



## Article

<https://doi.org/10.1038/s41588-025-02317-9>

# Robust and accurate Bayesian inference of genome-wide genealogies for hundreds of genomes

Received: 2 April 2024

Yun Deng <sup>1</sup>, Rasmus Nielsen <sup>2,3,4</sup> & Yun S. Song <sup>2,5</sup>

Accepted: 25 July 2025

Published online: 8 September 2025

Check for updates

The Ancestral Recombination Graph (ARG), which describes the genealogical history of a sample of genomes, is a vital tool in population genomics and biomedical research. Recent advancements have substantially increased ARG reconstruction scalability, but they rely on approximations that can reduce accuracy, especially under model misspecification. Moreover, they reconstruct only a single ARG topology and cannot quantify the considerable uncertainty associated with ARG inferences. Here, to address these challenges, we introduce SINGER (sampling and inferring of genealogies with recombination), a method that accelerates ARG sampling from the posterior distribution by two orders of magnitude, enabling accurate inference and uncertainty quantification for hundreds of whole-genome sequences. Through extensive simulations, we demonstrate SINGER's enhanced accuracy and robustness to model misspecification compared to existing methods. We demonstrate the utility of SINGER by applying it to individuals of British and African descent within the 1000 Genomes Project, identifying signals of population differentiation, archaic introgression and strong support for ancient polymorphism in the human leukocyte antigen region shared across primates.

Many problems in genomics rely on computationally inferring genealogical relationships from large collections of DNA sequences and interpreting the reconstructed trees. In particular, genealogical approaches have been instrumental in understanding human genetic variation<sup>1–3</sup> and underpin numerous computational methods used in biomedical research. In species with recombination, such as humans, the genealogical history cannot be represented by a single tree. Instead, each genomic position has its own tree that minimally differs from those at neighboring positions, resulting in millions of trees across the genome. The collection of these trees, along with recombination points, is represented by the ARG, and the associated generative model is known as ‘the coalescent with recombination’<sup>4,5</sup>.

Although simulating under the coalescent with recombination is straightforward<sup>6–8</sup>, inferring ARGs from genetic variation data remains

a major challenge owing to the enormous space of possible ARGs. ARGs can be built iteratively by determining where the  $n^{\text{th}}$  lineage joins the partial ARG for the first  $n - 1$  genomes, a process referred to as ‘threading’. By using an approximation known as the sequentially Markov coalescent (SMC)<sup>9–11</sup> and formulating the threading problem as a hidden Markov model (HMM)<sup>12–14</sup>, in combination with a clever Markov chain Monte Carlo (MCMC) method, ARGweaver<sup>15</sup> can sample genome-wide ARGs from the approximate posterior distribution for tens of whole-genome sequences. However, it is computationally intensive, rendering it impractical for larger sample sizes.

Recently, significant advances have been made in scaling up ARG reconstruction to tens or hundreds of thousands of genomes. Relate<sup>16</sup> and tsinfer + tsdate<sup>13</sup> use an efficient HMM (namely the Li–Stephens model)<sup>17</sup> to infer local tree topologies along the genome, followed by

<sup>1</sup>Center for Computational Biology, University of California, Berkeley, CA, USA. <sup>2</sup>Department of Statistics, University of California, Berkeley, CA, USA.

<sup>3</sup>Department of Integrative Biology, University of California, Berkeley, CA, USA. <sup>4</sup>Center for GeoGenetics, University of Copenhagen, Copenhagen, Denmark. <sup>5</sup>Computer Science Division, University of California, Berkeley, CA, USA. e-mail: [rasmus\\_nielsen@berkeley.edu](mailto:rasmus_nielsen@berkeley.edu); [yss@berkeley.edu](mailto:yss@berkeley.edu)

branch length estimation. ARG-Needle<sup>18</sup> uses a threading approach similar to ARGweaver but incorporates several heuristics to enhance scalability. These improvements have enabled new research directions<sup>19–23</sup> and have facilitated diverse ARG-based applications in population and statistical genetics<sup>24–35</sup>.

Despite the progress, there are substantial limitations in current ARG inference methods that impede ARG-based analyses of whole-genome sequencing (WGS) data. First, improved scalability usually comes at the cost of accuracy in key ARG features, such as coalescence times<sup>36</sup> and recombination events<sup>37</sup>. For instance, Relate and tsinfer + tsdate perform sub-optimally for ancient coalescence times, diminishing their effectiveness in applications involving ancient times, such as detecting balancing selection. Second, most scalable methods reconstruct only a single ARG topology and overlook estimation uncertainty. As we demonstrate in this article, accurate sampling of ARGs improves statistical inference, especially for local gene tree analysis, in which point estimates are often noisy. Applications such as local ancestry inference, introgression detection and selection analysis would be challenging without proper confidence intervals. For instance, CLUES<sup>34</sup>, a method for inferring selection and allele frequency trajectories, performs better with a sample of local trees. Third, current methods typically assume simple priors, such as a constant-size panmictic population and neutrality, and are not robust against violations of these assumptions. This limitation has consequential implications for applications, as many human populations have undergone complex demographic changes, including bottlenecks and recent expansions<sup>12,38</sup>. Similarly, the influence of background selection is pervasive, profoundly shaping the diversity landscape<sup>39,40</sup>.

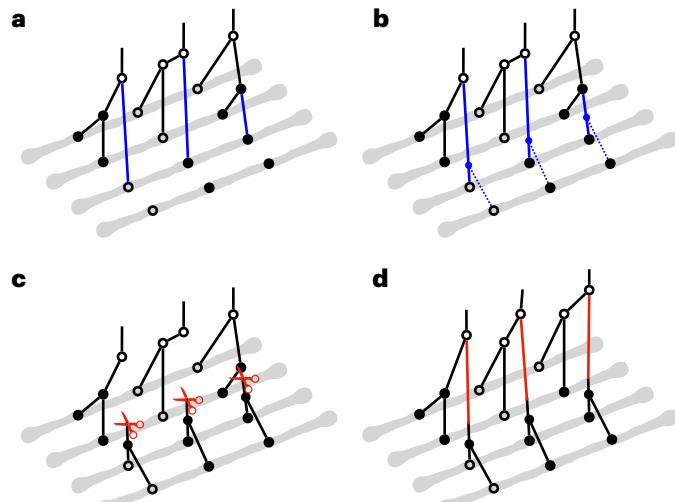
To address these challenges, we introduce SINGER, a Bayesian method for ARG inference. SINGER retains all functionalities of ARGweaver—including MCMC-based posterior sampling, topology exploration, tracking of recombination events and so on—while being at least an order of magnitude faster. Through extensive simulations, we demonstrate that SINGER attains higher accuracy than competing methods in several crucial aspects of ARG inference. It also exhibits greater robustness to various model misspecifications. We highlight the utility of our method by applying it to British and African individuals within the 1000 Genomes Project, revealing signals of population differentiation in coalescence times and archaic introgression and providing strong evidence of trans-species polymorphism and balancing selection in human leukocyte antigen (HLA) regions.

## Results

### An overview of the SINGER algorithm

SINGER takes in phased WGS data and samples ARGs iteratively by adding one haplotype at a time through an operation called threading<sup>15</sup>. Conditioned on a partial ARG for the first  $n - 1$  haplotypes, the threading operation samples the points at which the lineage for the  $n^{\text{th}}$  haplotype joins the partial ARG. SINGER solves this by first building an HMM with branches as hidden states and sampling a sequence of joining branches along the genome from the posterior, using stochastic traceback (Fig. 1a). Then, SINGER builds another HMM with joining times as hidden states, conditioned on these sampled joining branches (Fig. 1b). We refer to these two steps as ‘branch sampling’ and ‘time sampling’, respectively (Methods). Although this two-step threading algorithm is approximative, by substantially reducing the number of hidden states, it is much faster than ARGweaver’s HMM, which treats every joining point in the tree as a hidden state.

To explore the space of ARG topology and branch lengths according to the posterior distribution, SINGER uses an MCMC proposal called ‘sub-graph pruning and re-grafting’ (SGPR). In brief, an SGPR operation first prunes a sub-graph by introducing a cut and then extends it leftwards and rightwards (Fig. 1c); in Supplementary Section B.4, we show that the pruning step is equivalent to the removal



**Fig. 1 | Method overview.** **a–d**, The gray lines represent haplotypes, and the circles indicate the allelic states of nodes in coalescent trees. Hollow circles correspond to ancestral alleles, and solid circles are derived alleles. In **a**, **b** and **c**, a partial ARG for the first three haplotypes has already been constructed, and a fourth haplotype is about to be threaded onto this partial ARG. **a**, The initial step in threading the fourth haplotype involves sampling the joining branch (highlighted in blue) in each marginal coalescent tree of the partial ARG, a process we call ‘branch sampling’. **b**, Following the determination of the joining branches, the next step is to sample the joining time for each of these joining branches. This step is referred to as ‘time sampling’. **c,d**, To propose an update to an ARG in MCMC, we first introduce cuts (illustrated by red scissors) to a sequence of marginal coalescent trees to prune subtrees. Then we re-graft them by solving the threading problem for the sub-ARG above the cuts. The branch length in the first two marginal trees and the topology of the third tree are altered after ‘sub-graph pruning and re-grafting’.

step in the so-called ‘Kuhner move’<sup>41,42</sup>. However, our re-graft step (Fig. 1d) differs substantially from the Kuhner move; the latter samples from the prior by simulation, whereas SGPR uses the threading algorithm to sample from the posterior. Given that the Kuhner move ignores data during re-grafting, it rarely improves likelihood, whereas SGPR favors data-compatible updates. Compared to the Kuhner move and ARGweaver, SGPR introduces large updates to the ARG with higher acceptance rates (Supplementary Section B.4 and Supplementary Fig. 1), yielding a better convergence rate and mixing of the MCMC.

Lastly, to mitigate biases introduced by algorithmic approximations, SINGER performs ‘ARG re-scaling’ through a monotonic transformation of node times that aligns the inferred mutation density with branch lengths (Methods). This is conceptually similar to the ‘ARG normalization’ procedure introduced in ARG-Needle<sup>18</sup>, but ARG-Needle uses a provided demographic prior, whereas SINGER learns the transformation from the inferred ARG without external information. As long as the relative ordering of node ages is accurate, ARG re-scaling can calibrate the overall time distribution. In simulation benchmarks, we show that this greatly improves robustness against model misspecification (for example, population size changes) even though the HMMs assume a constant population size. This approach parallels site frequency spectra-based demography methods<sup>43–45</sup> but incorporates explicit tree topologies inferred from SINGER, offering greater robustness to changing population sizes.

### Performance benchmarks on simulated data

We first benchmarked the performance of several ARG inference methods (SINGER, ARGweaver, Relate, tsinfer + tsdate, ARG-Needle) using data simulated with msprime<sup>46</sup>. The simulation setup and benchmarking procedures are detailed in the Methods.

**Coalescence time accuracy.** To evaluate coalescence time estimation, we compared the ground truth and the inferred pairwise coalescence times for 100 randomly chosen leaf-node pairs, following a previous publication<sup>36</sup>. Pairwise coalescence times are important for applications such as demography inference<sup>16</sup>, genome-wide association studies<sup>18,25,47</sup> and evolutionary studies<sup>35</sup>. For 50 haplotypes, SINGER was the most accurate; ARGweaver and Relate performed similarly, while tsinfer + tsdate was the least accurate (Fig. 2a). For 300 haplotypes, we compared only SINGER, Relate, tsinfer + tsdate and ARG-Needle, as this sample size is too large for ARGweaver. SINGER again performed best; Relate and ARG-Needle performed similarly and tsinfer + tsdate remained the least accurate (Extended Data Fig. 1). SINGER's improved performance over ARGweaver might reflect better MCMC mixing efficiency and more flexible time discretization.

We also compared against pairwise-coalescent methods that analyze each sequence pair independently, specifically considering the recently proposed method Gamma-SMC<sup>48</sup>. SINGER substantially outperforms Gamma-SMC (Fig. 2a and Supplementary Fig. 2), whereas Relate and tsinfer + tsdate show no improvement in either mean squared error or correlation.

We also evaluated the genome-wide average number of lineages as a function of time in marginal trees, a statistic relevant to demography and selection inference. ARGweaver underestimates many recent coalescence times, causing the number of lineages to drop too fast (Fig. 2d and Supplementary Fig. 3); this aligns with Fig. 2a and corroborates a previous finding<sup>16</sup> that ARGweaver tends to underestimate times. On the other hand, tsdate substantially overestimates coalescence times (Fig. 2d). By contrast, Relate and SINGER agree well with the expectation (Fig. 2d).

**Tree topology accuracy.** To assess topology inference accuracy, we used the triplet distance, defined as the fraction of three-leaved subtrees with different topologies in a given pair of trees. This metric is relevant for applications such as imputation and local ancestry, which depend on the accuracy of local topologies. On average, SINGER achieved the lowest triplet distances to the ground truth (Fig. 2e). Again, ARGweaver was less accurate than SINGER, potentially owing to ARGweaver's less efficient MCMC and the presence of polytomies in its inferred trees. We also considered an evaluation metric related to the total variation distance introduced in a previous work<sup>18</sup>, and the results similarly favored SINGER over other methods (Supplementary Section C.2 and Supplementary Figs. 4 and 5).

**Robustness to model misspecification.** One advantage of SINGER is its robustness to model misspecification; specifically, it is less sensitive to incorrect effective population sizes ( $N_e$ ) and unmodeled population size changes. When using an  $N_e$  that is off by a factor of five, the coalescent times inferred by SINGER were less biased than Relate and tsinfer + tsdate, which showed systematic underestimation (Supplementary Fig. 6).

We simulated data under an inferred CEU population size history<sup>49,50</sup>, which contains a bottleneck and recent expansion. On these data, SINGER not only inferred the coalescence times more accurately than ARGweaver, Relate and tsinfer + tsdate (Fig. 2b) but also accurately captured the bi-modality in the pairwise coalescence time distribution caused by the bottleneck (Fig. 2c). Although Relate can incorporate population size changes, it requires running a separate module of estimating branch lengths and coalescent rates, which takes even longer than running Relate itself. ARG-Needle requires a user-specified size history to adjust its coalescence times and is not able to handle an unknown size history. By contrast, SINGER automatically adjusts branch lengths through ARG re-scaling, with little computational overhead.

**Accuracy of mutation and recombination inferences.** We also benchmarked allele age estimation using inferred ARGs, excluding ARG-Needle and ARGweaver because ARG-Needle does not map mutations to

branches, and ARGweaver's output is difficult to parse for this task. On simulated data with 50 sequences, SINGER noticeably outperformed Relate and tsinfer + tsdate (Fig. 3a). For 300 sequences, SINGER remained more accurate than Relate and tsinfer + tsdate (Extended Data Fig. 2).

We also compared the number of recombination breakpoints in 5 kb windows. Only ARGweaver and SINGER produced accurate estimates (Fig. 3b). Both Relate and tsinfer missed many recombination events, consistent with earlier studies<sup>37</sup>.

Finally, we assessed the accuracy of recombination inference by the distribution of pairwise identity-by-descent (IBD) lengths, which are shaped by recombination. ARGweaver and Relate were excluded from this analysis; the former because of difficulties extracting IBD information from its output, and the latter owing to a lack of node persistence across marginal trees. As illustrated in Fig. 3c, SINGER accurately captured the distribution of pairwise IBD lengths, while tsinfer substantially overestimated IBD lengths, consistent with previous findings<sup>37</sup>.

**Comparison of MCMC convergence.** In humans and many other organisms, the genome-wide average rates of recombination and mutation are similar, which leads to substantial uncertainty in ARG inference. Therefore, it is important to obtain samples from the posterior distribution and characterize uncertainties, rather than relying on point estimates. To assess MCMC convergence, we obtained 100 posterior MCMC samples from ARGweaver, Relate and SINGER, using the same burn-in and thinning intervals.

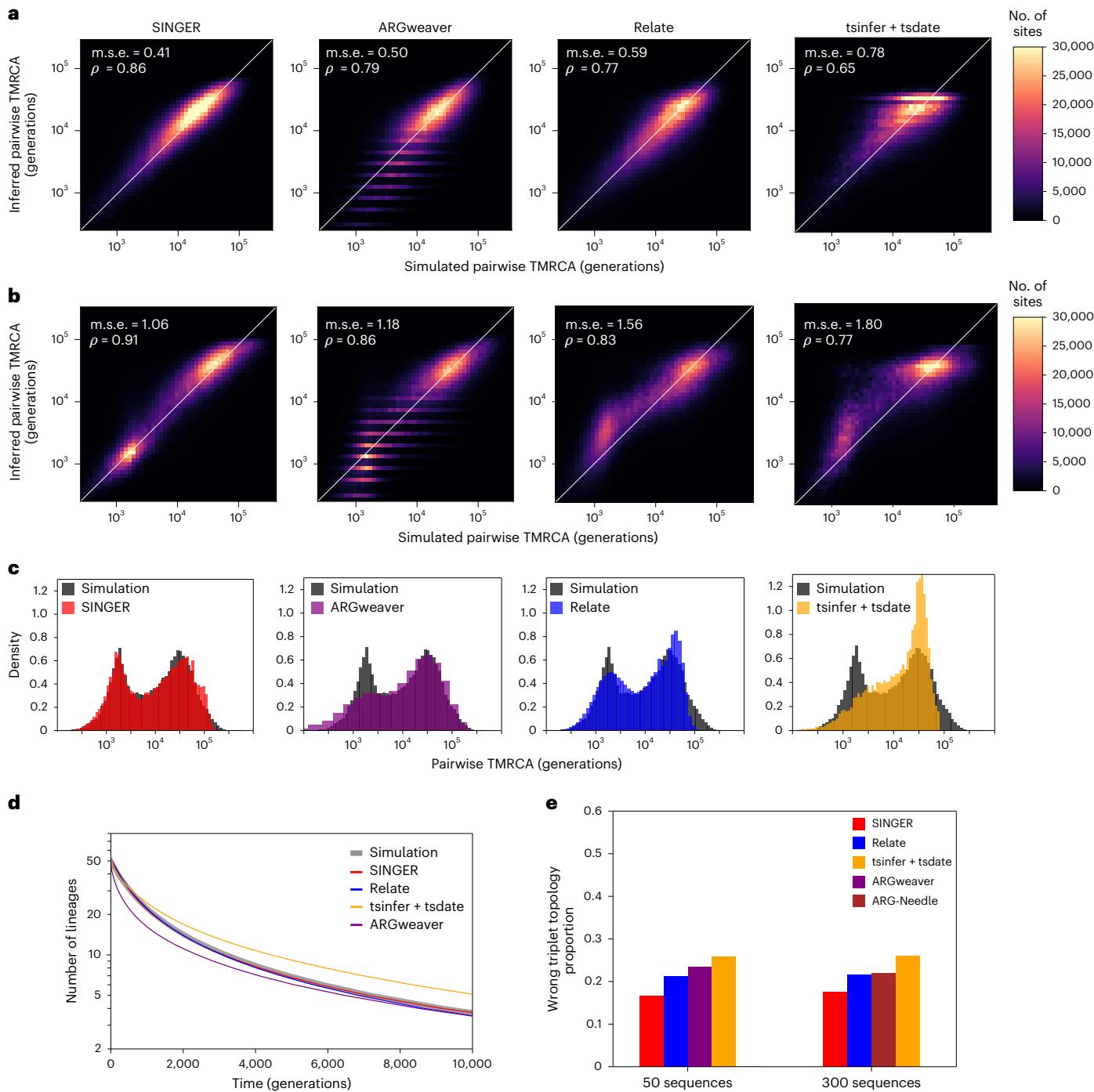
To assess the posterior sampling effectiveness, we used the same benchmark as in previous work<sup>36</sup>. This involved analyzing rank plots of pairwise coalescence times; a uniform distribution would be achieved by a perfect sampler from the posterior distribution<sup>51,52</sup>. A rank plot is a histogram of the rank of a parameter sampled from the prior relative to the posterior sample. Ideally, a converged and well-mixed MCMC should yield uniformly distributed ranks. By contrast, a U-shaped rank plot suggests sampling from an under-dispersed distribution<sup>51,52</sup>. Compared to ARGweaver and Relate, SINGER's rank plots are much closer to the uniform distribution (Fig. 4b).

The rank plot is closely related to the coverage property of empirical credible intervals (CIs). For each genomic position and pair of haplotypes, the empirical 90% CI is defined by the 5<sup>th</sup> to the 95<sup>th</sup> percentile of the sampled coalescence times (Fig. 4a). The same approach was applied to the 70% and 50% CIs. The 90% CI covered the ground truth in only 44% of instances for Relate and 54% for ARGweaver. By contrast, the coverage was substantially better for SINGER, at 85% (Fig. 4c). SINGER also compared favorably at other CI levels (Fig. 4c).

Furthermore, even with thinning intervals 40 times longer than SINGER, ARGweaver still underperforms in pairwise coalescence time inference and CI coverage (Extended Data Fig. 3). Combined with our faster threading algorithm, this suggests that ARGweaver would require hundreds to thousands of times longer to match SINGER's performance. For Relate, even with long thinning intervals, CI coverage remains substantially below nominal levels (Extended Data Fig. 3 and Supplementary Section C.4), probably because Relate samples only coalescence times under a fixed topology, whereas SINGER samples both topologies and coalescence times.

**Runtime comparison.** Given that both SINGER and ARGweaver use threading algorithms, we compared their threading runtimes as a function of the number of leaves in the partial ARG. SINGER's threading is approximately 10× faster than ARGweaver's (Fig. 4d).

**Other benchmarks.** We performed additional benchmarks<sup>53</sup> (Supplementary Sections C.2 and C.5 and Supplementary Fig. 7) and observed that SINGER outperforms other methods. We also note that a recent independent benchmarking study found that SINGER outperforms other ARG inference methods in reconstructing allele frequency trajectories and polygenic score histories<sup>54</sup>.



**Fig. 2 | Performance benchmarks on coalescence time and topology inference.**

**a**, Inferred pairwise coalescence times compared with the ground truth in simulations involving 50 sequences under a constant population size scenario. **b**, Similar to **a**, but for data simulated under an inferred population size history for the CEU population. **c**, Inferred distribution of pairwise coalescence times (colored) compared with the ground truth (dark gray) from simulations under

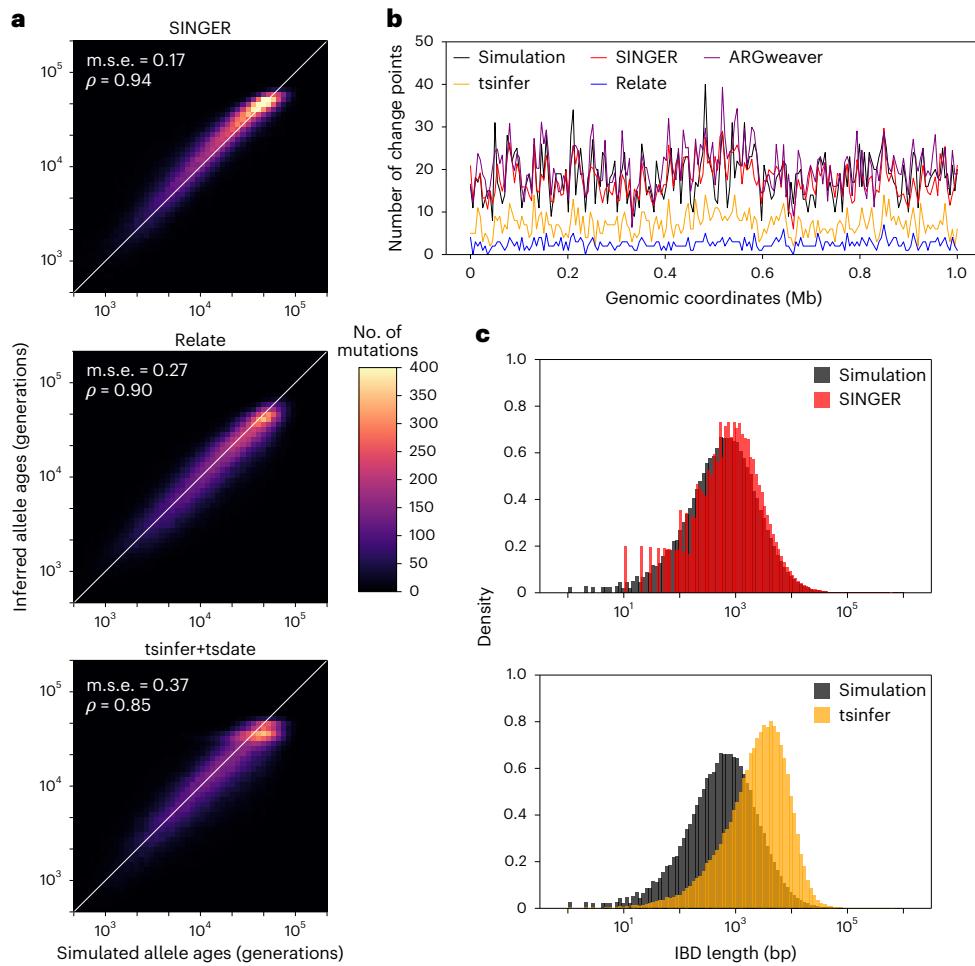
the same CEU demography as in **b**. **d**, Genome-wide average of the number of lineages as a function of time for 50 sequences under a constant population size history, compared with the ground truth in simulations. **e**, The proportion of triplet topologies that are incorrectly inferred for 50 and 300 sequences under a constant population size history. Owing to runtime constraints, ARGweaver is not benchmarked for 300 sequences.

SINGER supports using an input recombination map to account for the recombination rate variation along the genome<sup>55,56</sup>, which improves inference accuracy (Supplementary Section C.6 and Supplementary Figs. 8 and 9).

#### Applications to WGS data from the 1000 Genomes Project

We applied SINGER to 200 whole-genome sequences from five African indigenous populations (GWD, YRI, ESN, LWK and MSL) in the

1000 Genomes Project<sup>57</sup>, with 40 genomes randomly sampled per population (Supplementary Section D.1). To demonstrate the utility of SINGER, we analyzed population differentiation in coalescence times, trans-species polymorphism and archaic introgression. We also ran SINGER and Relate on the British (GBR) population data from the 1000 Genomes Project (Supplementary Section D.1) and used tsinfer + tsdate ARG from a previous publication<sup>3</sup> for comparison.



**Fig. 3 | Benchmarks on mutation and recombination inference for data simulated with 50 sequences and constant population size.** **a**, Inferred allele ages compared with the ground truth. **b**, Inferred number of recombination

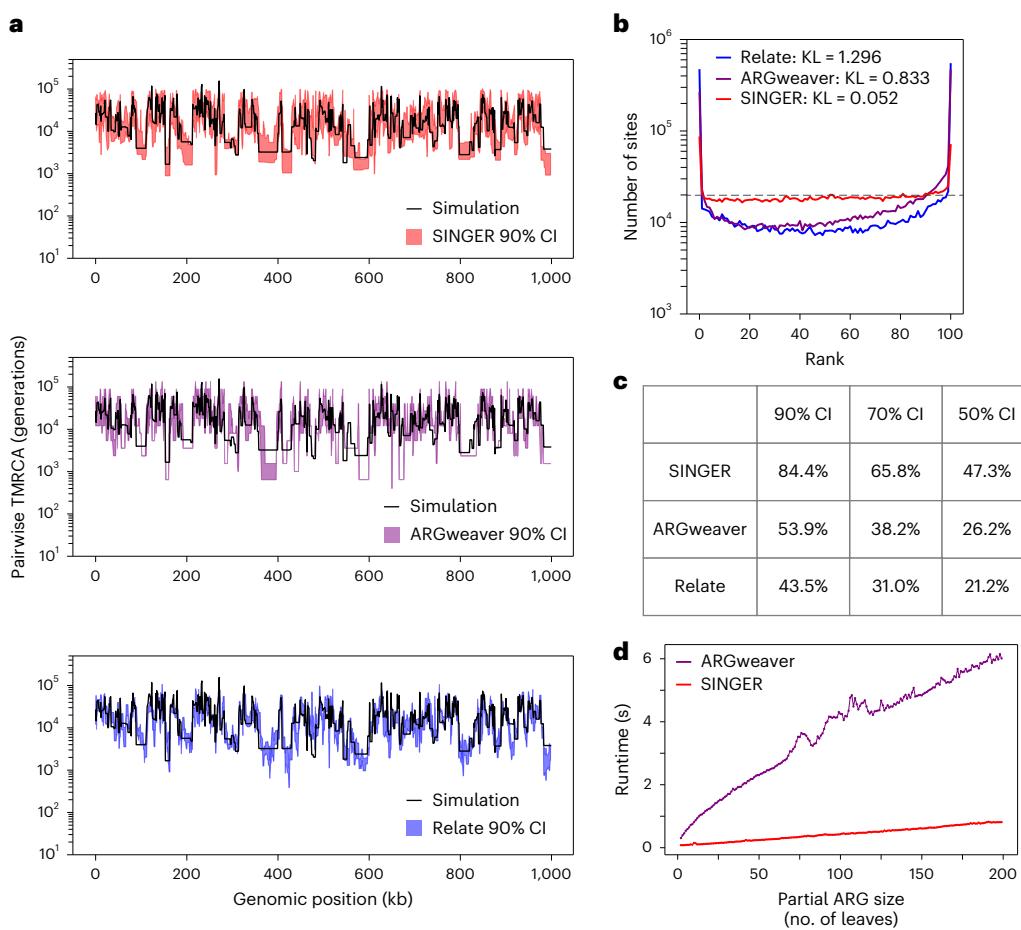
breakpoints in 5 kb genomic windows compared with the ground truth. **c**, The length distribution of pairwise IBD in the inferred ARGs compared with the ground truth.

**Diagnostics of the ARGs sampled by SINGER.** We examined the sampled ARGs to check MCMC convergence and to ensure that sampling for inference occurred past proper burn-in (Supplementary Section D.4). The chains generally converged well (Supplementary Fig. 10). Additionally, we validated the accuracy of the sampled ARGs by comparing the inferred average pairwise coalescence times (scaled by  $4N_e\mu$ ) with empirical single-nucleotide polymorphism (SNP)-based nucleotide diversities in 1 Mb windows, which showed high concordance. By contrast, Relate and tsinfer + tsdate<sup>3</sup> underestimated the genome-wide variation of diversity (Extended Data Fig. 4a). This is possibly because of the  $N_e$  variation as a result of background selection<sup>58,59</sup>. As shown earlier, SINGER is more robust to  $N_e$  misspecification. Additionally, tsinfer + tsdate has a very biased variant density prediction from inferred ARG compared to observed data (Extended Data Fig. 4b). This is probably caused by polytomies, which distort the total branch length. In addition, Relate and tsinfer + tsdate require allele polarization (that is, distinguishing ancestral vs derived alleles), but it is difficult for the HLA locus with high levels of trans-species polymorphisms.

**Population differentiation in coalescence times.** Population-level differentiation in coalescence times at the same genomic locus is often used to identify sites that warrant further evolutionary analysis<sup>60–62</sup>. Such differentiation could be a result of evolutionary forces such as local adaptations, which reduce diversity for the population experiencing selective sweeps. However, SNP-based diversity can be noisy at fine scales (Supplementary Fig. 11). On the other hand, with

accurately inferred ARGs, fine-scale diversity can be estimated more accurately. We observed that SINGER produces more accurate estimates of fine-scale diversity than Relate and tsinfer + tsdate (Supplementary Section D.5 and Supplementary Fig. 11). This improvement facilitates studying population-specific fine-scale differentiation in coalescence times. Many previously reported loci under positive selection in Europeans<sup>63</sup> appear as outliers when comparing the 1 kb-scale average pairwise time to the most recent common ancestor (TMRCA) between GWD and GBR (Supplementary Fig. 12), probably reflecting European-specific selection. They also exhibit long segments of reduced pairwise TMRCA among target allele carriers compared to the overall sample (Supplementary Section D.6 and Extended Data Fig. 5). Here, we focus on population-specific reduction in local diversity in African populations.

To find population-specific reduction in local diversity, we partitioned the genome into 1 kb windows and computed the ratio of the ARG-based diversity estimate for the combined sample to that for each of the five populations; reductions in local diversity would show up as peaks when these ratios are plotted along the genome. The full list of regions with elevated ratios for each population is available in the Data Availability section. We highlight a few interesting findings in Fig. 5 and Supplementary Fig. 13. For example, we found that the gene *MITF* has experienced a reduction in diversity in GWD relative to other populations (see also Supplementary Fig. 14); this gene has been reported to be related to skin<sup>64</sup>. Around *MITF*, we observed substantial differences in local diversity across the five populations, consistent with pigmentation variation within Africa<sup>61</sup>. In YRI, we found that *SPCS3*



**Fig. 4 | Properties of ARG samples and runtimes.** **a**, Empirical 90% CIs for pairwise coalescence times as inferred by SINGER, ARGweaver and Relate. **b**, Rank plots of pairwise coalescence times in MCMC samples. A perfect sampler from the posterior distribution would achieve the flat dashed line, corresponding to the uniform distribution. The Kullback–Leibler (KL) divergence is used to

quantify deviation from a uniform distribution. **c**, The empirical coverage of the

ground truth pairwise coalescence time by the CI for different nominal levels. **d**, The runtime of the threading algorithm as a function of the partial ARG size (measured by the number of leaves), for ARGweaver and SINGER.

has reduced diversity in YRI compared to other populations; this gene encodes an immune-related protein believed to impact virion production of flaviviruses such as yellow fever virus<sup>65</sup>. This is concordant with the report of the spread of these diseases in Nigeria<sup>66</sup>. Lastly, we found that *SCN9A*, which encodes a voltage-gated sodium channel involved in the perception of pain<sup>67</sup>, has substantially reduced diversity. These examples illustrate the utility of SINGER for exploratory evolutionary analysis, but additional studies are needed to investigate if, and how, selection is acting in these loci.

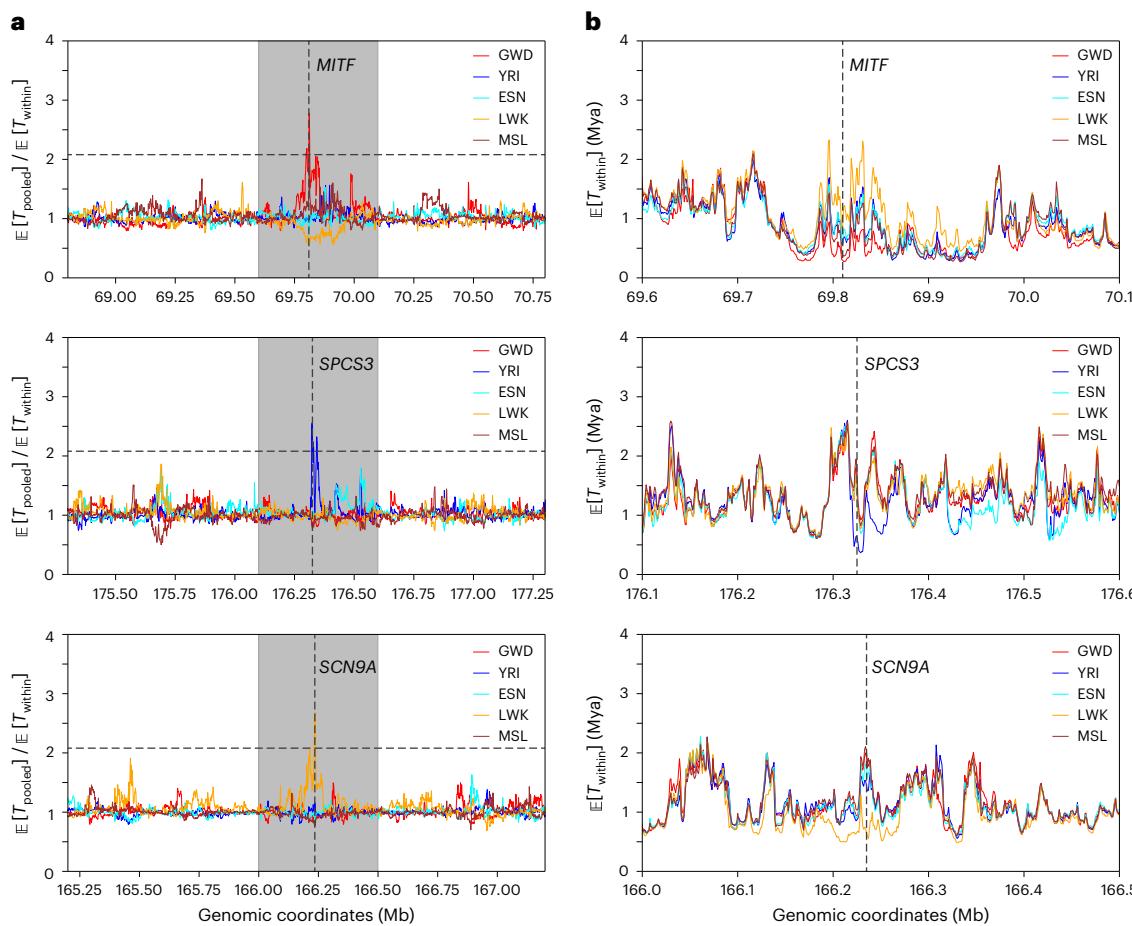
**Archaic introgression.** Evidence shows that modern humans carry DNA segments from Neanderthals and Denisovans<sup>68–70</sup> as well as unidentified hominid groups<sup>71</sup>. Identification of these introgressed genomic tracts is a challenging task, especially when there is little or no known genome of the source hominids. However, ARGs can facilitate this task by the following observation: for an introgressed tract in a given haplotype, its coalescence with other haplotypes will be depleted in the interval between the introgression time and the split time of modern humans from the ‘ghost’ population (Fig. 6a). This is similar to the ‘long branch’ signals described in previous work<sup>16</sup>, but expressed in the pairwise coalescence space.

However, the ‘long branch’ signals can be sensitive to topology inference errors (Supplementary Fig. 15); specifically, the introgressed lineage can group incorrectly with the ancestral lineages of non-introgressed sequences, thereby destroying the long branch

(Supplementary Section D.7). To mitigate this issue, we provide a technique based on the coalescence distribution heatmap. For each sequence, we plot the distribution of its pairwise coalescence time with the remaining sequences in 10 kb windows (Fig. 6c), in which each column corresponds to a 10 kb window. We found that using posterior samples of ARGs is helpful, as the coalescence distribution from a single ARG can be noisy (Supplementary Fig. 16). ARG samples with different topologies help smooth the heatmap (Supplementary Fig. 16). This is related to the visualization shown in previous work<sup>48</sup>.

To detect introgression tracts, we look for a depletion of probability mass in the aforementioned interval and an enrichment of mass above the interval. This is more robust than long branches because slight mis-grouping would still lead to probabilistic depletion in the interval, whereas the long branch would be disrupted completely. We demonstrated the feasibility of this approach by showing that the ARGs inferred by SINGER can recover the Neanderthal introgression tracts inferred by IBDmix<sup>72</sup>, a referenced-based method directly comparing Neanderthal and modern genomes (Supplementary Section D.7). The coalescence ratio slightly outperforms long branch signals for detecting archaic introgression in Relate, and SINGER compares favorably with Relate on this task (Fig. 6b).

Here, we highlight a potential 200 kb Neanderthal introgression tract in GBR (Fig. 6c), supported by IBDmix and the coalescence depletion signal from SINGER-inferred ARGs. We use 60 kya and 500 kya for introgression and split times, respectively, following previous work<sup>73</sup>,



**Fig. 5 | ARG-based detection of differentiated coalescence times in African populations.** **a**, The ratio of the average pairwise coalescence time in the pooled sample,  $T_{\text{pooled}}$  (combining all five populations), to the average population-specific pairwise coalescence time,  $T_{\text{within}}$ , for every 1 kb window. In each plot, the horizontal black dashed line denotes the genome-wide 99.99% quantile,

and the gray shaded area corresponds to a 50 kb window surrounding the peak. The positions of these peaks are marked by vertical dashed lines, and the genes overlapping with these signals are indicated. **b**, The average  $T_{\text{within}}$  for each population, zoomed into the gray regions highlighted in **a**.

and plot the ratio of coalescence probability above the split time to that in the interval between introgression and split times. This tract appears as distinctive peaks in the ARG-based analysis (Fig. 6d).

**Trans-species polymorphism in the HLA locus.** The HLA locus comprises a cluster of genes that encode transmembrane proteins that present antigen peptides to T cells. This region is known to be the most diverse region in the human genome, and it has been hypothesized to be under extreme balancing selection to maintain high diversity to handle various immune challenges<sup>74,75</sup>. There has been evidence of trans-species polymorphism for some alleles across primates, which otherwise is very rare<sup>74</sup>.

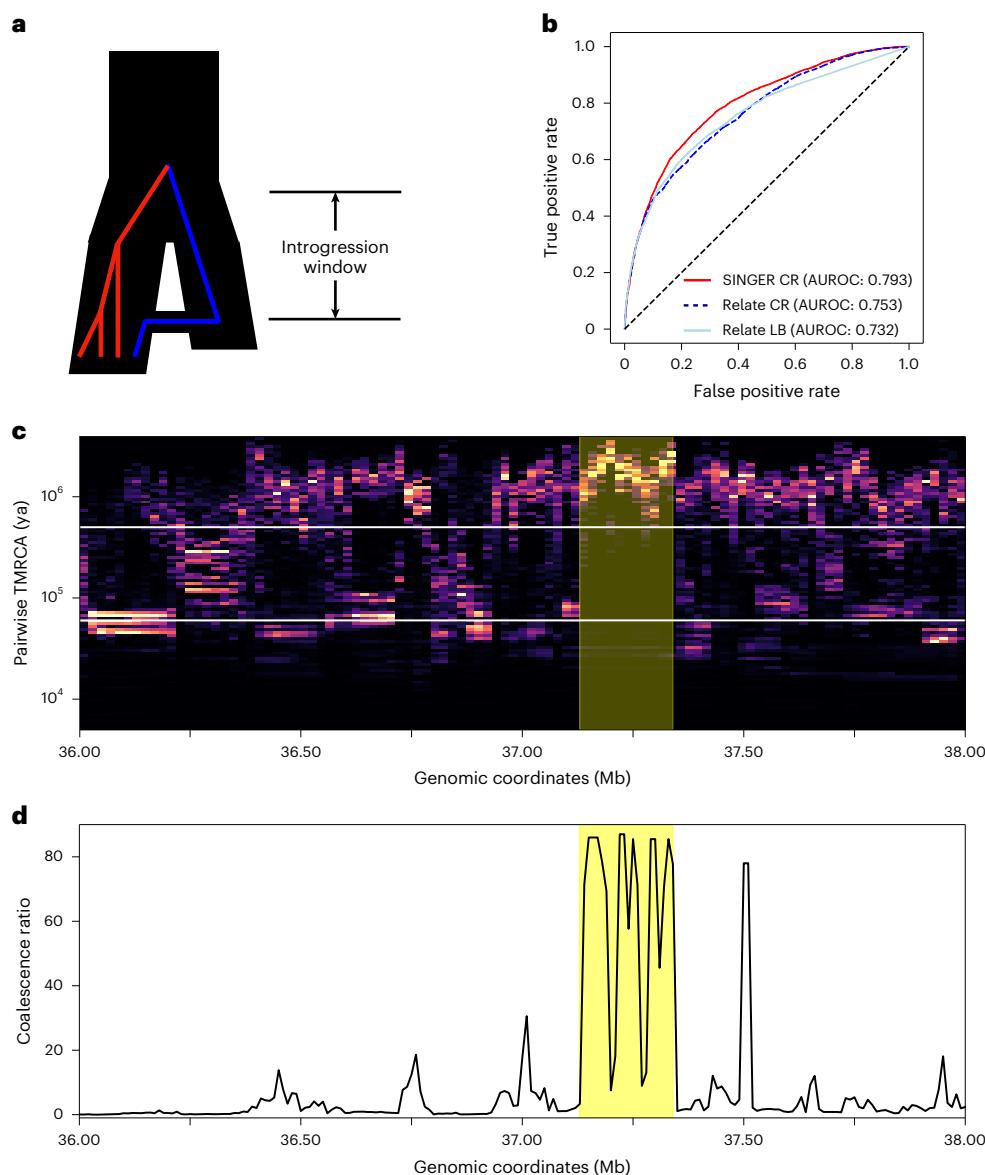
The ARGs inferred by SINGER show extremely ancient pairwise coalescence times in the HLA locus, with many regions harboring coalescence times older than the human–chimpanzee divergence time (Fig. 7a,b). In African individuals, we computed the average TMRCA in 1 kb windows on chromosome 6 and found that HLA is the only region with the average TMRCA above 10 million years (Supplementary Fig. 17), making erroneous ARG inference an unlikely explanation for these ancient coalescence times. This is consistent with the hypothesis of strong balancing selection in this locus and the known trans-species polymorphisms. The human–chimpanzee divergence time is estimated to be 5–12 Mya<sup>76</sup>. Although many genes in the HLA region do not show strong evidence of coalescence times older than the human–chimpanzee split (for example, *TAPI*, *TAP2* and *TAPBP*),

many do, including *HLA-A*, *HLA-DRB1* and *HLA-DRB6*. Unsurprisingly, there are no noticeable differences across the five populations, as the polymorphism has been maintained since ancient times. By contrast, in GBR, Relate and tsinfer + tsdate<sup>3</sup> do not recover such extreme coalescence times (Fig. 7b), probably owing to poor allele polarization in the HLA locus and model misspecification arising from deviations from selective neutrality. To validate our results, we compared the mutation densities in 10 kb windows from real data with predictions from the inferred ARG; SINGER provides a good fit, while Relate and tsinfer + tsdate underestimate substantially (Fig. 7c).

In addition to the HLA locus, we extended the analysis genome-wide to find other loci with exceptionally ancient coalescence times (Supplementary Fig. 18), some of which coincide with previous findings of long-term balancing selection, including *TRIM5* (ref. 77), *ABO*<sup>78</sup>, *IGFBP7* (ref. 79), *PKD1L1* and *DMBT1* (ref. 80).

## Discussion

In this article, we introduced SINGER, a Bayesian method for efficiently sampling ARGs from the posterior distribution. SINGER implements an improved MCMC algorithm to explore the ARG space, thereby enabling accurate uncertainty characterization in both coalescence times and ARG topologies. Our approach scales to at least hundreds of whole-genome sequences while performing full posterior sampling of both ARG branch lengths and topologies. Compared to ARGweaver, SINGER uses faster threading and more efficient MCMC algorithms.



**Fig. 6 | ARG-based detection of archaic introgression tracts.** **a**, The demography model involving introgression and the un-introgressed (red) and introgressed (blue) lineages under the model. The time interval from the introgression time to the divergence time of the two populations is called the ‘introgression window’. **b**, The receiver operating characteristic (AUROC) plot of using coalescence ratio (CR) and long branch proportion (LB) from Relate and SINGER to differentiate the inferred introgressed and un-introgressed tracts from IBDmix. **c**, Identification of potential archaic introgression tracts. For a given leaf node, its pairwise coalescence times with every other leaf node

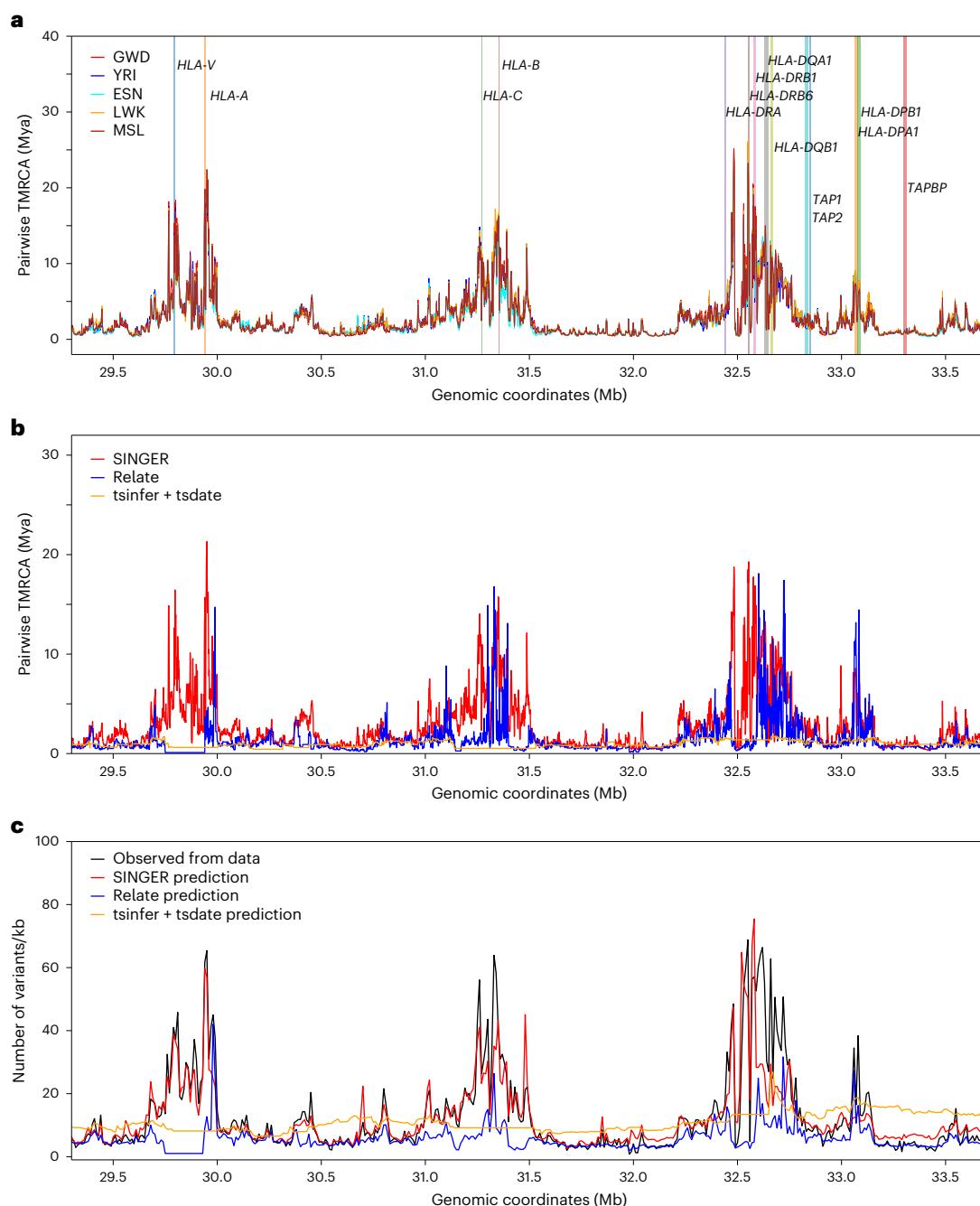
in the marginal tree are summarized as a distribution. In the plot, each column represents such a distribution from marginal trees within a 10 kb window. The two white horizontal lines delineate the interval between the introgression time and the split time. A tract indicative of introgression should exhibit a depletion of coalescence events within this interval and an enrichment of coalescence events above the split time. Regions shaded in red denote putative introgression tracts. **d**, The ratio of pairwise coalescence density above the split time to that within the interval between the introgression time and the split time.

In estimating key population genetic quantities—such as coalescence times, topologies, recombination densities and allele ages—SINGER compares favorably with existing methods, including ARGweaver, Relate, tsinfer + tsdate and ARG-Needle. As demonstrated in our benchmarks, using posterior samples can noticeably enhance inference accuracy and effectively quantify estimation uncertainty. Last but not least, SINGER exhibits greater robustness to model misspecification, such as population size changes and background selection.

We applied SINGER to data from individuals of British descent and individuals of African descent from the 1000 Genomes Project. Using ARG-based fine-scale diversity estimates, we identified genomic regions with exceptional population differentiation in coalescence times across African populations. In addition, we used a visualization

technique, the coalescence distribution heatmap, to identify genomic regions consistent with a specific model of archaic introgression. Lastly, we found strong evidence of trans-species polymorphisms in the HLA region and mapped genes associated with the peaks in these signals.

We note that our proposed approach for detecting archaic introgressed tracts requires a demographic model. However, there is ongoing debate regarding the timing, strength and even the existence of certain archaic introgression<sup>81</sup>. In this regard, the tract identification requires reasonably accurate introgression and split times. Moreover, the detection of introgressed tracts using sampled ARGs warrants further methodological development. Our proposed heatmap of coalescence distribution provides a basis for future methods.



**Fig. 7 | The fine-scale diversity landscape in the HLA region in Africans and GBR, and comparison between observed and inferred mutation density in GBR.** **a**, The average pairwise coalescence time in the HLA in the African sample, with a few genes highlighted (vertical bar colors have no meaning aside from

denoting different genes). **b**, The average pairwise coalescence time in the HLA in GBR, inferred from SINGER, Relate and tsinfer + tsdate. **c**, The observed mutation density from data compared with that predicted from ARGs, inferred from SINGER, Relate and tsinfer + tsdate.

Despite its strength, SINGER has some limitations and room for improvement. First, although it is substantially more scalable than alternative methods for posterior sampling of ARGs, real data applications often require a large number of MCMC iterations. More efficient ARG exploration strategies are therefore needed.

Second, although SINGER shows improved robustness to model misspecification compared to other methods, its inference accuracy might be further enhanced by jointly inferring population size history and branch lengths, similar to the approach used by Relate<sup>16</sup>. However, Relate's algorithm for inferring population size history and branch lengths cannot be readily applied to other ARG inference methods. On the other hand, tsdate is compatible with SINGER's data structure

but assumes a constant population size. More generally, SINGER could be extended to incorporate complex demographic models, as in ARGweaver-D<sup>82</sup>.

Third, SINGER may not be suitable for certain data. For example, it assumes an infinite-sites model, which may not be applicable to cross-species data. Incorporating a finite-sites model, as in ARGweaver<sup>15</sup>, could help. SINGER also requires WGS data and cannot analyze genotyping arrays directly. The locus-skipping algorithm<sup>49,83</sup> can facilitate the analysis of SNP genotyping array data, but applying ARG-Needle directly to genotype array data substantially reduces accuracy (Supplementary Section E.1 and Supplementary Fig. 19). By contrast, imputing genotypes using a reference panel before applying

SINGER yields higher accuracy (Supplementary Fig. 19). Therefore, when genome imputation is feasible, it may be preferable to the use of the locus-skipping algorithm.

Lastly, SINGER requires phased, contemporary genomes. Unfortunately, high-quality phasing is often challenging, especially for ancient DNA and non-model organisms. Therefore, supporting unphased data would increase SINGER's utility, especially for joint analysis of ancient and modern genomes, which are often poorly phased or completely unphased.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-025-02317-9>.

## References

1. Kelleher, J. et al. Inferring whole-genome histories in large population datasets. *Nat. Genet.* **51**, 1330–1338 (2019).
2. Rosenberg, N. A. & Nordborg, M. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat. Rev. Genet.* **3**, 380–390 (2002).
3. Wohns, A. W. et al. A unified genealogy of modern and ancient genomes. *Science* **375**, eabi8264 (2022).
4. Griffiths, R. Neutral two-locus multiple allele models with recombination. *Theor. Popul. Biol.* **19**, 169–186 (1981).
5. Hudson, R. R. Properties of a neutral allele model with intragenic recombination. *Theor. Popul. Biol.* **23**, 183–201 (1983).
6. Baumdicker, F. et al. Efficient ancestry and mutation simulation with msprime 1.0. *Genetics* **220**, iyab229 (2022).
7. Hudson, R. R. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337–338 (2002).
8. Kelleher, J., Thornton, K. R., Ashander, J. & Ralph, P. L. Efficient pedigree recording for fast population genetics simulation. *PLoS Comput. Biol.* **14**, e1006581 (2018).
9. Marjoram, P. & Wall, J. D. Fast “coalescent” simulation. *BMC Genet.* **7**, 16 (2006).
10. McVean, G. A. & Cardin, N. J. Approximating the coalescent with recombination. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360**, 1387–1393 (2005).
11. Wiuf, C. & Hein, J. Recombination as a point process along sequences. *Theor. Popul. Biol.* **55**, 248–259 (1999).
12. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
13. Paul, J. S., Steinrücken, M. & Song, Y. S. An accurate sequentially Markov conditional sampling distribution for the coalescent with recombination. *Genetics* **187**, 1115–1128 (2011).
14. Steinrücken, M., Paul, J. S. & Song, Y. S. A sequentially Markov conditional sampling distribution for structured populations with migration and recombination. *Theor. Popul. Biol.* **87**, 51–61 (2013).
15. Rasmussen, M. D., Hubisz, M. J., Gronau, I. & Siepel, A. Genome-wide inference of ancestral recombination graphs. *PLoS Genet.* **10**, e1004342 (2014).
16. Speidel, L., Forest, M., Shi, S. & Myers, S. R. A method for genome-wide genealogy estimation for thousands of samples. *Nat. Genet.* **51**, 1321–1329 (2019).
17. Li, N. & Stephens, M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* **165**, 2213–2233 (2003).
18. Zhang, B. C., Biddanda, A., Gunnarsson, Á. F., Cooper, F. & Palamara, P. F. Biobank-scale inference of ancestral recombination graphs enables genealogical analysis of complex traits. *Nat. Genet.* **55**, 768–776 (2023).
19. Brandt, D. Y., Huber, C. D., Chiang, C. W. & Ortega-Del Vecchio, D. The promise of inferring the past using the ancestral recombination graph. *Genome Biol. Evol.* **16**, evae005 (2024).
20. Harris, K. From a database of genomes to a forest of evolutionary trees. *Nat. Genet.* **51**, 1306–1307 (2019).
21. Harris, K. Using enormous genealogies to map causal variants in space and time. *Nat. Genet.* **55**, 730–731 (2023).
22. Lewanski, A. L., Grundler, M. C. & Bradburd, G. S. The era of the ARG: an introduction to ancestral recombination graphs and their significance in empirical evolutionary genomics. *PLoS Genet.* **20**, e1011110 (2024).
23. Nielsen, R., Vaughn, A. H. & Deng, Y. Inference and applications of ancestral recombination graphs. *Nat. Rev. Genet.* **26**, 47–58 (2025).
24. Fan, C. et al. A likelihood-based framework for demographic inference from genealogical trees. *Nat. Genet.* <https://doi.org/10.1038/s41588-025-02129-x> (2025).
25. Fan, C., Mancuso, N. & Chiang, C. W. A genealogical estimate of genetic relationships. *Am. J. Hum. Genet.* **109**, 812–824 (2022).
26. Grundler, M. C., Terhorst, J. & Bradburd, G. S. A geographic history of human genetic ancestry. *Science* **387**, 1391–1397 (2025).
27. Guo, F., Carbone, I. & Rasmussen, D. A. Recombination-aware phylogeographic inference using the structured coalescent with ancestral recombination. *PLoS Comput. Biol.* **18**, e1010422 (2022).
28. Hejase, H. A., Mo, Z., Campagna, L. & Siepel, A. A deep-learning approach for inference of selective sweeps from the ancestral recombination graph. *Mol. Biol. Evol.* **39**, msab332 (2022).
29. Ignatieva, A. et al. The distribution of branch duration and detection of inversions in ancestral recombination graphs. *Mol. Biol. Evol.* <https://doi.org/10.1093/molbev/msaf190> (2025).
30. Ignatieva, A., Hein, J. & Jenkins, P. A. Ongoing recombination in SARS-CoV-2 revealed through genealogical reconstruction. *Mol. Biol. Evol.* **39**, msac028 (2022).
31. Osmond, M. & Coop, G. Estimating dispersal rates and locating genetic ancestors with genome-wide genealogies. *Elife* **13**, e72177 (2024).
32. Salehi Nowbandegani, P. et al. Extremely sparse models of linkage disequilibrium in ancestrally diverse association studies. *Nat. Genet.* **55**, 1494–1502 (2023).
33. Speidel, L. et al. High-resolution genomic history of early medieval Europe. *Nature* **637**, 118–126 (2025).
34. Stern, A. J., Wilton, P. R. & Nielsen, R. An approximate full-likelihood method for inferring selection and allele frequency trajectories from DNA sequence data. *PLoS Genet.* **15**, e1008384 (2019).
35. Wang, S. & Coop, G. A complex evolutionary history of genetic barriers to gene flow in hybridizing warblers. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.11.14.516535> (2022).
36. YC Brandt, D., Wei, X., Deng, Y., Vaughn, A. H. & Nielsen, R. Evaluation of methods for estimating coalescence times using ancestral recombination graphs. *Genetics* **221**, iyac044 (2022).
37. Deng, Y., Song, Y. S. & Nielsen, R. The distribution of waiting distances in ancestral recombination graphs. *Theor. Popul. Biol.* **141**, 34–43 (2021).
38. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919–925 (2014).
39. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet.* **5**, e1000471 (2009).
40. Murphy, D. A., Elyashiv, E., Amster, G. & Sella, G. Broad-scale variation in human genetic diversity levels is predicted by purifying selection on coding and non-coding elements. *Elife* **12**, e76065 (2022).

41. Kuhner, M. K., Yamato, J. & Felsenstein, J. Maximum likelihood estimation of recombination rates from population data. *Genetics* **156**, 1393–1401 (2000).
42. Mahmoudi, A., Koskela, J., Kelleher, J., Chan, Y. B. & Balding, D. Bayesian inference of ancestral recombination graphs. *PLoS Comput. Biol.* **18**, e1009960 (2022).
43. Kamm, J. A., Terhorst, J. & Song, Y. S. Efficient computation of the joint sample frequency spectra for multiple populations. *J. Comput. Graph. Stat.* **26**, 182–194 (2017).
44. Liu, X. & Fu, Y. X. Exploring population size changes using SNP frequency spectra. *Nat. Genet.* **47**, 555–559 (2015).
45. Liu, X. & Fu, Y. X. Stairway plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biol.* **21**, 280 (2020).
46. Kelleher, J., Etheridge, A. M. & McVean, G. Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Comput. Biol.* **12**, e1004842 (2016).
47. Link, V. et al. Tree-based QTL mapping with expected local genetic relatedness matrices. *Am. J. Hum. Genet.* **110**, 2077–2091 (2023).
48. Schweiger, R. & Durbin, R. Ultrafast genome-wide inference of pairwise coalescence times. *Genome Res.* **33**, 1023–1031 (2023).
49. Palamara, P. F., Terhorst, J., Song, Y. S. & Price, A. L. High-throughput inference of pair-wise coalescence times identifies signals of selection and enriched disease heritability. *Nat. Genet.* **50**, 1311–1317 (2018).
50. Terhorst, J., Kamm, J. A. & Song, Y. S. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017).
51. Cook, S. R., Gelman, A. & Rubin, D. B. Validation of software for Bayesian models using posterior quantiles. *J. Comput. Graph. Stat.* **15**, 675–692 (2006).
52. Talts, S., Betancourt, M., Simpson, D., Vehtari, A. & Gelman, A. Validating Bayesian inference algorithms with simulation-based calibration. Preprint at <https://arxiv.org/abs/1804.06788> (2020).
53. Ralph, P., Thornton, K. & Kelleher, J. Efficiently summarizing relationships in large samples: a general duality between statistics of genealogies and genomes. *Genetics* **215**, 779–797 (2020).
54. Peng, D., Mulder, O. J. & Edge, M. D. Evaluating ARG-estimation methods in the context of estimating population-mean polygenic score histories. *Genetics* **229**, iyaf033 (2025).
55. Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321–324 (2005).
56. Spence, J. P. & Song, Y. S. Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. *Sci. Adv.* **5**, eaaw9206 (2019).
57. Byrska-Bishop, M. et al. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell* **185**, 3426–3440 (2022).
58. Charlesworth, B., Morgan, M. & Charlesworth, D. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303 (1993).
59. Hudson, R. R. & Kaplan, N. L. Deleterious background selection with recombination. *Genetics* **141**, 1605–1617 (1995).
60. Akey, J. M. et al. Tracking footprints of artificial selection in the dog genome. *Proc. Natl Acad. Sci. USA* **107**, 1160–1165 (2010).
61. Fan, S. et al. Whole-genome sequencing reveals a complex African population demographic history and signatures of local adaptation. *Cell* **186**, 923–939 (2023).
62. Whitlock, M. C. & Lotterhos, K. E. Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of  $F_{ST}$ . *Am. Nat.* **186**, S24–S36 (2015).
63. Mathieson, I. & Terhorst, J. Direct detection of natural selection in Bronze Age Britain. *Genome Res.* **32**, 2057–2067 (2022).
64. Levy, C., Khaled, M. & Fisher, D. E. MITF: master regulator of melanocyte development and melanoma oncogene. *Trends Mol. Med.* **12**, 406–414 (2006).
65. Zhang, R. et al. A CRISPR screen defines a signal peptide processing pathway required by flaviviruses. *Nature* **535**, 164–168 (2016).
66. Adogo, L. & Ogoh, M. Yellow fever in Nigeria: a review of the current situation. *Afr. J. Clin. Exp. Microbiol.* **21**, 1–13 (2020).
67. Reimann, F. et al. Pain perception is altered by a nucleotide polymorphism in SCN9A. *Proc. Natl Acad. Sci. USA* **107**, 5148–5153 (2010).
68. Green, R. E. et al. A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
69. Reich, D. et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2010).
70. Sankararaman, S. et al. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354–357 (2014).
71. Hammer, M. F., Woerner, A. E., Mendez, F. L., Watkins, J. C. & Wall, J. D. Genetic evidence for archaic admixture in Africa. *Proc. Natl Acad. Sci. USA* **108**, 15123–15128 (2011).
72. Chen, L., Wolf, A. B., Fu, W., Li, L. & Akey, J. M. Identifying and interpreting apparent Neanderthal ancestry in African individuals. *Cell* **180**, 677–687 (2020).
73. Skoglund, P. & Mathieson, I. Ancient genomics of modern humans: the first decade. *Annu. Rev. Genomics Hum. Genet.* **19**, 381–404 (2018).
74. Fortier, A. L. & Pritchard, J. K. Ancient trans-species polymorphism at the major histocompatibility complex in primates. Preprint at *eLife* <https://doi.org/10.7554/eLife.103547.2> (2025).
75. Liu, B., Shao, Y. & Fu, R. Current research status of HLA in immune-related diseases. *Immun. Inflamm. Dis.* **9**, 340–350 (2021).
76. Moorjani, P., Amorim, C. E. G., Arndt, P. F. & Przeworski, M. Variation in the molecular clock of primates. *Proc. Natl Acad. Sci. USA* **113**, 10607–10612 (2016).
77. Cagliani, R. et al. Long-term balancing selection maintains trans-specific polymorphisms in the human TRIM5 gene. *Hum. Genet.* **128**, 577–588 (2010).
78. Ségurel, L. et al. The ABO blood group is a trans-species polymorphism in primates. *Proc. Natl Acad. Sci. USA* **109**, 18493–18498 (2012).
79. Leffler, E. M. et al. Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science* **339**, 1578–1582 (2013).
80. Bitarello, B. D., Brandt, D. Y., Meyer, D. & Andrés, A. M. Inferring balancing selection from genome-scale data. *Genome Biol. Evol.* **15**, evad032 (2023).
81. Ragsdale, A. P. et al. A weakly structured stem for human origins in Africa. *Nature* **617**, 755–763 (2023).
82. Hubisz, M. J., Williams, A. L. & Siepel, A. Mapping gene flow between ancient hominins through demography-aware inference of the ancestral recombination graph. *PLoS Genet.* **16**, e1008895 (2020).
83. Paul, J. S. & Song, Y. S. Blockwise HMM computation for large-scale population genomic inference. *Bioinformatics* **28**, 2008–2015 (2012).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format,

as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended

use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025, corrected publication 2025

## Methods

### Branch sampling

To speed up the computation, we first partition the genome into equal-sized bins, in which the bin size is chosen to be about  $4 \times 10^{-3}/(4N_e r)$ . We then construct an HMM indexed by these bins (loci) to sample the branches at which the lineage for the  $n^{\text{th}}$  haplotype joins the partial ARG for the first  $n - 1$  haplotypes. The state space  $S_l$  for bin  $l$  comprises all the branches in the marginal tree for bin  $l$  in the partial ARG and some branches from earlier bins. The precise definition of the state space can be found in Supplementary Section B.1. For bin  $l$ , we use  $B_l$  to denote the branch onto which the lineage for the  $n^{\text{th}}$  haplotype joins. If the partial ARG does not already contain a recombination event between bins  $l - 1$  and  $l$ , then the transition probability of the HMM is defined as:

$$P(B_l = b_j | B_{l-1} = b_i) = (1 - r_i) \delta_{ij} + r_i q_j / \left( \sum_{k: b_k \in S_l} q_k \right)$$

where  $b_i \in S_{l-1}$ ,  $b_j \in S_l$  and  $r_i$  denotes the branch-specific recombination probability for branch  $b_i$ . The definition of  $r_i$  and  $q_j$ , which are computed under the assumption of a constant-sized panmictic population, can be found in Supplementary Section B.1.5. The structure of this transition probability is similar to that of the Li–Stephens model<sup>17</sup>, but with branch-specific recombination and re-joining probabilities. This allows us to reduce the HMM computational complexity to be linear with respect to the number of hidden states, as in the Li–Stephens model.

We restrict the final ARG to have at most one recombination event between adjacent bins. Therefore, if the partial ARG already contains a recombination event between bins  $l - 1$  and  $l$ , the threading operation is not allowed to introduce an additional recombination between these two bins, the transition probability in this case is defined similarly to that in previous work<sup>15</sup>; the details are provided in Supplementary Section B.1.4 and Supplementary Fig. 30.

### Time sampling

Conditioned on a sequence of joining branches along the genome resulting from the branch sampling algorithm, the time sampling step proceeds in a similar fashion as in (pairwise) PSMC<sup>12</sup> for a constant-sized panmictic population, but with the restriction that for each bin  $l$ , the coalescence time should reside between the two endpoints of the joining branch for bin  $l$ . To accelerate the computation, we implemented a previously published linearization technique<sup>84</sup>. The details are provided in Supplementary Section B.2.

### ARG re-scaling

Given an inferred ARG, we partition the time axis into non-overlapping windows such that the total branch length across all marginal trees (weighted by the span of each tree) in each time window is the same; by default, 100 windows are chosen. We then count the number of mutations falling into each of these windows. If a mutation falls on a branch striding multiple windows, then its contribution to the mutation count for each window is given by the proportion of the branch that overlaps with the window. We re-scale each window size such that the expected number of mutations for the window matches the empirical count (Extended Data Fig. 6). This is essentially a window-specific re-scaling to better match the mutation clock. Further details can be found in Supplementary Section B.3. ARG re-scaling is performed after the initialization and every thinning step.

### SGPR

In the MCMC algorithm, we propose updates to the current ARG by first removing some branches following a cut and then re-grafting from the breakpoint.

To remove a branch from a given marginal tree, we make a random cut on the tree; the probability that a given branch will be cut is

proportional to its length. We can extend the cut leftwards and rightwards along the genome, removing the partial branch from the cut to its upper endpoint of the branch. Typically, the cut will not extend over the entire chromosome; rather, the extension width will be the same as the span of the ancestral segment corresponding to the branch that was cut. The details can be found in Supplementary Section B.4 and Extended Data Fig. 7.

To re-graft the branch from the breakpoint, we use the same threading algorithm described in the Branch Sampling and Time Sampling sections above, with the only difference being that now we only consider the sub-ARG above the breakpoint. We show in Supplementary Section B.4.3 that, assuming the threading algorithm samples approximately from the posterior, the acceptance probability is typically much higher than that of previous proposals<sup>15,41</sup>, thereby improving convergence and mixing in MCMC.

### Simulation and benchmarking details

All coalescent simulations in this article were carried out using msprime<sup>46</sup> with  $r = \mu = 2 \times 10^{-8}$  and  $N_e = 1 \times 10^4$ . We simulated 50 datasets, each with 50 sequences over a 1 Mb region, and ten datasets, each with 300 sequences over 1 Mb. For simulations with 50 sequences, we also simulated with CEU population size history [https://github.com/PalamaraLab/ASMC\\_data/tree/main/demographies](https://github.com/PalamaraLab/ASMC_data/tree/main/demographies) estimated from SMC++ (ref. 50). We ran all inference methods with these true parameter values.

For Relate and tsinfer + tsdate, we use the posterior averages on their fixed estimated topology. ARG-Needle uses posterior averages of joining times when threading. SINGER and ARGweaver sample from the posterior over topologies, and we use 100 sampled ARGs for both.

As for MCMC sampling, we uniformized the number of iterations and the thinning scheme across all methods. We drew 100 samples with the thinning interval set to 20 for ARGweaver, Relate and SINGER, and used 1,000 iterations for burn-in.

For all simulation benchmarks involving ARGweaver and SINGER, posterior averages were taken for the statistics of interest, such as pairwise coalescence time, allele age and so on. Given that Relate outputs averages of MCMC iterations while tsdate results are averages from a probability table, we simply use their results from a single output, as they are effectively posterior averages.

### ARG inference methods benchmarked in this study

The version or the last-known maintenance date of the ARG inference methods we considered in this study were as follows: SINGER (v.0.1.8), Relate (v.1.1.9), tsinfer (v.0.2.1), tsdate (v.0.1.4), ARG-Needle (March 2024) and ARGweaver (Jan 2017). We note that the code for a newer version (v.0.2.1) of tsdate was released while the manuscript was under review, but it is unpublished work and therefore we did not use it in our study.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

We have uploaded the inferred ARG samples (100 samples) and genes with exceptional population differentiation in coalescence times to Zenodo at <https://doi.org/10.5281/zenodo.10437053> (ref. 86), <https://doi.org/10.5281/zenodo.10467284> (ref. 87), <https://doi.org/10.5281/zenodo.10467509> (ref. 88) and <https://zenodo.org/records/10828414> (ref. 89).

### Code availability

SINGER is written in C++ with Python helper scripts. The source code can be downloaded from <https://github.com/popgenmethods/SINGER> (ref. 85).

## References

84. Harris, K., Sheehan, S., Kamm, J. A., Song, Y. S. Decoding coalescent hidden Markov models in linear time. In *RECOMB 2014: Proc. 18th Annu. Int. Conf. Research in Computational Molecular Biology* (Ed. Sharan, R.) 100–114 (Springer, 2014).
85. Deng, Y., Nielsen, R. & Song, Y. S. Robust and accurate Bayesian inference of genome-wide genealogies for hundreds of genomes. SINGER beta version v.0.1.8. Zenodo <https://doi.org/10.5281/zenodo.15707558> (2025).
86. Deng, Y. SINGER-inferred ARG samples from Africans in 1000 Genomes Project (part 1). Zenodo <https://doi.org/10.5281/zenodo.10437053> (2024).
87. Deng, Y. SINGER-inferred ARG samples from Africans in 1000 Genomes Project (part 1). Zenodo <https://doi.org/10.5281/zenodo.10467284> (2024).
88. Deng, Y. SINGER-inferred ARG samples from Africans in 1000 Genomes Project (part 3). Zenodo <https://doi.org/10.5281/zenodo.10467509> (2024).
89. Deng, Y. SINGER-inferred targets for exceptional population differentiation in coalescence times in African populations in 1000 Genomes Project. Zenodo <https://zenodo.org/records/10828414> (2024).

## Acknowledgements

We thank M. Rasmussen and M. Hubisz for helpful correspondence regarding ARGweaver; L. Speidel for discussion regarding Relate; V. Buffalo for discussion on background selection and mutation rates; D. Brandt for discussion on the benchmark pipeline; Y. Zhang for discussion on introgression analysis; J. Schraiber for advice on data analysis; M. Slatkin and A. Fortier for discussion on HLA;

and J. Santangelo, C. Langley, N. Pope and J. Perez for testing the software. This research is supported in part by National Institutes of Health grants R56-HG013117, R01-HG013117, R35-GM153400 and R35-GM134922.

## Author contributions

Y.D., R.N. and Y.S.S. conceived the study. Y.D. implemented the algorithm, developed the software and performed the analysis under the supervision of R.N. and Y.S.S. Y.D., R.N. and Y.S.S. wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

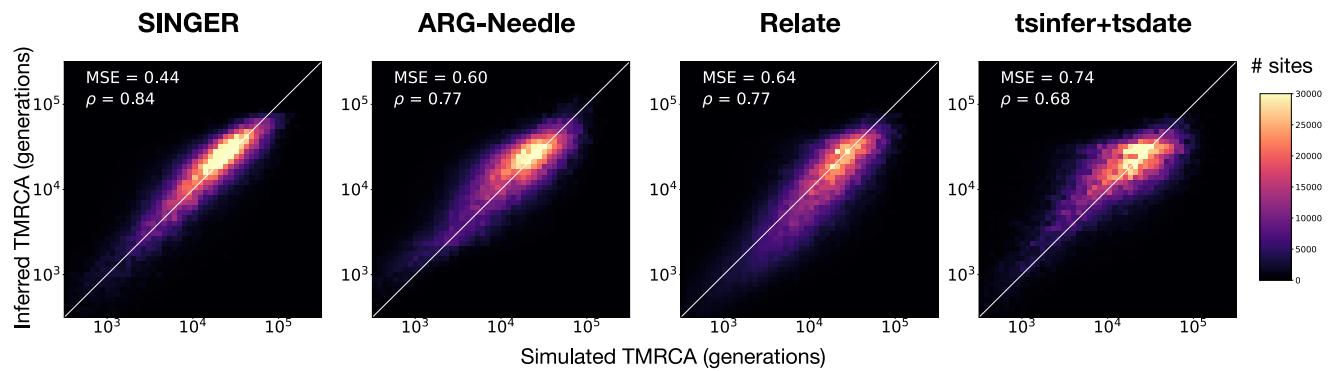
**Extended data** is available for this paper at <https://doi.org/10.1038/s41588-025-02317-9>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41588-025-02317-9>.

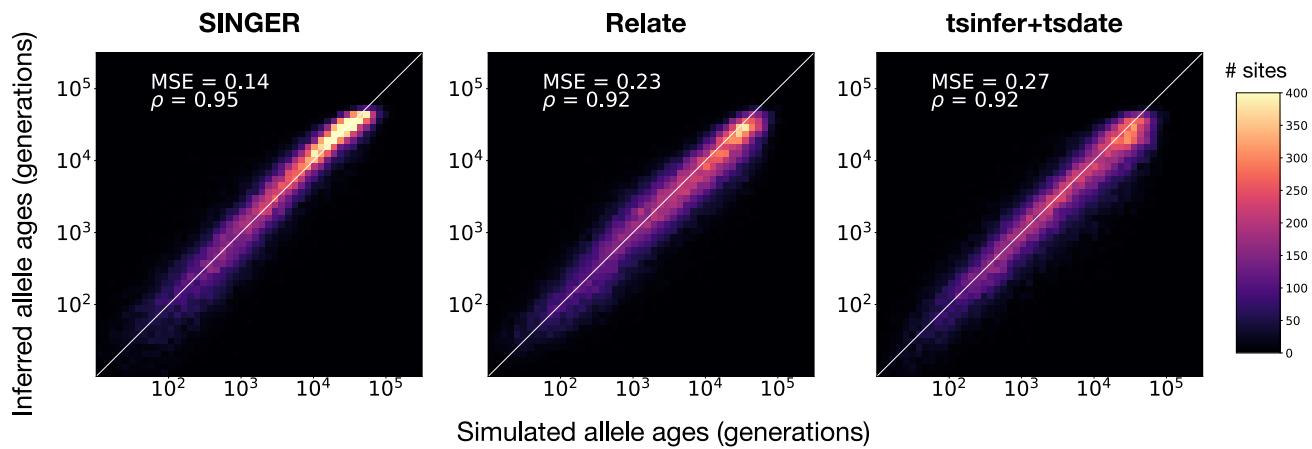
**Correspondence and requests for materials** should be addressed to Rasmus Nielsen or Yun S. Song.

**Peer review information** *Nature Genetics* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

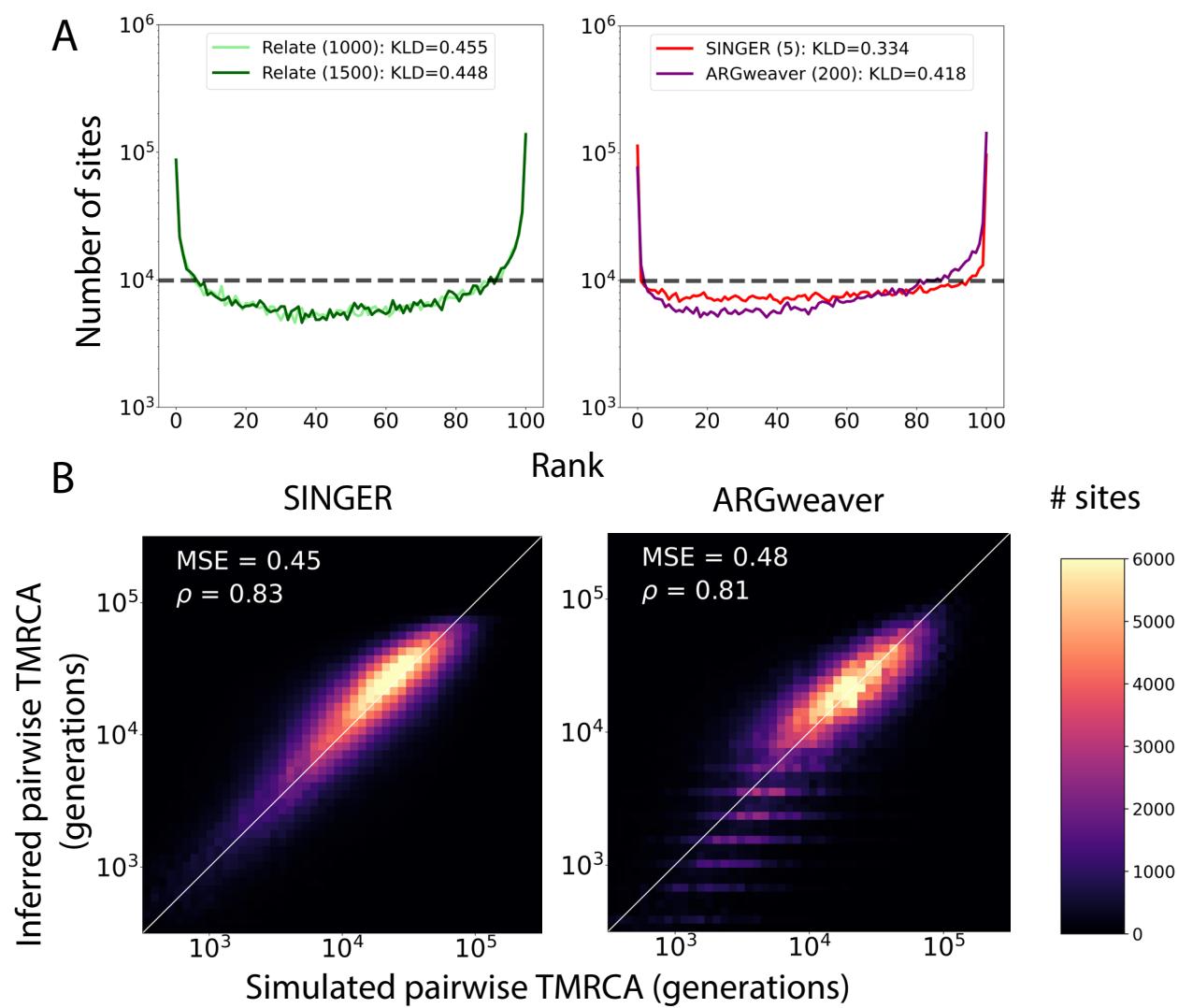
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



**Extended Data Fig. 1 | Inference of pairwise TMRCA on 300 sequences.** Performance of inferring pairwise TMRCA with SINGER, ARG-Needle, Relate and tsinfer+tsdate, with 300 sequences. SINGER performs the best, while ARG-Needle and Relate perform similarly, and tsinfer+tsdate performs the worst.

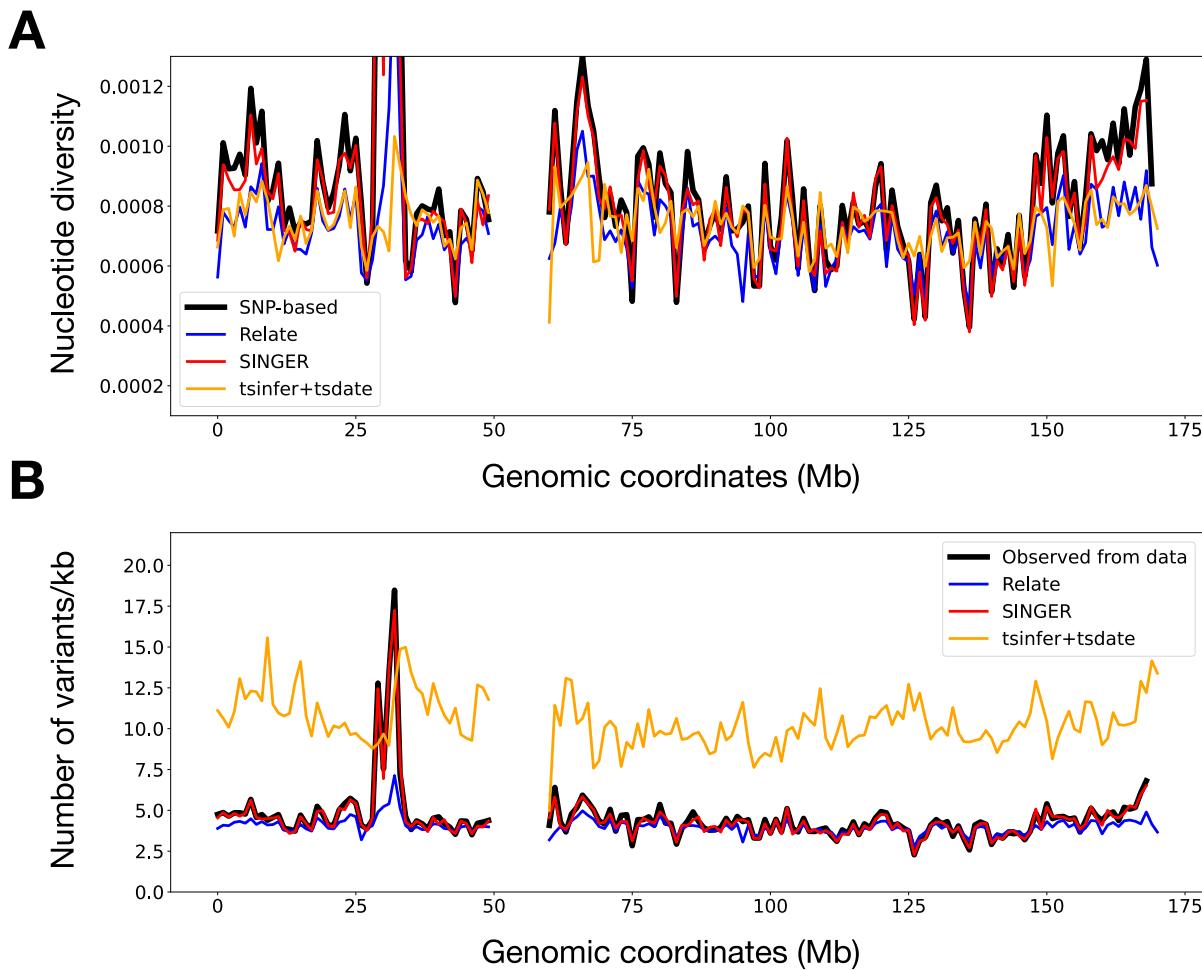


**Extended Data Fig. 2 | Inference of allele ages on 300 sequences.** Performance of inferring allele ages with SINGER, Relate and tsinfer+tsdate, with 300 sequences. SINGER performs the best, followed by Relate, and tsinfer+tsdate performs the worst. ARG-Needle is excluded because it does not map mutations to branches of the inferred ARG.

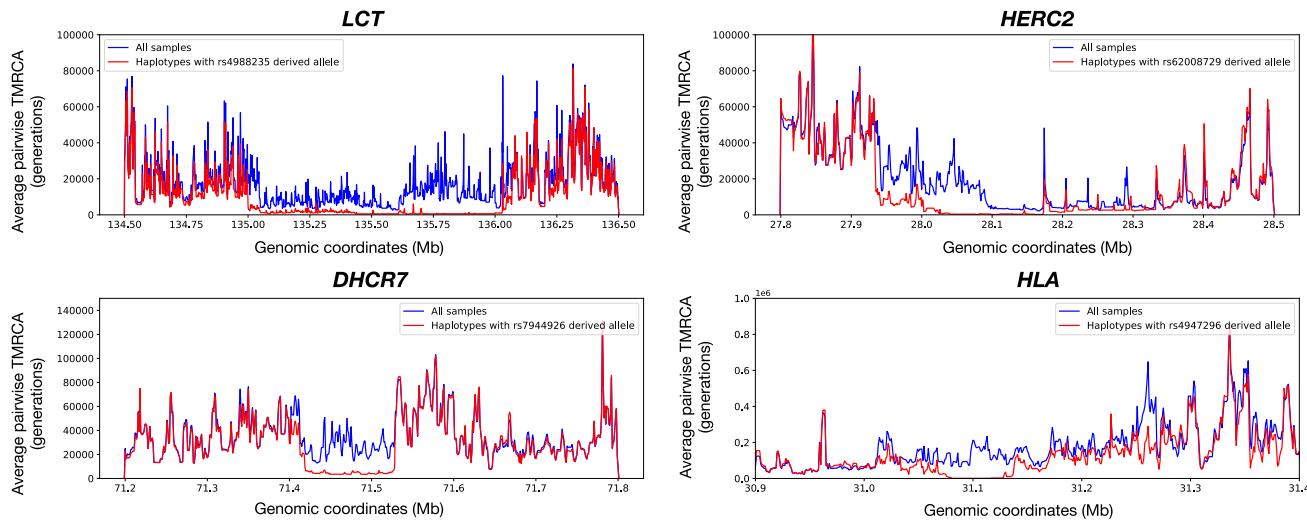


**Extended Data Fig. 3 | The influence of thinning on MCMC sampling.** (a) Rank plots for different methods. In these results, considerably longer thinning intervals, indicated in parentheses, were used for ARGweaver and Relate. Relate is close to saturated thinning because changing the thinning interval from 1,000 to 1,500 results in little change in the rank plot. (b) Pairwise TMRCA inference from

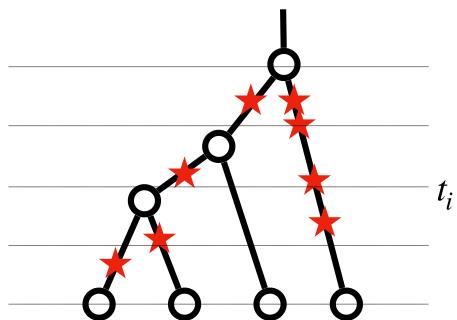
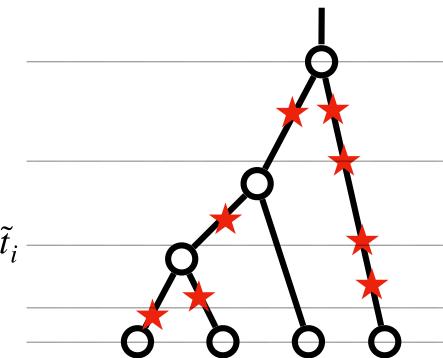
ARGweaver with thinning interval length 200 compared to that from SINGER with thinning interval length 5. ARGweaver still underperforms compared to SINGER in both rank plot and pairwise TMRCA, even with 40 times longer thinning than SINGER.



**Extended Data Fig. 4 | Consistency with Mb scale statistics.** The 1-Mb nucleotide diversity (a) and variant density (b) landscape for chromosome 6 in GBR, computed from real data (black) versus inferred from Relate (blue), SINGER (red), and tsinfer+tsdate (orange).

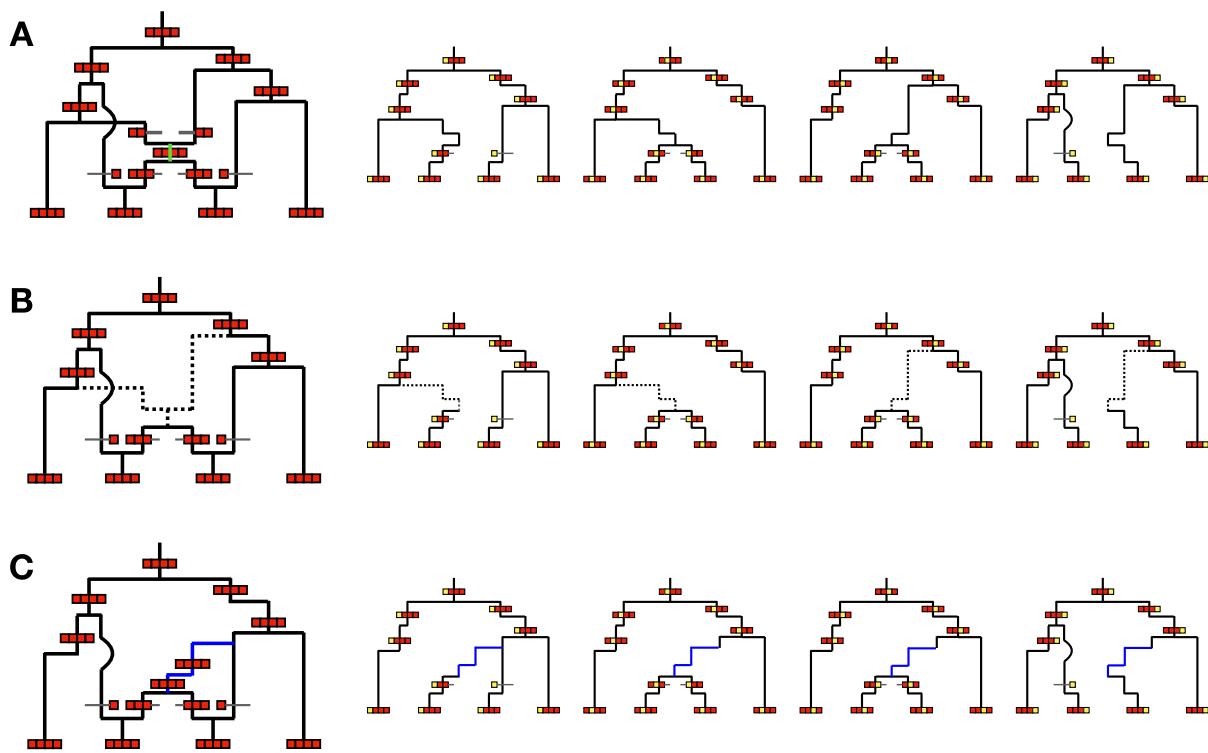


**Extended Data Fig. 5 | Examples of previously identified selection targets.** The comparison of the average pairwise TMRCA among carriers of potential target SNP and all samples, for the candidate genes in the selection scan.

**A****B**

**Extended Data Fig. 6 | ARG rescaling from original ARG (a) to a new ARG (b).** For simplicity we only show a single tree. We partition the ARG and count the number of mutations in each interval, and re-scale the interval length so that the expected

number of mutations match the observed mutation counts. For example, the highest interval has more mutations mapped to it than the lowest interval, so it is widened and the latter is narrowed.



**Extended Data Fig. 7 | The connection between the Kuhner move and SGPR.** ARG can be represented as the temporal network (left) or the spatial series of trees (right). (a) is the ARG before update in temporal network and spatial trees. In (b), the Kuhner move first picks a random branch (shown in green) in the network and removes any recombination or coalescence above from that branch, which is marked with dashed line (left). It is equivalent to choose a cut and trace

it to the flanking region in SGPR (right). The partial branches connecting the cut to its upper node in every marginal tree is removed. In (c), the sub-graph has been re-grafted back with a new coalescent history, leading to potential changes in topology and branch length. For example, the recombination separating the second and third tree in the previous ARG (a) is gone.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection No software was used to collect data.

Data analysis These softwares are used for simulation benchmarks and/or data analysis:  
Relate (v1.1.9): <https://myersgroup.github.io/relate/>  
tsinfer (v0.2.1): <https://tskit.dev/tsinfer/docs/stable/>  
tsdate (v0.1.4): <https://tsdate.readthedocs.io/en/latest/> (there is an unpublished version 0.2.1 which we didn't use)  
ARG-Needle (March 2024): <https://palamaralab.github.io/software/argneedle/manual/>  
ARGweaver (Jan 2017): <https://github.com/mdrasmus/argweaver>  
SINGER (v0.1.8): <https://github.com/popgenmethods/SINGER>

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The 1000 Genomes Project data can be accessed at: <https://www.internationalgenome.org/data-portal/data-collection/30x-grch38>

The inferred ARGs we produced using SINGER is publicly released at Zenodo: <https://doi.org/10.5281/zenodo.10437053>, <https://doi.org/10.5281/zenodo.10467284>, <https://doi.org/10.5281/zenodo.10467509>. The targets for local adaptation is publicly released at Zenodo: <https://zenodo.org/records/10828414>.

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	n/a
Data exclusions	n/a
Replication	n/a
Randomization	n/a
Blinding	n/a

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

**Materials & experimental systems**

n/a	Involved in the study
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology and archaeology
<input checked="" type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	Dual use research of concern

**Methods**

n/a	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	MRI-based neuroimaging