File S1

Supporting Information

Coalescence time distributions at recombination sites

Here, we derive the joint, marginal (i.e., one-locus), and conditional distributions of coalescence times at recombination sites where the coalescence time changes under the ARG, SMC', and SMC. The distributions related to the ARG and SMC' are derived from analysis of the continuous-time Markov chains representing coalescence times at such recombination sites under these models (Figure 7). Under the ARG and SMC', the joint density function of coalescence times at recombination sites that change the coalescence time (i.e., the joint density of S and T) is

$$f_{S,T}(s,t) = \begin{cases} \frac{3}{4}\left(1-e^{-2s}\right)e^{-t} & s < t \\ \frac{3}{4}\left(1-e^{-2t}\right)e^{-s} & s > t, \end{cases} \tag{S1}$$

and the marginal density function of S (or T) is

$$\pi(s) = \frac{3}{8}e^{-s}\left(2s+1-e^{-2s}\right). \tag{S2}$$

The conditional distribution of T given S is

$$f_{T|S}(t|s) = \frac{f_{S,T}(s,t)}{\pi(s)} = \begin{cases} \frac{2\left(1-e^{-2t}\right)}{1-e^{-2s}+2s} & t < s \\ \frac{2e^{-(t-s)}\left(1-e^{-2s}\right)}{1-e^{-2s}+2s} & t > s. \end{cases} \tag{S3}$$

Equations (S1), (S2), and (S3) hold marginally at recombination sites where the coalescence time changes under both the ARG and SMC'. Equations (S2) and (S3) were derived for the SMC' by Carmi *et al.* (2014, see eqns. (8) and (9), respectively), confirming our derivation.

Under the SMC the process for generating coalescence times at recombination sites is equivalent to the continuous-time Markov chain in Figure 7B with the transition rates from $\boldsymbol{R_1}$ to $\boldsymbol{C_L}$ and $\boldsymbol{C_R}$ equal to 1 instead of 3/2. Under this model for the SMC, the joint density of coalescence times on either side of a recombination event is

$$f_{S,T}(s,t) = \begin{cases} e^{-t}(1-e^{-s}) & s < t \\ e^{-s}(1-e^{-t}) & s > t \end{cases} \tag{S4}$$

and the marginal density of S (or T) is

$$\pi(s) = se^{-s}. \tag{S5}$$

The conditional distribution of T given S under the SMC is

$$f_{T|S}(t|s) = \frac{f_{S,T}(s,t)}{\pi(s)} = \begin{cases} \frac{1-e^{-t}}{s} & t < s \\ \frac{e^{-(t-s)}(1-e^{-s})}{s} & t > s, \end{cases} \tag{S6}$$

which confirms the derivation of Li and Durbin (2011, cf. their Eq. (S6)).

Pairwise ARG is ergodic

Here we show that the pairwise ARG is sequentially ergodic. Let $\{t(x)\}_{x\geq 0}$ represent the random pairwise coalescence time at point x along two aligned, continuous, infinitely-long chromosomes modeled by the ARG. Let time be scaled such that the marginal distribution of $t(x)$ is exponential with rate 1 for all $x \geq 0$, and thus $\mathrm{E}[t(x)] = 1$. Let the distance across the chromosome be measured such that a segment of length l

recombines apart back in time at rate $l/2$. (Equivalently, a recombination event happens in the chromosome interval $(x, x+dx)$ in the time interval $(t, t+dt)$ with infinitesimal probability $dx\,dt$.)

One useful property of $t(x)$ is that it is strongly stationary. That is, the joint distribution of $\{t(x)\}_{a\leq x\leq b}$ is the same as the joint distribution of $\{t(x)\}_{a+h\leq x\leq b+h}$ for all $0 \leq a < b$ and $h > 0$. To see that this is the case, consider the Wiuf and Hein (1999) algorithm for constructing an ARG sequentially across the chromosome: at a given point, a genealogy is drawn from the marginal distribution of genealogies, and then the algorithm proceeds along the chromosome generating recombination events and genealogies, where at each point along the chromosome, such events are drawn from the conditional distribution given all previous coalescence and recombination events. The initial point from which the marginal genealogy is drawn has no effect on the resulting joint distribution of genealogies.

A stationary process $t(x)$ is ergodic if the covariance function $r(x)$ converges to zero as x goes to infinity (Karlin and Taylor, 1975). Under the ARG, the covariance function is

$$r(x) = \frac{x+18}{x^2+13x+18}, \tag{S7}$$

which satisfies this condition. Thus the pairwise ARG is sequentially ergodic: the mean coalescence time across a long chromosome converges to the mean coalescence time at a single point. A similar proof could be given for the discrete-locus ARG with evenly spaced loci, which has a covariance function of the same form as the continuous-chromosome ARG.

Supplementary Figures

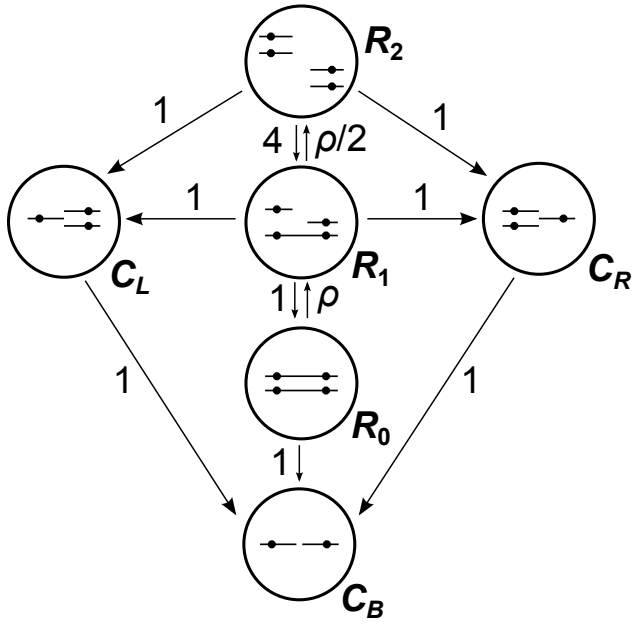

Figure S1: Schematic of the ARG back-in-time Markov process for two loci. The process starts in state $\boldsymbol{R_0}$, and transitions to other states occur with the rates indicated by arrows between states.

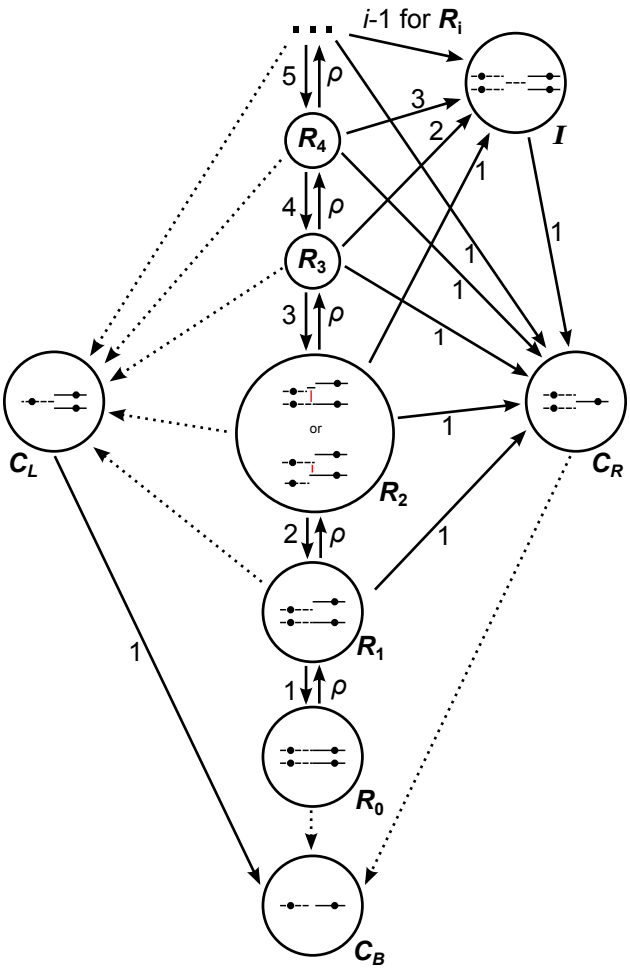

Figure S2: Back-in-time Markov process for generating a coalescence time T_2 at the right locus conditional on the time $T_1 = t_1$ at the left locus under the SMC'. Starting at time zero in state $\boldsymbol{R_0}$, the process follows the transitions indicated by the solid arrows at the rates accompanying these arrows. Transitions indicated by dotted arrows are followed instantaneously at time t_1. See Hobolth and Jensen (2014) for analogous processes for the ARG and SMC models.

References

Carmi, S., P. R. Wilton, J. Wakeley, and I. Peer, 2014 A renewal theory approach to IBD sharing. Theoretical Population Biology **97**: 35–48.

Hobolth, A., and J. L. Jensen, 2014 Markovian approximation to the finite loci coalescent with recombination along multiple sequences. Theoretical Population Biology **48**: 48–58.

Karlin, S., and H. M. Taylor, 1975 *A first course in stochastic processes*. Academic Press.

Li, H., and R. Durbin, 2011 Inference of human population history from individual whole-genome sequences. Nature **475**: 493–496.

Wiuf, C., and J. Hein, 1999 Recombination as a point process along sequences. Theoretical Population Biology **55**: 248–259.