Supplemental Material

S1: The full posterior distribution

Following the notation in the main paper's Methods section, the model parameters consist of θ, ε, λ, and $Z(x)$. We represent the transitions of IBD states $Z(x)$ along chromosomes by transition location x_k and the resulting state $z_k = Z(x_k)$ for k = 1, ..., K, where an IBD transition occurs between nucleotide sites at $x_k - 1$ and x_k for $k \geq 2$. For convenience we set $x_1 = 1$ and $x_{K+1} = \ell + 1$, where ℓ is the length of the chromosome in base pairs (bp). Let $\mathbf{x} = \{x_k\}_{k=1..K+1}$ and $\mathbf{z} = \{z_k\}_{k=1..K}$ denote the vectors of transition points and IBD partitions. Let $\mathbf{s} = \{s_i\}_{i=1..m}$ and $\boldsymbol{\pi} = \{\pi_i\}_{i=1..m}$, denote the vectors of SNP sites and their minor allele frequencies. These allele frequencies $\boldsymbol{\pi}$ and SNP locations $\mathbf{s}$ are assumed to be known. Finally, let $\mathbf{y}_i = \{y_{ij}\}_{j=1..n}$ denote the vector of observed alleles at SNP site i over gametes j, and $\mathbf{y} = \{y_{ij}\}_{i=1..m,j=1..n}$ the complete data over all gametes and sites. Note that subscript i indexes the SNP sites, j the gametes, and k the IBD transition locations.

The full posterior distribution is given by

$$p(\theta, \varepsilon, \lambda, K, \mathbf{x}, \mathbf{z}|\mathbf{y}, \boldsymbol{\pi}, \mathbf{s}) \quad \propto \quad p(\mathbf{y}|\boldsymbol{\pi}, \mathbf{s}, K, \mathbf{x}, \mathbf{z}, \varepsilon)\ p(K, \mathbf{x}, \mathbf{z}|\theta, \lambda)\ p(\theta)\ p(\lambda)\ p(\varepsilon),$$

where each term is explained as follows. First, the SNPs are assumed to be independent given the latent IBD state, so that the likelihood term is a product over SNP sites:

$$p(\mathbf{y}|\boldsymbol{\pi}, \mathbf{s}, K, \mathbf{x}, \mathbf{z}, \varepsilon) \quad = \quad \prod_{i=1}^{m} p(\mathbf{y}_i \mid Z(s_i), \pi_i, \theta, \varepsilon).$$

Additionally, since we assume independence of the allelic type of non-IBD DNA, each term in the product over SNPs is again a product over the IBD subsets of gametes. Second, the IBD process along the chromosome is modeled as a continuous time Markov process, with the prior distribution given by

$$p(K, \mathbf{x}, \mathbf{z}|\theta, \lambda) \quad = \quad p(\mathbf{z}|K, \theta)\ p(K, \mathbf{x}|\lambda).$$

The probability of the vector $\mathbf{z}$ of IBD partitions is given by

$$p(\mathbf{z} \mid K, \theta) \;=\; p(z_1 \mid \theta) \prod_{i=1}^{K-1} p(z_{i+1} \mid z_i, \theta),$$

where the distribution for the initial IBD state z_1 is given by the ESF (main paper Equation 1), and the transition probability $p(z_{i+1} \mid z_i, \theta)$ can be calculated from the modified Chinese restaurant processes (MCRP) (main paper Equations 2 and 3).

Since $\lambda \ll 1$ and thus $K \ll \ell$, the geometrically distributed discrete inter-transition base-pair counts are approximated by exponential distributions for the inter-transition distances. That is, if K is not constrained,

$$p(K, \mathbf{x} \mid \lambda) \;\propto\; (1-\lambda)^{x_{K+1}-x_K-1} \prod_{i=1}^{K-1} ((1-\lambda)^{x_{i+1}-x_i-1}\lambda) \;\approx\; \lambda^{K-1} e^{-\lambda\ell}.$$

If K is bounded by $K_c < \infty$, the distribution of K will involve a normalization constant that depends on λ. That is

$$p(K, \mathbf{x} \mid \lambda) \;\propto C(\lambda)\lambda^{K-1} e^{-\lambda\ell},$$

where $C(\lambda) = \Gamma(K_c, \lambda\ell)/\Gamma(K_c)$, and the numerator is the incomplete Gamma function:

$$\Gamma(a, b) \;= \textstyle\int_b^\infty t^{a-1} e^{-t} dt.$$

(Note $C(\lambda) = 1$ if $K_c = \infty$.) Thus $\mathbf{x}$ is uniform on the space $1 = x_1 < x_2 < .... < x_{K+1} = \ell+1$ and $\int d\mathbf{x} = \ell^{K-1}/(K-1)!$. Then

$$p(K \mid \lambda) \;=\; \int p(K, \mathbf{x} \mid \lambda) d\mathbf{x} \;=\; \begin{cases} C(\lambda)(\lambda\ell)^{K-1} e^{-\lambda\ell}/(K-1)! & \text{if } K \le K_c \\ 0, & \text{if } K > K_c \end{cases}.$$

That is, the prior distribution for $(K-1)$ is a truncated Poisson distribution with mean $\lambda\ell$.

The prior distributions for θ and λ are Gamma distributions, where that for θ is bounded below by θ_c. Thus if $G[u \mid \alpha,\ \beta] \;=\; (\Gamma(\alpha)\beta^\alpha)^{-1} u^{\alpha-1} e^{-u/\beta}$ denotes the gamma probability density on $u > 0$ with shape parameter α and scale parameter β, the prior distribution of θ is

$$p(\theta) \;\propto\; \begin{cases} G[\theta \mid \alpha_\theta, \beta_\theta] & \text{if } \theta \ge \theta_c \\ 0 & \text{if } \theta < \theta_c, \end{cases}$$

and the prior distribution of λ is

$$p(\lambda) \;=\; \Gamma[\lambda \mid \alpha_\lambda, \beta_\lambda]$$

The prior distribution of ε is the uniform distribution in $[0, 1]$:

$$p(\varepsilon) \;=\; \begin{cases} 1 & \text{if } 0 \le \varepsilon \le 1 \\ 0 & \text{otherwise.} \end{cases} .$$

In general λ must be sampled via a Metropolis algorithm, but if $K_c = \infty$ the full conditional distribution for λ is the gamma distribution $G(\lambda \mid (K + \alpha_\lambda - 1), (\beta_\lambda^{-1} + \ell)^{-1})$. In this case λ can be integrated out to obtain the posterior distribution on the other parameters:

$$p(\theta, \varepsilon, K, \mathbf{x}, \mathbf{z} \mid \mathbf{y}, \boldsymbol{\pi}, \mathbf{s}) \;\propto\; p(\mathbf{y} \mid \boldsymbol{\pi}, \mathbf{s}, K, \mathbf{x}, \mathbf{z}, \varepsilon)\; p(\mathbf{z} \mid K, \theta)\; p(K, \mathbf{x})\; p(\theta)\; p(\varepsilon), \quad \text{(S1)}$$

where

$$p(K, \mathbf{x}) \;=\; \int p(K, \mathbf{x} \mid \lambda)\; p(\lambda)\; d\lambda \;\propto\; \Gamma(K + \alpha_\lambda - 1)\; (\beta_\lambda^{-1} + \ell)^{-(K + \alpha_\lambda - 1)}.$$

S2: Possible transitions between two IBD states $\mathbf{z_A}$ and $\mathbf{z_B}$

In this section we list all the transformations between two IBD states that differ by at most two steps ($|z_A - z_B| \le 2$). In describing these transformations, we use lower case letters a, b, c and d to denote gametes, and upper case X, Y, P and Q to denote IBD subsets. The notation $\{a, X\}$ will denote the subset $\{a\} \bigcup X$. Note that any specified gamete such as a is not in any specified subset such as X. We denote the size of subset X by $|X|$, and group the transformations by the pair of sizes $(|z_A|, |z_B|)$ for the numbers of subsets in the two IBD states that are involved in the transformation. Note that the IBD subsets shared between z_A and z_B are irrelevant.

Case $|z_A - z_B| = 0$: If $z_A = z_B$ no transformation is needed.

Case $|z_A - z_B| = 1$: Recall that one step of our process can move one gamete a from a source subset S to a target set T. This move results from proposing the new gamete in set T, and then deleting a from S, and the transformation is denoted $(a : S \to T)$. In Table S1, we give both the transformation from z_A to z_B and the transformation from z_B to z_A.

Case $|z_A - z_B| = 2$: In Table S2, we denote the intermediate state by z_I, and give the transformations from z_A to z_I and from z_B to z_I.

# subsets in z_A, z_B	Subsets in z_A	Subsets in z_B	Transformation $z_A \to z_B$	Transformation $z_B \to z_A$	Condition
1, 2	$\{a, b\}$	$\{a\}, \{b\}$	$(a : \{a, b\} \to \{\})$ $(b : \{a, b\} \to \{\})$	$(a : \{a\} \to \{b\})$ $(b : \{b\} \to \{a\})$	
	$\{a, X\}$	$\{a\}, X$	$(a : \{a, X\} \to \{\})$	$(a : \{a\} \to X)$	$\lvert X\rvert \geq 2$
2, 2	$\{a, X\}, Y$	$X, \{a, Y\}$	$(a : \{a, X\} \to Y)$	$(a : \{a, Y\} \to X)$	$\lvert X\rvert \geq 1$ $\lvert Y\rvert \geq 1$

Table S1: List of transformations for $|z_A - z_B| = 1$.

# subsets in z_A, z_B	Subsets in z_A	Subsets in z_B	Transformation $z_A \to z_I$	Transformation $z_B \to z_I$	Subsets in z_I	Condition
1, 2	$\{a,b,c,d\}$	$\{a,b\},\{c,d\}$	$(a:\{a,b,c,d\}\to\{\})$	$(b:\{a,b\}\to\{c,d\})$	$\{a\},\{b,c,d\}$	
			$(b:\{a,b,c,d\}\to\{\})$	$(a:\{a,b\}\to\{c,d\})$	$\{b\},\{a,c,d\}$	
			$(c:\{a,b,c,d\}\to\{\})$	$(d:\{c,d\}\to\{a,b\})$	$\{c\},\{a,b,d\}$	
			$(d:\{a,b,c,d\}\to\{\})$	$(c:\{c,d\}\to\{a,b\})$	$\{d\},\{a,b,c\}$	
	$\{a,b,X\}$	$\{a,b\},X$	$(a:\{a,b,X\}\to\{\})$	$(b:\{a,b\}\to X)$	$\{a\},\{b,X\}$	$\lvert X\rvert \geq 3$
			$(b:\{a,b,X\}\to\{\})$	$(a:\{a,b\}\to X)$	$\{b\}.\{a,X\}$	
1, 3	$\{a,b,c\}$	$\{a\},\{b\},\{c\}$	$(a:\{a,b,c\}\to\{\})$	$(b:\{b\}\to\{c\})$	$\{b,c\},\{a\}$	
				$(c:\{c\}\to\{b\})$		
			$(b:\{a,b,c\}\to\{\})$	$(a:\{a\}\to\{c\})$	$\{a,c\},\{b\}$	
				$(c:\{c\}\to\{a\})$		
			$(c:\{a,b,c\}\to\{\})$	$(a:\{a\}\to\{b\})$	$\{a,b\},\{c\}$	
				$(b:\{b\}\to\{a\})$		
	$\{a,b,X\}$	$\{a\},\{b\},X$	$(a:\{a,b,X\}\to\{\})$	$(b:\{b\}\to X)$	$\{b,X\},\{a\}$	$\lvert X\rvert \geq 2$
			$(b:\{a,b,X\}\to\{\})$	$(a:\{a\}\to X)$	$\{a,X\},\{b\}$	
2, 2	$\{a,b,c\},\{d\}$	$\{c\},\{a,b,d\}$	$(a:\{a,b,c\}\to\{d\})$	$(b:\{a,b,d\}\to\{c\})$	$\{b,c\},\{a,d\}$	
			$(b:\{a,b,c\}\to\{d\})$	$(a:\{a,b,d\}\to\{c\})$	$\{a,c\},\{b,d\}$	
			$(c:\{a,b,c\}\to\{d\})$	$(d:\{a,b,d\}\to\{c\})$	$\{a,b\},\{c,d\}$	
			$(c:\{a,b,c\}\to\{\})$	$(d:\{a,b,d\}\to\{\})$	$\{a,b\},\{c\},\{d\}$	
			$(d:\{d\}\to\{a,b,c\})$	$(c:\{c\}\to\{a,b,d\})$	$\{a,b,c,d\}$	
	$\{a,b,X\},Y$	$X,\{a,b,Y\}$	$(a:\{a,b,X\}\to Y)$	$(b:\{a,b,Y\}\to X)$	$\{b,X\},\{a,Y\}$	$\lvert X\rvert \geq 1$
			$(b:\{a,b,X\}\to Y)$	$(a:\{a,b,Y\}\to X)$	$\{a,X\},\{b,Y\}$	$\lvert Y\rvert \geq 2$

	$\{a,b\},\{c,d\}$	$\{a,c\},\{b,d\}$	$(a:\{a,b\}\to\{c,d\})$ $(b:\{a,b\}\to\{c,d\})$ $(c:\{c,d\}\to\{a,b\})$ $(d:\{c,d\}\to\{a,b\})$	$(d:\{b,d\}\to\{a,c\})$ $(c:\{a,c\}\to\{b,d\})$ $(b:\{b,d\}\to\{a,c\})$ $(a:\{a,c\}\to\{b,d\})$	$\{b\},\{a,c,d\}$ $\{a\},\{b,c,d\}$ $\{d\},\{a,b,c\}$ $\{c\},\{a,b,d\}$	
	$\{a,X\},\{b\}$	$\{b,X\},\{a\}$	$(a:\{a,X\}\to\{b\})$ $(b:\{b\}\to\{a,X\})$ $(a:\{a,X\}\to\{\})$	$(b:\{b,X\}\to\{a\})$ $(a:\{a\}\to\{b,X\})$ $(b:\{b,X\}\to\{\})$	$X,\{a,b\}$ $\{a,b,X\}$ $X,\{a\},\{b\}$	$\lvert X\rvert\geq 3$
	$\{a,X\},\{b,Y\}$	$\{b,X\},\ \{a,Y\}$	$(a:\{a,X\}\to\{b,Y\})$ $(b:\{b,Y\}\to\{a,X\})$	$(b:\{b,X\}\to\{a,Y\})$ $(a:\{a,Y\}\to\{b,X\})$	$X,\{a,b,Y\}$ $\{a,b,X\},Y$	$\lvert X\rvert\geq 1$ $\lvert Y\rvert\geq 2$ or $\lvert X\rvert\geq 2$ $\lvert Y\rvert\geq 1$
2, 3	$\{a,b,c\},\{d\}$	$\{a\},\{c\},\{b,d\}$	$(a:\{a,b,c\}\to\{\})$ $(b:\{a,b,c\}\to\{d\})$ $(c:\{a,b,c\}\to\{\})$	$(b:\{b,d\}\to\{c\})$ $(a:\{a\}\to\{c\})$ $(c:\{c\}\to\{a\})$ $(b:\{b,d\}\to\{a\})$	$\{b,c\},\{a\},\{d\}$ $\{a,c\},\{b,d\}$ $\{a,b\},\{c\},\{d\}$	
	$\{a,b,X\},Y$	$X,\{a\},\{b,Y\}$	$(a:\{a,b,X\}\to\{\})$ $(b:\{a,b,X\}\to Y)$	$(b:\{b,Y\}\to X)$ $(a:\{a\}\to X)$	$\{a\},\{b,X\},Y$ $\{a,X\},\{b,Y\}$	$\lvert X\rvert\geq 2$ $\lvert Y\rvert\geq 1$
	$\{a,c\},\{b,d\}$	$\{a\},\{b,c\},\{d\}$	$(a:\{a,c\}\to\{\})$ $(c:\{a,c\}\to\{\})$ $(b:\{b,d\}\to\{\})$ $(d:\{b,d\}\to\{\})$ $(b:\{b,d\}\to\{a,c\})$ $(c:\{a,c\}\to\{b,d\})$	$(b:\{b,c\}\to\{d\})$ $(c:\{b,c\}\to\{a\})$ $(a:\{a\}\to\{b,c\})$ $(d:\{d\}\to\{b,c\})$	$\{a\},\{c\},\{b,d\}$ $\{b\},\{d\},\{a,c\}$ $\{a,b,c\},\{d\}$ $\{b,c,d\},\{a\}$	

	$\{a,c\},\{b,Y\}$	$\{a\},\{b,c\},Y$	$(b:\{b,Y\}\to\{\})$ $(b:\{b,Y\}\to\{a,c\})$ $(a:\{a,c\}\to\{\})$ $(c:\{a,c\}\to\{\})$	$(c:\{b,c\}\to\{a\})$ $(a:\{a\}\to\{b,c\})$ $(b:\{b,c\}\to Y)$	$\{a,c\},\{b\},Y$ $\{a,b,c\},Y$ $\{a\},\{c\},\{b,Y\}$	$\lvert Y\rvert\geq 2$
	$\{a,X\},\{b,Y\}$	$\{a\},\{b,X\},Y$	$(a:\{a,X\}\to\{\})$ $(b:\{b,Y\}\to\{a,X\})$	$(b:\{b,X\}\to Y)$ $(a:\{a\}\to\{b,X\})$	$\{a\},X,\{b,Y\}$ $\{a,b,X\},Y$	$\lvert X\rvert\geq 2$ $\lvert Y\rvert\geq 2$
	$\{a,X\},\{b,Y\}$	$\{a,b\},X,Y\}$	$(a:\{a,X\}\to\{\})$ $(b:\{b,Y\}\to\{\})$	$(b:\{a,b\}\to Y)$ $(a:\{a,b\}\to X)$	$\{a\},X,\{b,Y\}$ $\{a,X\},\{b\},Y$	$\lvert X\rvert\geq 2$ $\lvert Y\rvert\geq 2$
2, 4	$\{a,b\},\{c,d\}$	$\{a\},\{b\},\{c\},\{d\}$	$(a:\{a,b\}\to\{\})$ $(b:\{a,b\}\to\{\})$ $(c:\{c,d\}\to\{\})$ $(d:\{c,d\}\to\{\})$	$(c:\{c\}\to\{d\})$ $(d:\{d\}\to\{c\})$ $(a:\{a\}\to\{b\})$ $(b:\{b\}\to\{a\})$	$\{a\},\{b\},\{c,d\}$ $\{a,b\},\{c\},\{d\}$	
	$\{a,b\},\{c,Y\}$	$\{a\},\{b\},\{c\},Y$	$(a:\{a,b\}\to\{\})$ $(b:\{a,b\}\to\{\})$ $(c:\{c,Y\}\to\{\})$	$(c:\{c\}\to Y)$ $(a:\{a\}\to\{b\})$ $(b:\{b\}\to\{a\})$	$\{a\},\{b\},\{c,Y\}$ $\{a,b\},\{c\},Y$	$\lvert Y\rvert\geq 2$
	$\{a,X\},\{c,Y\}$	$\{a\},X,\{c\},Y$	$(a:\{a,X\}\to\{\})$ $(c:\{c,Y\}\to\{\})$	$(c:\{c\}\to Y)$ $(a:\{a\}\to X)$	$\{a\},X,\{c,Y\}$ $\{a,X\},\{c\},Y$	$\lvert X\rvert\geq 2$ $\lvert Y\rvert\geq 2$
3,3	$\{a,X\},\{b\},Y$	$\{a\},X,\{b,Y\}$	$(a:\{a,X\}\to\{\})$ $(b:\{b\}\to Y)$	$(b:\{b,Y\}\to\{\})$ $(a:\{a\}\to X)$	$\{a\},X,\{b\},Y$ $\{a,X\},\{b,Y\}$	$\lvert X\rvert\geq 1$ $\lvert Y\rvert\geq 1$
	$\{a,X\},\{b,Y\},P$	$X,Y,\{a,b,P\}$	$(a:\{a,X\},P)$ $(b:\{b,Y\}\to P)$	$(b:\{a,b,P\}\to Y)$ $(a:\{a,b,P\}\to X)$	$X,\{a,P\},\{b,Y\}$ $\{a,X\},\{b,P\},Y$	$\lvert X\rvert\geq 1$ $\lvert Y\rvert\geq 1$ $\lvert P\rvert\geq 1$

	$\{a,X\},\{b,Y\},P$	$\{b,X\},Y,\{a,P\}$	$(a:\{a,X\}\to P)$ $(b:\{b,Y\}\to\{a,X\})$	$(b:\{b,X\}\to Y)$ $(a:\{a,P\}\to\{b,X\})$	$X,\{b,Y\},\{a,P\}$ $\{a,b,X\},Y,P$	$\lvert X\rvert\geq 1$ $\lvert Y\rvert\geq 1$ $\lvert P\rvert\geq 1$
3, 4	$\{a,X\},Y,\{b,c\}$	$X,\{a,Y\},\{b\},\{c\}$	$(a:\{a,X\}\to Y)$ $(b:\{b,c\}\to\{\})$ $(c:\{b,c\}\to\{\})$	$(b:\{b\}\to\{c\})$ $(c:\{c\}\to\{b\})$ $(a:\{a,Y\}\to X)$	$X,\{a,Y\},\{b,c\}$ $\{a,X\},Y,\{b\},\{c\}$	$\lvert X\rvert\geq 1$ $\lvert Y\rvert\geq 1$
	$\{a,X\},Y,\{b,P\}$	$X,\{a,Y\},\{b\},P$	$(a:\{a,X\}\to Y)$ $(b:\{b,P\}\to\{\})$	$(b:\{b\}\to P)$ $(a:\{a,Y\}\to X)$	$X,\{a,Y\},\{b,P\}$ $\{a,X\},Y,\{b\},P$	$\lvert X\rvert\geq 1$ $\lvert Y\rvert\geq 1$ $\lvert P\rvert\geq 2$
4,4	$\{a,X\},Y,$ $\{b,P\},Q$	$X,\{a,Y\},$ $P,\{b,Q\}$	$(a:\{a,X\}\to Y)$ $(b:\{b,P\}\to Q)$	$(b:\{b,Q\}\to P)$ $(a:\{a,Y\}\to X)$	$X,\{a,Y\},\{b,P\},Q$ $\{a,X\},Y,P,\{b,Q\}$	$\lvert X\rvert\geq 1$ $\lvert Y\rvert\geq 1$ $\lvert P\rvert\geq 1$ $\lvert Q\rvert\geq 1$

Table S2: List of transformations for $|z_A - z_B| = 2$.

S3: The proposal distributions of an IBD state

We define three IBD proposal distributions used in the Metropolis type sampling from the full posterior distribution (section S4).

(A) *One-side distribution.* Let $q(z|z_A)$ be the proposal distribution where the IBD state z is proposed from z_A according to the MCRP.

(B) *Two-side distribution.* Let $q(z|z_A, z_B)$ be a proposal distribution for z as an intermediate state between z_A and z_B. Thus $|z_A - z_B| \leq 2$, and the proposed z must satisfy $|z - z_A| \leq 1$ and $|z - z_B| \leq 1$. We define $q(z|z_A, z_B)$ for three cases:

- $|z_A - z_B| = 0$ ($z_A = z_B$). We sample z from $q(z|z_A)$.
- $|z_A - z_B| = 1$. We sample $z = z_A$ with probability 1/4, $z = z_B$ with probability 1/4, and otherwise generate the proposal by the following three steps:
  - Insert a new gamete into a subset of z_A of size j with probability $j/(n + \theta)$, or insert it as a new singleton with probability $\theta/(n + \theta)$.
  - Delete the gamete that is deleted in a randomly chosen transformation from z_A to z_B (see Table S1).
  - Label the new gamete as the deleted one.
- $|z_A - z_B| = 2$. We list all the possible intermediate states, and randomly choose one of them (Table S2).

(C) *Propagation distribution.* Suppose that z_A and z_B are two consecutive IBD states along the chromosome ($|z_A - z_B| \leq 1$) and also that $|z_A - z_C| = 1$. Let $q(z|z_A, z_B, z_C)$ be the proposal distribution of an IBD state z satisfying $|z - z_B| \leq 1$ and $|z - z_C| \leq 1$. Thus there are four cases as shown in Figure S1.

- I: $|z_A - z_B| = 0$ and $|z_B - z_C| = 1$. Set $z = z_B$.
- II: $|z_A - z_B| = 1$ and $|z_B - z_C| = 0$. Set $z = z_B$.
- III: $|z_A - z_B| = 1$ and $|z_B - z_C| = 1$. Set $z = z_B$.

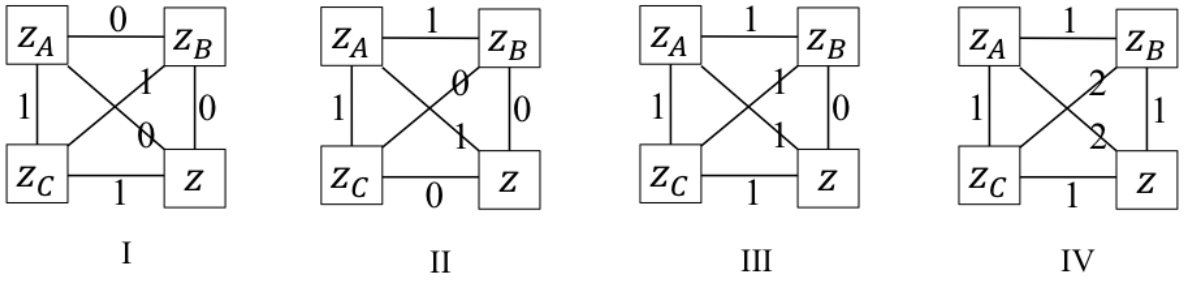

Figure S1: The possible scenarios for state z under $q(z|z_A, z_B, z_C)$. The distance between each pair of states is shown on the joining lines.

- IV: $|z_A - z_B| = 1$ and $|z_B - z_C| = 2$. Sample z uniformly from all the possible IBD states satisfying the distance constraints (Table S2). If there is no such IBD state, reject the proposal

Note that $q(z_B|z_C, z, z_B)$ is the reverse proposal distribution to $q(z|z_A, z_B, z_C)$. Cases I and II are paired in this reversal, while cases III and IV remain unchanged.

S4 Sampling the posterior distribution via MCMC

We estimate the model parameters θ, λ, ε, K, $\mathbf{x}$, and $\mathbf{z}$ by MCMC, using several versions of the Metropolis algorithm. Our general notation for the target distribution of state variables ω, is $p(\omega)$, and the proposal distribution for a new state given current state ω_t is $q(\cdot|\omega_t)$. The acceptance probability of proposed state ω^* is

$$\min\left(1, \frac{p(\omega^*)}{p(\omega_t)} \Big/ \frac{q(\omega^*|\omega_t)}{q(\omega_t|\omega^*)}\right). \tag{S2}$$

If the proposal is accepted $\omega_{t+1} = \omega^*$, and otherwise $\omega_{t+1} = \omega_t$. We refer to the ratio $p(\omega^*)/p(\omega_t)$ in equation S2 as the *target ratio*, and the ratio $q(\omega^*|\omega_t)/q(\omega_t|\omega^*)$ as the *proposal ratio*. In the case of reversible jump MCMC, the acceptance probability also includes a Jacobian factor (Green 1995).

The full conditional distributions of θ, ε and λ are given by

$$\begin{aligned} p(\theta \mid \cdot) &\propto p(\mathbf{z} \mid K, \theta)\, p(\theta) \\ p(\varepsilon \mid \cdot) &\propto p(\mathbf{y} \mid \pi, \mathbf{s}, K, \mathbf{x}, \mathbf{z}, \varepsilon)\, p(\varepsilon) \\ p(\lambda \mid \cdot) &\propto p(K, \mathbf{x} \mid \lambda)\, p(\lambda) \qquad\qquad \text{in general} \\ \text{and } p(\lambda \mid \cdot) &= \Gamma(\lambda | K + \alpha_\lambda - 1,\ 1/\left(\beta_\lambda^{-1} + \ell\right)) \quad \text{if } K_c = \infty \quad \text{(equation S1).} \end{aligned}$$

To sample θ, ε and λ from their full conditional distributions, we use random walk Metropolis algorithms (Gelman *et al.* 1996) for each parameter separately (except in the case where λ can be sampled directly). In this case the *proposal ratio* in equation S2 is equal to 1, and only the *target ratio* is required. In each case the proposal distribution is a normal distribution centered at the current value with the variance adjusted to give an acceptance ratio around 0.44 (Gelman *et al.* 2004).

Since the insertion or deletion of IBD change points involves a change in the dimension of the parameter space, we update K, $\mathbf{x}$, and $\mathbf{z}$ by reversible jump MCMC (Green 1995). We use six move types in the sampler: (A) update a transition location, (B) update an IBD state, (C) update an IBD state with downstream modification, (D) insert an IBD transition, (E) delete an IBD transition, and (F) update segments of IBD states by swapping their gamete labels. We denote by φ_i $(i = A, ..., F)$ the sampling probability for move type i. In move types (D-E), the parameter dimensions change by a transition location and an IBD state at the location, and the Jacobian factor is 1.

In the following, we describe the proposals and give the proposal ratios for all the move types (A-F), the target ratios for move types (D-E), and the full conditional distributions for move types (A-C) and (F) from which the target ratio can be obtained.

(A) *Single update of a transition location.* First randomly choose $2 \leq k \leq K$, and then sample a proposal value x_k^* from a discrete uniform distribution in the range from $x_{k-1} + 1$ to $x_{k+1} - 1$. The full conditional posterior distribution is

$$p(x_k|\cdot) \propto \prod_{\{i|x_{k-1}<s_i<x_{k+1}\}} p(\mathbf{y}_i \mid Z(s_i), \pi_i, \theta, \varepsilon).$$

The proposal ratio is 1 as the proposal distribution is symmetric.

(B) *Single update of an IBD state.* First randomly choose $1 \leq k \leq K$. If there are no IBD transitions ($K = 1$), a proposal state is sampled from $q\left(z_1^*|z_1\right)$. If the focal IBD state is at an end of the chromosome ($k = 1$ or K), a proposal state is sampled from $q\left(z_1^*|z_2\right)$ for $k = 1$ and from $q\left(z_K^*|z_{K-1}\right)$ for $k = K$. Otherwise, a proposal state z_k^* is sampled from $q\left(z_k^*|z_{k-1}, z_{k+1}\right)$. The full conditional posterior distribution is

$$p\left(z_k|\cdot\right) \propto \prod_{\{i|x_k \leq s_i < x_{k+1}\}} p(\mathbf{y}_i|Z(s_i), \pi_i, \theta, \varepsilon)\; p(z_{k+1}|z_k, \theta)\; p(z_k|z_{k-1}, \theta),$$

where $p(z_{K+1}|z_K, \theta) = p(z_1|z_0, \theta) = 1$ for the cases $k = 1$ or K.
The proposal ratio is

$$\begin{aligned} q(z_1^*|z_1)/q(z_1|z_1^*) \quad & \text{if } K = 1, \text{ and otherwise } q(z_1^*|z_2)/q(z_1|z_2) \text{ if } k = 1, \\ q(z_K^*|z_{K-1})/q(z_K|z_{K-1}) \quad & \text{if } k = K, \text{ and otherwise } q\left(z_k^*|z_{k-1}, z_{k+1}\right)/q\left(z_k|z_{k-1}, z_{k+1}\right). \end{aligned}$$

(C) *Single update of an IBD state with downstream modification.* First randomly choose k, $1 \leq k \leq K$. If $K = 1$ or $k = K$, update using move type (B). If $k = 1$, sample z_1^* from $q(z_1^*|z_1)$. Otherwise for $k > 1$, set $z_l^* = z_l$ for $l = 1 \ldots k-1$, and sample z_k^* sampled from $q(z_k^*|z_{k-1}, z_k)$. If $z_k^* = z_k$ we set $l' = k+1$. Otherwise we iteratively sample z_l^* from $q(z_l^*|z_{l-1}, z_l, z_{l-1}^*)$ from $l = k+1$ until there exists $l' \geq k+1$ such that $z_{l'}^* = z_{l'}$ or $l' = K+1$. If $l' < K$, we set $z_l^* = z_l$ for $l = l'+1 \ldots K$. The full conditional distribution is

$$p(\{Z(x)\}_{x_k \leq \mathrm{x} < x_{l'}}|\cdot) \propto \prod_{\{i|x_k \leq s_i < x_{l'}\}} p(\mathbf{y}_i|Z(s_i), \pi_i, \theta, \varepsilon) \prod_{l=k-1}^{l'-1} p(z_{l+1}|z_l, \theta),$$

where $p(z_{K+1}|z_K, \theta)$ is set to be 1 for $l' = K+1$. The proposal ratio is given by

$$\frac{q(z_k^*|z_k)}{q(z_k|z_k^*)} \frac{\prod_{l=k}^{l'-1} q(z_l^*|z_{l-1}, z_l, z_{l-1}^*)}{\prod_{l=k}^{l'-1} q(z_l|, z_{l-1}^*, z_l^*, z_{l-1})}$$

if $k = 1$, and otherwise ($1 < k < K$) by

$$\frac{q(z_k^*|z_k, z_{k-1})}{q(z_k|z_k^*, z_{k-1}^*)} \frac{\prod_{l=k}^{l'-1} q(z_l^*|z_{l-1}, z_l, z_{l-1}^*)}{\prod_{l=k}^{l'-1} q(z_l|, z_{l-1}^*, z_l^*, z_{l-1})}$$

(D) *Insert one IBD transition.* First randomly choose k, $1 \leq k \leq K$. Then insert an IBD transition location into $\mathbf{x}$ to give $\mathbf{x}^*$ with x^*_{k+1}, sampled from the discrete uniform distribution in the range from $x_k + 1$ to $x_{k+1} - 1$. Set $z^*_l = z_l$ for $l = 1 \ldots k$, and insert z^*_{k+1} sampled from $q(z^*_{k+1}|z_k)$. If $k = K$ we set $l' = K + 2$, and if $z^*_{k+1} = z_k$ we set $l' = k + 2$. Otherwise we iteratively sample z^*_l from $q(z^*_l|z_{l-2}, z_{l-1}, z^*_{l-1})$ from $l = k + 2$ until there exists $l' \geq k + 2$ such that $z^*_{l'} = z_{l'-1}$ or $l' = K + 2$. If $l' < K + 1$, we set $z^*_l = z_{l-1}$ for $l = l' + 1 \ldots K + 1$. The target ratio is

$$\frac{p(K+1, \mathbf{x}^*, \mathbf{z}^*|\cdot)}{p(K, \mathbf{x}, \mathbf{z}|\cdot)} = \frac{\prod_{\{i|x^*_k < s_i < x^*_{l'}\}} p(\mathbf{y}_i|Z^*(s_i), \pi_i, \theta, \varepsilon)}{\prod_{\{i|x_k < s_i < x_{l'-1}\}} p(\mathbf{y}_i|Z(s_i), \pi_i, \theta, \varepsilon)} \frac{\prod_{l=k}^{l'-1} p(z^*_{l+1}|z^*_l, \theta)}{\prod_{l=k}^{l'-2} p(z_{l+1}|z_l, \theta)} \frac{p(K+1, \mathbf{x}^*)}{p(K, \mathbf{x})},$$

where the term $p(K + 1, \mathbf{x}^*)/p(K, \mathbf{x})$ is replaced by $p(K + 1, \mathbf{x}^*|\lambda)/p(K, \mathbf{x}|\lambda)$ if λ is sampled (equation S1), $p(z_{K+1}|z_K, \theta)$ and $p(z^*_{K+2}|z^*_{K+1}, \theta)$ are set to be 1 for $l' = K+2$. The proposal ratio is given by

$$\frac{\varphi_D K^{-1}(x_{k+1} - x_k - 1)^{-1} q(z^*_{k+1}|z_k) \prod_{l=k+2}^{l'-1} q(z^*_l|z_{l-2}, z_{l-1}, z^*_{l-1})}{\varphi_E K^{-1} \prod_{l=k+1}^{l'-2} q(z_l|, z^*_l, z^*_{l+1}, z_{l-1})}.$$

(E) *Delete one IBD transition.* First randomly choose k, $2 \leq k \leq K$. (If $K = 1$, we do not change anything.) Set $\mathbf{x}^*$ by deleting x_k from $\mathbf{x}$, and set $z^*_l = z_l$ for $l = 1 \ldots k - 1$. If $k = K$ we set $l' = K$, and if $z^*_{k-1} = z_k$ we set $l' = k$. Otherwise, we iteratively sample z^*_l from $q(z^*_l|z_l, z_{l+1}, z^*_{l-1})$ from $l = k$ until there exists $l' \geq k$ so that $z^*_{l'} = z_{l'+1}$ or $l' = K$. If $l' < K - 1$, we set $z^*_l = z_{l+1}$ for $l = l' + 1 \ldots K - 1$. The target ratio is

$$\frac{p(K-1, \mathbf{x}^*, \mathbf{z}^*|\cdot)}{p(K, \mathbf{x}, \mathbf{z}|\cdot)} = \frac{\prod_{\{i|x^*_{k-1} < s_i < x^*_{l'}\}} p(\mathbf{y}_i|Z^*(s_i), \pi_i, \theta, \varepsilon)}{\prod_{\{i|x_{k-1} < s_i < x_{l'+1}\}} p(\mathbf{y}_i|Z(s_i), \pi_i, \theta, \varepsilon)} \frac{\prod_{l=k-1}^{l'-1} p(z^*_{l+1}|z^*_l, \theta)}{\prod_{l=k-1}^{l'} p(z_{l+1}|z_l, \theta)} \frac{p\,(K-1, \mathbf{x}^*)}{p(K, \mathbf{x})},$$

where the term $p(K - 1, \mathbf{x}^*)/p(K, \mathbf{x})$ is replaced by $p(K - 1, \mathbf{x}^*|\lambda)/p(K, \mathbf{x}|\lambda)$ if λ is sampled (equation S1), $p(z_{K+1}|z_K, \theta)$ and $p(z^*_K|z^*_{K-1}, \theta)$ are set to be 1 for $l' = K$. The proposal ratio is given by

$$\frac{\varphi_E (K-1)^{-1} \prod_{l=k}^{l'-1} q(z^*_l|z_l, z_{l+1}, z^*_{l-1})}{\varphi_D (K-1)^{-1}\ (x^*_k - x^*_{k-1} - 1)^{-1} q(z_k|z^*_{k-1}) \prod_{l=k+1}^{l'} q(z_l|z^*_{l-2}, z^*_{l-1}, z_{l-1})}.$$

(F) *Update segments of IBD states.* We first randomly choose one pair of gametes and partition them into IBD and non-IBD segments. Independently for each non-IBD

segment, we propose IBD states by swapping the labels for the pair of gametes. Let k and l ($l > k$) be the two ends of the segment so that z_k and z_l are IBD for the pair of gametes. We set $k = 0$ for the first segment, and $l = K + 1$ for the last segment. The full conditional distribution is

$$p(\{Z(x)\}_{x_{k+1} \leq \mathrm{x} < x_l} | \cdot) \quad \propto \prod_{\{i | x_{k+1} \leq s_i < x_l\}} p(\mathbf{y}_i | Z(s_i), \pi_i, \theta, \varepsilon),$$

which does not depend on the IBD transition probabilities. The proposal ratio is 1 for this symmetric proposal distribution.

In each iteration of a single MCMC, we update θ, λ, ε one by one, update $Z(x)$ $10^{-5}\ell$ times with move types (A-E), and update IBD states n/2 times with move type (F). To improve the mixing of the MCMC, with probability 0.5 we reverse the direction of the chromosome in every iteration. When λ is not sampled (equation S1) move types (A-E) are sampled with probabilities $\varphi = ((1-2c)/3, (1-2c)/3, (1-2c)/3, c, c)$, respectively. Here c is a tunable constant, and it is set to be 0.2. When λ must also be sampled due to the bounding of K by K_c, we set c to be a small value of 0.05, as the number of IBD transitions is distributed sharply around K_c due to the LD in the founder genomes.

We run two independent groups of MCMC chains. In each group there are four MCMC chains, and parallel algorithms are used where the full conditional distribution is raised to the power σ,$0 < \sigma \leq 1$ (Metropolis-coupled MCMC, (Geyer 1991)). The power σ is set to 1 for the coldest chain, and decreases with equal interval $\Delta\sigma$, which is adjusted so that the accept probability for swapping a pair of chains is 0.5. Only the coldest chain in each group is saved.

S5 IBD estimates from large data sets

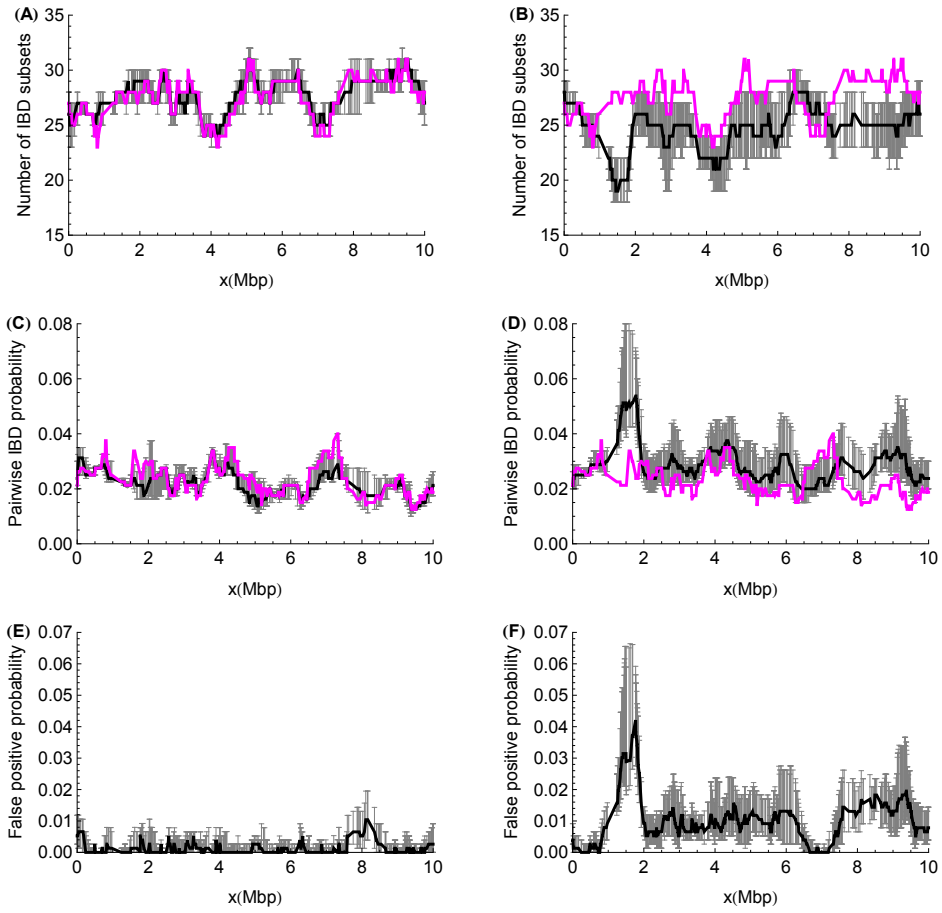

Figure S2

Figure S2: Overall recovery of IBD states from long data sets. Estimated IBD states along gametes obtained from the data sets *L-NoLD* (left panels) and *L-LD* (right panels). They are evaluated in terms of the number of IBD sets (A and B), the pairwise IBD probability (C and D) and the false positive probability (E and F). Error bars denote the 95posterior intervals with black lines connecting the medians. In panels A-D, magenta lines denote the true values. Compare with Figure 5 in the main text.

Literature Cited

References

Gelman, A., J. B. Carlin, H. S. Stern, *et al.,* 2004 *Bayesian Data Analysis* (2 nd ed.). Chapman & Hall/CRC press, Boca Raton, FL.

Gelman, A., G. O. Roberts, and W. R. Gilks, 1996 Efficient Metropolis jumping rules. Bayesian Statistics **5**: 599–607.

Geyer, C. J., 1991 . Markov chain Monte Carlo maximum likelihood. in *Computing Science and Statistics: Proceedings of the 23rd Conference on the Interface*, edited by Keramidas, pp. 156–163. Fairfax Station: Interface Foundation.

Green, P. J., 1995 Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika **82**: 711–732.