Haplotype Probabilities for Multiple-Strain Recombinant Inbred Lines

Friedrich Teuscher* and Karl W. Broman†,1

*Research Unit Genetics and Biometry, Research Institute for the Biology of Farm Animals (FBN), Dummerstorf, Germany 18196 and †Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland 21205

Manuscript received July 28, 2006
Accepted for publication November 26, 2006

ABSTRACT
Recombinant inbred lines (RIL) derived from multiple inbred strains can serve as a powerful resource for the genetic dissection of complex traits. The use of such multiple-strain RIL requires a detailed knowledge of the haplotype structure in such lines. Broman (2005) derived the two- and three-point haplotype probabilities for 2-way RIL; the former required hefty computation to infer the symbolic results, and the latter were strictly numerical. We describe a simpler approach for the calculation of these probabilities, which allowed us to derive the symbolic form of the three-point haplotype probabilities. We also extend the two-point results for the case of additional generations of intermating, including the case of 2-way intermated recombinant inbred populations (IRIP).

RECOMBINANT inbred lines (RIL) can serve as powerful tools for genetic mapping. An RIL is formed by crossing two inbred strains followed by repeated matings among relatives (e.g., selfing or sibling mating) to create a new inbred line whose genome is a mosaic of the parental genomes. As each RIL is an inbred strain and so can be propagated eternally, a panel of RIL 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, including measurements on a single invasive phenotype over time or in different environments; and, as the breakpoints in RIL are more dense than those that occur in any one meiosis, greater mapping resolution can be achieved.

Members of the Complex Trait Consortium have recently begun the development of a large panel of eight-way RIL 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 selfing or sibling mating to produce a new inbred line whose genome is a mosaic of the eight parental strains. (Figure 1, A and B, illustrates the production of eight-way RIL by selfing and sibling mating, respectively.) This panel will serve as a valuable community resource for mapping the loci that contribute to complex phenotypes in the mouse.

In general, one might consider the development of a panel of 2-way RIL, mixing the genomes of 2 different inbred lines. One might also consider an additional generation of interbreeding, preceding the process of inbreeding, to increase the density of breakpoints on the final RIL; we call this the RIL+ design. In 2-way RIL, inbreeding begins with individuals at generation 1; in 2-way RIL+, two G1 individuals from independent “funnels” (with initial crosses in the same order, but with no shared recombination events) are crossed, and inbreeding begins at generation n + 1. The production of eight-way RIL+ by selfing and sibling mating is shown in Figure 1, C and D, respectively. Note that in eight-way RIL+, one may mate cousins at generation G2, as these individuals have no shared recombination events. For higher-order RIL+, a more extensive set of matings will be required to ensure that the individuals at generation Gn−1 exhibit independent recombination events.

Further, it has been proposed to include some number of generations of random mating prior to inbreeding, a design that has been called an intermated recombinant inbred population (IRIP). Multiple designs for the formation of 2-way IRIP might be considered. First, one might create an unlimited population of individuals at generation n, each from a funnel having initial crosses in the same order, but with such crosses completely independent between individuals. Second, the individuals at generation n might each come from an independent, random funnel, with the order of the initial crosses completely randomized, though with all 2 parental strains represented. We focus on the latter design, as it requires the formation of a single large population from which a panel of IRIP may be developed. The former design would require separate populations of intermating individuals for each line to be formed. Note that the use of random funnels makes the IRIP design distinct from the RIL+ design, which uses a fixed funnel.
The use of multiple-strain RIL panels will require a detailed understanding of the haplotype structure in such lines. At any given genomic position, an RIL will be homozygous for one of the $2^n$ possible parental alleles; a haplotype is the set of alleles at linked loci along a chromosome. We seek to understand the pattern of exchanges among the parental alleles along an RIL chromosome. In particular, the decision of whether to include additional generations of intermating should be based upon an understanding of the additional mapping precision that such intermating will provide.

The seminal article of Haldane and Waddington (1931) provided the basic results for the standard two-way RIL by selfing or by sibling mating: they derived both two- and three-point haplotype probabilities (i.e., the probabilities for all possible two- and three-locus haplotypes) for such two-way RIL. Winkler et al. (2003) calculated the two-point haplotype probabilities for the case of two-way IRIP. Broman (2005) derived the two- and three-point haplotype probabilities for four- and eight-way RIL, though with enormous computational effort. Only numerical results were provided for the three-point probabilities.

Here, we improve on the work of Haldane and Waddington (1931) and Broman (2005). We describe a simpler approach for the calculation of two- and three-point probabilities in $2^n$-way RIL, which allowed us to determine exact formulas for the three-point probabilities. We also extend the results on two-point haplotype probabilities for the case of $2^n$-way RIL+ and $2^n$-way IRIP.

Our results on the map expansion obtained in each design will provide a useful guide to investigators considering the development of $2^n$-way RIL and considering whether additional generations of intermating should be performed.
TWO POINTS

Here we derive the two-point haplotype probabilities on the fixed chromosome in 2-way RIL, RIL+, and IRIP. We consider both selfing and sibling mating, and we focus on the autosome. (Results for the X chromosome may be derived in a similar manner, but since the X chromosome recombines in females but not in males and so different alleles have different numbers of opportunities for recombination before they arrive at the four-chromosome bottleneck, single-point results are difficult to write down for the general 2-way case.) We also derive the quantity analogous to the recombination fraction, but for the fixed RIL chromosome. Note that in the case of sibling mating, we generally assume \( n \geq 2 \) (that is, \( 2^n \geq 4 \)).

Selfing: Two-way RIL: Haldane and Waddington (1931) derived the two-locus haplotype probabilities for two-way RIL by selfing. Here, we describe a simpler solution to the problem.

Let \( W_1 W_2 | X_1 X_2 \) denote the haplotypes for a \( G_k \) individual (for \( k > 0 \), with subscripts denoting the alleles at the two loci. Let \( p_1 \) denote the probability that the \( W_1 W_2 \) haplotype goes on to be fixed, and let \( p_2 \) denote the probability that the \( W_1 X_2 \) haplotype goes on to be fixed. By symmetry, \( \Pr(W_1 X_1 | X_2 X_2 ) = \Pr(W_1 W_2 | X_1 X_2 ) \) and \( \Pr(X_1 W_2 | X_1 X_2 ) = \Pr(W_1 X_2 | X_1 X_2 ) \) and so \( 2p_1 + 2p_2 = 1 \).

Further, if we condition on the first step, we have \( p_1 = (1 - r) / 2 \cdot p_1 \cdot 2 + \frac{1}{2} \cdot p_2 \cdot 2 \). That is, the probability that the \( W_1 W_2 \) haplotype goes on to be fixed is the probability that it is transmitted intact to the next generation (and this can occur in two ways) and then becomes fixed plus the probability that \( W_1 \) is transmitted to one gamete and \( W_2 \) is transmitted to the other gamete and then these are brought together at fixation (and this can occur in two ways). Substituting \( p_2 = (1 - 2p_1) / 2 \), we find \( p_1 = 1 / [2(1 + 2r)] \) and \( p_2 = r / (1 + 2r) \).

Finally, note that in the \( G_1 \) generation, \( W_1 = A \) and \( X_1 = B \). Thus, in a two-way RIL by selfing \( \Pr(AB \text{ fixed}) = p_1 = 1 / [2(1 + 2r)] \) and \( \Pr(AB \text{ fixed}) = p_2 = r / (1 + 2r) \). These are the haplotype probabilities derived by Haldane and Waddington (1931).

2-way RIL. The results for higher-order RIL by selfing may be immediately derived from the results for two-way RIL, due to the two-chromosome bottleneck at the start of inbreeding. We consider the generation of 2-way RIL via a funnel, in which the genomes are brought together as rapidly as possible, followed immediately by inbreeding (see Figure 1A). In the following, we assume \( n \geq 2 \). Let \( L^1, L^2, \ldots, L^{2^n} \) denote the parental lines, and consider the cross \( \{L^1 \times L^2 \times (L^1 \times L^2)\} \). We also use \( L^i \) to denote the allele from that line.

Let \( W_1 W_2 | X_1 X_2 \) denote the alleles on the two chromosomes in generation \( n \), at which inbreeding begins. We must have \( W_1 \in \{L^1, L^2, \ldots, L^{2^{n-1}}\} \) and \( X_1 \in \{L^1 \times 2^{n-1}, \ldots, L^{2^n}\} \).

To derive the haplotype probabilities for the fixed 2-way RIL chromosome, we first determine the haplotype probabilities at the start of inbreeding and then combine them with the results for two-way RIL. We begin with the calculation of the haplotype probabilities at the start of inbreeding. We consider the case that the \( L^i \) allele will be fixed at the first locus; other probabilities follow by symmetry.

To obtain \( \Pr(W_1 = W_2 = L^i) \), note that there must be no recombination at any of the initial mixing generations, and that the \( L^i L^j \) haplotype must be transmitted at each generation. Thus we see that \( \Pr(W_1 = W_2 = L^i) = [(1 - r) / 2]^{n-i} \). Similarly, \( \Pr(W_1 = L^j, W_2 = L^k) = (r/2) [(1 - r) / 2]^{n-i-1} \), as the two loci must recombine at the first generation but not at subsequent generations, and the \( L^i \) allele at the first locus must always be transmitted. Finally, for \( i = 0, 1, \ldots, n - 2 \) and \( j = 1, \ldots, 2^n \), we have \( \Pr(W_1 = L^j, W_2 = L^{j+2^i}) = (1/2)^{2i}(r/2)((1 - r)/2)^{n-i-2} \).

We now proceed to calculate the haplotype probabilities for the fixed RIL chromosome. The probability that the fixed haplotype is \( L^i L^j \), for \( j = 1, \ldots, 2^n \), is simply the probability \( \Pr(W_1 = L^j, W_2 = L^{j+i}) \) multiplied by the probability that the \( W_1 W_2 \) haplotype gets fixed. For \( k = 2^n+1, \ldots, 2^{2n} \), the probability that the fixed haplotype is \( L^i L^k \) is \( \Pr(W_1 = L^j, W_2 = L^{k+2^i}) = (1/2)^{2^n}(r/2)^{n-i} \), multiplied by the probability that the \( W_1 X_2 \) haplotype gets fixed. Thus the two-locus haplotype probabilities in a 2-way RIL by selfing are as follows:

\[
\begin{align*}
\Pr(L^1 L^i) &= \frac{(1 - r)^{n-i}}{2^{n}(1 + 2r)} \\
\Pr(L^1 L^{j+2^i}) &= \frac{r(1 - r)^{n-i-2}}{2^{n+1}(1 + 2r)} \quad \text{for } i = 0, \ldots, n - 2, j = 1, \ldots, 2^n \\
\Pr(L^1 L^k) &= \frac{r}{2^{2n-1}(1 + 2r)} \quad \text{for } k = 2^n+1, \ldots, 2^{2n}.
\end{align*}
\]
Calculation of the haplotype probabilities on the fixed RIL+ chromosome proceeds as before, but a particular allele may come from either chromosome. Thus, for example, the probability that the RIL+ is fixed at $L'L'$ is $Pr(W_1 = W_2 = L') = \Pr(W_1 = W_2 = L')$ times the chance that the $W_1W_2$ haplotype gets fixed, plus $Pr(X_1 = X_2 = L')$ times the chance that the $X_1X_2$ haplotype gets fixed, plus $Pr(W_1 = X_2 = L')$ times the chance that the $W_1X_2$ haplotype gets fixed, plus $Pr(X_1 = W_2 = L')$ times the chance that the $X_1W_2$ haplotype gets fixed. This gives $Pr(L'L') = 2[(1 - r)/2]^n/[2(1 + 2r)] + 2(1/2^n)r(1 + 2r)$. The other cases are similar, and so the two-locus haplotype probabilities in a 2ways RIL+ by selfing are as follows:

$$Pr(L_1L_1) = \frac{(1 - r)^n + r2^{1-n}}{2^n(1 + 2r)}$$
$$Pr(L_1L_{i+j}^2) = \frac{r(1 - r)^{n-i-1} + r2^{1-n}}{2^n(1 + 2r)}$$
$$for \ i = 0, \ldots, n - 2, j = 1, \ldots, 2^i$$
$$Pr(L_1L_k) = \frac{r}{2^{n-k}(1 + 2r)}$$
$$for \ k = 2^{n-1} + 1, \ldots, 2^n.$$  \quad \quad (2)

The probability that the RIL+ chromosome is fixed for different alleles is then $R = 1 - 2Pr(L'L') = 1 - [(1 - r)^n + r2^{1-n}]/(1 + 2r)$. The map expansion for the 2ways RIL+ design by selfing is then $n + 2 - 2^n$.

2ways IRIP(s): In the formation of 2ways IRIP(s) by selfing, one generates an unlimited population of $G_n$ individuals from random funnels, intermates them for $s$ generations, and then inbreeds, by selfing, a random individual from the $n + s$ generation.

At generation $G_n$ in the case of the funnel $[(L_1 \times L_2) \times (L_2 \times L')] \times \ldots,$ the haplotype probabilities for the first chromosome are $Pr(W_1 = W_2 = L') = [(1 - r)/2]^n$ and $Pr(W_1 = L', W_2 = L') = (1/2)^n(r/2)(1 - r)/2^{n-i-2}$ for $i = 0, \ldots, n - 2, j = 1, \ldots, 2^i$. The other chromosome has a similar structure, but for the other alleles.

In the IRIP, we consider individuals from random funnels. That is, each individual comes from a cross of the form $[(L_1 \times L_2) \times (L_1 \times L_2)] \times \ldots,$ where $(k_1, k_2, \ldots, k_n)$ is a random permutation of $(1, 2, 2^2)$. The haplotype probability for a random individual at generation $G_n$ then becomes $Pr(W_1 = W_2 = L') = (1/2^i)(1 - r)/2^{n-i-2}$, and, for $j \neq 1, Pr(W_1 = L', W_2 = L') = [1 - (1 - r)^{-1}]/[2^{n}(2^2 - 1)].$ We thus have complete symmetry among the 2ways alleles.

A random $G_{n+1}$ individual is formed from a cross between two random $G_n$ individuals. Using the results above, we then have the haplotype probabilities $Pr(W_1 = W_2 = L') = [(1 - r)/2]^n$. We can use the symmetry of alleles to conclude that, for $j \neq 1, Pr(W_1 = L', W_2 = L') = [1 - 2^2Pr(W_2 = W_2 = L')]/[2^{n}(2^2 - 1)] = [1 - (1 - r)^n]/[2^{n}(2^2 - 1)].$

The disequilibrium parameter at $G_{n+1}$ is $D = Pr(W_1 = W_2 = L') - [Pr(W_1 = L')]^2 = [(1 - r)/2]^n - 1/2^{2n}$. The disequilibrium parameter at $G_{n+1}$, for $s \geq 1$, after $s - 1$ additional generations of random mating, is $D(1 - r)^{-1}$, and so at this generation $Pr(W_1 = W_2 = L') = D(1 - r)^{-1} + 1/2^{2n} - (1 - r)^{-1}/2^{2n}$. For $j \neq 1$, we again have, by symmetry, $Pr(W_1 = L', W_2 = L') = [1 - 2^2Pr(W_2 = W_2 = L')]/[2^{n}(2^2 - 1)]$, which we neglect to write out.

Finally, we arrive at the haplotype probabilities on the fixed 2ways RIL(s) chromosome, which are derived as before:

$$Pr(L_1L_1) = \frac{1}{1 + 2r}\left[\frac{(1 - r)^n + 2^r + (1 - r)^{-1}}{2^{2n}}\right]$$
$$Pr(L_1L_j) = \frac{1 - 2^2Pr(L_1L_1)}{[2^{n}(2^2 - 1)]}, \quad for \ j \neq 1.$$  \quad \quad (3)

It follows that the probability that the 2ways RIL(s) is fixed at different alleles at the two loci is $R = 1 - 1/(1 + 2r)[(1 - r)^n + 1 + (2r - 1 - (1 - r)^{-1})]/2^n$, and so the map expansion is $n + (s + 1)(1 - 2^{-n})$.

Sibling mating: Two-way RIL: Haldane and Waddington (1931) derived the two-locus haplotype probabilities for two-way RIL by sibling mating. Their derivation involved the solution of a system of 22 linear equations. Here we describe a simpler solution to the problem.

Let $W_1W_2 | X_1X_2 \sim Y_1Y_2 | Z_1Z_2$ denote the haplotypes for the pair of individuals at generation $G_n$ (for $k = 0$). Let $q_1$, $q_2$, and $q_3$ denote the probabilities that the $W_1W_2$, $X_1X_2$, and $Y_1Y_2$ haplotypes, respectively, go on to be fixed. Others follow by symmetry, and so we have $4q_1 + 4q_2 + 8q_3 = 1$.

At $G_0$, $W_1 = X_1 = A$ and $Y_1 = Z_1 = B$, and so $Pr(AB-fixed) = 4q_3$. At $G_1$, $W_1 = Y_1 = A$ and $X_1 = Z_1 = B$, and so $Pr(AB-fixed) = 2q_1 + q_2$. Equating this to the previous equation, we see immediately that $q_3 = q_2$.

Looking forward one generation (with the same technique used for the case of two-way RIL by selfing), we find $Pr(AB-fixed) = (1 - r)/2 \cdot q_1 + 4q_1 + 4q_2 + 8q_3 + 12. Thus 2(q_1 + q_3) = 2(1 - r)q_1 + 3q_2$, and so $q_2 = nq_1$. Using the fact that $4q_3 + 4q_2 + 8q_3 = 4q_1 + 12q_2 = 1$, we obtain $q_1 = 1/(4(1 + 6r))$ and $q_2 = r/[2(1 + 6r)]$.

Finally, we obtain, for two-way RIL by sibling mating, $Pr(AB-fixed) = 2(q_1 + q_3) = (1 + 2r)/(2(1 + 6r))$ and $Pr(AB-fixed) = 4q_3 = 2r/(1 + 6r)$. These are the haplotype probabilities derived by Haldane and Waddington (1931).

Four-way RIL: Our method for calculating the two-locus haplotype probabilities for two-way RIL by sibling mating (above) included the results for four-way RIL. The $q$'s defined above are exactly the two-locus haplotype probabilities for four-way RIL by sibling mating. If we let $L_1$, ..., $L_4$ denote the four alleles, we have...
Pr(\(L_j, L_j\) fixed) = \(q_1 = 1/(4(1 + 6r))\) and Pr(\(L_j, L_j\) fixed) = \(q_2 = r/(2(1 + 6r))\) for \(i \neq j\). These results are the same as those obtained by Broman (2005).

2-way RIL: Derivation of the two-point haplotype probabilities on the fixed chromosome in a 2-way RIL by sibling mating is similar to the case of selfing, although we must consider the four chromosomes at the start of inbreeding. Let \(W_i W_j X_i X_j Y_i Y_j Z_i Z_j\) denote the two-locus haplotypes in the two individuals at generation \(G_{n-1}\), prior to inbreeding (see Figure 1B), and note that \(W_i \in \{L_1, \ldots, L_2^{n-2}\}, X_i \in \{L_1^{+2}, \ldots, L_2^{n-2}\}, Y_i \in \{L_1^{+2}, \ldots, L_2^{2n-2+2}\}, \) and \(Z_i \in \{L_1^{+2}, \ldots, L_2^{2n-2+2}\}\). To determine the haplotype probabilities on the fixed chromosome, we first determine the probabilities that particular alleles survive to the \(G_{n-1}\) generation and then multiply those by the probabilities that such alleles go on to be fixed.

For the first part, note that Pr(\(W_1 = W_2 = L_1\) = \((1-r)/2\))\(^{n-2}\). For \(i = 0, \ldots, n-3\) and \(j = 1, \ldots, 2^i\), we have Pr(\(W_1 = L_1, W_2 = L_2\) = \((1/2)\)\(^{2i}(r/2)(1-r)/2\))\(^{n-3-i}\). Further note that, for \(k = 1 + 2e, \ldots, 2e + 2\), we have Pr(\(W_1 = L_1, Y_2 = L_2\) = Pr(\(W_1 = L_1\) Pr(\(Y_2 = L_2\) = \((1/2)\)\(^{2e}\)).

Combining these results with the fact that Pr(\(W_1 W_2\) fixed) = \(1/(4(1 + 6r))\) and Pr(\(W_1 Y_2\) fixed) = Pr(\(W_1 Y_2\) fixed) = \(r/(2(1 + 6r))\), we obtain the following two-point haplotype probabilities for the fixed 2-way RIL:

\[
\begin{align*}
\text{Pr}(L_1 L_1) &= \frac{(1-r)n/2}{2^{n} + 6r} \\
\text{Pr}(L_1 L_2) &= \frac{r}{2^{n}(1 + 6r)} \\
\text{Pr}(L_2 L_2) &= \frac{r}{2^{n}(1 + 6r)}.
\end{align*}
\]

(4)

It follows that the probability that the 2-way RIL is fixed at different alleles at the two loci is \(R = 1 - ((1-r)^{n-1} + r(2^{n-1})^{-1})/(1 + 6r)\), and so the map expansion is \((n/4)\).

2-way RIL+: In 2-way RIL+ by sibling mating, \(G_n\) individuals from independent funnels are crossed to form the \(G_{n+1}\) generation, at which point inbreeding by sibling mating begins (see Figure 1D). At generation \(n\), we have that \(W_i, Y_i \in \{L_1, \ldots, L_2^{n-1}\}\) and \(X_i, Z_i \in \{L_1^{+2}, \ldots, L_2^{n-1}\}\). Derivation of the two-point haplotype probabilities proceeds with two changes: there is an additional generation of recombination prior to the start of inbreeding, and the \(L_1 L_1\) haplotype may now be fixed in four possible ways: \(W_1 = W_2 = L_1\) and the \(W_1 W_2\) haplotype is fixed, \(Y_1 = Y_2 = L_1\) and the \(Y_1 Y_2\) haplotype is fixed, \(W_1 = L_1\) and \(Y_2 = L_2\) and the \(W_1 Y_2\) haplotype is fixed, and finally \(W_2 = L_1\) and \(Y_1 = L_1\) and the \(Y_1 W_2\) haplotype is fixed.

We first look at the haplotype probabilities at generation \(n\). We have Pr(\(W_1 = W_2 = L_1\) = \([(1-r)/2]^{n-3}\). For \(i = 0, \ldots, n-2, j = 1, \ldots, 2^{i}\), Pr(\(W_1 = L_1, W_2 = L_2\) = \((1/2)^{2i}(r/2)[(1-r)/2]^{n-2-i}\)).

To obtain the two-point haplotype probabilities on the fixed 2-way RIL+ chromosome, we note that, for example, the probability that the \(L_1 L_1\) haplotype is fixed is \(2[(1-r)/2]^{n-1}\) Pr(\(W_1 W_2\) fixed) + \(2(1/2)^{2n}Pr(Y_1 Y_2\) fixed). The probability that the \(L_1 L_1\) haplotype is fixed, for \(k > 2^{n-1}\), is \((1/2)(2^{n-i})Pr(W_i X_j\) fixed), with the 4 coming from the fixation of \(W_i X_j, W_j Z_i, Y_i X_j,\) or \(Y_j Z_i\). Thus, the final results are as follows:

\[
\begin{align*}
\text{Pr}(L_1 L_1) &= \frac{(1-r)n-r2^{n-1}}{2^{n} + 6r} \\
\text{Pr}(L_1 L_2) &= \frac{r[(1-r)^{n-1} + (1/2)^{n-2}]}{2^{n}(1 + 6r)} \\
\text{Pr}(L_2 L_2) &= \frac{r}{2^{n}(1 + 6r)} \quad \text{for } k = 2^{n-1} + 1, \ldots, 2^n.
\end{align*}
\]

(5)

It follows that the probability that the 2-way RIL+ is fixed at different alleles at the two loci is \(R = 1 - ((1-r)^{n-1} + r(2^{n-1})^{-1})/(1 + 6r)\), and so the map expansion is \((n + 5 - 2^{n-1})\).

2-way IRIP(s): In the formation of 2-way IRIP(s) by sibling mating, one generates an unlimited population of \(G_n\) individuals from random funnels, intermates them for \(s\) generations, and then inbreeds. The haplotype probabilities at generation \(n + s\), at which inbreeding begins, are the same for the case of 2-way IRIP(s) by selfing, and so we have, at generation \(n + s\), Pr(\(W_1 = W_2 = L_1\) = \((1-r)^{n+s-1}/2^n + [1 - (1-r)^{-1}]/2^n\). The probabilities Pr(\(W_1 = L_1, W_2 = L_1\) for \(j \neq 1\) may be derived by symmetry.

Derivation of the haplotype probabilities on the fixed 2-way IRIP(s) chromosome proceeds as before, and so we obtain the following:

\[
\begin{align*}
\text{Pr}(L_1 L_1) &= \frac{1}{1 + 6r} \left[ \frac{(1-r)^n + 6r + 1 - (1-r)^{-1}}{2^n} \right] \\
\text{Pr}(L_1 L_2) &= \frac{2^n}{2^n} \quad \text{for } j \neq 1.
\end{align*}
\]

(6)

It follows that the probability that the 2-way IRIP(s) is fixed at different alleles at the two loci is \(R = 1 - ((1-r)^{n+s-1} + (6r + 1 - (1-r)^{-1})/2^n\), and so the map expansion is \((n + (s + 5) - 2^n)\).

Summary: Here we have derived the two-point haplotype probabilities for 2-way RIL, RIL+, and IRIP. Perhaps our most important results concern the map expansion in the different designs, as these indicate the increased mapping resolution that may be obtained. The RIL+ and IRIP designs require additional generations of mating, and this additional effort must be weighed against the improved precision provided.

The map expansions for 2-way RIL, RIL+, and IRIP(s) by selfing are assembled in Table 1. The map expansion in the RIL+ design is somewhat \(<1\) unit...
TABLE 1
Map expansion in 2^s-way RIL, RIL+, and IRIP

| Selfing | Sibling mating |
|---------|----------------|
| 2^s-way RIL | n + 1 | n + 4 |
| 2^s-way RIL+ | n + 2 − 2^s | n + 5 − 2^s |
| 2^s-way IRIP(s) | n + (s + 1)(1 − 2^s) | n + (s + 5)(1 − 2^s) |

n ≥ 1 for selfing and n ≥ 2 for sibling mating.

greater than that for the RIL. In the IRIP, one obtains a slightly <1 unit increase in the map expansion for each additional generation of intermatings.

THREE POINTS

A technique similar to that used above for the case of two points may be used to derive the three-point haplotype probabilities in RIL. Broman (2005) derived these quantities, but obtained only numerical solutions. By our approach, we may obtain exact formulas for the three-point haplotype probabilities.

We focus exclusively on the autosome in four- and eight-way RIL by sibling mating. Exact formulas for four- and eight-way RIL by selfing were presented in Broman (2005). Results for higher-order RIL, RIL+, and IRIP may be obtained from the results provided below, and a similar technique may be used to derive results for the X chromosome.

We consider three points and assume that the recombination fractions in the two intervals are the same, n_{12} = n_{23} = r. Let c denote the three-point coincidence at meiosis, c = Pr(double recombinant)/r^2. Note that c is generally a function of r, with, for most organisms, c = 0 for small r (indicating strong positive interference) and c = 1 for r = 2^{−1}. We define n_{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 − cr).

To simplify some of the notation in what follows, define n_00 = 1 − 2r + cr^2, the chance that a non-recombinant haplotype is transmitted; n_{01} = r(1 − cr), the chance that the second but not the first interval recombines; and n_{11} = cr^2, the chance that both intervals recombine.

Four-way RIL: We consider the case of four-way RIL by sibling mating. Let p_{ijk} denote the probability that the ijk haplotype is fixed. (For ease of notation, here we denote the four alleles as the integers 1, 2, 3, 4.) Taking account of the various symmetries, there are seven distinct haplotype probabilities, shown in Table 2. Note that we must have 4p_{111} + 8p_{112} + 4p_{121} + 16p_{131} + 8p_{131} + 16p_{123} + 8p_{132} = 1.

To derive these seven probabilities, we condition on the first step toward inbreeding. For example, we can write p_{111} = 2((n_{00}/2)p_{111} + 4(1 − r)/2)(1/2)p_{133} + 2(1 − r_{13}/2)(1/2)p_{313}. This is derived as follows: the 111 haplotype can be fixed if it is transmitted intact in the first generation and then that haplotype goes on to be fixed (and this can happen in two ways), or if the 1 alleles at two adjacent loci are transmitted from the first parent in one generation and from the other parent at the third locus and these are brought together at fixation (and this can occur in four different ways), or

TABLE 2
Three-point haplotype probabilities on an autosome in four-way RIL by sibling mating

| Example haplotype | No. haplotypes | Probability of each |
|-------------------|----------------|---------------------|
| WWW               | 4              | p_{111} = \frac{5 + 40r - 2(44 + 17c)r^2 + 8(5 + 13c)r^3 - 8c(9 + 2c)r^4 + 32c^2r^5}{4(1 + 6r)(1 + 12r - 12c^2)(5 + 10r - 4(2 + c)r^2 + 8c^3)} |
| WWY               | 8              | p_{112} = \frac{5r + (36 - 5c)r^2 - 4(7 + 13c)r^3 + 4r(15 + 4c)r^4 - 32c^2r^5}{2(1 + 6r)(1 + 12r - 12c^2)(5 + 10r - 4(2 + c)r^2 + 8c^3)} |
| WXY               | 4              | p_{121} = \frac{5r + (36 - 5c)r^2 - 20(1 + 2c)r^3 + 4(9 + 2c)r^4 - 16c^2r^5}{2(1 + 6r)(1 + 12r - 12c^2)(5 + 10r - 4(2 + c)r^2 + 8c^3)} |
| WYY               | 16             | p_{131} = \frac{5r + (32 - 5c)r^2 - 20(1 + 2c)r^3 + 4(9 + 2c)r^4 - 16c^2r^5}{2(1 + 6r)(1 + 12r - 12c^2)(5 + 10r - 4(2 + c)r^2 + 8c^3)} |
| WYW               | 8              | p_{131} = \frac{5r + (32 - 5c)r^2 - 20(1 + 2c)r^3 + 4(9 + 2c)r^4 - 16c^2r^5}{2(1 + 6r)(1 + 12r - 12c^2)(5 + 10r - 4(2 + c)r^2 + 8c^3)} |
| WXY               | 16             | p_{132} = \frac{4r^2 + 3(16 - c)r^3 - 2(12 + 26c + c^2)r^4 + 16c(3 + c)r^5 - 24c^2r^6}{(1 + 6r)(1 + 12r - 12c^2)(5 + 10r - 4(2 + c)r^2 + 8c^3)} |
| WYX               | 8              | p_{132} = \frac{4r^2 + 3(16 - c)r^3 - 2(12 + 26c + c^2)r^4 + 16c(3 + c)r^5 - 24c^2r^6}{(1 + 6r)(1 + 12r - 12c^2)(5 + 10r - 4(2 + c)r^2 + 8c^3)} |
finally that the 1 alleles at the first and third loci are
transmitted from one parent and at the middle locus
from the other parent and these are brought together at
fixation (and this can happen in two ways).

Similar arguments lead to the following additional six
equations:

\[
\begin{align*}
\pi_{12} &= n_1 \pi_{11} + \left(\frac{1}{2}\right) \pi_{13} + \left(\frac{r_1}{2}\right) \pi_{13} \\
\pi_{21} &= n_1 \pi_{11} + n_2 \pi_{13} + \left(1 - \frac{r_1}{2}\right) \pi_{13} \\
\pi_{13} &= \left(\frac{1 - r_1}{2}\right) (\pi_{12} + \pi_{13}) + (\pi_{23} + \pi_{32})/4 \\
\pi_{31} &= \left(\frac{1 - r_1}{2}\right) (\pi_{21} + \pi_{31}) + \pi_{23}/2 \\
\pi_{25} &= \left(\frac{r_1}{2}\right) (\pi_{12} + \pi_{13}) + (\pi_{32} + \pi_{23})/4 \\
\pi_{32} &= \left(\frac{r_1}{2}\right) (\pi_{21} + \pi_{31}) + \pi_{23}/2.
\end{align*}
\]

We have now defined a set of eight linear equations
in seven unknowns. The seven unknown probabilities
may be calculated from these equations, as

\[
C = (1 - 4\pi_{12} - 8\pi_{13} - 16\pi_{13})/R^2,
\]

which gives the following:

\[
\begin{align*}
C &= \frac{[1 + 6r][100 + 40r - 288r^2 + 5(5 - 20r - 208r^2 + 192r^3 - 16r^2(2 - 13r + 18r^2)r^3)]}{18(1 + 12r - 12r^2)[5 + 64r - 4(2 + 4r)^2 + 8r^2]} \\
&= \frac{[1 + 6r][100 + 40r - 288r^2 + 5(5 - 20r - 208r^2 + 192r^3 - 16r^2(2 - 13r + 18r^2)r^3)]}{18(1 + 12r - 12r^2)[5 + 64r - 4(2 + 4r)^2 + 8r^2]} \\
&= \frac{[1 + 6r][100 + 40r - 288r^2 + 5(5 - 20r - 208r^2 + 192r^3 - 16r^2(2 - 13r + 18r^2)r^3)]}{18(1 + 12r - 12r^2)[5 + 64r - 4(2 + 4r)^2 + 8r^2]} \\
&= \frac{[1 + 6r][100 + 40r - 288r^2 + 5(5 - 20r - 208r^2 + 192r^3 - 16r^2(2 - 13r + 18r^2)r^3)]}{18(1 + 12r - 12r^2)[5 + 64r - 4(2 + 4r)^2 + 8r^2]}
\end{align*}
\]

**Eight-way RIL:** The three-point haplotype probabilities
for the autosome in eight-way RIL by sibling mating
may be immediately derived from the results on four-
way RIL, using the equations in Table 7 of Broman
(2005). We neglect to write these out, but do derive the
quantity analogous to the three-point coincidence, for
the fixed eight-way RIL chromosome:

\[
C = \frac{1 + 6r)[100 + 40r - 288r^2 + 5(5 - 20r - 208r^2 + 192r^3 - 16r^2(2 - 13r + 18r^2)r^3)]}{18(1 + 12r - 12r^2)[5 + 64r - 4(2 + 4r)^2 + 8r^2]} \\
= \frac{1 + 6r)[100 + 40r - 288r^2 + 5(5 - 20r - 208r^2 + 192r^3 - 16r^2(2 - 13r + 18r^2)r^3)]}{18(1 + 12r - 12r^2)[5 + 64r - 4(2 + 4r)^2 + 8r^2]}
\]

In the work of Broman (2005), nearly 3 years of total
computer time were used to derive the above quantity,
although the results were strictly numerical and were for
the case of no interference (\(r = 1\)) and for a model of
strong positive crossover interference. Here, we have
shown a simpler method to derive the result, which
allowed us to obtain explicit formulas for the three-
point probabilities. The formulas in this section match
the numerical results of Broman (2005) to within
round-off error.
breakpoint frequencies themselves and not in the underlying recombination fractions. Moreover, an understanding of recombination at meiosis, particularly regarding crossover interference, might best be studied in a large backcross or intercross, rather than with RIL, as the process of inbreeding to develop RIL is subject to considerable selection, and so our understanding of recombination on the basis of breakpoint frequencies in RIL would likely be distorted.

Martin and Hospital (2006) viewed the term “map expansion” as misleading, as it really concerns an increased frequency of breakpoints and no real change in the genetic map. We, however, still prefer the phrase, and no useful alternatives have been proposed; it provides a useful shorthand for a more complex phenomenon. They further take issue with the treatment, in software, of RIL as a backcross through equations such as $R = 4r/(1 + 6r)$ and with an assumption of no crossover interference, as even if meiosis exhibits no interference, occurrences of breakpoints in adjacent intervals on an RIL chromosome are not independent. (Note that this lack of independence was identified by Haldane and Waddington in 1931.) To the contrary, however, as was stated in Broman (2005), the breakpoint process on an RIL chromosome, at least for the mouse, will be more closely approximated by a Poisson process than is the crossover process at meiosis, which in the mouse exhibits extremely strong positive crossover interference, as even if meiosis exhibits no interference, occurrences of breakpoints in adjacent intervals on an RIL chromosome are not independent.

The authors thank Volker Guiard for helpful discussions, Gudrun Brockmann for suggesting our collaboration, and two anonymous reviewers for helpful comments. This work was supported in part by National Institutes of Health grant GM074244 (to K.W.B.).

LITERATURE CITED

Broman, K. W. 2005 The genomes of recombinant inbred lines. Genetics 169: 1133–1146.

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.

Lander, E. S., and D. Botstein, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185–199.

Lander, E. S., P. Green, J. Abrahamsson, A. Barlow, M. J. Daly et al., 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174–181.

Martin, O. C., and F. Hospital, 2006 Two- and three-locus tests for linkage analysis using recombinant inbred lines. Genetics 173: 451–459.

Teuscher, F., V. Guiard, P. E. Rudolph and G. A. Brockmann, 2005 The map expansion obtained with recombinant inbred strains and intermated recombinant inbred populations for finite generation designs. Genetics 170: 875–879.

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, 2005 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.

Communicating editor: D. Houle

APPENDIX

Let $r$ denote the recombination fraction (at meiosis) for an interval, and let $R(r)$ denote the analogous quantity for an RIL (or RIL+ or IRIP). Several authors (e.g., Winkler et al. 2003; Broman 2005) have referred to the fact that the map expansion in an RIL is $dR/dr |_{r=0}$, but this has been stated without proof. The map expansion for a given RIL design is particularly important for our work, and so we wish, in this appendix, to provide a proof of the result. This expands on a comment in Teuscher et al. (2005).

Let $d$ denote map distance (that is, the average number of crossovers in an interval), and let $D(d)$ denote the corresponding average number of breakpoints in that interval on an RIL (or RIL+ or IRIP) chromosome. We wish to show that $D(d)$ is linear in $d$ and in particular that $D(d) = ad$ for some $a$.

First, note that $D(d_1 + d_2) = D(d_1) + D(d_2)$ for all $d_1$, $d_2$. This comes from the fact that $d_i$ is the average number of crossovers in an interval, while $D(d_i)$ is the average number of breakpoints in the interval on the RIL chromosome, and so both are additive.

It follows that $D(nd) = nD(d)$ for any nonnegative integer $n$. Further, $D(d) = D(nd/n) = nD(d/n)$, and so $D(d/n) = D(d)/n$ for any positive integer $n$. Thus $D(qd) = qD(d)$ for all nonnegative rationals $q$. Surely $D(d)$ is continuous, and so then $D(xd) = xD(d)$ holds for all nonnegative $x$.

Now pick any $d_0 > 0$ and let $a = D(d_0)/d_0$. Then for any $d$, $D(d) = D((d/d_0)d_0) = (d/d_0)D(d_0) = ad$. Thus we have shown $D(d) = ad$ for some $a$.

Now let $r$ denote the recombination fraction corresponding to the distance $d$ and let $R(r)$ denote the analogous quantity for the RIL, and note that for small $r$, $r \approx d$. It follows that $a = \lim_{d \to 0} D(d)/d = \lim_{r \to 0} R(r)/r = dR/dr |_{r=0}$.

Note that this result requires no assumption about crossover interference, but does require the existence of a map function: that there is a one-to-one relationship between the recombination fraction for an interval and the average number of crossovers in the interval.