

The Genomes of Recombinant Inbred Lines

Karl W. Broman¹

Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland 21205

Manuscript received August 20, 2004

Accepted for publication November 5, 2004

ABSTRACT

Recombinant inbred lines (RILs) can serve as powerful tools for genetic mapping. Recently, members of the Complex Trait Consortium proposed the development of a large panel of eight-way RILs in the mouse, derived from eight genetically diverse parental strains. Such a panel would be a valuable community resource. The use of such eight-way RILs will require a detailed understanding of the relationship between alleles at linked loci on an RI chromosome. We extend the work of Haldane and Waddington on two-way RILs and describe the map expansion, clustering of breakpoints, and other features of the genomes of multiple-strain RILs as a function of the level of crossover interference in meiosis.

RECOMBINANT inbred lines (RILs) can serve as powerful tools for genetic mapping. An RIL is formed by crossing two inbred strains followed by repeated selfing or sibling mating to create a new inbred line whose genome is a mosaic of the parental genomes (Figure 1). As each RIL is an inbred strain, and so can be propagated eternally, a panel of RILs has a number of advantages for genetic mapping: one need genotype each strain only once; one can phenotype multiple individuals from each strain to reduce individual, environmental, and measurement variability; multiple invasive phenotypes can be obtained on the same set of genomes; and, as the breakpoints in RILs are more dense than those that occur in any one meiosis, greater mapping resolution can be achieved.

Members of the Complex Trait Consortium recently proposed the development of a large panel of eight-way RILs in the mouse (THREADGILL *et al.* 2002; COMPLEX TRAIT CONSORTIUM 2004). An eight-way RIL is formed by intermating eight parental inbred strains, followed by repeated sibling mating to produce a new inbred line whose genome is a mosaic of the eight parental strains (Figure 2). Such a panel would serve as a valuable community resource for mapping the loci that contribute to complex phenotypes in the mouse.

The use of such a panel requires a detailed understanding of the relationship between alleles at linked loci on a recombinant inbred (RI) chromosome, particularly for the reconstruction of the parental origin of DNA (the haplotypes) on the basis of less-than-fully informative genetic markers [such as single-nucleotide polymorphisms (SNPs)]. The haplotype reconstruction will likely make use of the standard hidden Markov model (HMM) tech-

nology, and a primary ingredient to the HMM will be the two-point haplotype probabilities on an RIL chromosome, such as the probability that the RIL is fixed at allele *A* at one locus and allele *E* at a second locus, as a function of the recombination fraction (per meiosis) between the two loci. Also of interest are the three-point haplotype probabilities, which inform us regarding the clustering of breakpoints on the RIL chromosome and of the appropriateness of the Markov assumption used in the HMM.

HALDANE and WADDINGTON (1931; which we abbreviate H&W) studied the case of two-way RILs by selfing and sibling mating and, in an impressive feat of algebra, derived the relationship between the recombination fraction between two loci and the probability that the loci are fixed at different parental alleles in the RIL. They further showed that such two-point results may be used to derive three-point probabilities. Related work includes that of HOSPITAL *et al.* (1996), who developed an algorithm for calculating multilocus genotype probabilities in two-way RILs, and WINKLER *et al.* (2003), who considered the case of two-way RILs subjected to extra generations of outbreeding.

In this article, we extend the work of H&W to the case of eight-way RILs. We derive the algebraic relationship between the recombination fraction at meiosis and the analogous quantity for the RIL chromosome, for the case of multiple-strain RILs by selfing and sibling mating (including the X chromosome in the case of sibling mating). In the case of multiple-strain RILs by selfing, we also obtain exact results for three-point probabilities. However, with multiple-strain RILs by sibling mating, such symbolic results for the three-point probabilities continue to elude us, and so we must be satisfied with numerical results.

A number of other features of the genomes of multiple-strain RILs are of considerable interest and yet are not amenable to such symbolic or numerical analysis

¹Address for correspondence: Department of Biostatistics, Johns Hopkins University, 615 N. Wolfe St., Baltimore, MD 21205-2179.
E-mail: kbroman@jhsphe.edu

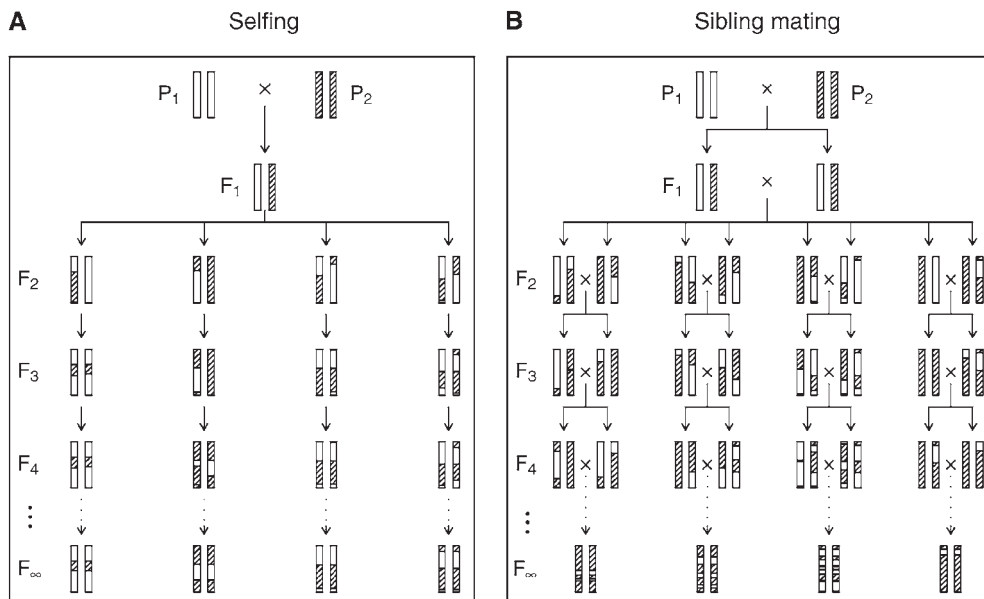


FIGURE 1.—The production of recombinant inbred lines by selfing (A) and by sibling mating (B).

(for example, the number of generations of inbreeding required to obtain complete homozygosity). We investigated such features via computer simulations.

TWO POINTS

We first consider the case of two loci. Let r denote the recombination fraction between the two loci, and let g_m denote the allele at locus m on a random RI chromosome at fixation. We seek the joint distribution of the g_m . We are particularly interested in $R = \Pr(g_1 \neq g_2)$, the quantity analogous to the recombination fraction, but on the fixed RI chromosome. Note that we assume no mutation and no selection. The alleles at each locus are denoted $A, B, C, D, \dots, H$. Two-locus

haplotypes are written, for example, AA, AB, BA, BB , where the first allele corresponds to the first locus and the second allele to the second locus. Two-locus diploypes (*i.e.*, phase-known genotypes) are written, for example, $AB|AB$ (an individual who is homozygous A at the first locus and homozygous B at the second locus). It might be better to write the diploype as

$$\begin{array}{c|c} A & A \\ \hline B & B \end{array}$$

but such notation is unwieldy.

We assume that two-way RILs are obtained with an initial cross of the form $(A \times B) \times (A \times B)$, four-way RILs by the cross $(A \times B) \times (C \times D)$, and eight-way

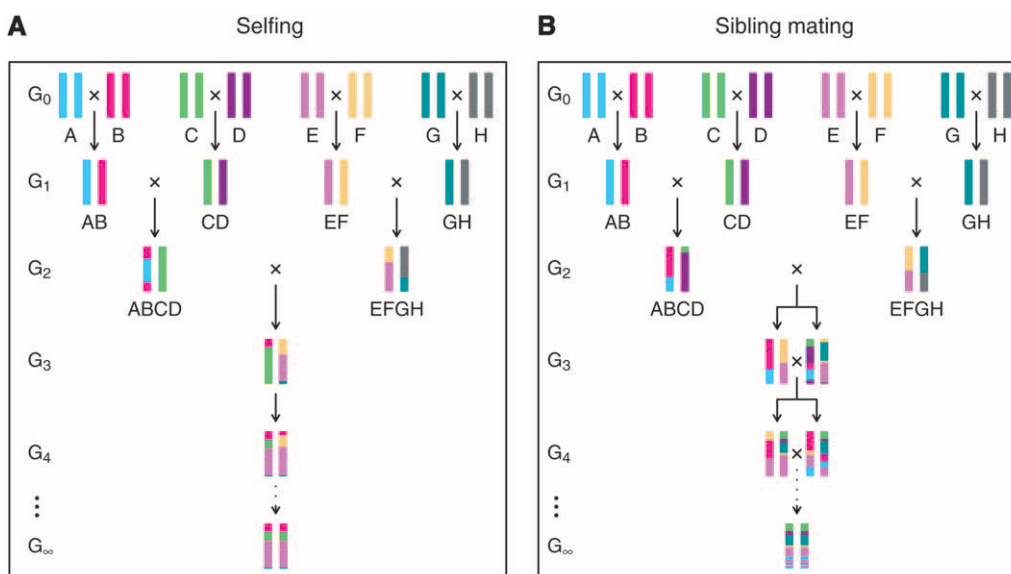


FIGURE 2.—The production of an eight-way recombinant inbred line by selfing (A) and by sibling mating (B).

TABLE 1

Equivalence classes of two-locus diplotypes states in the formation of two-way RILs by selfing

Prototype state	All possible states
AA AA	AA AA BB BB
AB AB	AB AB BA BA
AA AB	AA AB AA BA AB BB BA BB
AA BB	AA BB
AB BA	AB BA

RILs by the cross $[(A \times B) \times (C \times D)] \times [(E \times F) \times (G \times H)]$, with, in all cases, females listed first.

RILs by selfing: *Two-way RILs, selfing:* The results for two-way RILs by selfing were presented in H&W. By symmetry, it is clear that $\Pr(g_m = A) = \Pr(g_m = B) = 1/2$. They further showed that the two-point probabilities are

$$\Pr(g_1 = i, g_2 = j) = \begin{cases} \frac{1}{2(1+2r)} & \text{if } i = j \\ \frac{r}{1+2r} & \text{if } i \neq j. \end{cases}$$

Thus $R = \Pr(g_1 \neq g_2) = 2r/(1+2r)$.

We describe a general approach to obtain this result, as the technique is used in what follows and is most clear in this, the simplest case. Let X_n denote the two-locus diplotype for the individual at generation n . The $\{X_n\}$ form a Markov chain, as X_{n+1} is conditionally independent of $X_0, X_1, \dots, X_{n-1}$, given X_n . That is, the parental diplotype at a particular generation depends only on the diplotype at the preceding generation and not on the entire history. There are 10 possible diploypes, as the order of the two haplotypes may be ignored. (If haplotype order were taken into account, there would be $2^4 = 16$ states.) This number may be reduced further by accounting for further symmetries: The order of the two loci may be ignored, and the symbols A and B may be switched. Thus we form 5 distinct states, shown in Table 1. Let Y_n denote the state at generation n , among these 5 minimal states. The $\{Y_n\}$ also form a Markov chain.

Let $P_{ij} = \Pr(Y_{n+1} = j | Y_n = i)$, the transition matrix for the chain. Calculation of the P_{ij} deserves further explanation. Consider the state AA|BB, corresponding to heterozygosity at each locus, with the A alleles on the same haplotype. The possible meiotic products are AA, AB, BA, and BB, with probabilities $(1-r)/2$, $r/2$, $r/2$, and $(1-r)/2$, respectively. The probabilities for the states in the next generation may be obtained by calculating the Kronecker product of this vector with itself and then collapsing the 16 probabilities to give the probabilities of the five states in Table 1. Thus the probabilities of transition from state AA|BB to states AA|AA,

AB|AB, AA|AB, AA|BB, and AB|BA are $(1-r)^2/2$, $r^2/2$, $2r(1-r)$, $(1-r)^2/2$, and $r^2/2$, respectively.

The states AA|AA and AB|AB are absorbing; for these states, $P_{ii} = 1$. Our goal is to obtain the absorption probabilities starting at state AA|BB (for example, starting at the state AA|BB, the chance that the chain will eventually hit the state AA|AA). These absorption probabilities may be obtained as the solutions of sets of linear equations (NORRIS 1997, Sect. 1.3).

Let h_i denote the probability, starting at state i , that the chain is absorbed into the state AA|AA. Clearly $h_{AA|AA} = 1$ and $h_{AB|AB} = 0$. For the other three states, we condition on the first step and obtain $h_i = \sum_k P_{ik} h_k$. Thus we obtain a set of three linear equations in three unknowns, which may be solved to obtain $h_{AA|BB} = 1/(1+2r)$. Thus $\Pr(Y_n \rightarrow AA|AA | Y_0 = AA|BB) = 1/(1+2r)$. The states within an equivalence class are equally likely, and so $\Pr(X_n \rightarrow AA|AA | X_0 = AA|BB) = 1/[2(1+2r)]$.

Four-way RILs, selfing: The results for two-way RILs by selfing may be extended immediately to obtain those for four-way RILs by selfing, by considering one preceding generation of recombination, as the two-chromosome generation (in which inbreeding begins) is a bottleneck: Alleles that do not appear in this generation cannot appear on the final RI chromosome.

For example, the chance that the final haplotype in a four-way RIL by selfing is AA is the probability that in the initial cross of $AA \times BB$, the AA haplotype is transmitted, multiplied by the probability that a two-way RIL by selfing is fixed at AA. Thus, $\Pr(AA) = (1-r)/2 \times 1/[2(1+2r)]$. Similarly, the chance that the final haplotype is AB is the chance that the initial cross of $AA \times BB$ delivers AB, multiplied by the chance that a two-way RIL by selfing is fixed at AA, and so $\Pr(AB) = r/2 \times 1/[2(1+2r)]$. Finally, the chance that the final haplotype is AC is the chance that the initial cross of $AA \times BB$ delivers A at the first locus, multiplied by the chance that the cross of $CC \times DD$ delivers C at the second locus, multiplied by the chance that a two-way RIL by selfing is fixed at AB, and so $\Pr(AC) = 1/2 \times 1/2 \times r/(1+2r)$.

For four-way RILs, the marginal probabilities are of course $\Pr(g_m = i) = 1/4$ for $i = A, B, C, D$. The two-locus probabilities are

$$\Pr(g_1 = i, g_2 = j) = \begin{cases} \frac{1-r}{4(1+2r)} & \text{if } i = j \\ \frac{r}{4(1+2r)} & \text{if } i \neq j. \end{cases}$$

Thus $R = \Pr(g_1 \neq g_2) = 3r/(1+2r)$. When $r = 1/2$, $R = 3/4$.

Eight-way RILs, selfing: The results for eight-way RILs can be deduced from the results for four-way RILs, by the same technique that allowed us to obtain the results for four-way RILs by selfing from those for two-way RILs by selfing.

TABLE 2
Two-locus haplotype probabilities for eight-way RILs by selfing

	A	B	C	D	E	F	G	H
A	$\frac{(1-r)^2}{8(1+2r)}$	$\frac{r(1-r)}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$
B	$\frac{r(1-r)}{8(1+2r)}$	$\frac{(1-r)^2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$
C	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{(1-r)^2}{8(1+2r)}$	$\frac{r(1-r)}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$
D	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r(1-r)}{8(1+2r)}$	$\frac{(1-r)^2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$
E	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{(1-r)^2}{8(1+2r)}$	$\frac{r(1-r)}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$
F	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r(1-r)}{8(1+2r)}$	$\frac{(1-r)^2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$
G	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{(1-r)^2}{8(1+2r)}$	$\frac{r(1-r)}{8(1+2r)}$
H	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r/2}{8(1+2r)}$	$\frac{r(1-r)}{8(1+2r)}$	$\frac{(1-r)^2}{8(1+2r)}$

For eight-way RILs by selfing, the marginal probabilities are $\Pr(g_m = i) = 1/8$ for $i = A, B, \dots, H$. The two-locus probabilities are shown in Table 2. It is especially interesting in Table 2 that the off-diagonal elements are not all the same. The results in Table 2 give $R = r(4-r)/(1+2r)$. When $r = 1/2$, $R = 1/8$.

One can easily go backward, from eight-way RILs to four-way RILs, by taking $A = B$, $C = D$, $E = F$, $G = H$, and collapsing the joint probabilities. Similarly, taking $A = B = C = D$ and $E = F = G = H$, one can collapse to obtain the results for two-way RILs.

X chromosome for RILs by sibling mating: *Two-way RILs, X chromosome:* H&W derived the connection between r and R for the X chromosome for two-way RILs by sibling mating. The full two-point distribution may be obtained from their result, using the marginal distribution $\Pr(g_m = A) = 2/3$, $\Pr(g_m = B) = 1/3$, and that $\Pr(AB) = \Pr(BA)$. The two-locus probabilities are shown

TABLE 3

Two-locus haplotype probabilities for the X chromosome in two-way RILs by sibling mating

	A	B
A	$\frac{2(1+2r)}{3(1+4r)}$	$\frac{4r}{3(1+4r)}$
B	$\frac{4r}{3(1+4r)}$	$\frac{1}{3(1+4r)}$

in Table 3. These results give $R = (8/3)r/(1+4r)$. When $r = 1/2$, $R = 1/9$.

Four-way RILs, X chromosome: The case of four-way RILs by sibling mating cannot be deduced from the above, but we were able to calculate the results symbolically using a combination of R (IHAKA and GENTLEMAN 1996) and Mathematica (WOLFRAM RESEARCH 2003), by the approach described above, in *Two-way RILs by selfing*.

Let X_n denote the parental type at the n th generation, with $X_0 = AA|BB \times CC$. There are 405 such parental types, but they may be reduced to 116 distinct states by taking account of two symmetries: The order of the two loci may be reversed, and the A and B alleles may be exchanged. There are four absorbing states: $AA|AA \times AA$, $AB|AB \times AB$, $AC|AC \times AC$, and $CC|CC \times CC$. The determination of the absorption probabilities again requires the solution of systems of linear equations, in this case 112 equations in 112 unknowns, since there are a total of 116 states, of which 4 are absorbing.

The marginal probabilities are $\Pr(g_m = i) = 1/3$ for $i = A, B, C$. The two-locus probabilities are the following:

$$\Pr(g_1 = i, g_2 = j) = \begin{cases} \frac{1}{3(1+4r)} & \text{if } i = j \\ \frac{2r}{3(1+4r)} & \text{if } i \neq j. \end{cases}$$

Thus $R = 4r/(1+4r)$. When $r = 1/2$, $R = 2/3$.

Eight-way RILs, X chromosome: The case of eight-way RILs can be deduced from the case of four-way RILs (due to the bottleneck at the four-chromosome stage),

TABLE 4
Two-locus haplotype probabilities for the X chromosome in eight-way RILs by sibling mating

	A	B	C	E	F
A	$\frac{1-r}{6(1+4r)}$	$\frac{r}{6(1+4r)}$	$\frac{2r}{6(1+4r)}$	$\frac{r}{6(1+4r)}$	$\frac{r}{6(1+4r)}$
B	$\frac{r}{6(1+4r)}$	$\frac{1-r}{6(1+4r)}$	$\frac{2r}{6(1+4r)}$	$\frac{r}{6(1+4r)}$	$\frac{r}{6(1+4r)}$
C	$\frac{2r}{6(1+4r)}$	$\frac{2r}{6(1+4r)}$	$\frac{2}{6(1+4r)}$	$\frac{2r}{6(1+4r)}$	$\frac{2r}{6(1+4r)}$
E	$\frac{r}{6(1+4r)}$	$\frac{r}{6(1+4r)}$	$\frac{2r}{6(1+4r)}$	$\frac{1-r}{6(1+4r)}$	$\frac{r}{6(1+4r)}$
F	$\frac{r}{6(1+4r)}$	$\frac{r}{6(1+4r)}$	$\frac{2r}{6(1+4r)}$	$\frac{r}{6(1+4r)}$	$\frac{1-r}{6(1+4r)}$

by the technique described above, for the case of RILs by selfing. For example, the chance that an eight-way RIL is fixed at AB on the X chromosome is equal to the chance that, in the $AA \times BB$ cross, the AB haplotype is transmitted, times the chance that a four-way RIL is fixed at AA , giving $r/2 \times 1/[3(1+4r)]$.

The marginal probabilities are $\Pr(g_m = A) = \Pr(g_m = B) = \Pr(g_m = E) = \Pr(g_m = F) = 1/6$ and $\Pr(g_m = C) = 1/3$. The joint two-locus probabilities are shown in Table 4. It follows that $R = (14/3)r/(1+4r)$. When $r = 1/2$, $R = 7/6$.

Autosomes for RILs by sibling mating: *Two-way RILs, autosomes by sib mating:* H&W provided the results for the autosome in two-way RILs by sibling mating. The marginal distribution is $\Pr(g_m = A) = \Pr(g_m = B) = 1/2$. The two-locus joint probabilities are

$$\Pr(g_1 = i, g_2 = j) = \begin{cases} \frac{1+2r}{2(1+6r)} & \text{if } i = j \\ \frac{2r}{1+6r} & \text{if } i \neq j. \end{cases}$$

Thus $R = 4r/(1+6r)$. When $r = 1/2$, $R = 1/2$.

Four-way RILs, autosomes by sib mating: The case of four-way RILs cannot be deduced from the above. Let X_n denote the parental type at generation n , with $X_0 = AA|BB \times CC|DD$. There are 9316 such states, which reduce to 700 distinct states after we take account of several symmetries: reversing the order of the two loci, exchanging the A and B alleles, exchanging the C and D alleles, exchanging A for C and B for D , and any combination of these. There are three distinct absorbing states, $AA|AA \times AA|AA$, $AB|AB \times AB|AB$, and $AC|AC \times AC|AC$. The determination of the absorption probabilities requires the solution of a system of 697 linear equations in 697 unknowns.

This system of equations proved too large to solve symbolically; however, the numerical solution for any particular value of the recombination fraction r was relatively

simple to obtain, and we could infer the algebraic forms of the equations. Insertion of the inferred solution back into the system of equations demonstrated that our algebraic results are correct.

The marginal distribution is $\Pr(g_m = i) = 1/4$ for $i = A, B, C, D$. The two-locus joint probabilities are

$$\Pr(g_1 = i, g_2 = j) = \begin{cases} \frac{1}{4(1+6r)} & \text{if } i = j \\ \frac{r}{2(1+6r)} & \text{if } i \neq j. \end{cases}$$

Thus $R = 6r/(1+6r)$. When $r = 1/2$, $R = 3/4$.

Eight-way RILs, autosomes by sib mating: The case of eight-way RILs can be deduced from the results for four-way RILs. The marginal distribution is $\Pr(g_m = i) = 1/8$ for $i = A, B, \dots, H$. The two-locus joint probabilities are

$$\Pr(g_1 = i, g_2 = j) = \begin{cases} \frac{1-r}{8(1+6r)} & \text{if } i = j \\ \frac{r}{2(1+6r)} & \text{if } i \neq j. \end{cases}$$

Thus $R = 7r/(1+6r)$. (This was the key target of all of our efforts.) When $r = 1/2$, $R = 7/8$. Note that here, all off-diagonal elements are the same.

Higher-order RILs: The two-point probabilities for 2^n -way RILs for any n may be derived from the results presented above, although here we sought to obtain only the relationship between r and R . These results are assembled in Table 5. The case of the X chromosome proved cumbersome and a closed-form solution elusive. Omitting some rather tedious algebra, we have, for the X chromosome of a 2^n -way RIL by sibling mating with $n \geq 2$,

$$R = 1 - \frac{1}{3(1+4r)} \sum_{k=1}^n x_{nk} \left(\frac{1-r}{2} \right)^{k-1},$$

TABLE 5
Crossover probabilities on recombinant inbred
line chromosomes

	Selfing	Sibling mating	
		X chromosome	Autosome
2-way	$\frac{2r}{1+2r}$	$\frac{(8/3)r}{1+4r}$	$\frac{4r}{1+6r}$
4-way	$\frac{3r}{1+2r}$	$\frac{4r}{1+4r}$	$\frac{6r}{1+6r}$
8-way	$\frac{r(4-r)}{1+2r}$	$\frac{(14/3)r}{1+4r}$	$\frac{7r}{1+6r}$
16-way	$\frac{r(5-3r+r^2)}{1+2r}$	$\frac{r(16-r)}{3(1+4r)}$	$\frac{r(8-r)}{1+6r}$
2 ⁿ -way	$1 - \frac{(1-r)^{n-1}}{1+2r}$	See text	$1 - \frac{(1-r)^{n-2}}{1+6r}$

where the coefficients x_{nk} may be defined inductively, as follows. Begin with $x_{21} = 3$, $x_{31} = 1$, and $x_{32} = 4$ and with $x_{2k} = 0$ for $k \geq 2$ and $x_{3k} = 0$ for $k \geq 3$. Then, for $n \geq 4$, let $x_{n1} = 0$ and let $x_{nk} = x_{n-1,k-1} + x_{n-2,k-1}$ for $k \geq 2$. The complexity of these formulas is due to the fact that the X chromosome recombines in females but not males, and so different alleles have different numbers of opportunities for recombination before they arrive at the four-chromosome bottleneck, at which inbreeding begins.

Of special interest is the map expansion in RILs. For two-way RILs by selfing, note that $R \approx 2r$ for small r , indicating a $2\times$ map expansion. In general, one may define the map expansion as $dR/dr|_{r=0}$. The map expansion for 2ⁿ-way RILs by selfing is then $(n+1)$ for $n \geq 1$. In 2ⁿ-way RILs by sibling mating (for $n \geq 2$), the map expansion is $(n+4)$ for the autosomes and $(\frac{2}{3}n+4)$ for the X chromosome.

THREE POINTS

We consider the case of three loci. We assume the recombination fractions in the two intervals are the same, $r_{12} = r_{23} = r$. Results for the case of separate recombination fractions could also be obtained, but the expressions can be much more complex, and they provide essentially no further insight.

Let c denote the three-point coincidence at meiosis, $c = \text{Pr}(\text{double recombinant})/r^2$, which may also be written as $\text{Pr}(\text{recn in 2-3}|\text{recn in 1-2})/\text{Pr}(\text{recn in 2-3})$. Note that c is generally a function of r , with $c = 0$ for small r (indicating strong positive crossover interference) and $c = 1$ for $r = \frac{1}{2}$. We define r_{13} to be the recombination fraction between the first and third loci, so that $c = (2r - r_{13})/(2r^2)$ and so $r_{13} = 2r(1 - c)$.

In the case of no crossover interference, we have, of

course, $c = 1$ for all r . We are particularly interested in the case of positive crossover interference. BROMAN *et al.* (2002) studied crossover interference in the mouse and showed that the gamma model (MCPECK and SPEED 1995) provided a good fit to available data. The gamma model involves a single parameter, ν , which indicates the strength of crossover interference; $\nu = 1$ corresponds to no interference, and $\nu > 1$ corresponds to positive crossover interference. BROMAN *et al.* (2002) obtained the estimate $\hat{\nu} = 11.3$ for the mouse, indicating especially strong crossover interference.

ZHAO and SPEED (1996) derived the map function for stationary renewal models of the recombination process at meiosis. Their results may be used to calculate the three-point coincidence for the gamma model, as a function of r and the interference parameter, ν . The map function for the gamma model is

$$M_\nu(d) = \int_0^d \int_x^\infty f(t; \nu) dt dx,$$

where $f(t; \nu) = e^{-2\nu x}(2\nu)^\nu x^{\nu-1}/\Gamma(\nu)$, the density of the gamma distribution with shape parameter ν and rate parameter 2ν .

We thus have, for the gamma model, $r_{13} = M_\nu[2M_\nu^{-1}(r)]$, and we can obtain the three-point coincidence by $c = (2r - r_{13})/(2r^2)$. While $M_\nu(d)$ cannot be obtained in closed form, it can be calculated by numerical integration. Further, $M_\nu^{-1}(r)$ cannot be calculated directly, but can be obtained by solving $r = M_\nu(d)$ for d by Newton's method. This was done in R (IHAKA and GENTLEMAN 1996).

The three-point coincidence for the gamma model with $\nu = 11.3$ is displayed as the dashed curve in Figure 3A. For $r < 0.1$, the coincidence is essentially 0, indicating that if the first pair of loci recombine, the second pair will not. As r approaches $\frac{1}{2}$, the coincidence approaches 1.

RILs by selfing: *Two-way RILs, selfing:* H&W showed that the two-point probabilities for two-way RILs by selfing are sufficient to determine the three-point probabilities. Note that the equation $R = 2r/(1+2r)$ applies to each interval between loci, and so, because $r_{13} = 2r(1 - c)$, we have $R_{13} = 2r_{13}/(1+2r_{13}) = 4r(1 - c)/[1 + 4r(1 - c)]$. Thus, for example, to calculate the probability of the haplotype AAA on the RIL, we note that $\text{Pr}(AAA) + \text{Pr}(AAB) = \text{Pr}(AA-)$, $\text{Pr}(ABB) + \text{Pr}(ABA) = \text{Pr}(AB-)$, and $\text{Pr}(AAA) + \text{Pr}(ABA) = \text{Pr}(A-A)$, and thus, as $\text{Pr}(ABB) = \text{Pr}(AAB)$, we have

$$\begin{aligned} \text{Pr}(AAA) &= \frac{1}{2}[\text{Pr}(AA-) - \text{Pr}(AB-) + \text{Pr}(A-A)] \\ &= \frac{1}{2}[(1 - R)/2 - R/2 + (1 - R_{13})/2]. \end{aligned}$$

Plugging in $R = 2r/(1+2r)$, and using a similar approach for the other two cases, we obtain the distribution for the three-locus haplotype on the RI chromosome, in two-way RILs by selfing:

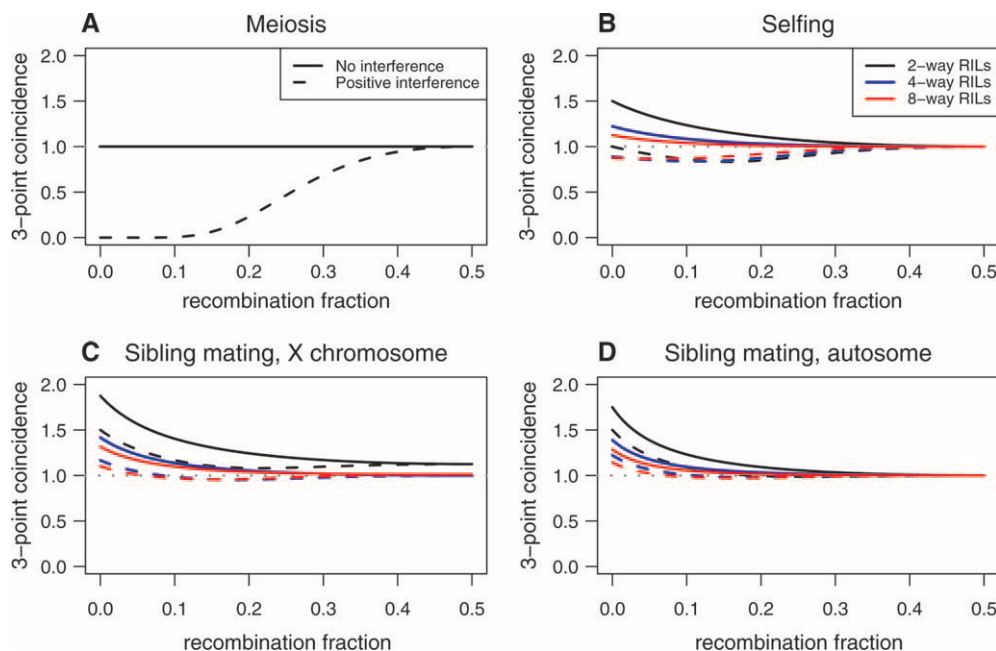


FIGURE 3.—Three-point coincidence in meiosis (A), RILs by selfing (B), the X chromosome for RILs by sibling mating (C), and autosomes for RILs by sibling mating (D). Solid curves are for the case of no interference; dashed curves correspond to strong positive cross-over interference (according to the gamma model with $\nu = 11.3$, as estimated for the mouse genome). (B–D) Black, blue, and red curves correspond to two-way, four-way, and eight-way RILs, respectively. Note that coincidence on the RIL chromosome is displayed as a function of the recombination fraction at meiosis.

$$x_1 = \Pr(AAA) = \Pr(BBB) = \frac{1 + 2r - 4r^2 - 2cr^2 + 4cr^3}{2(1 + 2r)(1 + 4r - 4cr^2)}$$

$$x_2 = \Pr(AAB) = \Pr(BBA) = \Pr(ABB) = \Pr(BAA) = \frac{r - cr^2}{1 + 4r - 4cr^2}$$

$$x_3 = \Pr(ABA) = \Pr(BAB) = \frac{2r^2 + cr^2 - 2cr^3}{(1 + 2r)(1 + 4r - 4cr^2)}$$

We are especially interested in the quantity analogous to the coincidence for the RI chromosome, $C = [\Pr(ABA) + \Pr(BAB)]/R^2$, which gives the following:

$$C = \frac{2 + c + 4r - 4cr^2}{2 + 8r - 8cr^2} = \frac{1 + (1 + c)(1 - 2R)}{2[1 - (1 + c)R^2]}.$$

Note that in the case $r = 0$, we have $C = (2 + c)/2$; with no interference ($c = 1$), $C = 3/2$, and with strong positive interference ($c = 0$), $C = 1$. In the case that $r = 1/2$ and $c = 1$, we have, of course, $C = 1$.

The coincidence is plotted as the black curves in Figure 3B, with the solid and dashed curves corresponding to no interference and strong positive interference (the gamma model with $\nu = 11.3$), respectively. Note that in the case of no interference, the coincidence is entirely >1 , indicating clustering of breakpoints: if the first two loci have recombined on the RIL chromosome, the second two loci are *more likely* to have recombined. In the case of strong positive interference, the coincidence is ≤ 1 for all r .

Four-way RILs, selfing: In the case of four-way RILs by selfing, the joint three-locus genotype probabilities may be derived from the above results, by the technique used for the case of two loci. There are 64 possible three-locus genotypes, which collapse into seven distinct cases. These cases and the corresponding probabilities are shown in Table 6.

The coincidence-type quantity for the RI chromo-

some is then $C = (4a_3 + 8a_5 + 16a_6 + 8a_7)/R^2$, which gives the following:

$$C = \frac{(1 + 2r)[8(1 + r) + 3c(1 - 2r - 4r^2) - 2c^2r^2(1 - 2r)]}{9(1 + 4r - 4cr^2)}$$

$$= \frac{8(3 - 2R)^2(3 - R) + 3c(27 - 72R + 48R^2 - 8R^3) - 2c^2(3 - 4R)R^2}{3(3 - 2R)^2[9 - 4(1 + c)R^2]}.$$

Thus, for $r = 0$, $C = (8 + 3c)/9$, and in the case of no interference, $C = 11/9$; with strong positive interference, $C = 8/9$. With $r = 1/2$ and $c = 1$, $C = 1$. The coincidence is displayed as the blue curves in Figure 3B and is generally smaller than that for the case of two-way RILs by selfing.

Eight-way RILs, selfing: The case of eight-way RILs by selfing can be derived directly from the case of four-way RILs by selfing. There are $8^3 = 512$ three-locus haplotypes, but they collapse to 13 distinct classes. The joint three-locus haplotype probabilities are presented in Table 7. Here, we present only one genotype pattern for each of the 13 classes, but also list the number of genotypes that fall into each class.

We thus have $C = [1 - 8(b_1 + 2b_2 + 4b_4 + 8b_6)]/R^2$, which gives the following:

$$C = \frac{(1 + 2r)[2(7 + 8r - 8r^2) + 4c(1 - 3r - 8r^2 + 8r^3) - 2c^2r^4(1 - 2r) - c^2r^2(3 - 18r + 20r^2)]}{(4 - r)^2(1 + 4r - 4cr^2)}.$$

It is difficult to write this in terms of R , as $r = (2 - R) - \sqrt{(2 - R)^2 - R}$, and so we neglect to do so. Note that when $r = 0$, $C = (7 + 2c)/8$, which takes value $7/8$ under no interference and $7/8$ under strong positive interference.

The coincidence is displayed as the red curves in Figure 3B and is generally smaller than that for the case of four-way RILs by selfing.

X chromosome for RILs by sibling mating: *Two-way RILs, X chromosome:* In the case of the X chromosome

TABLE 6
Three-locus haplotype probabilities for four-way RILs by selfing

Three-locus haplotypes	Probability of each
AAA BBB CCC DDD	$a_1 = x_1(1 - 2r + cr^2)/2$
AAB BBA CCD DDC ABB BAA CDD DCC	$a_2 = x_1r(1 - cr)/2$
ABA BAB CDC DCD	$a_3 = x_1cr^2/2$
AAC AAD BBC BBD CCA CCB DDA DDB ACC ADD BCC BDD CAA CBB DAA DBB	$a_4 = x_2(1 - r)/4$
ACA ADA BCB BDB CAC CBC DAD DBD	$a_5 = x_3(1 - r_{13})/4$
ABC ABD BAC BAD CDA CDB DCA DCB ACD ADC BCD BDC CAB CBA DAB DBA	$a_6 = x_2r/4$
ACB ADB BCA BDA CAD CBD DAC DBC	$a_7 = x_3r_{13}/4$

The x_i are the three-locus probabilities for two-way RILs by selfing.

for two-way RILs by sibling mating, the two-point probabilities are not sufficient to determine the full three-locus probabilities, due to the difference in the frequency of the A and B alleles on the X chromosome, which prevents us from making use of symmetries, such as the relation $\Pr(ABB) = \Pr(AAB)$, which held in the case of two-way RILs by selfing.

However, the two-point probabilities are sufficient to determine the coincidence-type quantity, which has the form $C = [\Pr(ABA) + \Pr(BAB)]/R^2$. By the approach used for three points in two-way RILs by selfing, we obtain the following:

$$C = \frac{3(1 + 4r)(4 + c - 4cr)}{8(1 + 8r - 8cr^2)} = \frac{3[2 + (2 + c)(1 - 3R)]}{8 - 9(2 + c)R^2}.$$

With $r = 0$, $C = 3(4 + c)/8$, so that with no interference, $C = 15/8$, while with strong positive interference, $C = 3/2$. When $r = 1/2$ and $c = 1$, $C = 9/8$. The coincidence is displayed as the black curves in Figure 3C and is greater than that for two-way RILs by selfing. For both no interference and strong positive interference, the coincidence is entirely >1 .

TABLE 7
Three-locus haplotype probabilities for eight-way RILs by selfing

Prototype	No. cases	Probability of each
AAA	8	$b_1 = a_1(1 - 2r + cr^2)/2$
AAB	16	$b_2 = a_1r(1 - cr)/2$
ABA	8	$b_3 = a_1cr^2/2$
AAC	32	$b_4 = a_2(1 - r)/4$
ACA	16	$b_5 = a_3(1 - r_{13})/4$
AAE	64	$b_6 = a_4(1 - r)/4$
AEA	32	$b_7 = a_5(1 - r_{13})/4$
ABC	32	$b_8 = a_2r/4$
ACB	16	$b_9 = a_3r_{13}/4$
ABE	64	$b_{10} = a_4r/4$
AEB	32	$b_{11} = a_5r_{13}/4$
ACE	128	$b_{12} = a_6/8$
AEC	64	$b_{13} = a_7/8$

The a_i are the three-locus probabilities for four-way RILs by selfing (Table 6).

The full distribution of the three-locus haplotypes may be obtained by the approach we used to calculate the two-locus probabilities. There are 288 parental types, which reduce to 168 distinct states after accounting for symmetries, of which 6 states are absorbing. Thus, the absorption probabilities may be obtained by solving a set of 162 linear equations. Alternatively, one may collapse the results for the more complex case of four-way RILs, derived below, to obtain the results for two-way RILs.

Four-way RILs, X chromosome: In the case of four-way RILs, the two-point probabilities are not sufficient to determine the three-point probabilities, or even the three-point coincidence. Thus we must return to the technique used to calculate the two-point probabilities, calculating the absorption probabilities of a Markov chain, here defined by the parental types at three loci at each generation of inbreeding.

There are 10,206 parental types, of the form $AAA|BBB \times CCC$ (the three-locus, X chromosome diplotype of the female parent and the three-locus, X chromosome haplotype of the male parent). These reduce to 2690 distinct states after accounting for symmetries (exchange alleles A and B and invert the order of the three loci), of which 10 states are absorbing. The transition matrix contains 65,612 nonzero elements (that is, $\sim 1\%$ of the matrix).

The absorption probabilities could, conceivably, be obtained by solving a system of 2680 linear equations, but the scale and complexity of this system was too unwieldy in practice. We thus took a different approach. We also abandoned the effort to obtain symbolic solutions, seeking instead numerical solutions. (The absorption probabilities are ratios of polynomials, but we hypothesize that, in this case, the expressions are extremely complex.)

Let $\boldsymbol{\pi}^{(n)}$ denote the distribution (as a row vector) of the Markov chain at generation n , with $\boldsymbol{\pi}^{(0)}$ denoting the starting distribution, for which the state $AAA|BBB \times CCC$ has probability 1 and all other states have probability 0. Let P denote the transition matrix for the chain. Then $\boldsymbol{\pi}^{(n)} = \boldsymbol{\pi}^{(0)}P^n$. We seek $\lim_{n \rightarrow \infty} \boldsymbol{\pi}^{(n)}$, which we calculated numerically. For each value of the recombination fraction, r , and the three-point coincidence at meiosis, c , we iterated across generations until the maximum dif-

TABLE 8

Three-locus haplotype probabilities for the X chromosome in four- and eight-way RILs by sibling mating

A. Four-way RILs			B. Eight-way RILs		
Prototype	No. cases	Probability of each	Prototype	No. cases	Probability of each
AAA	2	c_1	AAA/EEE	4	$d_1 = c_1(1 - 2r + cr^2)/2$
AAB	4	c_2	AAB/EEF	8	$d_2 = c_1r_{13}/4$
ABA	2	c_3	ABA/EFE	4	$d_3 = c_1cr^2/2$
AAC	4	c_4	CCC	1	$d_4 = c_1$
ACC	4	c_5	AAC	4	$d_5 = c_2(1 - r)/2$
ACA	2	c_6	CCA	4	$d_6 = c_2/2$
CAC	2	c_7	ACA	2	$d_7 = c_3(1 - r_{13})/2$
ABC	4	c_8	CAC	2	$d_8 = c_3/2$
ACB	2	c_9	ABC	4	$d_9 = c_3r/2$
CCC	1	c_{10}	ACB	2	$d_{10} = c_3r_{13}/2$
			AAE	8	$d_{11} = c_4(1 - r)/4$
			AEE	8	$d_{12} = c_5(1 - r)/4$
			AEA	4	$d_{13} = c_6(1 - r_{13})/4$
			EAE	4	$d_{14} = c_7(1 - r_{13})/4$
			ABE	8	$d_{15} = c_4r/4$
			AEF	8	$d_{16} = c_5r/4$
			AEB	4	$d_{17} = c_6r_{13}/4$
			EAF	4	$d_{18} = c_7r_{13}/4$
			CEE	4	$d_{19} = c_5(1 - r)/2$
			CCE	4	$d_{20} = c_4/2$
			CEC	2	$d_{21} = c_6/2$
			ECE	2	$d_{22} = c_7(1 - r_{13})/2$
			EFC	4	$d_{23} = c_5r/2$
			ECF	2	$d_{24} = c_7r_{13}/2$
			ACE	16	$d_{25} = c_8/4$
			AEC	8	$d_{26} = c_9/4$

ference between $\pi^{(n)}$ and $\pi^{(n+1)}$ was small ($<10^{-14}$). Approximately 150 generations were required.

The most difficult part of the calculation was the construction of the transition matrix, and the most difficult part of that construction was the reduction of the full set of 10,206 parental types to the minimal set of 2690 states, to account for symmetries. This was done by first creating a look-up table. (Because the central task concerned this collapse of states by symmetry, we performed these calculations via a pair of short Perl programs. Such text manipulation is most conveniently accomplished in Perl.) Rather than construct the entire transition matrix in advance, each row of the transition matrix was constructed anew at each generation, and only those rows that were needed were so constructed (rows i for which $\pi_i^{(n)} > 10^{-16}$). This approach, which saves memory but requires considerably more computation, was used in anticipation of the case of autosomes for four-way RILs by sibling mating, in which the transition matrix contains 73 million nonzero elements.

The 10 absorbing states for the X chromosome in four-way RILs by sibling mating are presented in Table 8A. The haplotype probabilities, denoted c_i , must be calculated numerically. The symbols are presented in Table 8A, as they are used to calculate the haplotype

probabilities for eight-way RILs (below and Table 8B). It turns out that $c_1 = \Pr(\text{AAA}) \equiv \Pr(\text{CCC}) = c_{10}$, which could not have been anticipated in advance. (Note that the c_i here are not at all related to the coincidence, c .)

The coincidence-type quantity is $C = [2c_3 + 2c_6 + 2c_7 + 4c_8 + 2c_9]/R^2$. This is displayed as the blue curves in Figure 3C.

Eight-way RILs, X chromosome: The three-point probabilities for the X chromosome in eight-way RILs by sibling mating may be obtained from those for four-way RILs, by the same approach used in the case of selfing. There are 29 distinct absorbing states (see Table 8B), although three pairs may be collapsed due to the fact that, for the X chromosome four-way RILs by sibling mating, $\Pr(\text{AAA}) = \Pr(\text{CCC})$.

The coincidence-type quantity is then $C = [1 - (4d_1 + 8d_2 + d_4 + 4d_5 + 4d_6 + 8d_{11} + 8d_{12} + 4d_{19} + 4d_{20})]/R^2$. This is displayed as the red curves in Figure 3C, with the solid curve corresponding to no interference and the dashed curve corresponding to the case of strong positive crossover interference.

Autosomes for RILs by sibling mating: *Two-way RILs, autosomes by sib mating:* The results for autosomes in two-way RILs by sibling mating can be derived immediately from the results of H&W, by the technique used for

two-way RILs by selfing. The three-point distribution is the following:

$$\begin{aligned}\Pr(AAA) &= \Pr(BBB) = \frac{1 + 10r - 8cr^2}{2(1 + 6r)(1 + 12r - 12cr^2)} \\ \Pr(AAB) &= \Pr(BBA) = \Pr(ABB) \\ &= \Pr(BAA) = \frac{2r(1 - cr)}{1 + 12r - 12cr^2} \\ \Pr(ABA) &= \Pr(BAB) = \frac{2r^2(6 + c - 6cr)}{(1 + 6r)(1 + 12r - 12cr^2)}.\end{aligned}$$

Thus the coincidence for the autosome in two-way RILs by sibling mating is the following:

$$C = \frac{(1 + 6r)(6 + c - 6cr)}{4(1 + 12r - 12cr^2)} = \frac{3 + (3 + c)(1 - 3R)}{4 - 3(3 + c)R^2}.$$

With $r = 0$, $C = (6 + c)/4$, so that with no interference, $C = 7/4$, while with strong positive interference, $C = 3/2$. With $r = 1/2$ and $c = 1$, $C = 1$. This is displayed as the black curves in Figure 3D, with the solid curve corresponding to no interference and the dashed curve corresponding to the case of strong positive crossover interference. Note that the coincidence is yet smaller than that for the X chromosome in two-way RILs by sibling mating.

Four-way RILs, autosome by sib mating: The three-point probabilities for autosomes in four-way RILs by sibling mating may be calculated by the approach used for the X chromosome, although the scale of the problem is greatly increased. There are 2,164,240 parental types (of the form $AAA|BBB \times CCC|DDD$), which reduce to 137,488 distinct states after accounting for symmetries; seven states are absorbing. The transition matrix contains 73,022,406 nonzero elements (that is, $\sim 0.4\%$ of the matrix).

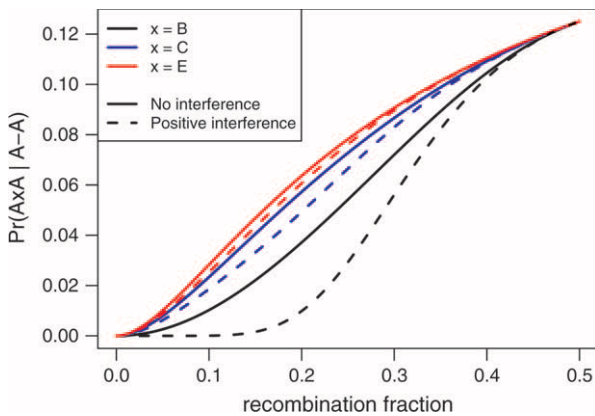


FIGURE 4.—Assessment of symmetry in the three-point probabilities on the autosomes of eight-way RILs by sibling mating. The conditional probabilities $\Pr(AxA|A-A)$ are displayed as a function of the recombination fraction between adjacent loci, with the solid curves corresponding to no interference and the dashed curves corresponding to strong positive interference.

Calculation of the three-point probabilities for a single (r, c) pair took ~ 1 min for the X chromosome, but required ~ 1.5 days for the autosome. Thus, the three-point coincidence curves (one for no interference, one for strong positive interference), displayed in blue in Figure 3D and containing 250 points each, required ~ 750 days of computation time. (Spread across 12 computers, that is just 2 months.)

Eight-way RILs, autosome by sib mating: The three-point probabilities for autosomes in eight-way RILs by sibling mating may be obtained from those for four-way RILs, by the same approach used in the case of selfing. The equations in Table 7 apply, with the a_i now representing the probabilities for the autosome in four-way RILs by sibling mating.

The three-point coincidence for the autosomes in eight-way RILs by sibling mating is displayed as the red curves in Figure 3D, with the solid curve corresponding to no interference and the dashed curve corresponding to the case of strong positive crossover interference. Note that, again, the three-point coincidence is entirely > 1 in the case of no interference and is very near 1 in the case of strong positive crossover interference.

The autosomes in eight-way RILs by sibling mating are of particular interest to us, and so it is valuable to study the probabilities more thoroughly. The three-point coincidence is most informative for the two-way RILs; here they give information largely on the clustering of breakpoints, rather than on the dependence between alleles at adjacent loci.

First, we consider the symmetry of the alleles. In the two-point probabilities, complete symmetry was observed: The chance of switching from A to x across an interval on an eight-way RIL autosome was identical for all $x \neq A$. An inspection of the three-point probabilities, however, indicates that such symmetry does not continue to hold.

For example, consider the case of haplotypes of the form AxA for $x \neq A$. In Figure 4, we plot the conditional probabilities $\Pr(AxA|A-A)$ for $x = B, C, E$. (The alleles C and D are equivalent in this context, as are the alleles E, F, G , and H .) A stretch of A alleles is much more likely to contain a small segment of E than of B , especially in the case of strong positive interference.

Most interesting is an assessment of deviation from the Markov property for the alleles at three points on an eight-way RIL autosome. If the process along an RIL chromosome were Markov, then the allele at the first of three loci would provide no further information about the allele at the third locus, given knowledge of the allele at the central locus. In other words, if the Markov property held, we would have $\Pr(xyA|xy-) = \Pr(-yA|-y-)$ for all x and y . Thus we consider $\log_2 \{\Pr(xyA|xy-)/\Pr(-yA|-y-)\}$, which would be strictly 0 if the Markov property held. These are displayed in Figure 5 for all distinct cases (x, y) . (We make careful use of

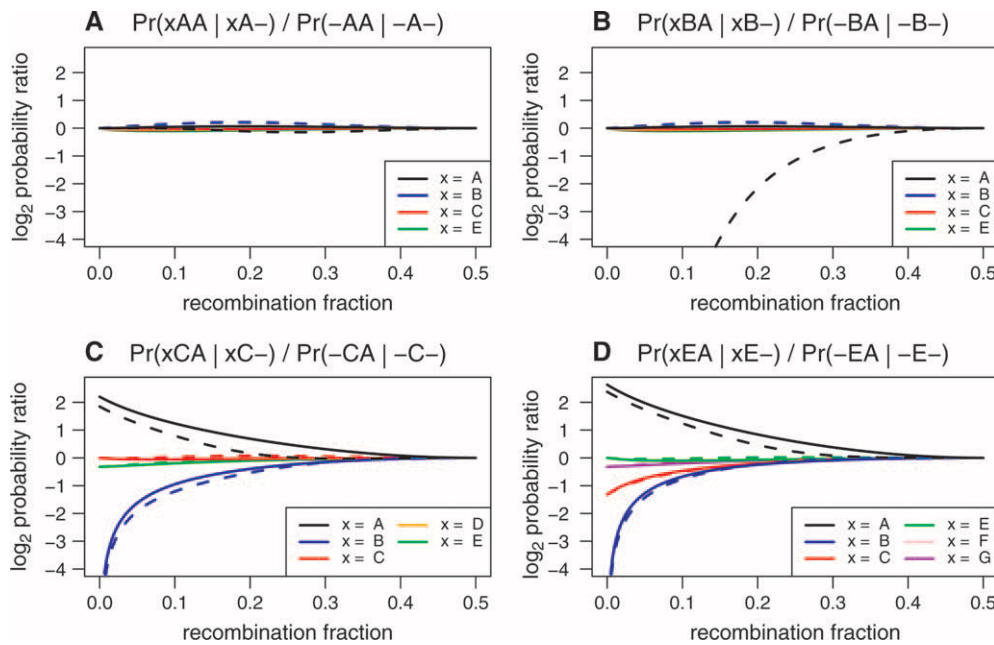


FIGURE 5.—Assessment of the Markov property in the three-point probabilities on autosomes of eight-way RILs by sibling mating. $\log_2[\Pr(xyA|xy-)/\Pr(-yA|-y-)]$ is displayed for each distinct case of x, y , with the solid and dashed curves corresponding to no interference and strong positive crossover interference, respectively.

the symmetries of the problem to reduce the possible cases to 19.) The solid and dashed curves correspond to no interference and strong positive crossover interference, respectively.

In the case of no crossover interference (the solid curves), the probabilities are largely Markov-like, though with important exceptions: If the first two loci are *AC* or *AE*, the third locus is more likely to be *A* than one would expect in the absence of information about the allele at the first locus (see the black curves in Figure 5, C and D), while if the first two loci are *BC* or *BE*, the third locus is less likely to be *A* (see the blue curves in Figure 5, C and D). In the case of strong positive crossover interference (the dashed curves), we also see that *AB* is considerably less likely to be followed by another *A* (see the black dashed curve in Figure 5B). These observations are closely connected to the lack of symmetry in the three-point probabilities seen in Figure 4: Small segments of *C* or *E* will be inserted within longer stretches of *A*.

As it turns out, the cases *CCA* and *DCA* give identical probability ratios, although this could not be anticipated in advance. (These are the red and orange curves in Figure 5C, although only the red curves may be seen, as they overlap.) The same is true for the cases *EEA* and *FEA* (the green and pink curves in Figure 5D, of which only the green curves may be seen).

THE WHOLE GENOME

A number of interesting questions about multiple-strain RILs cannot easily be answered by analytic means and so require large-scale computer simulations. For example, we are interested in the number of generations of breeding that will be required to achieve ge-

nome-wide fixation and the density of genotypes that will be required to identify all breakpoints. We are most interested in the eight-way RILs formed by sibling mating, but we also considered two-way RILs by selfing and sibling mating, to serve as benchmarks.

The simulated genome was modeled after the mouse, with the genetic lengths of the chromosomes taken from the Mouse Genome Database (see Table 9). The total genetic length was 1665 cM. We considered solely the case of strong positive crossover interference, as has

TABLE 9
Chromosomal lengths (in centimorgans) used
in the computer simulations

Chromosome	Length
1	127.0
2	114.0
3	119.2
4	84.0
5	92.0
6	75.0
7	74.0
8	82.0
9	79.0
10	77.0
11	80.0
12	66.0
13	80.0
14	69.0
15	81.0
16	72.0
17	81.6
18	60.0
19	55.7
X	96.5

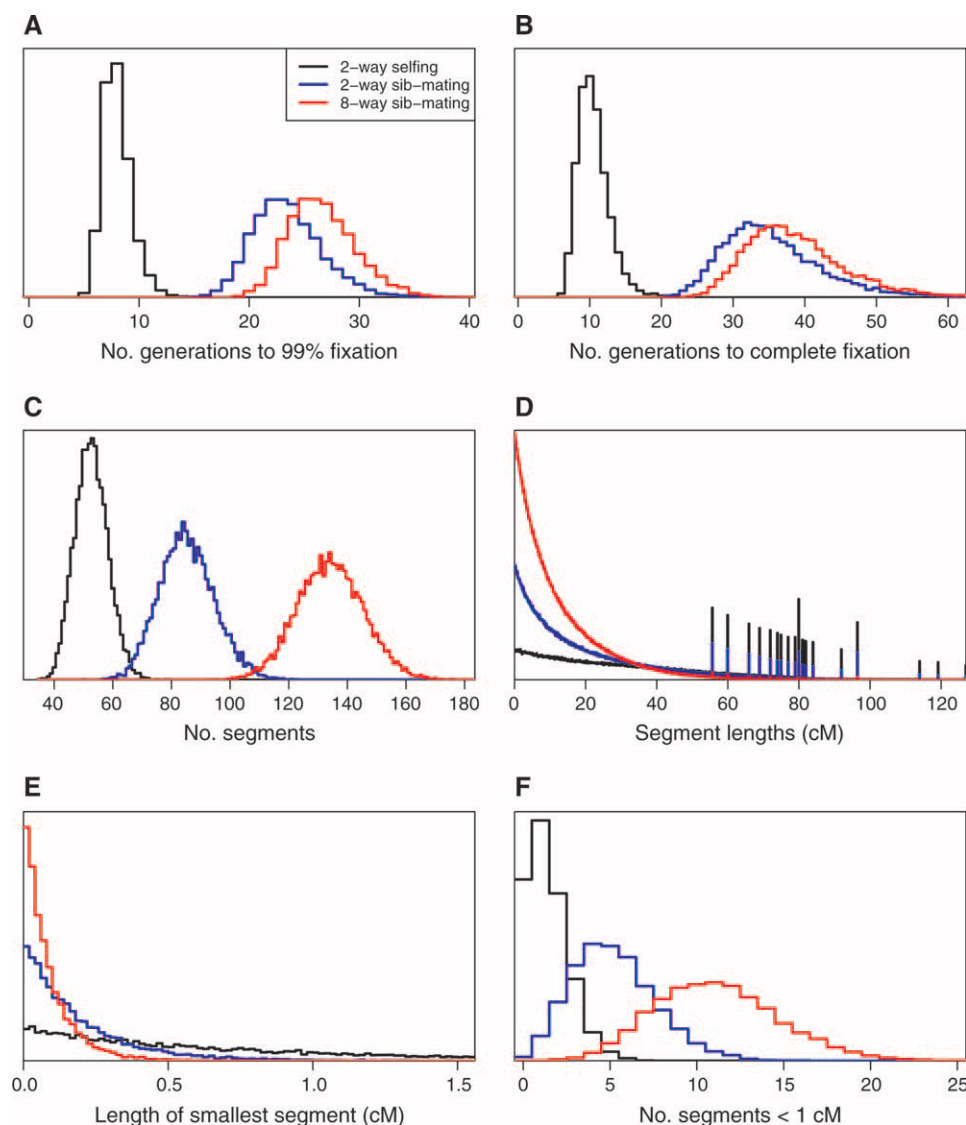


FIGURE 6.—Results of 10,000 simulations of two-way RILs by selfing (black), two-way RILs by sibling mating (blue), and eight-way RILs by sibling mating (red), with a mouse-like genome of length 1665 cM and exhibiting strong crossover interference. (A) Distribution of the number of generations of breeding to achieve 99% fixation. (B) Distribution of the number of generations of breeding to achieve complete, genome-wide fixation. (C) Distribution of the total number of segments, genome-wide. (D) Distribution of the lengths of segments. (E) Distribution of the length of the smallest segment, genome-wide. (F) Distribution of the number of segments < 1 cM in length.

been observed in the mouse (BROMAN *et al.* 2002). To simulate meiosis, we used the χ^2 -model (ZHAO *et al.* 1995), which is a special case of the gamma model (considered in the previous two sections), for which the interference parameter $\nu = m + 1$, for a nonnegative integer m . (The χ^2 -model is more convenient for computer simulation.) We used $m = 10$, corresponding to $\nu = 11$, close to the estimate obtained by BROMAN *et al.* (2002). We used 10,000 simulation replicates and bred until complete fixation of the entire genome.

A variety of summaries of the results of the whole-genome simulations are displayed in Figure 6. The distribution of the number of generations of breeding required to achieve fixation at 99% of the genome is displayed in Figure 6A. It is important to note that this includes the initial mixing generations of breeding in addition to the many generations of inbreeding. (One additional mixing generation is required for eight-way RILs than for two-way RILs; see Figures 1 and 2.) Two-way RILs by selfing required an average of 8 generations, while two- and eight-way RILs by sibling mating required

23.5 and 26.7 generations, on average, respectively, to achieve 99% fixation. Thus, eight-way RILs require an additional 3 generations of breeding (including the additional generation of mixing). There is considerable variation in the number of generations required to achieve this level of fixation.

The distribution of the number of generations to achieve *complete* fixation is displayed in Figure 6B. For two-way RILs by selfing, 10.5 generations are required, on average. For two- and eight-way RILs by sibling mating, complete fixation requires an average of 35.6 and 38.9 generations, respectively. Eight-way RILs again require only 3 or so additional generations of breeding to achieve the same level of fixation as two-way RILs.

The distribution of the total number of segments (with segments defined as the regions between breakpoints on the RIL chromosomes) is displayed in Figure 6C. Two-way RILs by selfing have an average of 53 segments; two- and eight-way RILs by sibling mating have an average of 85 and 134 segments, respectively.

The marginal distribution of the lengths of the seg-

ments is displayed in Figure 6D. It should be no surprise that eight-way RILs have shorter segments. The spikes in the right tails in Figure 6D are whole chromosomes that were inherited intact. The higher spike at 80 cM corresponds to chromosomes 11 and 13, which had identical lengths (see Table 9). Another unusually high spike occurs at 96.5 cM and corresponds to the X chromosome (which, as was seen above, behaves differently than autosomes). Two-way RILs have a good chance of inheriting an intact chromosome, while for eight-way RILs this is relatively rare. The median segment length for two-way RILs by selfing is 25.6 cM; the median segment lengths for two- and eight-way RILs by sibling mating are 12.9 and 8.5 cM, respectively. The chance that a two-way RIL by selfing will have at least one intact chromosome is ~99%; the average number of intact chromosomes is 3.8. For two-way RILs by sibling mating, the chance of at least 1 intact chromosome is 79%, and the average number of intact chromosomes is 1.5. For eight-way RILs by sibling mating, the chance of at least 1 intact chromosome is 12%, and the chance of 2 intact chromosomes is just 0.5%.

The high frequency of small segments seen in Figure 6D raises the question: How small is the smallest segment? The distribution of the smallest segment in the genome of an RIL is displayed in Figure 6E. For eight-way RILs by sibling mating, the smallest segment is generally quite small, which suggests that an extremely high density of genetic markers will be required to identify all segments in a panel of eight-way RILs. The 95th percentile of the length of the smallest segment in two-way RILs by selfing was 2.2 cM, while for two- and eight-way RILs by sibling mating, it was 0.58 and 0.27 cM, respectively. That is, 95% of the time, an eight-way RIL will have at least one segment that is <0.27 cM long.

Finally, the distribution of the number of small segments (that is, segments <1 cM in length) is displayed in Figure 6F. The average number of such small segments is just 1.4 for two-way RILs by selfing, but is 5.2 and 11.2 for two- and eight-way RILs by sibling mating, respectively.

DISCUSSION

Our aim in this work was to characterize the genomes of multiple-strain recombinant inbred lines. We were particularly interested in the case of eight-way RILs by sibling mating, as the development of a large panel of such RILs has been proposed for the mouse (THREADGILL *et al.* 2002; COMPLEX TRAIT CONSORTIUM 2004). We have extended the work of HALDANE and WADDINGTON (1931) on two-way RILs to the case of 2^n -way RILs.

Our key result is the equation $R = 7r/(1 + 6r)$, connecting the recombination fraction at meiosis to the analogous quantity for the autosome of an eight-way RIL derived by sibling mating, which indicates a $7\times$ map expansion in the RIL. (Compare this to the $2\times$ map expansion in two-way RILs by selfing and the $4\times$ map expansion in two-

way RILs by sibling mating). Perhaps more important is the two-point transition matrix, which is critical for the analysis of eight-way RILs. An essential component of the use of RILs for genetic mapping is the reconstruction of the parental origin of DNA on the RIL chromosomes (the haplotypes), on the basis of less-than-fully informative genotype data. Such RILs, if developed, will likely be genotyped at a dense set of diallelic markers, such as SNPs. Application of the standard HMM technology for the haplotype reconstruction will make critical use of this transition matrix.

In advance, we had hypothesized that the two-point transition matrix would have a complex structure, with the *A* and *B* alleles found together more often than the *A* and *H* alleles. And so we were surprised to see that all off-diagonal elements in the transition matrix are the same. We discovered this initially by computer simulation. Indeed, the fundamental equation, $R = 7r/(1 + 6r)$, was initially identified by simulation: We hypothesized the form $R = ar/(1 + br)$ and used nonlinear regression to identify *a* and *b*. Nonlinear regression was rather dissatisfying, especially in comparison to the work of H&W, and so we pursued symbolic and numeric approaches and eventually the intense computation of three-point probabilities. The key breakthrough was the observation that an understanding of four-way RILs is sufficient for an understanding of eight-way RILs.

The three-point coincidence function is especially interesting. That it generally exceeds 1 indicates a clustering of breakpoints on RIL chromosomes. Such clustering is often seen in illustrations of RIL chromosomes (*e.g.*, see SILVER 1995, Figure 9.4), but the cause of such clustering was not immediately obvious. Our explanation is the following: Closely spaced breakpoints had their origin in different generations, but breakpoints in later generations can occur only in regions that have not yet been fixed. Thus, regions that are not fixed early have a tendency to become saturated with breakpoints late.

The three-point coincidence function does not tell the full story regarding multiple-strain RIL chromosomes. That the coincidence is near 1 indicates that the locations of breakpoints follow something like a Poisson process. However, the three-point probabilities clearly indicate that the alleles along an RIL chromosome do not follow a Markov chain. Nevertheless, an RIL chromosome is more like a Markov chain than is the product of a single meiosis, and so one may be confident in the successful use of the HMM technology for haplotype reconstruction in eight-way RILs.

The symmetry in the two-point transition matrix for eight-way RILs implies that the genealogy of an RIL (the order of the eight parental lines in the initial crosses) will not generally need to be considered in the effort to reconstruct haplotypes from genotype data, even though the three-point probabilities indicate clear (and interesting) lack of symmetry. Nevertheless, in the analysis of the X chromosome, such genealogy information

will be important, as the transition matrix for the X chromosome has considerable structure.

It should be emphasized that H&W considered sex-specific recombination fractions in their analysis of two-way RILs and showed that the two-point probabilities for the RIL depended only on the sex-averaged recombination fraction. We neglected such niceties in our analysis of higher-order RILs. We did examine the effect of sex differences in recombination on the results of our simulation study (data not shown). As sex-specific genetic maps are not available for the mouse, we simply assumed that the recombination rate in females was 50% greater than that in males, uniformly across the genome. The results (with 10,000 simulation replicates) were identical (to within sampling error) to the results in the case of no sex difference in recombination.

That our results assume no mutation and no selection is even more worthy of emphasis. Selection against particular alleles or combinations of alleles during the process of inbreeding will clearly have important influences on the structure and content of the RILs that survive the process. However, the study of the effects of selection by the analytic means we have pursued here would be extremely difficult, and our results ignoring selection will no doubt be of considerable value, just as the work of H&W, who also ignored selection, has been of great value. Exploration of the effects of selection on the products of inbreeding, likely by computer simulation, may be of practical importance.

While we considered only the case of multiple-strain RILs obtained via a "funnel," in which no more intercrossing is used than is required to bring all of the alleles together, one might consider the addition of extra generations of outbreeding prior to the start of inbreeding. WINKLER *et al.* (2003) tackled this problem for two-way RILs; their approach, combined with our results, could be used to derive haplotype probabilities for multi-way RILs with extra outbreeding.

The software that was written to accomplish this work is available at the author's web site (www.biostat.jhsph.edu/~kbroman/software) as an add-on package, R/ricalc, for

the freely available statistical software, R (IHAKA and GENTLEMAN 1996). The most intensive computations, of the three-point probabilities for four-way RILs derived by sibling mating, were performed via a pair of Perl programs; these programs are included within the R/ricalc package, although they have no connection to R. In addition, the package includes a set of Mathematica notebooks that document much of the algebraic details for the simpler results.

The author is grateful to David Levin and Dan Naiman for valuable advice; to an anonymous reviewer, Šaunak Sen, and Hongyu Zhao for comments to improve the manuscript; and to the Department of Biostatistics, Johns Hopkins University, for putting up with his intense use of the Department's computing resources for this work.

LITERATURE CITED

- BROMAN, K. W., L. B. ROWE, G. A. CHURCHILL and K. PAIGEN, 2002 Crossover interference in the mouse. *Genetics* **160**: 1123–1131.
- COMPLEX TRAIT CONSORTIUM, 2004 The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat. Genet.* **36**: 1133–1137.
- HALDANE, J. B. S., and C. H. WADDINGTON, 1931 Inbreeding and linkage. *Genetics* **16**: 357–374.
- HOSPITAL, F., C. DILLMANN and A. E. MELCHINGER, 1996 A general algorithm to compute multilocus genotype frequencies under various mating systems. *Comput. Appl. Biosci.* **12**: 455–462.
- IHAKA, R., and R. GENTLEMAN, 1996 R: A language for data analysis and graphics. *J. Comp. Graph. Stat.* **5**: 299–314.
- MCPEEK, M. S., and T. P. SPEED, 1995 Modeling interference in genetic recombination. *Genetics* **139**: 1031–1044.
- NORRIS, J. R., 1997 *Markov Chains*. Cambridge University Press, Cambridge, UK.
- SILVER, L. M., 1995 *Mouse Genetics: Concepts and Applications*. Oxford University Press, Oxford.
- THREADGILL, D. W., K. W. HUNTER and R. W. WILLIAMS, 2002 Genetic dissection of complex and quantitative traits: from fantasy to reality via a community effort. *Mamm. Genome* **13**: 175–178.
- WINKLER, C. R., N. M. JENSEN, M. COOPER, D. W. PODLICH and O. S. SMITH, 2003 On the determination of recombination rates in intermated recombinant inbred populations. *Genetics* **164**: 741–745.
- WOLFRAM RESEARCH, 2003 *Mathematica*, Version 5. Wolfram Research, Champaign, IL.
- ZHAO, H., and T. P. SPEED, 1996 On genetic map functions. *Genetics* **142**: 1369–1377.
- ZHAO, H., T. P. SPEED and M. S. MCPEEK, 1995 Statistical analysis of crossover interference using the chi-square model. *Genetics* **139**: 1045–1056.

Communicating editor: J. WAKELEY