Brief Population Bottlenecks Are Beyond The Genetic Streetlight

Jack Elliott-Higgins

Washington University School of Medicine

S. Joshua Swamidass (✉ swamidass@gmail.com)

Washington University School of Medicine <https://orcid.org/0000-0003-2191-0778>

Article

Keywords: population bottlenecks, population genetics, Lineage Time Inference (LiTI)

Posted Date: August 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-817056/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Brief Population Bottlenecks Are Beyond The Genetic Streetlight

Jack Elliott-Higgins¹, S. Joshua Swamidass^{1,*},

1 Department of Pathology and Immunology, Washington University School of Medicine

* swamidass@wustl.edu

Abstract

Inferring human demographic history from extant genomes is an important goal of population genetics. To date, the sensitivity of coalescence-based methods in detecting population bottlenecks has not been well characterized. In this study, we find that brief bottlenecks, of just a few generations, are undetectable by current methods. A new approach to population inference, Lineage Time Inference (LiTI), uses data-derived windows to demarcate the limits of the genetic data. We find that a sharp population bottleneck at the time of the Youngest Toba Eruption. At more ancient timepoints in the human lineage, brief bottlenecks would also be outside the genetic streetlight.

As more genomes are sequenced, especially ancient genomes, we are gaining unprecedented insight into the history of human populations. In recent years, several new methods have been developed and applied to infer demographic information in the past using genomic-scale data.

Though the theory behind these methods is well established, the practical limitations of their application to human data are not fully understood. In particular, no studies have examined the sensitivity of population demography inference in detecting sharp population bottlenecks of just a few generations, such as those that might have occurred at founding or speciation events, or following natural disasters.

This gap in knowledge has bearing on important debates on human origins. For example, the Youngest Toba Eruption, 74,000 years ago, has been thought, by some, to have caused a global bottleneck as low as 3,000 individuals, followed by rapid growth. [2, 14] Some have contested this hypothesis, in part, because this bottleneck is not visible in the genetic data [7] and due to a lack of corroborative geological evidence. [17, 21] It is beyond our scope to fully review this debate or summarize all the relevant lines of evidence. Failure to detect a drop in population size, however, should only be relevant if the methods employed are capable, in the first place, of detecting a brief drop in population size at the relevant time period.

A brief and sharp bottleneck is simultaneously a severe and mild bottleneck. The impact of a bottleneck on the genetic data is greater the sharper the drop in population and greater the longer the reduced population is maintained. Sharp and brief bottlenecks reduce a population to very small numbers for only a handful of generations. Along the dimension of size, they are severe demographic events. Along the dimension of duration, they are very mild events.

Using simulations in which the true population history was known, we studied the sensitivity of several of the most advanced demography inference methods, including *MSMC* [16] and *Relate*. [18] We found that they could not detect bottlenecks of only a few generations. This limitation is due, in part, to the way population size estimates are computed, in fixed windows that are equally distributed in log space.

To address this limitation, we developed Lineage Time Inference (LiTI), based on the median coalescence times of dated tree sequences inferred using *ARGweaver*, [15] or *tsdate*, [20] or *Relate*. By using data-derived windows for its predictions, LiTI can, in principle, model tight bottlenecks. In practice, the uncertainty and noise inherent to inferring ancestral recombination graphs still limits sensitivity. Bottlenecks of just a few generations, even when down to two individuals, are not detectable by LiTI either. Improved inference of coalescent dates may improve performance.

Distinct from prior methods, LiTI also computes the *minimum* population size that could possibly be consistent with the data. In this way, LiTI can, in principle, rule out even brief bottlenecks. We find that a LiTI analysis of human genetic data cannot rule out even a very tight bottleneck of ten individuals at the time of the Youngest Toba Eruption, nor can it rule out bottlenecks at more ancient times relevant to human evolution.

Results

We tested the sensitivity of population demography inference algorithms in inferring the demographic history of simulated populations of known history. The simulations match the assumptions of the inference algorithms, excluding real world confounding factors such as population structure or positive and negative selection. If bottlenecks cannot be detected in these idealized tests, we do not expect them to be detectable in practice.

Efficient simulation of thousands of genomes has only recently become possible (Fig. 1A), enabled with a tree-sequence data structure that efficiently encodes the full ancestral recombination graph (ARG) of simulated populations (Fig. 1B). To simulate the ARGs of constant size populations, we used “backward” simulations of coalescence with *msprime*. [10] Coalescent simulations are efficient, but sharp bottlenecks violate the large-population assumption they rely upon. To simulate ARGs through sharp bottlenecks, we used “forward” simulations with *SLiM* of a Wright-Fisher population. [8,9] These simulations both record and output the simulation using a tree-sequence data structure. The simulation tree-sequences were stitched together and pruned to compute a final simulated ARG. From the simulated ARG, genetic haplotypes for large numbers of individuals can be efficiently sampled. Inference algorithms use these haplotypes to infer the demographic history of the population.

Population history is often estimated across windows equally distributed in log space, as the inverse of the Kingman-weighted coalescent rate (Fig 1C). We tested recent implementations of two best-in-class inference algorithms that employ this strategy: *MSMC2* and *Relate*. *MSMC2*, which infers population demographics without first inferring an ARG, is the older of the two methods and can only efficiently use the data from up to eight haplotypes. *Relate* is a more recent method that can work on hundreds of haplotypes. In contrast with *MSMC2*, *Relate* alternately and iteratively refines an explicit ARG of the samples and the population demography.

Figure 1. Simulating and inferring bottlenecks with tree sequences and ancestral recombination graphs (ARGs). Population bottlenecks were simulated by stitching together two backwards and one forward simulation. **(A)** Backwards simulations are very efficient, but sharp population drops violate their assumptions. The bottlenecks, therefore, are simulated using Wright-Fisher forward simulation. **(B)** Both types of simulations construct ancestral recombination graphs (ARGs), recorded as sequences of phylogenetic trees mapped to segments of the genome. Each ARG records the full history of coalescence and recombination of each segment in the genome. The tree sequences from each simulation are merged and pruned to form the final simulation dataset. **(C)** Statistics on the ARG, either simulated or inferred from data, are used to infer past population history. Commonly, coalescence rate is computed across the ARG at fixed-width time-windows, and these rates are used to infer the effective population size (N_e). On the simulation ARG, which is free of estimation error, this approach fails to detect a two individual bottleneck at 75 kya (dotted line).

Lineage Time Inference using data-derived estimation windows

There is a significant pitfall to estimating N_e across windows that are equally distributed in log space. Bottlenecks of just a few generations do not span a full window. So, even in principle, the resolution of the predictions will never be sufficient to pick up brief bottlenecks.

To overcome this limitation, we devised a new approach to inferring demographic history from an estimated ARG, Lineage Time Inference (LiTI, Figure 2). LiTI works off measurements of the median time of each coalescent across the whole genome. The “time to most recent x alleles” (TMRxA) is defined as the median time, weighted by the length of each segment, at which $x+1$ lineages coalesce to x lineages (Figures 2A). These data-derived times are used to delimit the boundaries of LiTI estimation windows. Effective population size (N_e) within each of these windows is estimated using the distances between adjacent TMRxA (Figure 2B). In addition to estimating N_e , LiTI also estimates the minimum population (N_{min}) that could be consistent with the data for a short period of time within the larger time period of each window.

When applied to simulated ARGs, which are free of estimation error, we find that LiTI nearly perfectly predicts N_e (Figures 2C and S1), although the power declines for more ancient bottlenecks (Figure S2). This test demonstrates that LiTI, in principle, is capable of inferring short and brief bottlenecks in noiseless data.

Using current methods, the coalescent dates of estimated ARGs are not determined

with high precision, and real-world performance will be reduced (Figure 2D). LiTI-based estimates of N_e were computed from dated ARGs estimated by three different algorithms, *ARGweaver*, *Relate*, and *tsdate*. *ARGweaver* is the oldest of the three methods and is computationally costly. *Relate* and *tsdate* are far more efficient. Each of the three methods has slightly different biases to their estimates.

Figure 2. Data-derived windows for Lineage Time Inference (LiTI) of population size enable detection of brief bottlenecks. The data in an ARG is not equally distributed in log-space, and can be used to define better estimation windows. **(A)** Each phylogenetic tree in a tree sequence encodes several coalescence times. **(B)** Across the whole genome there are many trees. The TMRxA is defined as the median time at which the trees across the genome coalesce to x lineages. Note that this definition follows the opposite number convention as commonly used, with the time of the most ancient coalescent designated TMR1A. **(C)** The width of each window is directly related to N_e by a simple formula, which both enables population inference and ensures increased resolution where coalescence is highest. LiTI also estimates N_{min} , the minimum population size that could be consistent with the TMRxA times. Applied to the simulation ARG, which is free of estimation error, LiTI accurately detects a two individual bottleneck at 75 kya. **(D)** In practice, however, LiTI's performance depends on the accuracy of the inferred ARG. Here, the plots compare the inferred and true "critical" TMRxA from simulations with bottlenecks at several different time points. The "critical" TMRxA fixes x at the maximum number of lineages possible at the bottleneck, i.e. two times the bottleneck size. Three different approaches to inferring ARGs (*ARGweaver*, *Relate*, *tsdate*, from left to right) all exhibit estimation error in inferring the TMRxA. *ARGweaver*'s estimates do not usually overestimate TMRxA, which means its N_{min} estimates have higher confidence.

Long bottlenecks are visible in N_e inferences

As a positive control, we first studied bottlenecks of 1,000 individuals which lasted for 1,000 or 2,000 generations, down from a stable population size of 10,000 (Figure S3). These bottlenecks are brief, and are moderately sharp. Surprisingly, *msmc2* does not detect these changes in demography 75,000 years ago (Figure 3A). *Relate* and two LiTI variants, however, do detect these events, with varying degrees of accuracy and precision (Figure 3B, 3C, and 3D). In general, *Relate* estimates are the best localized, however, these estimates have very high variance inferences at ancient time points where data is sparse or absent. The LiTI variants using *ARGweaver* and *tsdate* are less accurate in

localizing the bottleneck in this test, but make much higher precision predictions at ancient time points, rightly truncating predictions where data is absent.

Figure 3. Bottlenecks of long duration are detectable by existing methods, and LiTI increases the precision of these estimates. In these experiments we considered a 1,000 individual bottleneck lasting 1,000 generations (left) and 2,000 generations (right) at 75 kya, down from a stable population size of 10,000. In these graphs, twenty replicates are plotted alongside a sliding window average. **(A)** *MSMC2* bases its inference on just eight haplotypes, but does not detect either bottleneck. **(B)** *Relate* detects both bottlenecks, and reasonably captures the sharpness of each bottleneck too. However, the variance of estimates more ancient than the bottleneck is very high. **(C)** *ARGweaver* LiTI detects the bottlenecks, but underestimates the sharpness of the bottleneck. This may reflect its use of a fixed prior, which assumes constant population size. **(D)** *tsdate* LiTI detects both bottlenecks, but underestimates the sharpness of each bottleneck. Similar to *ARGweaver*, this may be due to *tsdate*'s prior. The LiTI estimation windows, however, increase in log space at ancient time points, which pools data more efficiently, dramatically reducing the variance of estimates. LiTI also truncates estimates where data is insufficient.

Brief bottlenecks are beyond the N_e streetlight

No method we tested accurately detected bottlenecks that last just a few generations. We considered bottlenecks of 2 and 8 individuals for one generation, followed by exponential growth (Figure 4). The simulation population size was not instantaneously increased to 10,000, so these events extended over several centuries. When simulated 75,000 years ago, nonetheless, these bottlenecks were invisible to *MSMC2* (Figure 4A). In contrast, *Relate* and LiTI estimates from *ARGweaver* and *tsdate* ARGs show a gentle dip that is orders of magnitude too long and thousands of individuals larger than the true bottleneck (Figure 4B, 4C, and 4D). Similar results were observed at other bottleneck sizes and times (Figures S4, S5, S6, S7, S8, S9, and S10); as the time point moves to more ancient than 250 kya, even a bottleneck down to two individuals is invisible to all methods.

The true history of the human lineage is far more complex than the idealized simulations on which this study is based. The simulations did not model population

structure and admixture, nor did they model selection in the genome. All these complexities deviate from the assumptions of the inference algorithms, and would be expected to reduce the sensitivity of detection even further.

Figure 4. Sharp and tight bottlenecks are essentially undetectable in practice. In these experiments, a brief bottleneck down to two (left) or eight (right) individuals is simulated at 75 kya. **(A)** *MSMC2* does not detect either bottleneck. **(B)** *Relate* estimates show a shallow dip for the two individual bottleneck, but cannot detect the eight individual bottlenecks. Similar results are observed for *ARGweaver* **(C)** and *tsdate* **(D)**. Once again, LiTI estimates use data-derived estimation windows, which result in much greater precision than do *Relate* estimates.

Improving *Relate* demography inference with LiTI

We also investigated two ways of improving *Relate* inferences with LiTI. First, LiTI estimates were computed from the ARG that *Relate* infers (Figure 5A). Second, LiTI windows computed from the ARG were used as the aggregation windows for *Relate*'s own estimates of population size (Figure 5B). Both approaches improve on *Relate* (1) by increasing the time-resolution of inferences where there is sufficient data, (2) by truncating inferences at ancient time-points where there is not sufficient data, and (3) by reducing the variance of inferences at ancient time points by pooling data into larger windows. Though beyond our scope, further improvements may be realized by integrating LiTI directly into the iterative refinement loop of *Relate*.

Figure 5. LiTI improves upon *Relate* inferences of population size. Unlike *tsdate* and *ARGweaver*, *Relate* does not use a prior that favors fixed population sizes in the past. Instead, *Relate* iteratively refines an ARG and its inference of the population history, alternating between using the ARG to refine an estimate of population history and using this inferred population history as a prior to refine the ARG. **(A)** LiTI on the final *Relate* ARG estimates population size with greater precision (compare with Figure 3B and 4B). Bottlenecks of a few generations are identifiable, but brief bottlenecks are not detected. **(B)** More accurate estimates are derived by re-windowing the *Relate* estimates to match LiTI windows. Integrating LiTI estimates of population history into the *Relate* refinement loop are beyond our scope, but may enable higher resolution and higher accuracy estimates.

Ruling out some bottlenecks with N_{min}

Detecting brief bottlenecks is beyond current methods, but particularly sharp bottlenecks can be ruled out at recent timepoints. LiTI estimates N_{min} , the minimum number of individuals required by the data at particular time points. This estimate arises from the fact that each individual can hold at most two allele lineages. For example, there must be at least five individuals when there are nine or ten uncoalesced alleles. We find that the N_{min} correctly bounds the minimum population size for an eight individual bottleneck at 75 kya for all three methods (Figure 6.) For a two individual bottleneck, the N_{min} is overestimated by just one to three individuals.

Guiding interpretation of LiTI estimates, there are several systematic limitations exposed by testing across different timepoints. Diploid populations can never yield a true TMRx A time more ancient than the bottleneck when x is two times the bottleneck size. ARG inference algorithms, however, systematically overestimated the coalescence times in simulations with recent bottlenecks (Figure 7A), though *ARGweaver* is least affected. Consequently, N_{min} values are overestimated by a few individuals at recent time points (Figure 7B). Though bottlenecks could be detected at the most recent time point, 10 kya, all methods overestimate the population size of the bottleneck by orders of magnitude (Figure 7C). LiTI computed on the simulation ARG itself, free of estimation error, is not subject to the same patterns. Consequently, improvements in dating and inferring ARGs may yield improvements in LiTI estimates. Development of these methods might benefit from using these simulations as benchmarks to improve upon.

LiTI on human genome data

In human data, the average of three LiTI estimates places the N_{min} of 2 individuals at 358 kya, with TMR3A and TMR4A at 422 kya and 358 kya, respectively (Figures 8 and

Figure 6. LiTI computes N_{min} , which can rule out the sharpest bottlenecks at recent time points. Each diploid individual can pass on a maximum of two haplotypes. Hence, the number of lineages active at particular time points puts a hard minimum on the population size at any point in a LiTI estimation window. In our simulations, N_{min} estimates are only rarely in conflict with the true population history. For the two individual bottlenecks, N_{min} is too aggressive in ruling out bottlenecks, but not by much. Notably, tight bottlenecks of tens or dozens of individuals can only be ruled out at very recent timepoints. These plots show simulations with bottlenecks of two (left) or eight (right) individuals at 75 kya. **(A)** *ARGweaver* LiTI inference of N_{min} rules out less than three individuals at the two individual bottleneck, and four individuals at the eight individual bottleneck. **(B)** *tsdate* LiTI inference of N_{min} rules out less than five and less than six individuals at the bottlenecks, respectively. **(C)** *Relate* LiTI inference of N_{min} rules out less than five individuals at both the two and eight individual bottlenecks.

S11, Table T1). LiTI estimates were computed across three human genome datasets: the 1000 Genomes Project and Simons Genetic Diversity Project compiled using *tsinfer*, [11] and *ARGweaver*-inferred “69 Genomes” data from Complete Genomics (CG) dataset. [15] This indicates that even a single couple bottleneck more ancient than 450 thousand years ago cannot be ruled out by estimates of N_e alone.

In recent human history, less severe bottlenecks cannot be ruled out either. At the time of the Youngest Toba Eruption, for example, we find an N_{min} range of just 6 to 16 individuals. Estimates of N_e from simulations in ideal conditions for detection, also, cannot detect a sharp bottleneck at this time. Hence, a global bottleneck of a few thousand, as has been proposed by others, is not excluded by the genetic data. Careful consideration of other lines of evidence is beyond our scope, but critical in making any inferences about human population demography at this time. The data in this analysis merely indicates the decrease in global population size hypothesized by others, if lasting only a few generations, would likely be outside the genetic streetlight.

Looking at more ancient time points, tight population bottlenecks and founding events could have been a mechanism of speciation at the origin of the *Homo* genus, *Homo sapiens*, or other points in the evolution of the human lineage. [6, 19] If these events were just a few generations long, in most cases they also would be outside the genetic streetlight.

Figure 7. Systematic errors in estimated ARGs guide interpretation of LiTI estimates. These plots summarize estimated ARGs across simulations with bottlenecks at different timepoints in the past, comparing them to the simulation ARG, which has no estimation error. **(A)** For very recent bottlenecks, the critical TMRxA is overestimated. *ARGweaver* is least susceptible to this bias. **(B)** Consistent with the bias in estimating critical TMRxA, at recent timepoints, N_{min} incorrectly rules out bottlenecks that are in fact compatible with the data. This effect is more pronounced with the two individual bottlenecks, and becomes less and less important as the bottleneck population is increased. **(C)** The more recent a bottleneck, the more likely it is to be detected, but the depth of the bottleneck is always underestimated.

Figure 8. LiTI analysis of human genome data does not rule out a bottleneck at 75 kya, locating the Youngest Toba Eruption outside the genetic streetlight. To limit computational costs, LiTI estimates were limited to genetic data from Chromosome 1. (A) *ARGweaver* is computed across the “69 Genomes” data from Complete Genomics (CG). *tsdate* (B) is computed across the 1000 Genomes Project (1000GP) and the Simons Genome Diversity Project (SGDP) and *Relate* (C) is computed across only the SGDP. Results for both N_e and N_{min} are broadly consistent across all approaches.

Discussion

These results demonstrate the critical importance of hypothesis testing in interpreting the results of population demography inference. Inferred population histories can be misunderstood as data-derived estimates of the past demography. We know, however, that the true history of populations is usually more complex than such simplified models. Inferred population size estimates are better understood, instead, as a measurement computed from genetic data, a measurement against which multiple alternative hypotheses should be simulated and tested. Noting the low time-resolution of these inferences, we may find that multiple distinct hypotheses are consistent with the data.

Brief bottlenecks of just a few generations can be important evolutionary events. Perhaps many species have passed through brief bottlenecks in founder events. [12] For this reason, the determining if there were bottlenecks in the human lineage is an important question in human evolution. The inferences of brief bottlenecks from extant genetic data, however, may not be capable of detecting brief bottlenecks.

This limitation can guide development of improved demography inference algorithms. LiTI is an improved approach that uses data-derived windows to ensure predictions are high resolution where data is sufficient, and that no predictions are made where no data is not sufficient. LiTI, in principle, can detect brief bottlenecks. In practice, however, LiTI is limited by error in estimating ARGs from extant data. Improved inference of ARGs may improve our confidence in ruling out bottlenecks of just a few individuals more recent than 100 kya.

In principle, trans-species variation, the shared variation between extant humans and apes, can rule out the sharpest bottlenecks of just a few individuals. [3] One report indicates about 20 HLA lineages may be shared between extant humans and great apes, which would establish a minimum population size of ten individuals in our lineage over the last six million years. [4] Later studies with better data and advanced methods, however, found far fewer than 20 distinct lineages with trans-species variation. [5] Moreover, HLA sequences show evidence of convergent evolution, which limits confidence of any inference of ancient shared history. [13]

Population bottlenecks need not be as sharp as just a few individuals to be important in explaining and understanding the evolutionary trajectory of a population.

Therefore, even with algorithmic improvements and more genetic data, we may not ever be able to rule out brief bottlenecks of tens or hundreds of individuals.

Conclusion

Genomes remain a powerful tool for elucidating the history of our species. The best methods of inferring population demography in the past from genetic data, however, cannot detect brief bottlenecks, even when they are very sharp. Improved methods can compute a minimum population size that could possibly be consistent with the data, but these estimates are quite low. For this reason, these estimates do not rule out sharp reductions in population size that could have taken place in human history and be important for understanding human evolution. Several important questions about human evolution, therefore, may remain beyond the genetic streetlight.

Methods

Simulating bottlenecks. We used a hybrid simulation that merged two backwards simulations and one forward simulation through the bottleneck. [1] The bottleneck and growth were simulated in a forward Wright-Fisher simulation, and the periods of stable population before and after the bottleneck were simulated by coalescence simulations.

We simulated sharp bottlenecks of 2, 4, and 8 individuals at 10, 75, 100, 150, 250, 500, 700, 1,000, 1,500, and 2,000 kya, with 20 replicates. From each replicate, 100 haplotype samples were drawn at random and used for demography inference. Longer bottlenecks were simulated using the same hybrid protocol, simulating an initial population of 10,000 individuals that drops to 1,000 individuals for 1,000 or 2,000 generations, before doubling every generation until reaching 10,000 individuals again.

We used *msprime* version 0.7.4 to backwards simulate a population of 10,000 individuals. Each replicate used a genome size of 2.2 million base pairs and a generation time of 25 years, matching parameters from Rasmussen et al. [15] Consistent with known rates in human populations, both forward and backward simulations used a mutation rate of 1.25×10^{-8} mutations per base pair per generation and a recombination rate of 6.25×10^{-9} per base pair per generation.

Using the mutation and recombination parameters, the forward simulation was modeled with *SLiM* version 3.5, using an instantaneous drop from a population of 10,000 individuals to a population of two, four, or eight individuals. This was immediately followed by rapid growth, with the population size doubling every generation until it reached the pre-bottleneck population size. This growth rate equates to a 2.8% growth rate per year with a generation length of 25 years. Under these parameters, a bottleneck of two individuals returns to 10,000 individuals in 14 generations, or about 350 years.

The forward and backwards simulations both recorded a tree-sequence data structure. These tree sequences were joined and pruned using the *pyslim* package version 0.501 to yield a final ARG for each simulation, from which 100 haplotype samples were drawn. [1]

Inferring Population Size. We used *MSMC2* version 2.1.1 to infer effective population size from haplotypes sampled from the simulation ARGs. *MSMC2* is not capable of analyzing 100 haplotypes at once, so eight samples were selected at random

for inference. *MSMC2* parameters were set to use a mutation to recombination ratio of 2:1, matching that of the simulation.

We used *Relate* version 1.1.6 to infer the effective population size from the simulated genomic data of each of the 20 runs. We used 100 random haplotype samples from each tree sequence. We ran *Relate* with a N_e of 10,000, and with mutation and recombination rates matching the simulation, following guidance in *Relate*'s documentation.

Estimating Ancestral Recombination Graphs. We employed three ARG inference methods to estimate ARGs: *ARGweaver*, *Relate*, and *tsdate*. ARGs were inferred from 100 sampled haplotypes from each simulation. Accuracy of ARG inference plateaus before 100 haplotypes, so there is little to gain by adding additional samples (**Figure S12**). All algorithms were run using mutation and recombination parameters that matched the simulations. We ran *ARGweaver* for 100 time steps with a maximum time of 120,000 generations or 3,000 kya.

Estimating TMRxA from Genome Data. LiTI inference requires us to estimate the median time of each coalescent across the genome, the Time to the Most Recent x Alleles (TMRxA), from an inferred ARG. *ARGweaver*, *Relate*, and *tsdate* all use discrete times to infer date coalescence, so the raw distribution of coalescent times is a discrete distribution. We computed this discrete distribution by weighting each coalescence time by the length of the genome the tree it was derived from covered. This weighting places higher emphasis on trees supported by more data. We estimated a smooth continuous distribution by adding normally distributed noise to each coalescence estimate in logarithmic space, with variance scaled to the size of the discretization window. The TMRxA was calculated as the median value of this continuous distribution. This process was repeated for each lineage time independently, from TMR1A, TMR2A, and so on, up to TMR100A (**Figure S13**). Due to uncertainty in estimating the median, these estimates can be nearly equal, but not strictly decreasing. We know, however, that the coalescence times must be strictly decreasing. To reduce the effect of this noise, the estimated TMRxA times were sorted and reassigned so as to ensure that they are strictly decreasing.

LiTI population demography estimates from TMRxA. The median TMRxA times defined windows in the past over which population estimates were computed. For each LiTI window, the estimated population size was estimated as,

$$N_e = \Delta T \cdot \frac{k(k-1)}{4}. \quad (1)$$

Where ΔT is the estimation window size in generations and k is the extant lineages present across that window. For example, in the LiTI window between TMR4A and TMR3A, the number of active lineages k is 4.

This estimator is derived by rearranging the formula $A \cdot K(k, N_e) \cdot \Delta T = C$. Here, A is the fraction of the genome with k lineages active across the genome in the estimation window. K , the rate at which k lineages coalesce to $k-1$ lineages, depends on N_e , and is given by the Kingman coalescent. ΔT is the time-width in generations of the estimation window. C is the amount of the genome that coalesces from k to $k-1$ lineages over the estimation window.

This formula simplifies to $K\Delta T = 1$ because A approximately equals C in LiTI estimation windows. The fraction of the genome that coalesces from k to $k-1$ lineages across the estimation window (i.e. C) exactly equals $D_{k-1}(t_a) - D_{k-1}(t_r)$, where t_r and t_a are, respectively, the time of the recent and ancient boundaries of the window and

D_k is the cumulative distribution of the k th TMRxA. The fraction of the genome with exactly k active lineages at time t_r is $D_k(t_r) - D_{k-1}(t_r)$. Near the median, this quantity is approximately equal to the average fraction of the genome with exactly k lineages across the whole estimation window (i.e. A). The estimation window is defined by the median coalescent times, so both $D_{k-1}(t_a)$ and $D_k(t_r)$ are 0.5, hence A is approximately equal to C . A more accurate estimate might use the empirically computed average across the full window.

There must be at least two individuals in a sexually reproducing population, and a population of k individuals has at most $2k$ haplotype lineages. For each LiTI window, we estimated the minimum possible population size over time as,

$$N_{min} = \max \left(\left\lceil \frac{k}{2} \right\rceil, 2 \right). \quad (2)$$

These estimates are easily implemented and work well in practice. We expect, however, that better estimates of both N_{min} and N_e are possible. For example, a refined estimate of N_{min} would take the estimation error of the TMRxA times into account. Further improvements might integrate other information, such as trans-species variation and population structure. Likewise, LiTI windows can be defined in ways which might increase the resolution of population estimates. Perhaps most important, portions of the genome that deviate from neutral evolution could be filtered out, reducing the error introduced by positive and negative selection on the genome. These refinements are promising, but they are beyond the scope of this study.

LiTI in Human Genome Data. For *ARGweaver*, we used dated tree sequences of the “69 Genomes” data from Complete Genomics (CG). We parsed the dated tree sequences computed by the *ARGweaver* team to determine the coalescence times for use with LiTI. For *tsdate*, we used undated sequence trees for the 1000 Genomes Project (1000GP) and the Simons Genome Diversity Project (SGDP), made available by the *tsinfer* team. These trees include 5,008 and 554 haplotypes for 1000GP and SGDP, respectively, and *tsdate* operated on these trees to infer dates for each coalescence. For *Relate*, we extracted the haplotypes from the *tsinfer* tree sequence of only the SGDP, as running *Relate* on 1000GP proved prohibitively computationally expensive, and ran *Relate* directly on these haplotypes. The ARG inferred by *Relate* was used to compute LiTI estimates. To limit computational costs, LiTI estimates were limited to genetic data from Chromosome 1.

References

1. pyslim, 2020.
2. S. H. Ambrose. Late Pleistocene human population bottlenecks, volcanic winter, and differentiation of modern humans. *Journal of Human Evolution*, 34(6):623–651, June 1998.
3. F. J. Ayala, A. Escalante, C. O’Hugin, and J. Klein. Molecular genetics of speciation and human origins. *Proceedings of the National Academy of Sciences of the United States of America*, 91(15):6787–6794, 1994.
4. F. J. Ayala and A. A. Escalante. The evolution of human populations: A molecular perspective. *Molecular Phylogenetics and Evolution*, 5(1):188–201, 1996.

-
5. T. F. Bergström, A. Josefsson, H. A. Erlich, and U. Gyllensten. Recent origin of HLA-DRB1 alleles and implications for human evolution. *Nature Genetics*, 18(3):237–242, 1998.
 6. W. M. Brown. Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proc. Natl. Acad. Sci. USA*, 77(6):3605–3609, 1980.
 7. F. Gathorne-Hardy and W. Harcourt-Smith. The super-eruption of Toba, did it cause a human bottleneck? *Journal of Human Evolution*, 45(3):227–230, Sept. 2003.
 8. B. C. Haller, J. Galloway, J. Kelleher, P. W. Messer, and P. L. Ralph. Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. *Molecular Ecology Resources*, 19(2):552–566, 2019.
 9. B. C. Haller and P. W. Messer. SLiM 3: Forward Genetic Simulations Beyond the Wright-Fisher Model. *Molecular Biology and Evolution*, 36(3):632–637, 2019.
 10. J. Kelleher, A. M. Etheridge, and G. McVean. Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. *PLoS Computational Biology*, 12(5):1–22, 2016.
 11. J. Kelleher, Y. Wong, A. W. Wohns, C. Fadil, P. K. Albers, and G. McVean. Inferring whole-genome histories in large population datasets. *Nature Genetics*, 51(9):1330–1338, Sept. 2019.
 12. E. Mayr. *Animal Species and Evolution*. Harvard University Press, 1963.
 13. A. R. Vahdati and A. Wagner. Parallel or convergent evolution in human population genomic data revealed by genotype networks. *BMC Evolutionary Biology*, 16(1):1–19, 2016. Publisher: BMC Evolutionary Biology.
 14. M. R. Rampino and S. H. Ambrose. Volcanic winter in the Garden of Eden: The Toba supereruption and the late Pleistocene human population crash. In *Volcanic Hazards and Disasters in Human Antiquity*. Geological Society of America, 2000.
 15. M. D. Rasmussen, M. J. Hubisz, I. Gronau, and A. Siepel. Genome-Wide Inference of Ancestral Recombination Graphs. *PLoS Genetics*, 10(5), 2014.
 16. S. Schiffels and R. Durbin. Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, 46(8):919–925, 2014. Publisher: Nature Publishing Group.
 17. E. I. Smith, Z. Jacobs, R. Johnsen, M. Ren, E. C. Fisher, S. Oestmo, J. Wilkins, J. A. Harris, P. Karkanias, S. Fitch, A. Ciravolo, D. Keenan, N. Clegghorn, C. S. Lane, T. Matthews, and C. W. Marean. Humans thrived in South Africa through the Toba eruption about 74,000 years ago. *Nature*, 555(7697):511–515, Mar. 2018.
 18. L. Speidel, M. Forest, S. Shi, and S. R. Myers. A method for genome-wide genealogy estimation for thousands of samples. *Nature Genetics*, 51(9):1321–1329, Sept. 2019.
-

-
19. A. C. Wilson, R. L. Cann, S. M. Carr, M. George, U. B. Gyllensten, K. M. Helm-Bychowski, R. G. Higuchi, S. R. Palumbi, E. M. Prager, R. D. Sage, and M. Stoneking. Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society*, 26(January):375–400, 1985.
 20. A. W. Wohns, Y. Wong, B. Jeffery, A. Akbari, S. Mallick, R. Pinhasi, N. Patterson, D. Reich, J. Kelleher, and G. McVean. A unified genealogy of modern and ancient genomes. preprint, Genomics, Feb. 2021.
 21. C. L. Yost, L. J. Jackson, J. R. Stone, and A. S. Cohen. Subdecadal phytolith and charcoal records from Lake Malawi, East Africa imply minimal effects on human evolution from the 74 ka Toba supereruption. *Journal of Human Evolution*, 116:75–94, Mar. 2018.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplement.pdf](#)