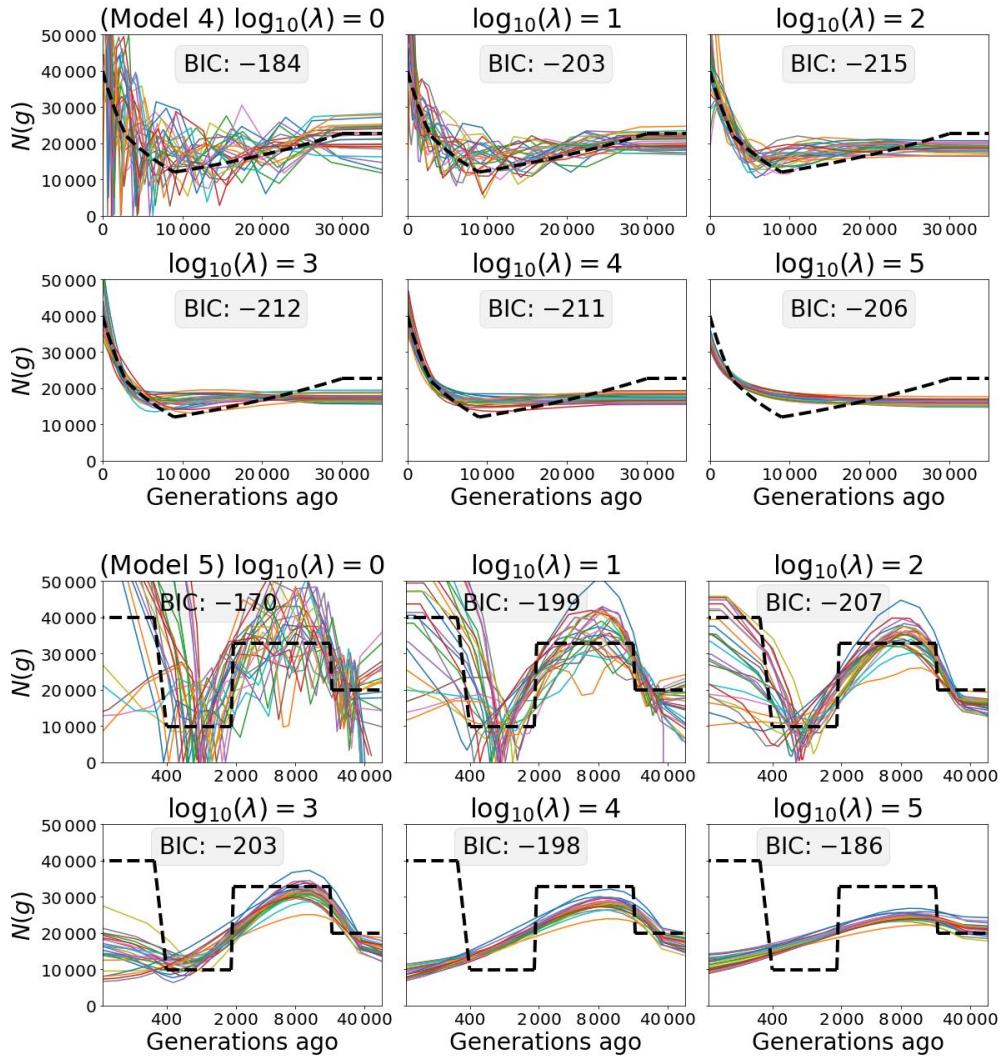


Figure S2: **Selecting smoothing parameter  $\lambda$  using BIC (in units of  $10^3$ ).** Estimates of population size  $N(g)$ , with  $n = 100$ , for Model 4 (top two rows) and Model 5 (bottom two rows, logarithmic time scale). The  $\lambda$  value is indicated above each panel. The minimum BIC is obtained for  $\lambda = 100$  in both cases.

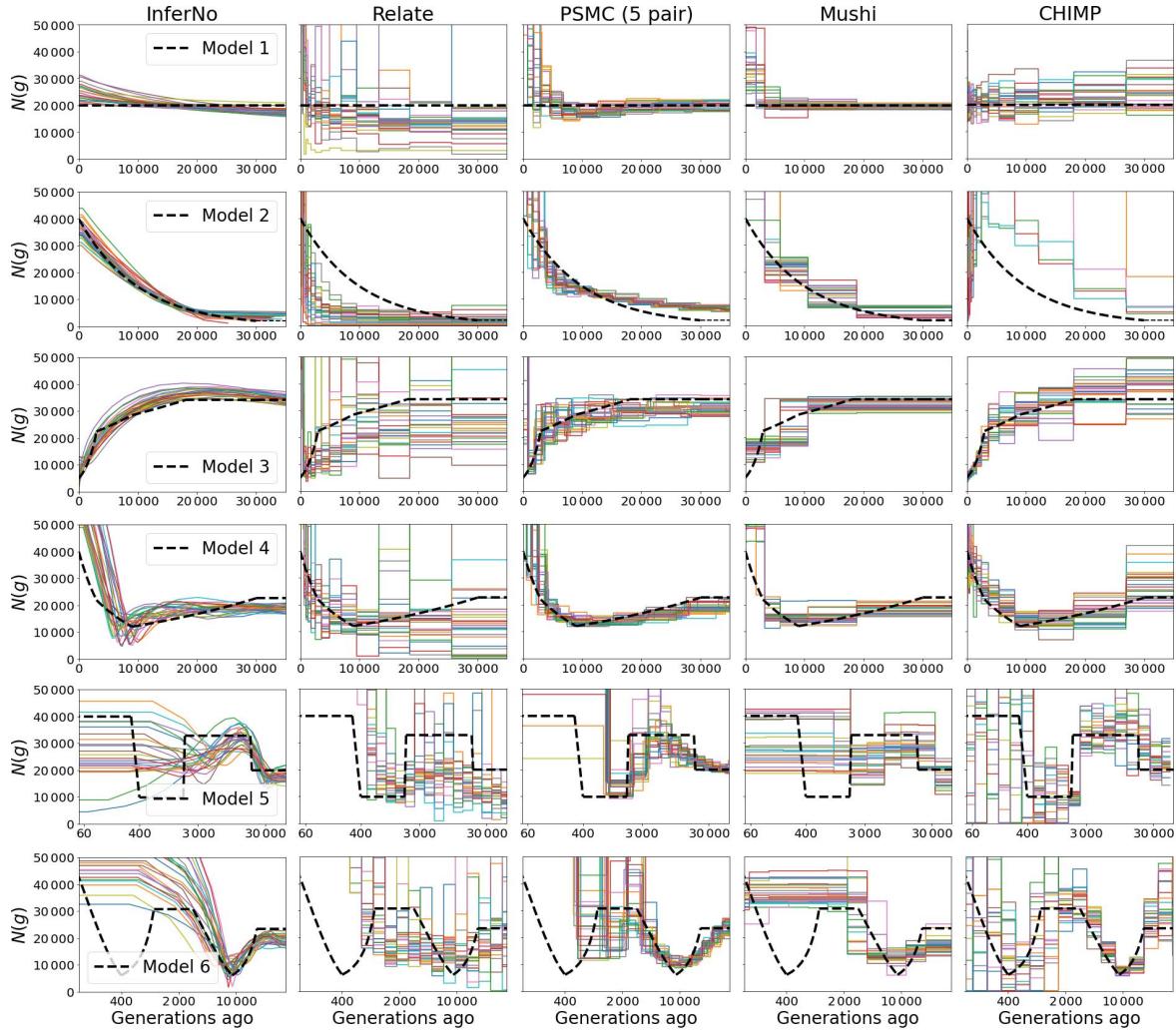


Figure S3: Comparison with other methods ( $n = 10$ ). Other details are the same as for Figure 2.

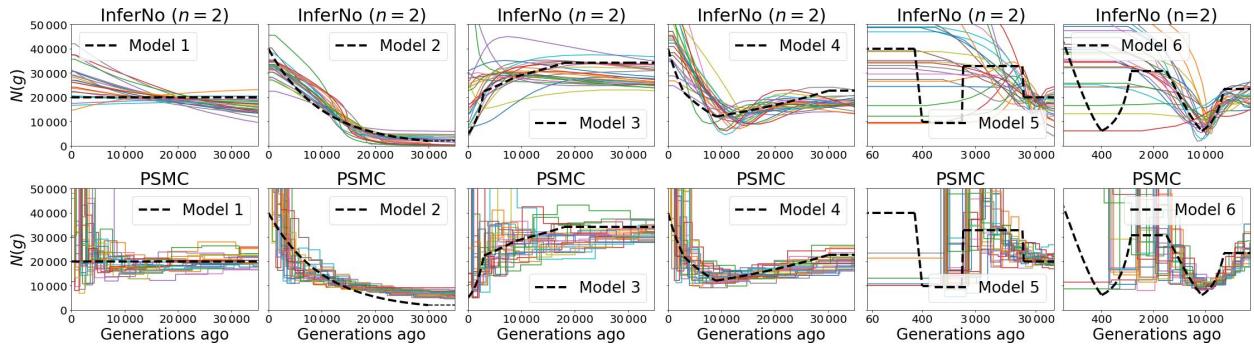
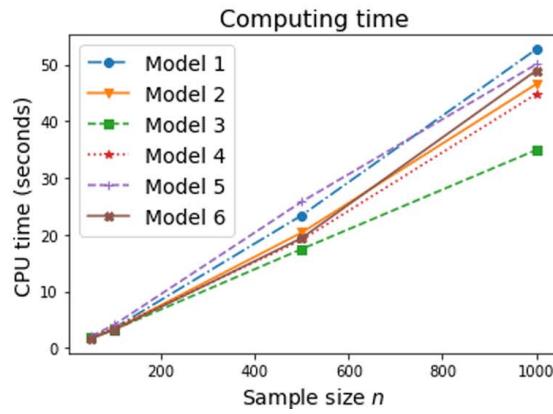
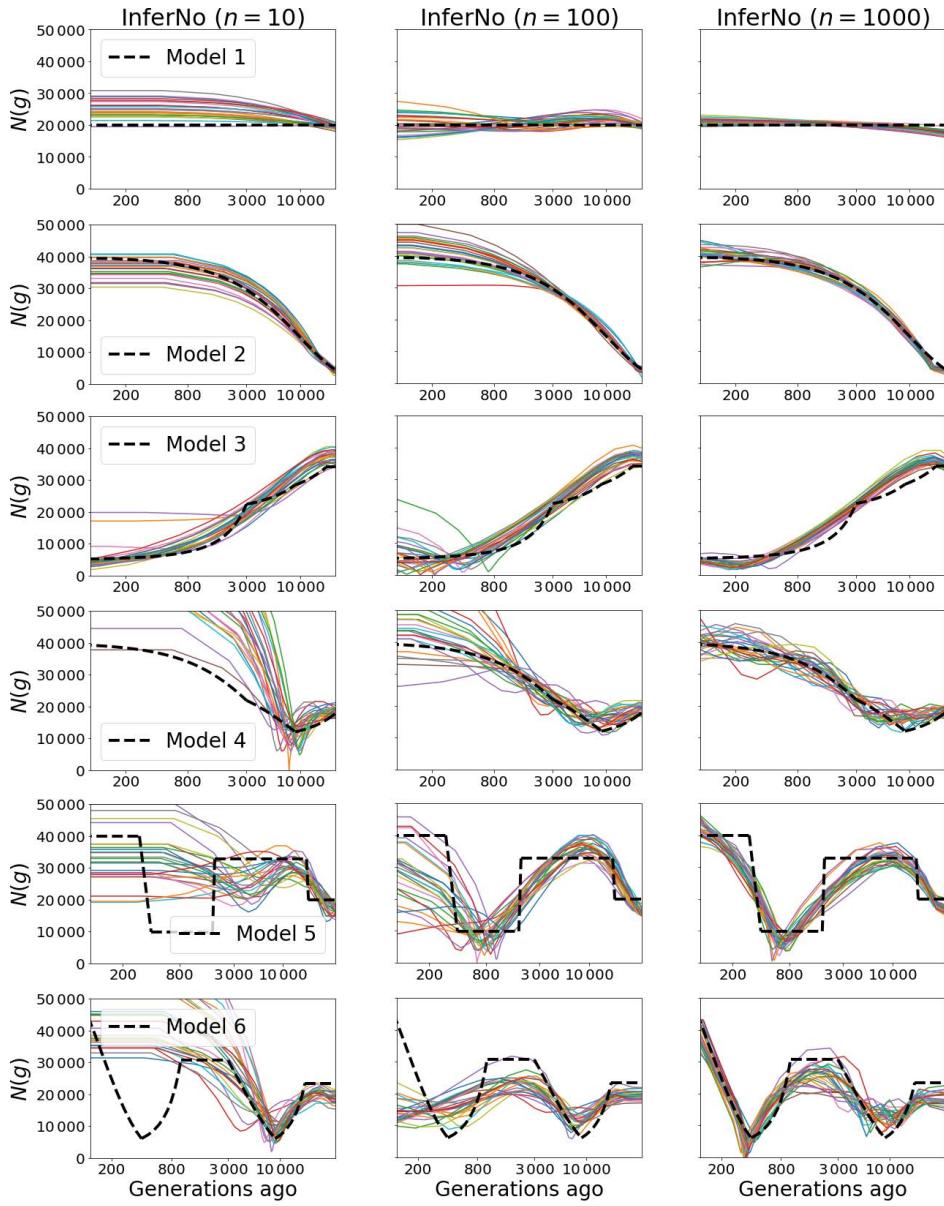
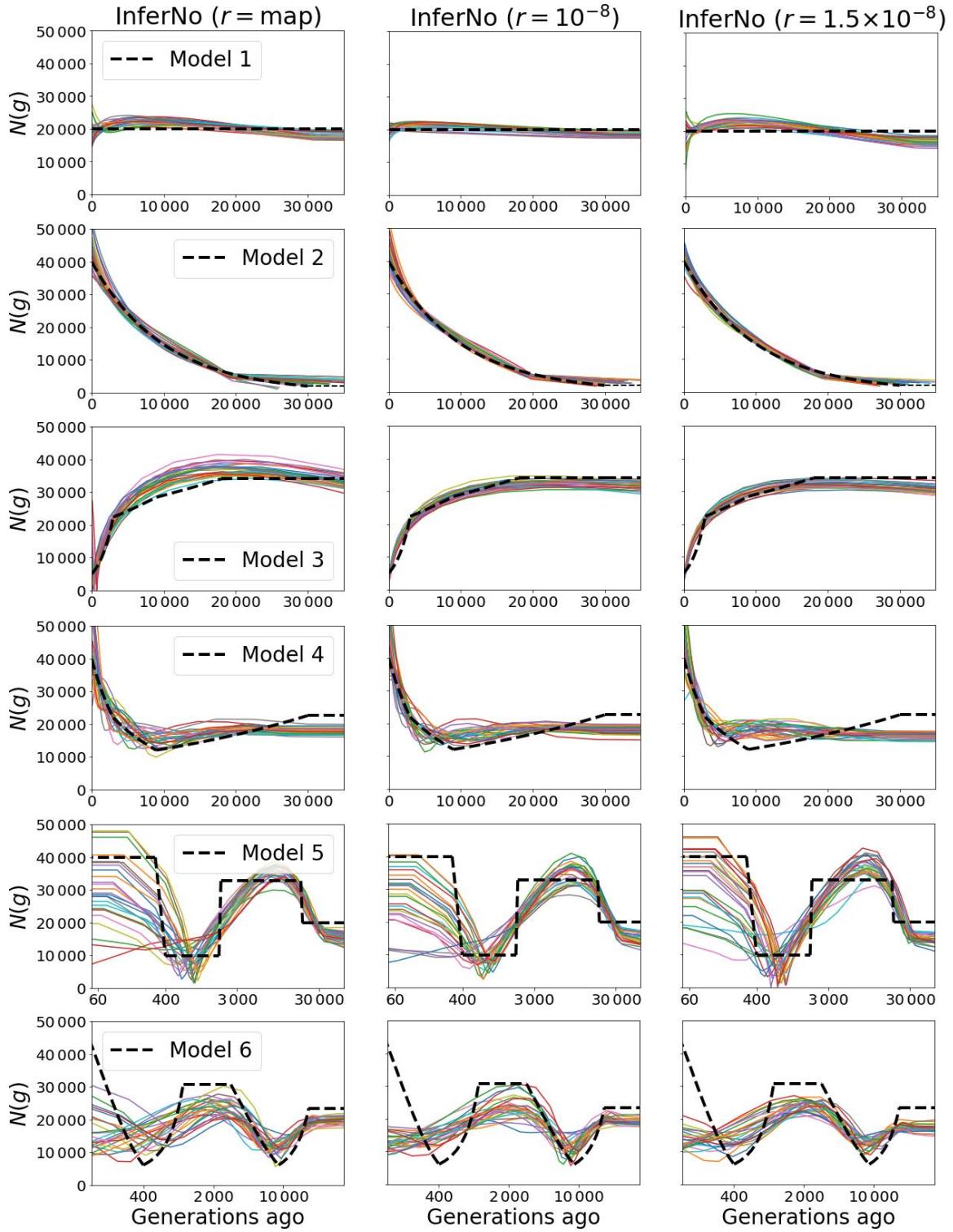


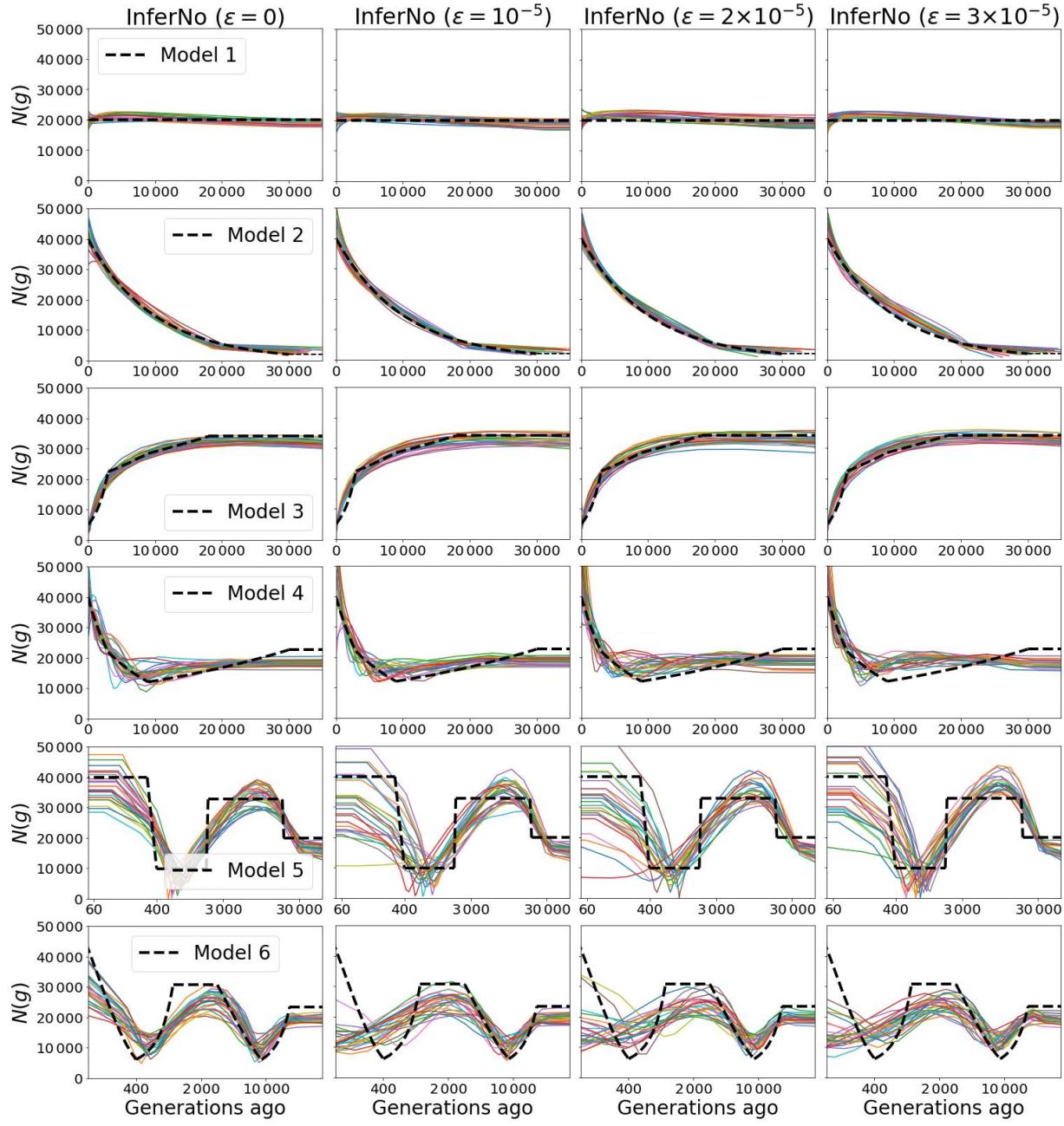
Figure S4: Comparison of InferNo with PSMC when  $n = 2$ . Sequence length  $\ell = 10^7$  sites. Time scale is logarithmic for Models 5 and 6. See Table 4 for corresponding RMISE values.



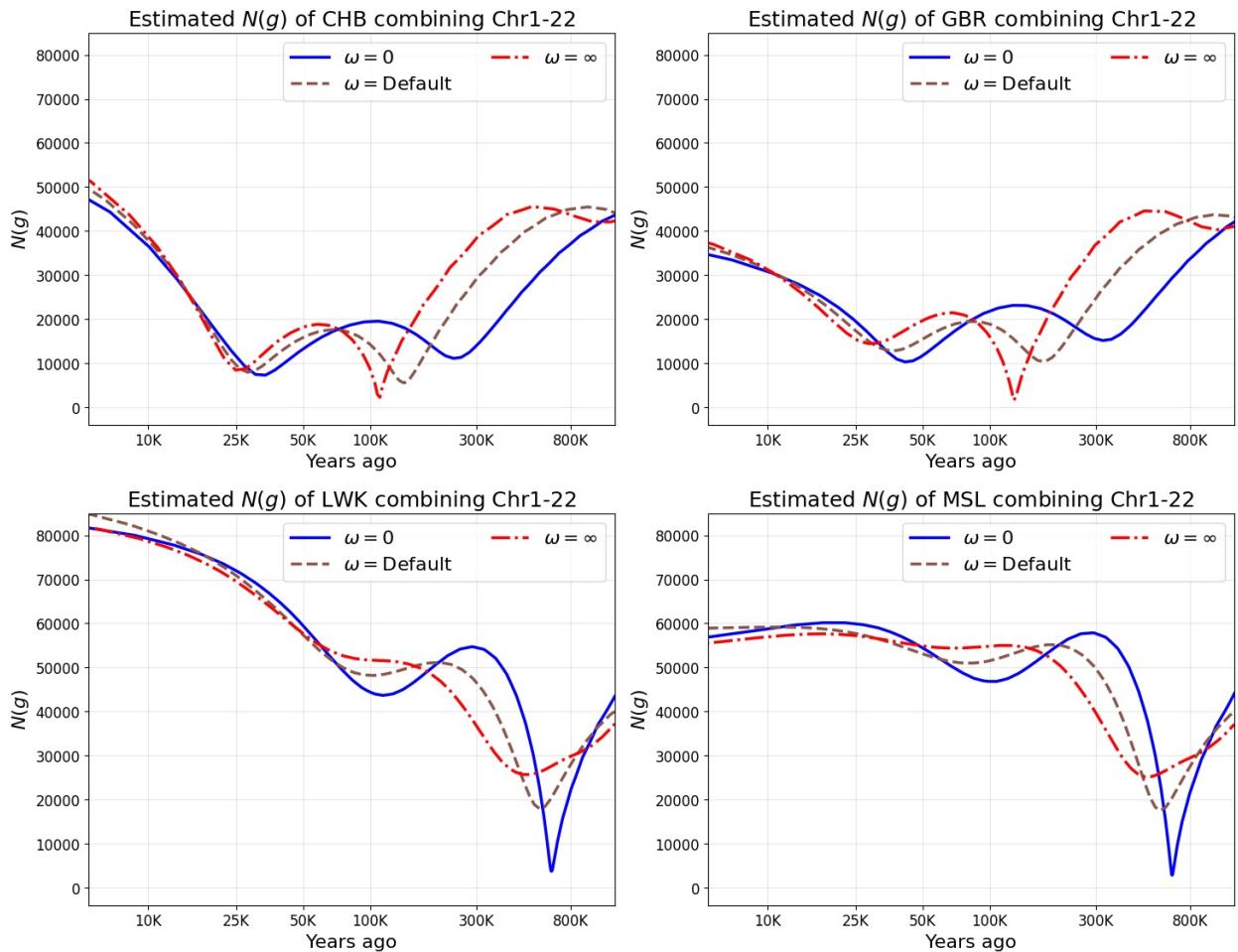
**Figure S5: Estimation and computational performance with varying sample sizes.**  
 Rows 1-6: InferNo estimates of population size  $N(g)$  for three values of  $n$  in Models 1-6, with  $g$  on a logarithmic scale. Bottom: run time with  $n = 10, 50, 100, 500$  and  $1\,000$ .



**Figure S6: 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  $r = 1 \times 10^{-8}$  (column 2) and  $r = 1.5 \times 10^{-8}$  (column 3). The middle column corresponds to the recombination model assumed by InferNo. Rows correspond to simulation Models 1 through 6.



**Figure S7: 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 S8: Components of InferNo inference for four human populations.** Estimates of population size  $N(g)$  using only AFS ( $\omega = 0$ ), only TMRCA estimates ( $\omega = \infty$ ) and the default  $\omega$  setting (used for all other plots).

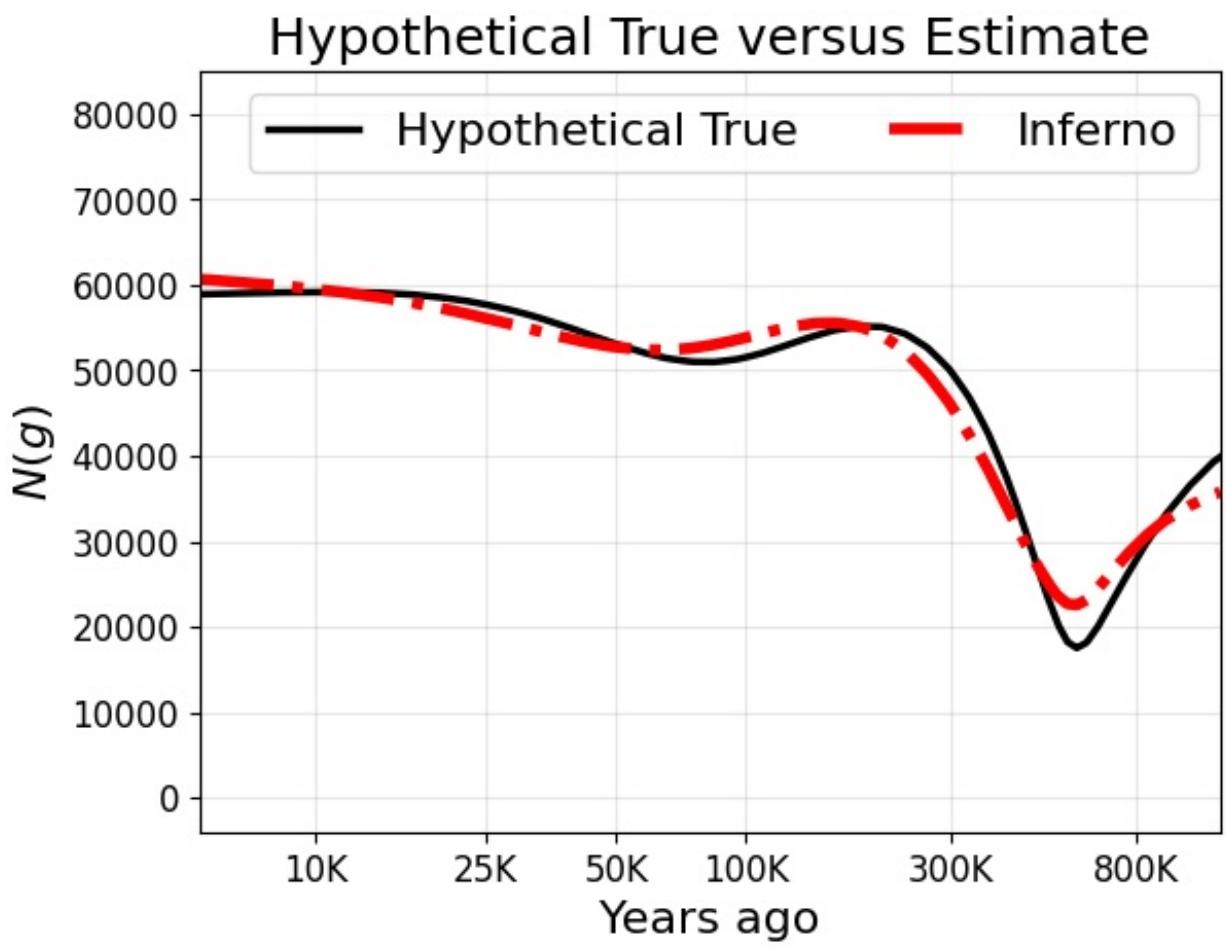


Figure S9: **Testing InferNo inference for an AFR population.** Similar to Fig. 5 (left) which shows inference for a non-AFR population. The red curve shows InferNo inference from  $n = 200$  human sequences simulated with  $N(g)$  as shown in the black curve, similar to the inferred  $N(g)$  curve for AFR populations (Fig. 4, top left).