G3
Genes | Genomes | Genetics

Recursive Algorithms for Modeling Genomic Ancestral Origins in a Fixed Pedigree

Chaozhi Zheng, Martin P. Boer, and Fred A. van Eeuwijk
Biometris, Wageningen University and Research, Wageningen, The Netherlands
ORCID ID: 0000-0001-6030-3933 (C.Z.)

ABSTRACT The study of gene flow in pedigrees is of strong interest for the development of quantitative trait loci (QTL) mapping methods in multiparental populations. We developed a Markovian framework for modeling ancestral origins along two homologous chromosomes within individuals in fixed pedigrees. A highly beneficial property of our method is that the size of state space depends linearly or quadratically on the number of pedigree founders, whereas this increases exponentially with pedigree size in alternative methods. To calculate the parameter values of the Markov process, we describe two novel recursive algorithms that differ with respect to the pedigree founders being assumed to be exchangeable or not. Our algorithms apply equally to autosomes and sex chromosomes, another desirable feature of our approach. We tested the accuracy of the algorithms by a million simulations on a pedigree. We demonstrated two applications of the recursive algorithms in multiparental populations: design a breeding scheme for maximizing the overall density of recombination breakpoints and thus the QTL mapping resolution, and incorporate pedigree information into hidden Markov models in ancestral inference from genotypic data; the conditional probabilities and the recombination breakpoint data resulting from ancestral inference can facilitate follow-up QTL mapping. The results show that the generality of the recursive algorithms can greatly increase the application range of genetic analysis such as ancestral inference in multiparental populations.

Many complicated experimental crosses have recently been produced for mapping QTL, particularly in plants (*e.g.*, Kover *et al.* 2009; Sannemann *et al.* 2015). In contrast to traditional biparental populations, multiple parents have been used to increase genetic diversity and thus QTL segregation probability, and many generations of intercross mating have been used to increase accumulated recombination breakpoints in sampled offspring in the final generation and thus QTL mapping resolution. The expected density of recombination points is one of key quantities for optimizing experimental designs prior to collecting genotypic and phenotypic data (Rockman and Kruglyak 2008). Furthermore, the identification of recombination breakpoints from genotypic data provides useful information for increasing detection power and mapping resolution (Xu 2013; Li *et al.* 2015). The primary aims of this paper are to develop the theory of gene flow in a fixed pedigree with arbitrary structure to calculate the prior distribution of recombination points, and to apply the theory for ancestral inference and detection of recombination points from genotypic data in multiparental populations.

The theory of one-locus gene flow in a pedigree has been well developed. The haploid genomes of pedigree founders are designated by distinct labels, called founder genome labels (FGLs). A set of genes are identical by descent (IBD) if they carry the same FGLs. The identity state for two genes is either IBD or non-IBD. The IBD probability for two distinct genes within an individual is the inbreeding coefficient, and it is the kinship coefficient for the two genes randomly sampled from two

Corresponding Author: Biometris, Wageningen University and Research, PO Box 16, 6700 AA Wageningen, The Netherlands. E-mail: chaozhi.zheng@wur.nl

distinct individuals. Rostron (1978) developed a recursive algorithm for calculating the two-gene coefficients of inbreeding and kinship. Karigl (1981) developed a recursive algorithm for calculating the coefficients of the fifteen gene identity states for four genes between two individuals. Thompson (1983) showed that a similar algorithm applies to the probabilities of joint descent of multiple genes from a specific FGL. Bauman *et al.* (2008) and Zhou *et al.* (2012) developed a recursive algorithm for calculating the ancestral coefficients such as the probability of one gene descending from a given FGL and the probability of two genes descending from two given FGLs (distinct or not).

The theory of gene flow at two linked loci in a pedigree has also been developed. Weir and Cockerham (1969) and Cockerham and Weir (1973) developed a recursive algorithm for calculating the two-locus inbreeding coefficient of an individual that is defined as a linear function of the probabilities of the fifteen identity states for four genes, two at each of two linked loci. Thompson (1988) developed a recursive algorithm for calculating the two-locus kinship between two individuals defined as the probability of IBD at both loci, see Hill and Weir (2007) for calculating multi-locus IBD probabilities in random mating populations. In contrast to the identity coefficients, the ancestral coefficient at two loci is defined as the two-locus diplotype probabilities, and there are 4^L possible diplotypes for four genes in an individual with respect to L distinct FGLs. Thus, the calculation of the two-locus ancestral inferences has been developed only for simple breeding systems including selfing, brother-sister (or sibling) mating, and parent-offspring mating (Haldane and Waddington 1931; Johannes and Colome-Tatche 2011; Broman 2012).

There has been much interest in developing the theory of gene flow in a pedigree with chromosomes assumed to be a genomic continuum. Donnelly (1983) developed a formal mathematical framework, where the crossover processes in a chromosome pedigree follow jointly a continuous time Markov random walk on the vertices of a hypercube, the time parameter being position along homologous chromosomes. Here a vertex represents one of possible gene transmissions from founders to all non-founders, and the number of vertices in a d-dimensional hypercube is 2^d, d being the number of non-founders. This framework has been used in many kinds of small pedigree systems (*e.g.*, Bickeböller and Thompson 1996; Stefanov 2000; Martin and Hospital 2011). However, it is impractical to apply Donnelly's Markovian framework to a large complex pedigree since the number of states (hypercube vertices) increases exponentially with the pedigree size.

Alternatively, the theory of junctions has been developed to study the gene flow for a genomic continuum (Fisher 1949, 1954). A junction is defined as a boundary point (recombination breakpoint) on a chromosome where two distinct FGLs meet. Fisher (1949, 1954) and Bennett (1953, 1954) developed the theory of junctions for simple breeding systems including selfing, brother-sister mating, and parent-offspring mating. It has been extended to populations (Stam 1980; Baird *et al.* 2003; Chapman and Thompson 2003; MacLeod *et al.* 2005; Zheng *et al.* 2014; Zheng 2015). In particular, Zheng *et al.* (2014) and Zheng (2015) developed a Markovian framework for modeling ancestral origins within an individual with the state space being the possible pairs of FGLs, and thus the number of states is much smaller than that of Donnelly's Markovian framework (Donnelly 1983). On the other hand, our Markovian framework (Zheng *et al.* 2014; Zheng 2015) is restricted by two assumptions: the mating scheme from one generation to the next is random mating, and the FGLs are assumed to be exchangeable so that the probability distribution of ancestral origins in offspring is invariant to all possible permutations of the FGLs. However, the exchangeability generally does not hold for a mapping population produced via an arbitrary breeding pedigree.

In this paper, we relax the restriction of our previous Markovian framework by extending it to a fixed pedigree. We first describe a recursive algorithm (denoted by EXCH) under the assumption of FGLs being exchangeable, which is applicable for the construction of Markovian framework in simple breeding schemes. Then we develop a recursive algorithm (denoted by NON-EXCH) for modeling ancestral origins to relax the exchangeability assumption. The two recursive algorithms apply to both autosomes and sex chromosomes if they exist. The results of both recursive algorithms are compared with those from extensive simulations on a classical pedigree. We first apply the two algorithms to compare different population designs (or breeding pedigrees) prior to experiments, for example, in terms of the overall density of junctions (recombination breakpoints), an important factor of QTL mapping resolution. In addition, the non-exchangeability can be illustrated under various breeding designs. Then we apply the two algorithms to incorporate pedigree information for ancestral inference from genotypic data in simulated and real collaborative cross (CC) populations, resulting in conditional probabilities that are necessary for downstream QTL mapping.

RECURSIVE ALGORITHM EXCH

Notations and overview

The symbols are briefly explained in Table 1, and some of them are illustrated in Figure 1. Pedigree members with unspecified mother and father are called the founders of the pedigree, and the other members are non-founders. The recursive algorithm presupposes that pedigree members are ordered under the constraint that parents always precede children. We denote by subscripts a, b, c the members of a pedigree, and denote by $a > b$ individual a comes after individual b ($\neq a$). We denote by superscripts m the maternally derived genes or chromosomes, and p for the paternally derived. We denote by superscripts o, o_1, o_2, and o_3 the unspecified parental origins (m or p) of genes or chromosomes. For the sake of brevity, $f(a) = \{AA, XX, XY\}$ denotes a horizontal piecewise equation, so that $f(a) = AA$, XX, and XY for the autosomes of individual a, XX chromosomes of female a, and XY chromosomes of male a, respectively.

In the derivations of recurrence relations of a quantity, we will focus on a particular ordering $a > b$ if the quantity concerns genes in two individuals; we will always trace the maternally derived gene or chromosome within non-founder a back to its two parental genes or chromosomes if the maternally derived gene or chromosome concerns the quantity, and otherwise trace the paternally derived gene or chromosome. Throughout this paper, ♂ and ♀ always denote the father and mother of individual a, rather than any other, respectively.

The identity coefficient $\phi_{ab}^{o_1 o_2}(12)$ denotes that two genes at a single locus have identity state (12), where the first gene is in individual a and has parental origin o_1, and the second gene is in individual b and has parental origin o_2 (Figure 1A). There are only two two-gene identity states: IBD (11) or non-IBD (12), and $\phi_{ab}^{o_1 o_2}(11) = 1 - \phi_{ab}^{o_1 o_2}(12)$. Similarly, we denote by $\phi_{abc}^{o_1 o_2 o_3}(123)$ the three-gene identity coefficient with the identity state being the non-IBD state (123), which is required together with two-gene identity coefficients for deriving the recursive relations for junction densities.

An expected identity junction density is defined as the expected number of the specified type of recombination breakpoints per Morgan along two homologous chromosomes. Here expectation concerns the stochasticity of gene flow from founders to descendants on a fixed pedigree. And identity indicates that only the identity patterns (not specific FGLs) on the two sides of junctions matter. In the recursive

■ **Table 1 List of symbols and their brief descriptions**

Symbol	Description
1_S	An indicator function that equals 1 if statement S is true and 0 otherwise
0_{S1}^{S2}	An indicator function that equals 0 if $S1$ and $S2$ are true and 1 otherwise.
o, o_1, o_2, o_3	A gene or haplotype is maternally ($= m$) or paternally ($= p$) derived
a, b, c	Pedigree members
♂, ♀	Father of a, mother of a
$\phi_{ab}^{mp}(11)$	Probability of IBD between maternal gene of a and paternal gene of b
$\phi_{ab}^{o_1 o_2}(11)$	Probability of IBD between two genes of ab with parental origins $o_1 o_2$
$\phi_{ab}^{o_1 o_2}(12)$	Probability of non-IBD between two genes of ab with parental origins $o_1 o_2$
$\phi_{abc}^{o_1 o_2 o_3}(123)$	Probability of non-IBD among three genes of abc with parental origins $o_1 o_2 o_3$
R_a^m	Expected junction density (per Morgan) on the maternal chromosome of a
R_a^o	On the chromosome of a with parental origin o
$(g_1 g_2 g_3 g_4)$	A junction type denoted by two-locus four-gene identity state: $g_1 g_2$ ($g_3 g_4$) are on the left-hand (right-hand) side of junction, and $g_1 g_3$ and $g_2 g_4$ are two haplotypes. Seven types: 1112, 1121, 1122, 1211, 1213, 1222, and 1232
$J_{ab}^{o_1 o_2}(g_1 g_2 g_3 g_4)$	Expected density of junction type ($g_1 g_2 g_3 g_4$), where haplotype $g_1 g_3$ of a has parental origin o_1 and haplotype $g_2 g_4$ of b has parental origin o_2
ρ_{aa}^{mp}	Expected overall junction density for a. $\rho_{aa}^{mp} = R_a^m + R_a^p - J_{aa}^{mp}(1122)$
Ω	A set of distinct FGLs assigned to the founders of a pedigree, $i, j, k \in \Omega$
$i \neq j \neq k$	FGLs i, j, k differ from each other so that $i \neq j$, $j \neq k$, and $i \neq k$
$\phi_a^m(i)$	Probability that the maternal gene of a has FGL i
$\phi_a^o(i)$	Probability that the gene of a with parental origin o has FGL i. $\sum_{i \in \Omega} \phi_a^o(i) = 1$
$\phi_{ab}^{mp}(ij)$	Probability that the maternal gene of a has FGL i and the paternal gene of b has FGL j.
$\phi_{ab}^{o_1 o_2}(ij)$	Probability that two genes of ab with parental origins $o_1 o_2$ have FGLs ij $\phi_{ab}^{o_1 o_2}(12) = \sum_{i,j \in \Omega, i \neq j} \phi_{ab}^{o_1 o_2}(ij)$
$\phi_{abc}^{o_1 o_2 o_3}(ijk)$	Probability that three genes of abc with parental origins $o_1 o_2 o_3$ have FGLs ijk. $\phi_{abc}^{o_1 o_2 o_3}(123) = \sum_{i,j,k \in \Omega, i \neq j \neq k} \phi_{abc}^{o_1 o_2 o_3}(ijk)$
$R_a^o(ij)$	Expected junction density (per Morgan) on chromosome of a with parental origin o, where the left-hand (right-hand) side of junction has FGL $i(j \neq i)$. $R_a^o = \sum_{i,j \in \Omega, i \neq j} R_a^o(ij)$
$J_{ab}^{o_1 o_2}(ijkj)$	Expected junction density: the four genes have FGLs $ijkj$, haplotype ik of a has parental origin o_1, and haplotype jj of b has parental origin o_2. $J_{ab}^{o_1 o_2}(1232) = \sum_{i,j,k \in \Omega, i \neq j \neq k} J_{ab}^{o_1 o_2}(ijkj)$. Similarly for $J_{ab}^{o_1 o_2}(iiij)$, $J_{ab}^{o_1 o_2}(iiji)$, $J_{ab}^{o_1 o_2}(iijj)$, $J_{ab}^{o_1 o_2}(ijii)$, $J_{ab}^{o_1 o_2}(ijik)$, $J_{ab}^{o_1 o_2}(ijjj)$

algorithm NON-EXCH, we will introduce ancestral junction densities where the specific FGLs do matter. For example, $J_{ab}^{o_1 o_2}(1232)$ denotes that the four genes, two at each side of a breakpoint along two chromosomes, have genetic identity state (1232) (Figure 1B). Here the first and third genes are on the left and right sides of a breakpoint, respectively, and they are in individual a and have parental origin o_1. And the second and fourth genes are on the left and right sides, respectively, and they are in individual b and have parental origin o_2. We adopt the notation of Nadot and Vayssiex (1973) for an identity state: the genes are labeled by natural integers starting 1, and the same integer is assigned to the gene that is IBD with a previous gene.

Following the previous framework (Zheng *et al.* 2014; Zheng 2015), we model ancestral origins along two homologous chromosomes within an individual by a continuous time Markov process, which can be described by an initial distribution π of ancestral origins and a rate matrix Q. Under the assumption of exchangeability among FGLs, the initial distribution π of individual a is determined by the identity coefficient $\phi_{aa}^{mp}(11)$, and the rate matrix Q of individual a is determined by the five expected identity junction densities $J_{aa}^{mp}(1122)$, $J_{aa}^{mp}(1211)$, $J_{aa}^{mp}(1213)$, $J_{aa}^{mp}(1222)$, and $J_{aa}^{mp}(1232)$; see Zheng (2015) for an illustrative construction of rate matrix Q from the five junction densities.

In the following, we describe a recursive algorithm for calculating the identity coefficients and the expected junction densities for an individual in a given pedigree. In short, let 1_S be an indicator function that equals 1 if statement S is true and 0 otherwise, and let $0_{S1}^{S2} = 1 - 1_{S1} 1_{S2}$ be an indicator function that equals 0 if statements $S1$ and $S2$ are true and 1 otherwise.

Identity coefficients

The recurrence relations for the two- and three-gene identity coefficients are similar to the recursive algorithm of Karigl (1981). However, the calculation proceeds by tracing back genes instead of individuals, see also *e.g.*, Jacquard (1974), Nadot and Vayssiex (1973), Denniston (1974), and Garcia-Cortes (2015). Thus, we can account for the asymmetry between maternally and paternally derived chromosomes, and account for autosomes and sex chromosomes simultaneously. Throughout this paper, we assume that there are no crossovers between X and Y sex chromosomes within a male. In addition, we focus on non-IBD state (12) instead of IBD state (11) to simplify the recurrence relations.

We derive the recurrence relations according to the Mendelian inheritance rules: (1) a maternally (or paternally) derived autosomal gene descends from the two genes within the mother (or father, respectively) with equal probability $1/2$, (2) similar to (1) for a maternally derived X-linked gene, and (3) a paternally derived sex-linked gene within a female (or male) must descend from the X-linked (or Y-linked, respectively) gene of the father. The recurrence relations of the two-gene identity coefficient $\phi_{ab}^{o_1 o_2}(12)$ for non-founder a are given by

$$\phi_{ab}^{mo}(12) = \frac{1}{2}\left[\phi_{♀b}^{mo}(12) + \phi_{♀b}^{po}(12)\right]0_{a=b}^{m=o}, a \geq b$$

$$\phi_{ab}^{po}(12) = \left\{\frac{1}{2}\left[\phi_{♂b}^{mo}(12) + \phi_{♂b}^{po}(12)\right], \phi_{♂b}^{mo}(12),\right.$$

$$\left.\phi_{♂b}^{po}(12)\right\}0_{a=b}^{p=o}, a \geq b,$$

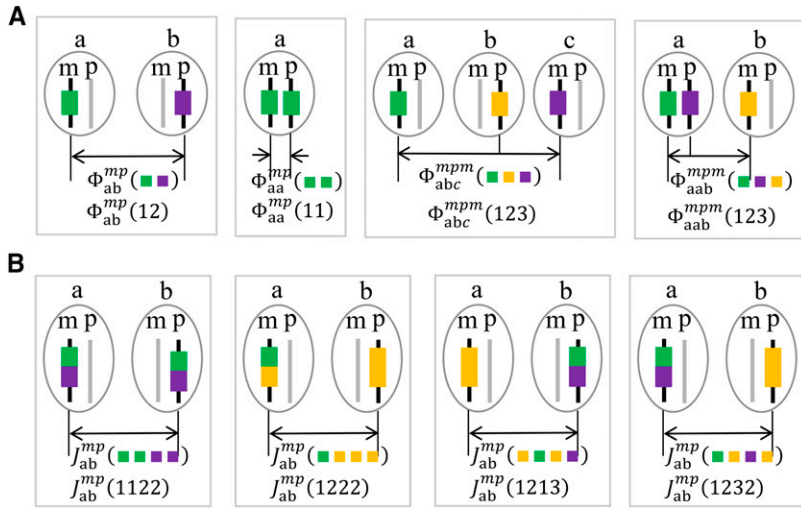$$\phi_{ab}^{o_1 o_2}(12) = \phi_{ba}^{o_2 o_1}(12), a < b,$$

Figure 1 Illustration of some quantities. An identity state may correspond to many ancestral states. Different FGLs are shown by different colors, and the irrelevant chromosomes are shown as gray. (A) Some two- or three-gene ancestral coefficients and their corresponding identity coefficients. (B) Some expected ancestral junction densities and their corresponding expected identity junction densities. See Table 1 for brief explanations of these quantities.

for $o, o_1, o_2 \in \{m, p\}$. Here the first equation follows from the rules 1 and 2 for the maternally derived gene within individual a, the second equation follows from the rules 1 and 3 for the paternally derived gene within individual a, and the last equation is obtained by reversing the ordering of two genes. The symmetries such as $\phi_{abc}^{o_1 o_2 o_3}(123) = \phi_{acb}^{o_1 o_3 o_2}(123) = \ldots = \phi_{cba}^{o_3 o_2 o_1}(123)$ apply to three-gene identity coefficients by permuting the three genes, and thus we consider only a particular ordering $a \geq b \geq c$, and we have

$$\phi_{abc}^{mo_1 o_2}(123) = \frac{1}{2} \left[\phi_{\female bc}^{mo_1 o_2}(123) + \phi_{\female bc}^{po_1 o_2}(123) \right] 0_{a=b}^{m=o_1} 0_{a=c}^{m=o_2} 0_{b=c}^{o_1=o_2},$$

$$\phi_{abc}^{po_1 o_2}(123) = \left\{ \frac{1}{2} \left[\phi_{\male bc}^{mo_1 o_2}(123) + \phi_{\male bc}^{po_1 o_2}(123) \right], \phi_{\male bc}^{mo_1 o_2}(123), \right.$$

$$\left. \phi_{\male bc}^{po_1 o_2}(123) \right\} 0_{a=b}^{p=o_1} 0_{a=c}^{p=o_2} 0_{b=c}^{o_1=o_2},$$

for $o_1, o_2 \in \{m, p\}$, where the first (second) equation traces the maternally (or paternally, respectively) derived gene within individual a.

The boundary conditions of these recurrence relations are given according to the assignment of FGLs to the founders in a pedigree. If we assign distinct FGLs to each haploid genome of an outbred founder, all multi-gene non-IBD probabilities are 1 if any pair of genes is not from the same haploid genome, and 0 otherwise. Four-gene or higher order identity coefficients may be derived similarly, but they are not required for the derivations of junction densities.

Expected identity junction densities

Let R_a^m (R_a^p) be the maternal (paternal) map expansion, the expected density of recombination breakpoints on the maternally (paternally) derived chromosome of individual a. The implicit two-gene identity state for the map expansions is (12). It holds that

$$R_a^m = J_{aa}^{mp}(1122) + J_{aa}^{mp}(1121) + J_{aa}^{mp}(1222) + J_{aa}^{mp}(1232),$$
$$R_a^p = J_{aa}^{mp}(1122) + J_{aa}^{mp}(1112) + J_{aa}^{mp}(1211) + J_{aa}^{mp}(1213),$$

where $J_{aa}^{mp}(1121) = J_{aa}^{mp}(1222)$ and $J_{aa}^{mp}(1112) = J_{aa}^{mp}(1211)$ according to the reversibility of chromosome direction (Zheng 2015). The expected overall junction density for individual a is given by $\rho_{aa}^{mp} = R_a^m + R_a^p - J_{aa}^{mp}(1122)$. In the following, we derive the recurrence relations for the five expected junction densities R_a^m, R_a^p, $J_{aa}^{mp}(1122)$, $J_{aa}^{mp}(1213)$, and $J_{aa}^{mp}(1232)$, from which $J_{aa}^{mp}(1211)$ and $J_{aa}^{mp}(1222)$ can be derived according to the above equations of map expansions.

The recurrence relations for the two map expansions of non-founder a are relatively simple, and they are given by (Zheng 2015)

$$R_a^m = \frac{1}{2} \left(R_\female^m + R_\female^p \right) + \phi_{\female \female}^{mp}(12),$$

$$R_a^p = \left\{ \frac{1}{2} \left(R_\male^m + R_\male^p \right) + \phi_{\male \male}^{mp}(12), R_\male^m, R_\male^p \right\},$$

according to the theory of junctions that a new identity junction is formed whenever a recombination event occurs between two homologous chromosomes that are non-IBD at the location of a crossover (Fisher 1954).

Since chromosomes are modeled as a genomic continuum, the shared junction type (1122) between two chromosomes can be formed only by duplicating the chromosome segment harboring the recombination breakpoint. For the expected junction density $J_{aa}^{o_1 o_2}(1122)$ of non-founder a, we have

$$J_{aa}^{oo}(1122) = R_a^o,$$

$$J_{aa}^{mp}(1122) = J_{aa}^{pm}(1122) = \frac{1}{2} \left[J_{a\female}^{pm}(1122) + J_{a\female}^{pp}(1122) \right],$$

$$J_{ab}^{mo}(1122) = \frac{1}{2} \left[J_{\female b}^{mo}(1122) + J_{\female b}^{po}(1122) \right], a > b,$$

$$J_{ab}^{po}(1122) = \left\{ \frac{1}{2} \left[J_{\male b}^{mo}(1122) + J_{\male b}^{po}(1122) \right], J_{\male b}^{mo}(1122), J_{\male b}^{po}(1122) \right\},$$
$$a > b,$$

$$J_{ab}^{o_1 o_2}(1122) = J_{ba}^{o_2 o_1}(1122), a < b,$$

for $o, o_1, o_2 \in \{m, p\}$, where and the last equation is obtained by reversing the ordering of the two haplotypes. In the first equation, $J_{aa}^{oo}(1122)$ refers to the same chromosome in individual a, and thus equals R_a^o by definition. We trace the maternally derived haplotype within non-founder a back to its two parental haplotypes if the maternally derived haplotype concerns the junction density, and otherwise track the paternally derived haplotype.

We derive the recurrence relations for $J_{ab}^{o_1 o_2}(1213)$ and $J_{ab}^{o_1 o_2}(1232)$ jointly. The recombination breakpoint for (1232) occurs on the first chromosome of a homologous pair, and it is on the second chromosome for (1213). The junction type (1213) becomes (1232) when reversing the ordering of two haplotype of junction type (1213). We have

$$J_{ab}^{mo}(1232) = \frac{1}{2}\left[J_{\female b}^{mo}(1232) + J_{\female b}^{po}(1232)\right]0_{a=b}^{m=o} + \phi_{\female\female b}^{mpo}(123)0_{a=b}^{m=o},$$
$$a \geq b,$$

$$J_{ab}^{po}(1232) = \left\{\frac{1}{2}\left[J_{\male b}^{mo}(1232) + J_{\male b}^{po}(1232)\right] + \phi_{\male b}^{mpo}(123),\right.$$
$$\left. J_{\male b}^{mo}(1232), J_{\male b}^{po}(1232)\right\}0_{a=b}^{p=o}, a \geq b,$$

$$J_{ab}^{o_1 o_2}(1232) = J_{ba}^{o_2 o_1}(1213), a < b,$$

$$J_{ab}^{mo}(1213) = \frac{1}{2}\left[J_{\female b}^{mo}(1213) + J_{\female b}^{po}(1213)\right]0_{a=b}^{m=o}, a \geq b,$$

$$J_{ab}^{po}(1213) = \left\{\frac{1}{2}\left[J_{\male b}^{mo}(1213) + J_{\male b}^{po}(1213)\right], J_{\male b}^{mo}(1213),\right.$$
$$\left. J_{\male b}^{po}(1213)\right\}0_{a=b}^{p=o}, a \geq b,$$

$$J_{ab}^{o_1 o_2}(1213) = J_{ba}^{o_2 o_1}(1232), a < b,$$

for $o, o_1, o_2 \in \{m, p\}$. There are no contributions from the three-gene non-IBD probabilities in the equations for $J_{ab}^{mo}(1213)$ and $J_{ab}^{po}(1213)$, because crossovers between the two parental haplotypes of the tracing haplotype within individual a are invisible.

The boundary conditions are given by the assignment of FGLs to the founders. The within- and between-founder identity junction densities are 0 if a single FGL is assigned to the whole haploid genome of each founder.

RECURSIVE ALGORITHM NON-EXCH

Notations and overview

Let Ω denote the set of distinct FGLs assigned to the founders of a pedigree. We adopt the same notations in the recursive algorithm with FGL exchangeability except for the following changes. As shown in Figure 1, we replace identity states by ancestral states: $(1122) \rightarrow (iijj)$, $(1211) \rightarrow (ijii)$, $(1213) \rightarrow (ijik)$, $(1222) \rightarrow (ijjj)$, and $(1232) \rightarrow (ijkj)$ for the two-chromosome expected junction densities, and $R_a^m \rightarrow R_a^m(ij)$ and $R_a^p \rightarrow R_a^p(ij)$ explicitly for the map expansions, where $i, j, k \in \Omega$ are different from each other (denoted by $i \neq j \neq k$).

We represent a two-gene ancestral state by (ij), and a three-gene ancestral state by (ijk), where $i, j, k \in \Omega$ are not necessary different from each other. In addition we introduce one-gene ancestral coefficient $\phi_a^o(i)$, the probability that the maternally ($o = m$) or paternally ($o = p$) derived gene in individual a carries FGL $i \in \Omega$. The identity coefficients can be obtained by summing the corresponding ancestral coefficients. For example, $\phi_{ab}^{o_1 o_2}(11) = \sum_{i \in \Omega} \phi_{ab}^{o_1 o_2}(ii)$. Similar relationships hold between the expected identity junction densities and the expected ancestral junction densities, for example, $J_{ab}^{o_1 o_2}(1122) = \sum_{i, j \in \Omega, i \neq j} J_{ab}^{o_1 o_2}(iijj)$ and $R_a^o = \sum_{i, j \in \Omega, i \neq j} R_a^o(ij)$.

Similarly, the initial distribution π of the Markov chain can be constructed from the two-gene ancestral coefficients, and the rate matrix Q can be constructed from the expected ancestral junction densities, without assuming the exchangeability of FGLs.

Ancestral coefficients

Bauman et al. (2008) and Zhou et al. (2012) derived the recurrence relations of the one- and two-gene ancestral coefficients for autosomes by tracing two distinct genes simultaneously in their parents. We derive equivalent recurrence relations for both autosomes and sex chromosomes by tracing one gene once in its parent, instead of simultaneously tracing two genes within an individual. In addition, we derive the recurrent relations of the three-gene ancestral coefficients that are required in the recurrence relations of the expected ancestral junction densities.

The relations of the ancestral coefficients are very similar to those of the identity coefficients, and they are based on the same rules of Mendelian inheritance. For example for the one-gene ancestral coefficient $\phi_a^o(i)$ of non-founder a, we have

$$\phi_a^m(i) = \frac{1}{2}\left[\phi_\female^m(i) + \phi_\female^p(i)\right],$$

$$\phi_a^p(i) = \left\{\frac{1}{2}\left[\phi_\male^m(i) + \phi_\male^p(i)\right], \phi_\male^m(i), \phi_\male^p(i)\right\}.$$

See Appendix A for the recurrence relations of the two-gene ancestral coefficient $\phi_{abc}^{o_1 o_2}(ij)$ and three-gene ancestral coefficient $\phi_{abc}^{o_1 o_2 o_3}(ijk)$. The boundary conditions of the recurrence relations of the ancestral coefficients are given by the FGLs assigned to the founders in a pedigree.

Expected ancestral junction densities

The ancestral map expansions can be expressed in terms of the expected ancestral junction densities as follows

$$R_a^m(ij) = J_{aa}^{mp}(iijj) + J_{aa}^{mp}(iiji) + J_{aa}^{mp}(ijjj) + \sum_{k \in \Omega, i \neq j \neq k} J_{aa}^{mp}(ikjk),$$

$$R_a^p(ij) = J_{aa}^{mp}(iijj) + J_{aa}^{mp}(iiij) + J_{aa}^{mp}(ijjj) + \sum_{k \in \Omega, i \neq j \neq k} J_{aa}^{mp}(kikj).$$

Here $R_a^o(ij)$, $o \in \{m, p\}$ refers to the junctions on the single chromosome of a with parental origin o, while the terms on the right-hand side refer to the junctions on the two homologous chromosomes of a. We have $J_{aa}^{mp}(iiji) = J_{aa}^{mp}(jiii) \neq J_{aa}^{mp}(ijjj)$ and $J_{aa}^{mp}(iiij) = J_{aa}^{mp}(ijii) \neq J_{aa}^{mp}(jijj)$, where the equal signs are based on the reversibility of chromosome direction, and the unequal signs refer to the non-exchangeability of FGLs i and j.

The recurrence relations for the ancestral map expansions are given by

$$R_a^m(ij) = \frac{1}{2}\left[R_\female^m(ij) + R_\female^p(ij)\right] + \frac{1}{2}\left[\phi_{\female\female}^{mp}(ij) + \phi_{\female\female}^{pm}(ij)\right]1_{i \neq j},$$

$$R_a^p(ij) = \left\{\frac{1}{2}\left[R_\male^m(ij) + R_\male^p(ij)\right] + \frac{1}{2}\left[\phi_{\male\male}^{mp}(ij) + \phi_{\male\male}^{pm}(ij)\right]1_{i \neq j},\right.$$
$$\left. R_\male^m(ij), R_\male^p(ij)\right\},$$

where the contributions of the two-gene ancestral coefficients account for the asymmetry of FGLs i and j, for example, $\phi_{\female\female}^{mp}(ij) = \phi_{\female\female}^{pm}(ji) \neq \phi_{\female\female}^{pm}(ij)$. A new ancestral junction is formed whenever a recombination event occurs between two homologous chromosomes that have the unordered FGLs i and j at the location of a crossover.

The recurrence relations for $J_{ab}^{o_1 o_2}(iijj)$ are the same as those for $J_{ab}^{o_1 o_2}(1122)$, except that identity states are replaced by ancestral states. We have

$$J_{aa}^{oo}(iijj) = R_a^o(ij),$$

$$J_{aa}^{mp}(iijj) = J_{aa}^{pm}(iijj) = \frac{1}{2}\left[J_{a\female}^{pm}(iijj) + J_{a\female}^{pp}(iijj)\right],$$

$$J_{ab}^{mo}(iijj) = \frac{1}{2}\left[J_{\female b}^{mo}(iijj) + J_{\female b}^{po}(iijj)\right], a > b,$$

$$J_{ab}^{po}(iijj) = \left\{\frac{1}{2}\left[J_{\male b}^{mo}(iijj) + J_{\male b}^{po}(iijj)\right], J_{\male b}^{mo}(iijj), J_{\male b}^{po}(iijj)\right\}, a > b,$$

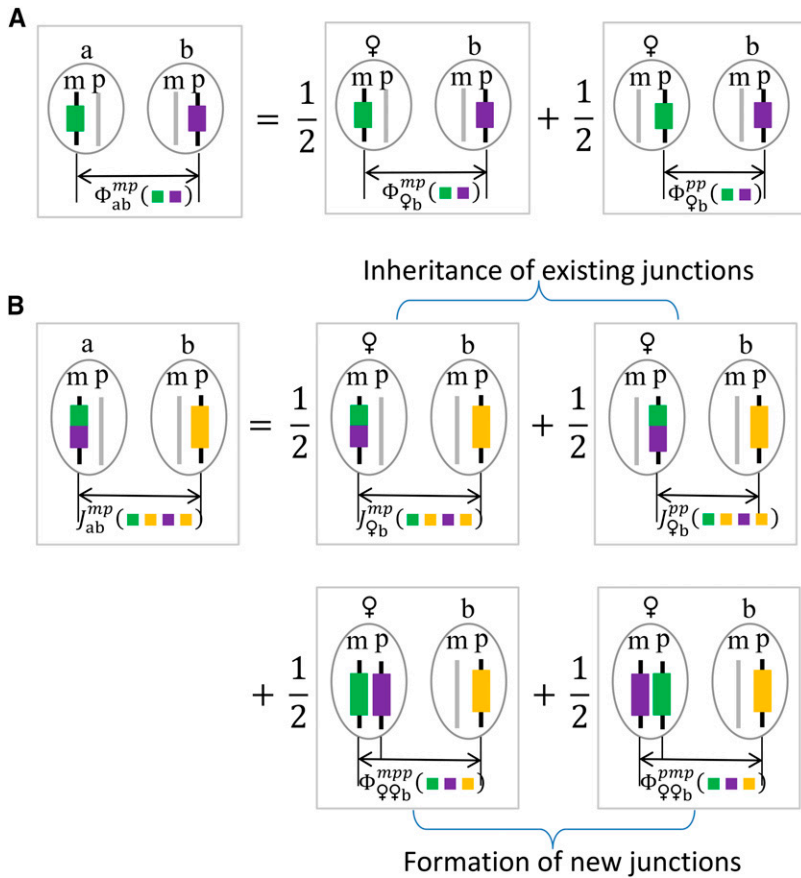$$J_{ab}^{o_1 o_2}(iijj) = J_{ba}^{o_2 o_1}(jjii), a < b,$$

Figure 2 Pictorial representation of recurrence equations. (A) The equation for ancestral coefficient $\phi_{ab}^{mp}(ij)$ $(a > b,)$. (B) The equation for ancestral junction density $J_{ab}^{mp}(ijkj)$ $(a > b)$. The FGLs i, j, and k are shown by different colors, and the irrelevant chromosomes are shown as gray. The symbol ♀ denotes the mother of individual a, and see Table 1 for other notations.

for $o, o_1, o_2 \in \{m, p\}$, where the last equation is obtained by reversing the ordering of the two haplotypes. See Appendix B for the recurrence relations of $J_{ab}^{o_1 o_2}(ijkj)$, $J_{ab}^{o_1 o_2}(ijik)$ $J_{ab}^{o_1 o_2}(ijjj)$, and $J_{ab}^{o_1 o_2}(ijii)$.

As before, the within- and between-founder ancestral junction densities are 0 if a single FGL is assigned to the whole haploid genome of each founder.

Data availability

Ancestral inference using the two recursive algorithms is implemented in the previous developed RABBIT software, where the function magicReconstruct is extended to incorporate pedigree information. RABBIT is available at https://github.com/chaozhi/RABBIT.git, and it is offered under the GNU Affero general public license, version 3 (AGPL-3.0). Example scripts for using the recursive algorithms, simulating genotypic data, and ancestral inference are included. The real CC data have been described by Durrant *et al.* (2011).

APPLICATION TO MULTIPARENTAL POPULATIONS

Simulation evaluation

Before applying the two recursive algorithms EXCH and NON-EXCH to multiparental populations, we evaluate their accuracy by forward simulations on the classical pedigree of Native Americans (Figure 3) (Chapman and Jacquard 1971). The pedigree has been previously used for evaluating recursive algorithms (Jacquard 1974; Karigl 1981; Garcia-Cortes 2015). We simulate two linkage groups: one pair of homologous autosomes and one pair of sex chromosomes. We assign FGLs to the haploid or diploid genomes of the founders. Each descendant gamete is specified as a list of FGL segments determined by

chromosomal crossovers. The number of crossovers in a linkage group of a gamete follows a Poisson distribution with mean 1. We assume no genetic interference so that the positions of crossovers are independently and randomly distributed across the chromosomes. We obtain simulated results for the pedigree member "M22" (Figure 3) by averaging over 10^6 simulations.

Table 2 shows the comparisons between the numerical results from the recursive algorithm EXCH and the simulated results. A unique FGL is assigned to each haploid genome of each founder, so that in total twelve distinct FGLs are assigned to the fixed founders. The differences between the numerical and simulated results are less than 0.002, which is very likely due to the stochasticity of gene flow from founders to descendants. The identity coefficient $\phi_{aa}^{mp}(11)$ for autosomes is in agreement with the previous result (Karigl 1981).

Table 3 evaluates the recursive algorithm NON-EXCH by the forward simulations. A unique FGL is assigned to the whole diploid genome of each founder, so that in total six distinct FGLs are assigned to the fixed founders. Here we use a different assignment of FGLs because Table 3 was otherwise too large. The results show that founder "J" does not contribute to the maternally derived autosome and X chromosome in offspring "M22", which is relatively straightforward from the pedigree structure in Figure 3. The differences between the numerical and simulated results, including those for $\phi_{aa}^{mp}(ij)$, $J_{aa}^{mp}(iijj)$, $J_{aa}^{mp}(ijii)$, $J_{aa}^{mp}(ijik)$, $J_{aa}^{mp}(ijjj)$, and $J_{aa}^{mp}(ijkj)$, are less than 0.001.

We confirmed that the numerical results from the algorithm NON-EXCH are reduced to those from the algorithm EXCH, by summing over possible FGLs under the given identity pattern. In addition, we confirmed the consistency between the two algorithms in the multiparental populations shown in Figure 4.
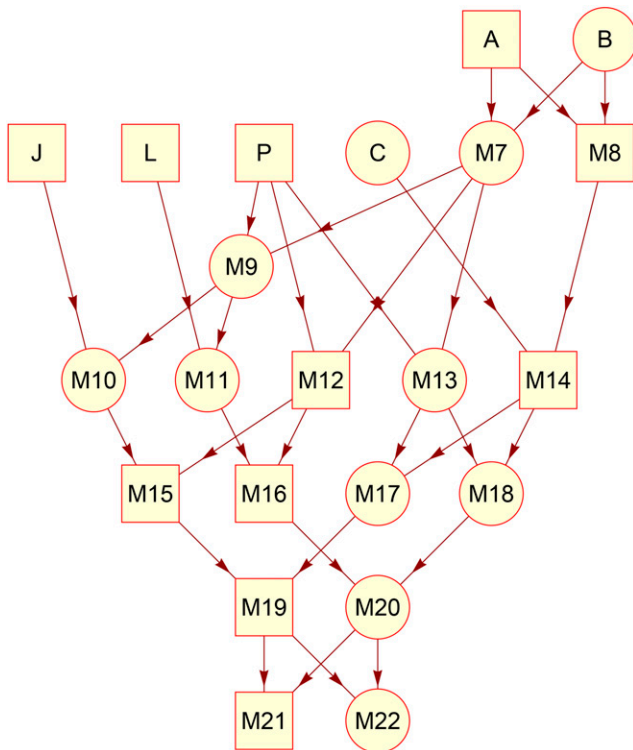
Figure 3 The pedigree of Native Americans. It consists of 6 founders and 16 non-founders. Circles denote females, and rectangles for males.

■ **Table 2 The identity coefficients and the expected identity junction densities for pedigree member "M22"**

Quantity [a]	AA autosomes		XX autosomes	
	Numerical	Simulated	Numerical	Simulated
$\phi_{aa}^{mp}(11)$	0.18359	0.18324	0.23437	0.23459
$J_{aa}^{mp}(1122)$	0.18359	0.18371	0.17969	0.17969
$J_{aa}^{mp}(1211)$	0.61572	0.61516	0.35938	0.35978
$J_{aa}^{mp}(1213)$	2.93652	2.93462	1.47656	1.47587
$J_{aa}^{mp}(1222)$	0.61572	0.61605	0.57031	0.57047
$J_{aa}^{mp}(1232)$	2.93652	2.93824	1.71094	1.71227

[a] A unique FGL is assigned to each haploid genome of each founder in Figure 3.

generations to reach complete fixation of genome as homozygous, and showed that the number of generation increases from the alternating backcross to the father-daughter backcross and to the RIL by sibling. The differences in the number of generation can be partially because the differences of the IBD probabilities for sex chromosomes, and partially because the complete fixation refers to one individual for the alternating backcross and two individuals for the father-daughter backcross.

Figure 5D and F show that the expected overall junction densities ρ_{aa}^{mp} for the alternating backcross and the father-daughter backcross are lower than those of the RIL by sibling for autosomes and sex chromosomes, and that the junction densities for the alternating backcross are slightly smaller than those of the father-daughter backcross for autosomes. These results are consistent with those of Welsh and McMillan (2012): the number of chromosome segments in the final inbred lines increases from the alternating backcross to the father-daughter backcross and to the RIL by sibling.

FGL exchangeability

We apply the recursive algorithm NON-EXCH to study the prior FGL exchangeability for the multi-way funnel breeding schemes (Figure 4). We assume that all founder parents are fully inbred, and assign FGLs from A to H to the eight inbred parents in order from left to right. To examine the FGL exchangeability, we calculate the ancestral coefficient $\phi_{aa}^{mp}(ij)$ and the expected ancestral junction density $J_{aa}^{mp}(iijj)$, the latter being the only ancestral junction type for complete inbred individuals.

Figure 6 shows the exchangeability in terms of $\phi_{aa}^{mp}(ij)$ and $J_{aa}^{mp}(iijj)$ for the females in generation $t = 6$, 11, and 22 for the RIL by sibling. The left panels show the results for autosomes. The FGL non-exchangeability confirms the pedigree inconsistency introduced by Liu *et al.* (2010): each of the four mating pairs of founder parents is impossible at a single locus in an individual, that is, $\phi_{aa}^{mp}(ij) = 0$ for (ji) or $(ij) = (AB)$, (CD), (EF), and (GH). The FGL non-exchangeability shows a different pattern for $J_{aa}^{mp}(iijj)$: there are three levels of expected junction densities and the four mating pairs of founder parents have the highest values. For both $\phi_{aa}^{mp}(ij)$ and $J_{aa}^{mp}(iijj)$, the non-exchangeability diminishes in generation 22 with almost complete inbreeding.

The right panels of Figure 6 show the exchangeability patterns for sex chromosomes of the RIL by sibling. The founder parents D, G, and H cannot pass their X chromosomes beyond F1, and thus their FGLs are impossible in generation $t \geq 2$. Similar to autosomes, $\phi_{aa}^{mp}(ij) = 0$ for (ji) or $(ij) = (AB)$ and (EF). For $\phi_{aa}^{mp}(ij)(i = j)$ or $\phi_a^m(i)$, the probability of FGL C is around twice as large as that of A, B, E or F in generation $t \geq 2$, because the X chromosome carrying FGL C in generation 1 is inherited with probability 1, whereas each of the X chromosomes carrying FGLs A, B, E, and F in generation 1 is inherited with probability 1/2. Similarly, the values of $J_{aa}^{mp}(iijj)$ involving FGL

Quantity [a]	AA autosomes		XX autosomes	
	Numerical	Simulated	Numerical	Simulated
$\phi_a^m(i)$				
A	0.21875	0.21885	0.125	0.12507
B	0.21875	0.21874	0.125	0.12467
J	0	0	0	0
L	0.125	0.12487	0.25	0.25078
P	0.3125	0.31241	0.25	0.24939
C	0.125	0.12513	0.25	0.25009
$R_a^m(ij)$				
AB	0.31250	0.31272	0.14062	0.14098
AJ	0	0	0	0
AL	0.09375	0.09348	0.09375	0.09364
AP	0.28906	0.28880	0.15625	0.15602
AC	0.11719	0.11768	0.09375	0.09330
BJ	0	0	0	0
BL	0.09375	0.09405	0.09375	0.09384
BP	0.28906	0.28890	0.15625	0.15626
BC	0.11719	0.11742	0.09375	0.09331
JL	0	0	0	0
JP	0	0	0	0
JC	0	0	0	0
LP	0.15625	0.15644	0.18750	0.18748
LC	0.03125	0.03124	0.12500	0.12422
PC	0.10938	0.10922	0.18750	0.18735

[a] The label of each founder is used as FGL that is assigned to its diploid genomes.

C increase with inbreeding generation, they are about twice as large as others in generation 22.

We examine the exchangeability patterns for the other multi-way funnel breeding schemes, and focus on autosomes since the FGL exchangeability does not hold in general for sex chromosomes because of the asymmetry between X and Y chromosomes. The non-exchangeability of $\phi_{aa}^{mp}(ij)$ disappears in a completely inbred individual for all 4- 8- and 16-way breeding schemes, because of random chromosomal segregations over many inbreeding generations. However, the FGL exchangeability of $J_{aa}^{mp}(iijj)$ often does not hold even after for a completely inbred individual, for example, see Figure 7 for the breeding schemes after 100 generations of selfing inbreeding.

Ancestral inference

Either one of the two recursive algorithms can be used as a prior to incorporate pedigree information for ancestral inference in multiparental populations from genotypic data, which will be illustrated as follows by simulated and real CC populations.

Simulated CC: We simulate a CC population (Churchill *et al.* 2004) with 100 independent funnels, and take the female at the last generation. The SNP data for the eight founder mouse strains are from Collaborative Cross Consortium (2012), with marker density ∼ 5 SNPs per cM. The genotypic data of a sample female are obtained by simulating the FGLs forwardly and then combining them with the founder SNP data. Allelic errors are assumed to occur independently, and we simulate genotypic data of founders and sampled individuals with the same allelic error probability 0.005.

We analyze four simulated data sets: F11-AA, F11-XX, F22-AA, and F22-XX, where the first part denotes the generation, and the second part denotes the 19 pairs of autosomes (AA) or the sex chromosomes (XX). We perform two analyses for each data set by applying each of the two recursive algorithms as the prior for modeling ancestral origins along two homologous chromosomes within an individual. See Figure 6 for the prior distributions of $\phi_{aa}^{mp}(ij)$ and $J_{aa}^{mp}(iijj)$ as a part of results obtained from the algorithm NON-EXCH. The true allelic error probability is used for the genotypic data of founders and sampled individuals.

Table 4 shows the evaluation of the prior FGL exchangeability on ancestral inference in the simulated CC population. The two analyses of the data set F22-AA are almost the same because there is approximately no prior FGL exchangeability for the autosomes in generation $t = 22$,
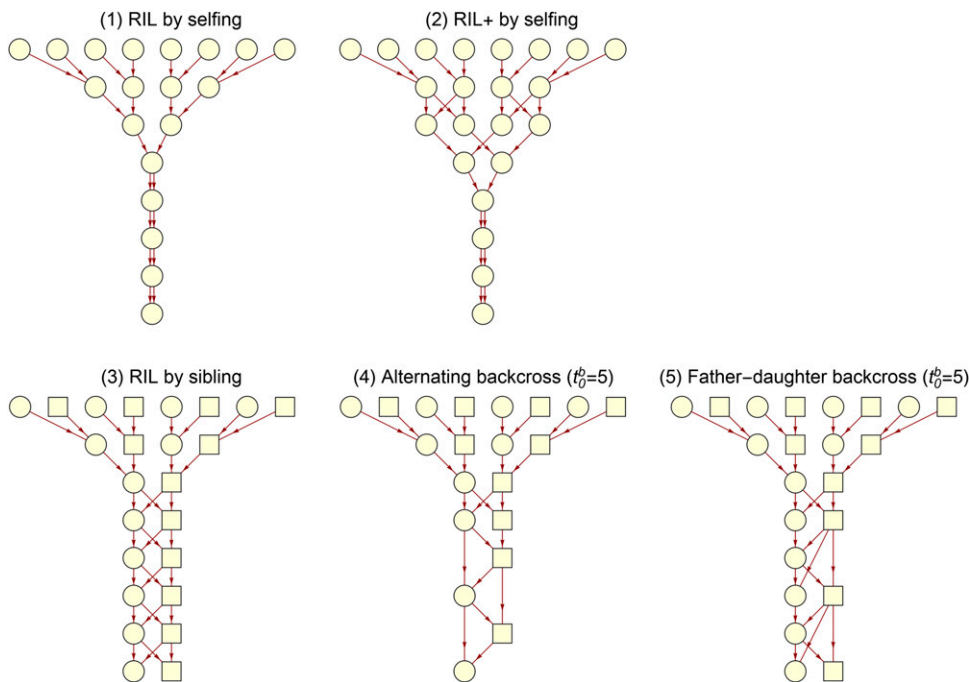

(1) RIL by selfing
(2) RIL+ by selfing
(3) RIL by sibling
(4) Alternating backcross (t_0^b=5)
(5) Father–daughter backcross (t_0^b=5)

Figure 4 Illustration of different 8-way funnel breeding schemes. The alternating backcross scheme alternates between mother-son and father-daughter matings in generation $t \geq t_0^b$, and the father-daughter backcross alternates between father-daughter and random sibling matings in generation $t \geq t_0^b$. Circles denote females, and rectangles for males.
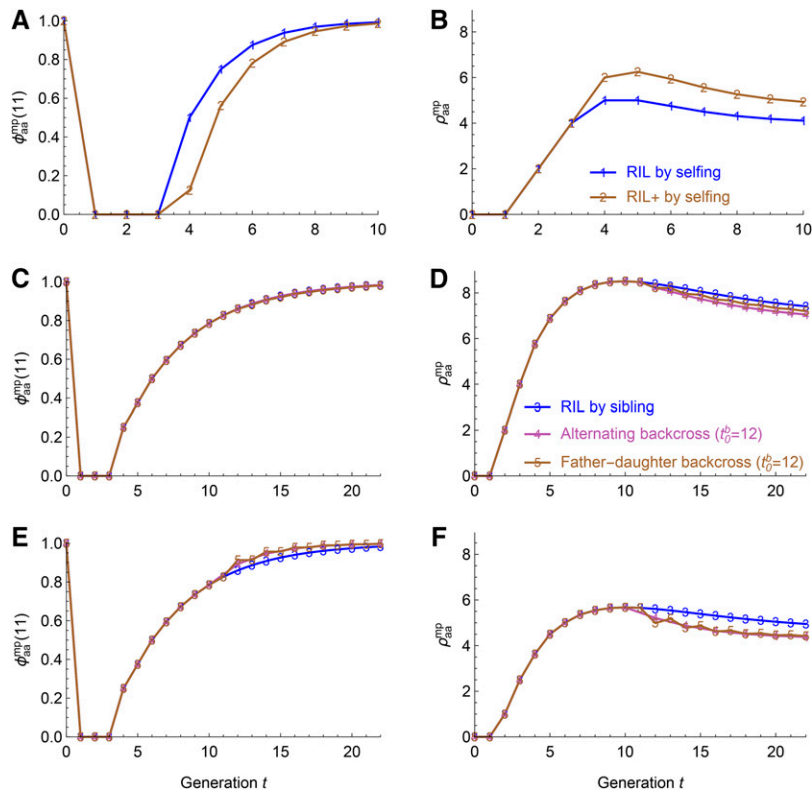
Figure 5 The identity coefficient $\phi_{aa}^{mp}(11)$ and the expected overall junction density ϕ_{aa}^{mp} for the breeding schemes in Figure 4. (A-B) The results for selfing mating schemes 1-2 in Figure 4. (C-D) The results for the autosomes of mating schemes 3-5 with backcross starting from $t_0^b = 12$ in Figure 4. (E-F) The results for the sex chromosomes.

and the improvement of NON-EXCH over EXCH for F11-AA is ∼ 3% because of the FGL non-exchangeability such as that of $J_{aa}^{mp}(iijj)$ in Figure 6. The improvements for sex chromosomes are larger than those for autosomes because of the more pronounced FGL non-exchangeability (Figure 6). The improvement for the data set F22-XX is ∼ 6%, and it increases to ∼ 9% for F11-XX.
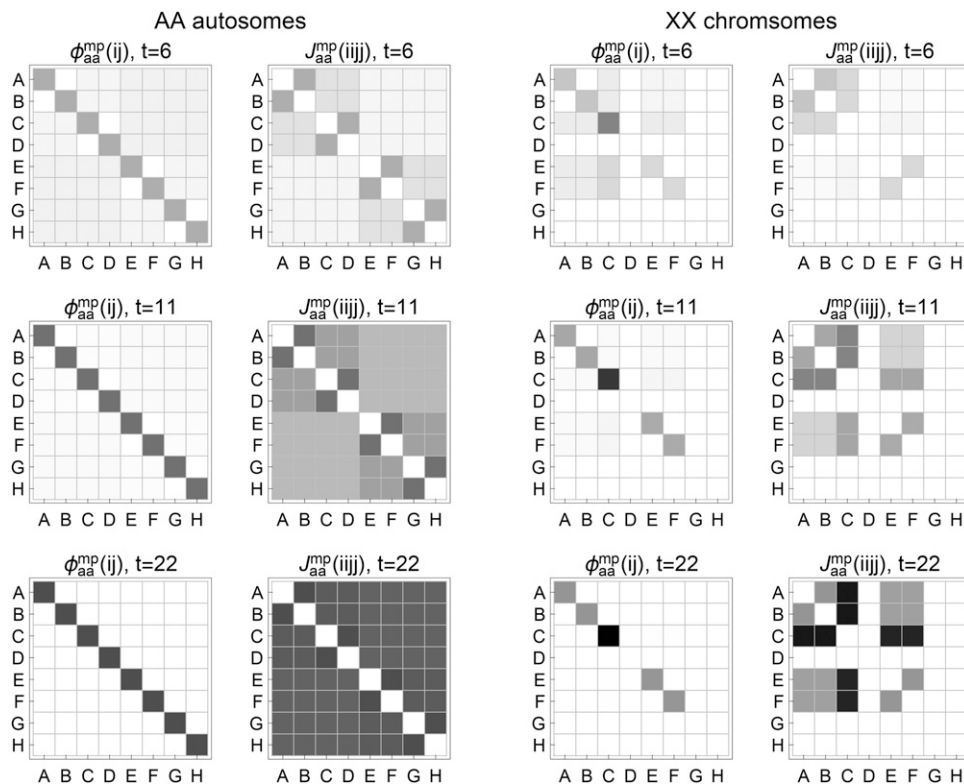

Figure 6 The FGL non-exchangeability patterns for the ancestral co-efficient $\phi_{aa}^{mp}(ij)$ and the expected ancestral junction density $J_{aa}^{mp}(iijj)$ in the eight-way RIL by sibling. The FGLs of eight founders are A-H from left to right.
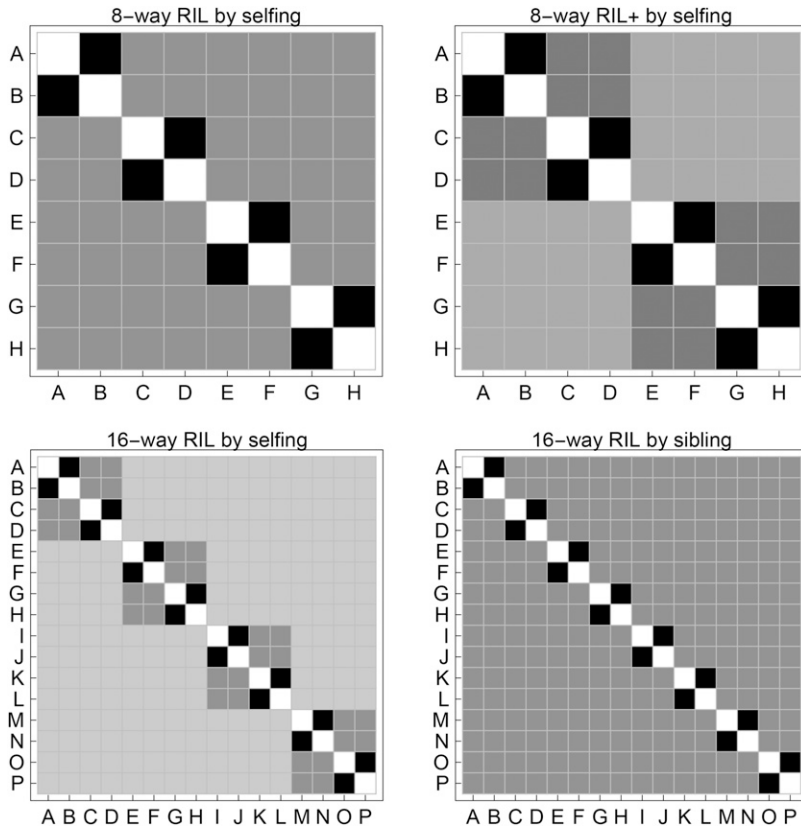
Figure 7 The FGL non-exchangeability patterns of the expected ancestral junction density $J_{aa}^{mp}(iijj)$. The results are for autosomes of the multi-way RIL by 100 generations of selfing or sibling. The FGLs for founder parents are letters starting from A up to P from left to right.

Real CC: The real CC population consists of 120 lines that are sampled at generation t in the range of $8 \leq t \leq 14$ (Durrant *et al.* 2011). For each line, we estimate the sampling generation by using the algorithm EXCH and taking the generation with the maximum likelihood, where the funnel code is not required. Then we estimate the funnel code by using the algorithm NON-EXCH and an iterative maximum likelihood, where a funnel code is proposed by slightly disturbing the current funnel code and it is accepted if the likelihood is increased. We compare the algorithms EXCH and NON-EXCH with GAIN (Liu *et al.* 2010), the latter being specifically developed for the CC.

To study the effect of marker densities on ancestral inference, we analyze only the first pair of homologous autosomes and thin the full dataset by taking every second SNP markers, and repeat to obtain nested sub datasets. The data fractions or the relative marker densities are given by $\rho_M = 1, 2^{-1}, ..., 2^{-7}$. The absolute maximum marker density is

145 SNPs per cM, ~ 28 times higher than that of simulated CC. From the full dataset $\rho_M = 1$, we calculate the marginal posterior probabilities by NON-EXCH, EXCH, and GAIN, and set the pseudo-true values to the most probable ancestral origins if they are the same among the three methods. Overall the pseudo-true ancestral origins for 98.4% observed genotypes are obtained.

Figure 8 shows that the results on ancestral inferences are only slightly different among the three methods. The results of GAIN and NON-EXCH contain no pedigree inconsistencies, whereas NON-EXCH assigns ancestral origins to the impossible mating pairs of founder parents with probability around 0.005. The wrongly assigned probability for GAIN is larger than those of NON-EXCH and EXCH, and the difference increases with the decreasing marker density. The ranking of the wrong called probability from lowest to highest is NON-EXCH, EXCH, and GAIN, and the improvement of NON-EXCH over

■ **Table 4 Evaluation of the prior FGL exchangeability on ancestral inference in the simulated CC population consisting of 100 funnels**

Probability	Simulated data set	EXCH	NON-EXCH	Improvement (%)[a]
Wrongly assigned[b]	F11-AA	0.02779	0.02715	2.3
	F11-XX	0.01821	0.01650	9.4
	F22-AA	0.02484	0.02476	0.3
	F22-XX	0.01540	0.01452	5.7
Wrongly called[c]	F11-AA	0.01962	0.01888	3.8
	F11-XX	0.01284	0.01175	8.5
	F22-AA	0.01766	0.01771	−0.3
	F22-XX	0.01170	0.01084	7.4

[a] The percentage decrease of the wrongly assigned (or called) probability for the analysis using the algorithm NON-EXCH, relative to the algorithm EXCH.
[b] One minus the posterior probability of being true ancestral states.
[c] One minus the fraction of called ancestral states being true ancestral states. At each SNP location within an individual, the ancestral state is called by its maximum posterior probability.

EXCH is around 4.2% and the improvement of NON-EXCH over GAIN is around 7.2% when relative density $\rho_M \geq 1/8$. The similar performances between algorithms NON-EXCH and EXCH indicate that the prior FGL exchangeability is a reasonable approximation for autosomes in the CC.

DISCUSSION

We have developed two novel recursive algorithms for modeling genomic ancestral origins along two homologous chromosomes in a pedigree with arbitrary but known structure. The algorithms apply to both autosomes and sex chromosomes when they exist, and allow selfing in the pedigree. The extensive simulations on a real example pedigree show that the numerical results from the two recursive algorithms are consistent with the simulated results, apart from the stochasticity of gene flow. The Markov property is assumed for the ancestral origin process along two chromosomes, and thus genetic interference is assumed to be absent. The assumption does not affect the expected junction densities calculated from the two recursive algorithms, although it does affect the distribution of inter-junction distances including the variance of junction densities.

One important application of the recursive algorithms is to design a breeding scheme for accelerating the inbreeding process to obtain immortal lines and for maximizing the resolution of mapping QTL. Welsh and McMillan (2012) studied inbreeding processes among different types of multiparental crosses by simulations, and it takes approximately 5.5 hr to complete 100,000 simulations of eight-way RILs. In contrast, it takes less than one second for our recursive algorithms. Rockman and Kruglyak (2008) compared different intercross breeding designs in multiparental RILs by simulations, aiming at increasing the density of junctions (recombination breakpoints) and thus fine-mapping resolution. Our previous recursive algorithms (Zheng *et al.* 2014; Zheng 2015) can be used for calculating the junction density in random mating schemes, and the two new algorithms extend the calculation for any breeding schemes with fixed pedigrees. The two recursive algorithms can also be used in random mating schemes by applying them to many pedigrees that are simulated according to specified mating schemes and averaging the results, which would still require less number of replicates and less computational time than simulation studies.

The second important application of the two recursive algorithms is to provide an appropriate way of incorporating pedigree information for analyzing genotypic data in bi- or multiparental populations. Specifically, the two recursive algorithm can be used to calculate the process parameter values of hidden Markov models for genotypic data, that is, the prior probability distribution of ancestral origins (FGLs) at an initial site and the prior transition probability matrix describing how ancestral origins change along two homologous chromosomes within an individual. See Figure 1 of Zheng (2015) for an example. The application to the ancestral inference in the CC shows that the new algorithms implemented in RABBIT performs only slightly better than GAIN (Figure 8). However, GAIN is specifically designed for the CC, and our previous algorithm applies only to breeding schemes with stage-wise random mating. The new recursive algorithms have pronounced advantages of generality and computational efficiency, and they apply to arbitrary breeding pedigrees that are far away from random mating. For example, one of the multiparental barley populations consists of backcrossing, half-diallel crossing, and selfing, where one inbred founder is much stronger represented in the offspring lines (Liller *et al.* 2017).

Furthermore, we have applied the two recursive algorithms to incorporate pedigree information for genotype imputation in multiparental populations (Zheng *et al.* 2018), and the application for
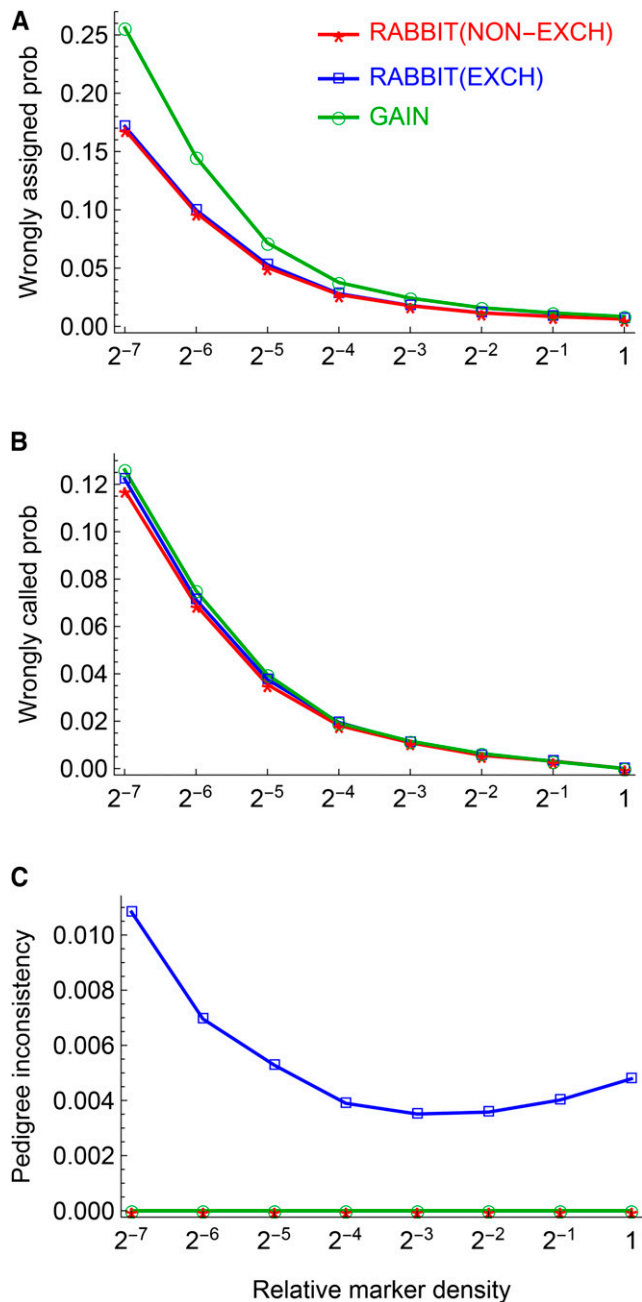

Figure 8 Evaluation of the recursive algorithms that are applied in the ancestral inference for the first pair of autosomes of the 120 real CC lines. Panels (A-C) show the results for the wrongly assigned probability, the wrongly called probability, and the pedigree inconsistency, respectively. The pedigree inconsistency is measured by the sum of the posterior probabilities over the four mating pairs of founder parents.

genetic linkage map construction in multiparental populations is under development.

ACKNOWLEDGMENTS

LITERATURE CITED

Baird, S., N. Barton, and A. Etheridge, 2003   The distribution of surviving blocks of an ancestral genome. Theor. Popul. Biol. 64: 451–471. https://doi.org/10.1016/S0040-5809(03)00098-4

Bauman, L. E., J. S. Sinsheimer, E. M. Sobel, and K. Lange, 2008   Mixed effects models for quantitative trait loci mapping with inbred strains. Genetics 180: 1743–1761. https://doi.org/10.1534/genetics.108.091058

Bennett, J. H., 1953   Junctions in inbreeding. Genetica 26: 392–406. https://doi.org/10.1007/BF01690623

Bennett, J. H., 1954   The distribution of heterogeneity upon inbreeding. J. R. Stat. Soc. Series B Stat. Methodol. 16: 88–99.

Bickeböller, H., and E. A. Thompson, 1996   Distribution of genome shared ibd by half-sibs: approximation by the poisson clumping heuristic. Theor. Popul. Biol. 50: 66–90. https://doi.org/10.1006/tpbi.1996.0023

Broman, K. W., 2012   Genotype probabilities at intermediate generations in the construction of recombinant inbred lines. Genetics 190: 403–412. https://doi.org/10.1534/genetics.111.132647

Chapman, A., and A. Jacquard, 1971   Un isolat d'Amerique centrale: les indiens Jicaques du Honduras, *Génétique et Populations*, Presses Universitaires de France, Paris.

Chapman, N., and E. Thompson, 2003   A model for the length of tracts of identity by descent in finite random mating populations. Theor. Popul. Biol. 64: 141–150. https://doi.org/10.1016/S0040-5809(03)00071-6

Churchill, G. A., D. C. Airey, H. Allayee, J. M. Angel, A. D. Attie *et al.*, 2004   The collaborative cross, a community resource for the genetic analysis of complex traits. Nat. Genet. 36: 1133–1137. https://doi.org/10.1038/ng1104-1133

Cockerham, C. C., and B. S. Weir, 1973   Descent measures for 2 loci with some applications. Theor. Popul. Biol. 4: 300–330. https://doi.org/10.1016/0040-5809(73)90013-0

Collaborative Cross Consortium, 2012   The genome architecture of the collaborative cross mouse genetic reference population. Genetics 190: 389–401. https://doi.org/10.1534/genetics.111.132639

Denniston, C., 1974   Extension of probability approach to genetic relationship - one locus. Theor. Popul. Biol. 6: 58–75. https://doi.org/10.1016/0040-5809(74)90031-8

Donnelly, K. P., 1983   The probability that related individuals share some section of genome identical by descent. Theor. Popul. Biol. 23: 34–63. https://doi.org/10.1016/0040-5809(83)90004-7

Durrant, C., H. Tayem, B. Yalcin, J. Cleak, L. Goodstadt *et al.*, 2011   Collaborative cross mice and their power to map host susceptibility to aspergillus fumigatus infection. Genome Res. 21: 1239–1248. https://doi.org/10.1101/gr.118786.110

Fisher, R., 1949   *The Theory of Inbreeding*, Oliver and Boyd, London.

Fisher, R., 1954   A fuller theory of junctions in inbreeding. Heredity 8: 187–197. https://doi.org/10.1038/hdy.1954.17

Garcia-Cortes, L. A., 2015   A novel recursive algorithm for the calculation of the detailed identity coefficients. Genet. Sel. Evol. 47: 33.

Haldane, J., and C. Waddington, 1931   Inbreeding and linkage. Genetics 16: 357–374.

Hill, W. G., and B. S. Weir, 2007   Prediction of multi-locus inbreeding coefficients and relation to linkage disequilibrium in random mating populations. Theor. Popul. Biol. 72: 179–185. https://doi.org/10.1016/j.tpb.2006.05.006

Huang, B. E., A. W. George, K. L. Forrest, A. Kilian, M. J. Hayden *et al.*, 2012   A multiparent advanced generation inter-cross population for genetic analysis in wheat. Plant Biotechnol. J. 10: 826–839. https://doi.org/10.1111/j.1467-7652.2012.00702.x

Jacquard, A., 1974   *The Genetic Structure of Populations*, Springer-Verlag, New York. https://doi.org/10.1007/978-3-642-88415-3

Johannes, F., and M. Colome-Tatche, 2011   Quantitative epigenetics through epigenomic perturbation of isogenic lines. Genetics 188: 215–227. https://doi.org/10.1534/genetics.111.127118

Karigl, G., 1981   A recursive algorithm for the calculation of identity coefficients. Ann. Hum. Genet. 45: 299–305. https://doi.org/10.1111/j.1469-1809.1981.tb00341.x

Kover, P. X., W. Valdar, J. Trakalo, N. Scarcelli, I. M. Ehrenreich *et al.*, 2009   A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. PLoS Genet. 5: e1000551. https://doi.org/10.1371/journal.pgen.1000551

Li, C. H., Y. X. Li, P. J. Bradbury, X. Wu, Y. S. Shi *et al.*, 2015   Construction of high-quality recombination maps with low-coverage genomic sequencing for joint linkage analysis in maize. BMC Biol. 13: 78. https://doi.org/10.1186/s12915-015-0187-4

Liller, C. B., A. Walla, M. P. Boer, P. Hedley, M. Macaulay *et al.*, 2017   Fine mapping of a major QTL for awn length in barley using a multiparent mapping population. Theor. Appl. Genet. 130: 269–281. https://doi.org/10.1007/s00122-016-2807-y

Liu, E. Y., Q. Zhang, L. McMillan, F. Pardo-Manuel de Villena, and W. Wang, 2010   Efficient genome ancestry inference in complex pedigrees with inbreeding. Bioinformatics 26: i199–i207. https://doi.org/10.1093/bioinformatics/btq187

Mackay, I. J., P. Bansept-Basler, T. Barber, A. R. Bentley, J. Cockram *et al.*, 2014   An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: creation, properties, and validation. G3 (Bethesda) 4: 1603–1610. https://doi.org/10.1534/g3.114.012963

MacLeod, A. K., C. S. Haley, and J. A. Woolliams, 2005   Marker densities and the mapping of ancestral junctions. Genet. Res. 85: 69–79. https://doi.org/10.1017/S0016672305007329

Martin, O. C., and F. Hospital, 2011   Distribution of parental genome blocks in recombinant inbred lines. Genetics 189: 645–654. https://doi.org/10.1534/genetics.111.129700

Nadot, R., and G. Vayssiex, 1973   Algorithme du calcul des coefficients d'identite. Biometrics 29: 347–359. https://doi.org/10.2307/2529397

Pascual, L., N. Desplat, B. E. Huang, A. Desgroux, L. Bruguier *et al.*, 2015   Potential of a tomato magic population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. Plant Biotechnol. J. 13: 565–577. https://doi.org/10.1111/pbi.12282

Rockman, M. V., and L. Kruglyak, 2008   Breeding designs for recombinant inbred advanced intercross lines. Genetics 179: 1069–1078. https://doi.org/10.1534/genetics.107.083873

Rostron, J., 1978   Computation of inbreeding coefficients. Ann. Hum. Genet. 41: 469–475. https://doi.org/10.1111/j.1469-1809.1978.tb00918.x

Sannemann, W., B. E. Huang, B. Mathew, and J. Leon, 2015   Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept. Mol. Breed. 35: 86. https://doi.org/10.1007/s11032-015-0284-7

Stam, P., 1980   The distribution of the fraction of the genome identical by descent in finite random mating populations. Genet. Res. 35: 131–155. https://doi.org/10.1017/S0016672300014002

Stefanov, V., 2000   Distribution of genome shared identical by descent by two individuals in grandparent-type relationship. Genetics 156: 1403–1410.

Teuscher, F., and K. W. Broman, 2007   Haplotype probabilities for multiple-strain recombinant inbred lines. Genetics 175: 1267–1274. https://doi.org/10.1534/genetics.106.064063

Thompson, E. A., 1983   A recursive algorithm for inferring gene origins. Ann. Hum. Genet. 47: 143–152. https://doi.org/10.1111/j.1469-1809.1983.tb00981.x

Thompson, E. A., 1988   2-locus and 3-locus gene identity by descent in pedigrees. IMA J. Math. Appl. Med. Biol. 5: 261–279. https://doi.org/10.1093/imammb/5.4.261

Weir, B. S., and C. C. Cockerham, 1969   Pedigree mating with 2 linked loci. Genetics 61: 923–940.

Welsh, C. E., and L. McMillan, 2012   Accelerating the inbreeding of multiparental recombinant inbred lines generated by sibling matings. G3 (Bethesda) 2: 191–198. https://doi.org/10.1534/g3.111.001784

Xu, S. Z., 2013   Genetic mapping and genomic selection using recombination breakpoint data. Genetics 195: 1103–1115. https://doi.org/10.1534/genetics.113.155309

Zheng, C., 2015   Modeling X-linked-linked ancestral origins in multiparental populations. G3 (Bethesda) 5: 777–801. https://doi.org/10.1534/g3.114.016154

Zheng, C., M. P. Boer, and F. A. van Eeuwijk, 2014   A general modeling framework for genome ancestral origins in multiparental populations. Genetics 198: 87–101. https://doi.org/10.1534/genetics.114.163006

Zheng, C., M. P. Boer, and F. A. van Eeuwijk, 2018   Accurate genotype imputation in multiparental populations from low-coverage sequence. Genetics genetics.300885.2018. https://doi.org/10.1534/genetics.118.300885

Zhou, J. J., A. Ghazalpour, E. M. Sobel, J. S. Sinsheimer, and K. Lange, 2012   Quantitative trait loci association mapping by imputation of strain origins in multifounder crosses. Genetics 190: 459–473. https://doi.org/10.1534/genetics.111.135095

Communicating editor: D. J. de Koning

APPENDIX A: RECURRENCE RELATIONS OF ANCESTRAL COEFFICIENTS

We derive the recurrence relations of one-, two- and three-gene ancestral coefficients. For the one-gene ancestral coefficient $\phi_a^o(i)$ of non-founder a, we have

$$\phi_a^m(i) = \frac{1}{2}\left[\phi_\female^m(i) + \phi_\female^p(i)\right],$$

$$\phi_a^p(i) = \left\{\frac{1}{2}\left[\phi_\male^m(i) + \phi_\male^p(i)\right], \phi_\male^m(i), \phi_\male^p(i)\right\}.$$

And for two-gene ancestral coefficient $\phi_{ab}^{o_1 o_2}(ij)$, we have

$$\phi_{aa}^{oo}(ij) = 1_{i=j}\phi_a^o(i),$$

$$\phi_{aa}^{mp}(ij) = \phi_{aa}^{pm}(ji) = \frac{1}{2}\left[\phi_{a\female}^{pm}(ji) + \phi_{a\female}^{pp}(ji)\right],$$

$$\phi_{ab}^{mo}(ij) = \frac{1}{2}\left[\phi_{\female b}^{mo}(ij) + \phi_{\female b}^{po}(ij)\right], a > b,$$

$$\phi_{ab}^{po}(ij) = \left\{\frac{1}{2}\left[\phi_{\male b}^{mo}(ij) + \phi_{\male b}^{po}(ij)\right], \phi_{\male b}^{mo}(ij), \phi_{\male b}^{po}(ij)\right\}, a > b,$$

$$\phi_{ab}^{o_1 o_2}(ij) = \phi_{ba}^{o_2 o_1}(ji), a < b,$$

for $o, o_1, o_2 \in \{m, p\}$, where the last equation is obtained by reversing the ordering of the two genes. See Figure 2A for a pictorial representation of the equation for $\phi_{ab}^{mp}(ij)$.

The symmetries such as $\phi_{abc}^{o_1 o_2 o_3}(ijk) = \phi_{abc}^{o_1 o_3 o_2}(ikj) = \ldots = \phi_{cba}^{o_3 o_2 o_1}(kji)$ apply to three-gene ancestral coefficients by permuting the three genes, and thus we consider only a particular ordering $a > b > c$, and we have

$$\phi_{aaa}^{ooo}(ijk) = \phi_a^o(i)1_{i=j=k},$$

$$\phi_{aaa}^{o_1 o_1 o_2}(ijk) = \phi_{aa}^{o_1 o_2}(jk)1_{i=j}, o_1 \neq o_2,$$

$$\phi_{aaa}^{o_1 o_2 o_1}(ijk) = \phi_{aaa}^{o_1 o_1 o_2}(ikj), o_1 \neq o_2,$$

$$\phi_{aaa}^{o_2 o_1 o_1}(ijk) = \phi_{aaa}^{o_1 o_1 o_2}(jki), o_1 \neq o_2,$$

$$\phi_{aab}^{o_1 o_1 o_2}(ijk) = \phi_{ab}^{o_1 o_2}(jk)1_{i=j},$$

$$\phi_{aab}^{mpo}(ijk) = \phi_{aab}^{pmo}(jik) = \frac{1}{2}\left[\phi_{a\female b}^{pmo}(jik) + \phi_{a\female b}^{ppo}(jik)\right],$$

$$\phi_{abb}^{o_1 o_2 o_2}(ijk) = \phi_{ab}^{o_1 o_2}(ij)1_{j=k},$$

$$\phi_{abb}^{mo_1 o_2}(ijk) = \frac{1}{2}\left[\phi_{\female bb}^{mo_1 o_2}(ijk) + \phi_{\female bb}^{po_1 o_2}(ijk)\right], o_1 \neq o_2,$$

$$\phi_{abb}^{po_1 o_2}(ijk) = \left\{\frac{1}{2}\left[\phi_{\male bb}^{mo_1 o_2}(ijk) + \phi_{\male bb}^{po_1 o_2}(ijk)\right], \phi_{\male bb}^{mo_1 o_2}(ijk), \phi_{\male bb}^{po_1 o_2}(ijk)\right\}, o_1 \neq o_2,$$

$$\phi_{abc}^{mo_1 o_2}(ijk) = \frac{1}{2}\left[\phi_{\female bc}^{mo_1 o_2}(ijk) + \phi_{\female bc}^{po_1 o_2}(ijk)\right],$$

$$\phi_{abc}^{po_1 o_2}(ijk) = \left\{\frac{1}{2}\left[\phi_{\male bc}^{mo_1 o_2}(ijk) + \phi_{\male bc}^{po_1 o_2}(ijk)\right], \phi_{\male bc}^{mo_1 o_2}(ijk), \phi_{\male bc}^{po_1 o_2}(ijk)\right\},$$

for $o, o_1, o_2 \in \{m, p\}$.

APPENDIX B: RECURRENCE RELATIONS OF EXPECTED ANCESTRAL JUNCTION DENSITIES

The joint recurrence relations of $J_{ab}^{o_1 o_2}(ijkj)$ and $J_{ab}^{o_1 o_2}(ijik)$ are the same as those of $J_{ab}^{o_1 o_2}(1232)$ and $J_{ab}^{o_1 o_2}(1213)$, except that identity states are replaced by ancestral states and that the contributions of three-gene ancestral coefficients account for the asymmetry of FGLs. We have

$$J_{ab}^{mo}(ijkj) = \frac{1}{2}\left[J_{\female b}^{mo}(ijkj) + J_{\female b}^{po}(ijkj)\right]0_{a=b}^{m=o} + \frac{1}{2}\left[\phi_{\female\female b}^{mpo}(ikj) + \phi_{\female\female b}^{pmo}(ikj)\right]1_{i\neq j\neq k}0_{a=b}^{m=o}, a \geq b,$$

$$J_{ab}^{po}(ijkj) = \left\{\frac{1}{2}\left[J_{\male b}^{mo}(ijkj) + J_{\male b}^{po}(ijkj)\right] + \frac{1}{2}\left[\phi_{\male\male b}^{mpo}(ikj) + \phi_{\male\male b}^{pmo}(ikj)\right]1_{i\neq j\neq k},\right.$$
$$\left. J_{\male b}^{mo}(ijkj), J_{\male b}^{po}(ijkj)\right\}0_{a=b}^{p=o}, a \geq b,$$

$$J_{ab}^{o_1 o_2}(ijkj) = J_{ba}^{o_2 o_1}(jijk), a < b,$$

$$J_{ab}^{mo}(ijik) = \frac{1}{2}\left[J_{\female b}^{mo}(ijik) + J_{\female b}^{po}(ijik)\right]0_{a=b}^{m=o}, a \geq b,$$

$$J_{ab}^{po}(ijik) = \left\{\frac{1}{2}\left[J_{\male b}^{mo}(ijik) + J_{\male b}^{po}(ijik)\right], J_{\male b}^{mo}(ijik), J_{\male b}^{po}(ijik)\right\}0_{a=b}^{p=o}, a \geq b,$$

$$J_{ab}^{o_1 o_2}(ijik) = J_{ba}^{o_2 o_1}(jiki), a < b,$$

for $o, o_1, o_2 \in \{m, p\}$, where $J_{ba}^{o_2 o_1}(jijk)$ in the 3rd equation and $J_{ba}^{o_2 o_1}(jiki)$ in the 6th equation can be transformed into the form of $J_{ba}^{o_2 o_1}(ijik)$ and $J_{ba}^{o_2 o_1}(ijkj)$, respectively, by swapping i and j. There are no contributions of the three-gene ancestral coefficients in the equations for $J_{ab}^{mo}(ijik)$ and $J_{ab}^{po}(ijik)$ because crossovers between the two parental haplotypes of the tracing haplotype within individual a are invisible. See Figure 2B for a pictorial representation of the equation for $J_{ab}^{mp}(ijkj)$.

The joint recurrence relations of $J_{ab}^{o_1 o_2}(ijjj)$ and $J_{ab}^{o_1 o_2}(ijii)$ are the same to those of $J_{ab}^{o_1 o_2}(ijkj)$ and $J_{ab}^{o_1 o_2}(ijik)$, except that the contributions of the three-gene ancestral coefficients have slightly different forms. We have

$$J_{ab}^{mo}(ijjj) = \frac{1}{2}\left[J_{\female b}^{mo}(ijjj) + J_{\female b}^{po}(ijjj)\right]0_{a=b}^{m=o} + \frac{1}{2}\left[\phi_{\female\female b}^{mpo}(ijj) + \phi_{\female\female b}^{pmo}(ijj)\right]1_{i\neq j}0_{a=b}^{m=o}, a \geq b,$$

$$J_{ab}^{po}(ijjj) = \left\{\frac{1}{2}\left[J_{\male b}^{mo}(ijjj) + J_{\male b}^{po}(ijjj)\right] + \frac{1}{2}\left[\phi_{\male\male b}^{mpo}(ijj) + \phi_{\male\male b}^{pmo}(ijj)\right]1_{i\neq j},\right.$$
$$\left. J_{\male b}^{mo}(ijjj), J_{\male b}^{po}(ijjj)\right\}0_{a=b}^{p=o}, a \geq b,$$

$$J_{ab}^{o_1 o_2}(ijjj) = J_{ba}^{o_2 o_1}(jijj), a < b,$$

$$J_{ab}^{mo}(ijii) = \frac{1}{2}\left[J_{\female b}^{mo}(ijii) + J_{\female b}^{po}(ijii)\right]0_{a=b}^{m=o}, a \geq b,$$

$$J_{ab}^{po}(ijii) = \left\{\frac{1}{2}\left[J_{\male b}^{mo}(ijii) + J_{\male b}^{po}(ijii)\right], J_{\male b}^{mo}(ijii), J_{\male b}^{po}(ijii)\right\}0_{a=b}^{p=o}, a \geq b,$$

$$J_{ab}^{o_1 o_2}(ijii) = J_{ba}^{o_2 o_1}(jiii), a < b,$$

for $o, o_1, o_2 \in \{m, p\}$, where $J_{ba}^{o_2 o_1}(jijj)$ in the 3rd equation and $J_{ba}^{o_2 o_1}(jiii)$ in the 6th equation can be transformed into the form of $J_{ba}^{o_2 o_1}(ijii)$ and $J_{ba}^{o_2 o_1}(ijjj)$, respectively, by swapping i and j.