Supplemental Legends

Figure S1 True pairwise coalescence time from msprime simulations compared to inferred coalescence time from (A) ARGweaver (B) Relate (C) tsdate. Note that axes are in linear scale. See Figure 3A, D, G for this data plotted on a logarithmic scale. These results are for simulations with n=8 samples (haplotypes), mutation and recombination rates of 2×10^{-8}. Diagonal line shows x=y, points show the mean inferred coalescence time within a true coalescence time bin.

Figure S2 Mean (A,B) and mean squared error (C,D) of point estimates of pairwise coalescence times by ARGweaver, Relate and tsdate in each bin of size 0.1 of simulated coalescence times. Diagonal gray line in plots A and B show 1:1 line. These results are for simulations with n=8 samples, mutation and recombination rates of 2×10^{-8}. Plots B and D are in log scale to highlight small values of coalescence times, which are the most abundant. Note that estimates are best (i.e. means in plots a and b are closer to the simulated value) at values near the expected mean coalescence time under the coalescent (i.e. 1 in the coalescent units of 2Ne generations).

Figure S3 Histogram of the distribution of coalescence times in msprime simulations. Red line show expected exponential distribution with rate 1.

Figure S4 Distributions of pairwise coalescence times in Relate and tsdate without ARGweaver time discretization. These results are for simulations with n=8 samples, mutation and recombination rates of 2×10^{-8}. (a) Relate, (b) tsdate, both with 20 equal size bins.

Figure S5 Point estimates (A-C), distribution of coalescence times (D-F) and counts of ranks from simulation-based calibration (G,H) from ARGweaver (A,D,C), Relate (B,E,H) and tsinfer+tsdate (C,F). Simulations with reduced mutation rate ($\mu = 2 \times 10^{-9}$ and $\rho = 2 \times 10^{-8}$). Compared to simulations with mutation rate equal to recombination rate, mean square error (MSE) values are all larger (Figure 3), distributions of coalescence times deviate more from the theoretical expectation (Figure 4), and KLD is lower in ARGweaver but higher in Relate (Figure 5).

Figure S6 Point estimates (A-C), distribution of coalescence times (D-F) and counts of ranks from simulation-based calibration (G,H) from ARGweaver (A,D,C), Relate (B,E,H) and tsinfer+tsdate (C,F). Simulations with increased recombination rate ($\mu = 2 \times 10^{-8}$ and $\rho = 2 \times 10^{-7}$). Compared to simulations with mutation rate equal to recombination rate, Mean square error (MSE) values are all larger (Figure 3), distributions of coalescence times deviate more from the theoretical expectation (Figure 4), and KLD is lower in ARGweaver, but higher in Relate (Figure 5).

Figure S7 Point estimates (A-C), distribution of coalescence times (D-F) and counts of ranks from simulation-based calibration (G,H) from ARGweaver (A,D,C), Relate (B,E,H) and tsinfer+tsdate (C,F). Simulations with sample size of 8 haplotypes, $\mu = \rho = 2 \times 10^{-8}$, and input sequence length of 5Mb.

Figure S8 Point estimates (A-C), distribution of coalescence times (D-F) and counts of ranks from simulation-based calibration (G,H) from ARGweaver (A,D,C), Relate (B,E,H) and tsinfer+tsdate (C,F). Simulations with sample size of 8 haplotypes, $\mu = \rho = 2 \times 10^{-8}$, and input sequence length of 250kb.

Figure S9 ARGweaver likelihood traces (top) and autocorrelation between consecutive MCMC iterations (bottom, also showing effective sample sizes (N_{eff})) for the number of iterations used in the main text. Left column: simulations with 8 haplotypes, mutation rate equal to the recombination rate(2×10^{-8}). Potential scale reduction factor (PSRF) is 1.02, upper confidence interval (CI) is 1.05. Middle

column: simulations with recombination rate decreased to 2×10^{-9}. PSRF is 1.04, upper CI is 1.11. For both of these simulated datasets we used a burn in of 200 iterations (indicated by vertical line) and ran them for 1200 iterations in total, sampling every 10th iteration. Right column: simulations with mutation rate increased to 2×10^{-7}. PSRF is 1.01, upper CI is 1.02. For this dataset we used a burn in of 1200 iterations (indicated by vertical line) and ran them for 2200 iterations in total, sampling every 10^{th} iteration.

Figure S10 Similar to Figure S9, but running ARGweaver for 10 thousand iterations, with a burn in of 9 thousand applied before calculating effective sample sizes, to keep the same number of samples (1000). ARGweaver likelihood traces (A,B,C) and autocorrelation between consecutive MCMC iterations (D,E,F). Left column: simulations with 8 haplotypes, mutation rate equal to the recombination rate (2×10^{-8}). Middle column: simulations with recombination rate decreased to 2×10^{-9}. Right column: simulations with mutation rate increased to 2×10^{-7}.

Figure S11 ARGweaver likelihood traces (top) and autocorrelation between consecutive MCMC iterations (bottom). A,D: simulations with 4 haplotypes, mutation rate equal to recombination rate (2×10^{-8}). For this simulated dataset we used a burn in of 200 iterations (indicated by vertical line) and ran them for 1200 iterations in total, sampling every 10th iteration. B,E: simulations with 16 haplotypes. C,F: simulations with 32 haplotypes. For both of these datasets we used a burn in of 1200 iterations (indicated by vertical line) and ran them for 2200 iterations in total, sampling every 10th iteration.

Figure S12 Coalescence times for one pair of samples inferred by 5 independent runs of ARGweaver at 10 sites equally spaced sites along the 5Mb sequence. Simulations with 8 samples and mutation rate equal to recombination rate.

Figure S13 Coalescence times for one pair of samples inferred by 5 independent runs of Relate at 10 sites equally spaced sites along the 5Mb sequence. Simulations with 8 samples and mutation rate equal to recombination rate.

Figure S14 Tsdate results with a prior grid constructed with timepoints=100. (A) Comparisons of estimated and simulated point estimates of pairwise coalescence times. (B) Comparisons of the distribution of coalescence times to the expected exponential distribution, using ARGweaver time discretization bins. (C) Same as B, but without imposing ARGweaver time discretization.

Figure S15 Tsdate results with a prior grid constructed with a maximum value of 12. (A) Comparisons of estimated and simulated point estimates of pairwise coalescence times. (B) Comparisons of the distribution of coalescence times to the expected exponential distribution, using ARGweaver time discretization bins. (C) Same as B, but without imposing ARGweaver time discretization.

Figure S16 Acceptance rate from ARGweaver subtree sampling steps in one 5Mb region of each simulation.

Figure S17 Distribution of coalescence times in msprime simulations using the SMC (A) or SMC' model (B). ARGweaver inference is done using the same model used in the simulations.

Figure S18 Simulation-based calibration results in msprime simulations using the SMC (A) or SMC' model (B). ARGweaver inference is done using the same model used in the simulations.

Figure S19 Evaluation of ARGweaver point estimates (A,D), distribution of coalescence times (B,E) and posterior calibration (C,F) for simulations with mutation rate to recombination rate ratio of 2 (A-C, $\mu = 4 \times 10^{-8}$, $\rho = 2 \times 10^{-8}$) and mutation rate to recombination rate ratio of 4 (D-F, $\mu = 8 \times 10^{-8}$, $\rho = 2 \times 10^{-8}$)

Figure S20 Evaluation of ARGweaver point estimates (A,D), distribution of coalescence times (B,E) and posterior calibration (C,F) with simulations under the Jukes and Cantor (1969) mutational model. A-C: simulations with 8 haplotypes and $\mu = \rho = 2 \times 10^{-8}$. D-F: simulations with 8 haplotypes and $\mu = 2 \times 10^{-8}$ and $\rho = 2 \times 10^{-9}$.

Figure S21 Evaluation of ARGweaver point estimates (A,D,G), distribution of coalescence times (B,E,H) and posterior calibration (C,F,I) with simulations under the Jukes and Cantor (1969) mutational model. In all cases we simulated 8 haplotypes and used $\rho = 2 \times 10^{-8}$. A-C: $\mu = 4 \times 10^{-8}$. D-F: $\mu = 8 \times 10^{-8}$. G-I: $\mu = 2 \times 10^{-7}$.

Table S1 Potential scale reduction factor point estimates (PSRF), their upper confidence intervals (C.I.) and effective sample sizes (N_{eff}) for ARGweaver stats. μ: mutation rate, ρ: recombination rate.

Table S2 Potential scale reduction factor (PSRF) mean, variance and range for each of 200 coalescence times in ARGweaver, the multivariate PSRF (Plummer et al. 2006) and the number of coalescence times for each the effective sample size (Ne f f) is smaller than 100. Unless otherwise noted, mutation rate (μ) and recombination rate (ρ) are 2×10^{-8} and sample sizes (n) are 8 haplotypes.

Table S3 Potential scale reduction factor (PSRF) mean, variance and range for each of 200 coalescence times in Relate, the multivariate PSRF (Plummer et al. 2006) and the number of coalescence times for each the effective sample size (Ne f f) is smaller than 100. Unless otherwise noted, mutation rate (μ) and recombination rate (ρ) are 2×10^{-8} and sample sizes (n) are 8 haplotypes.

Table S4 Minimum and maximum acceptance rates of ARGweaver subtree sampling steps for each simulation.

Table S5 Comparison of ARGweaver results with simulations under infinite sites mutational model and Jukes-Cantor finite sites mutational model, including simulations with values of mutation to recombination rate ratio in between the ones shown in the main text. * indicate results shown in the main text and presented here again for comparison.