On the Determination of Recombination Rates in Intermated Recombinant Inbred Populations

Christopher R. Winkler,[1] Nicole M. Jensen, Mark Cooper, Dean W. Podlich and Oscar S. Smith

Pioneer Hi-Bred International, Johnston, Iowa 50131

Manuscript received November 8, 2002
Accepted for publication February 10, 2003

ABSTRACT

The recurrent intermating of F_2 individuals for some number of generations followed by several generations of inbreeding produces an intermated recombinant inbred (IRI) population. Such populations are currently being developed in the plant-breeding community because linkage associations present in an F_2 population are broken down and a population of fixed inbred lines is also created. The increased levels of recombination enable higher-resolution mapping in IRI populations relative to F_2 populations. Herein we derive relationships, under several limiting assumptions, for determining the expected recombination fraction in IRI populations from the crossover rate per meiosis. These relationships are applicable to situations where the inbreeding component of IRI population development is by either self-fertilization or full-sib mating. Additionally, we show that the derived equations can be solved for the crossover rate per meiosis if the recombination fraction is known for the IRI population. Thus, the observed recombination fraction in any IRI population can be expressed on an F_2 basis. The implications of this work on the expansion of genetic maps in IRI populations and limits for detecting linkage between markers are also considered.

HIGH-RESOLUTION genetic maps are an important resource for identifying genes associated with quantitative traits, for understanding gene regulation and expression, and for understanding allele diversity at the gene level. In principle these maps provide more power for detecting genes of interest than has traditionally been provided at the level of quantitative trait loci. The human, animal, and plant genetics communities are all pursuing a number of methods for creating such maps in an effort to progress in the above areas. The resolution of genetic maps is directly related to the number of opportunities for recombination events between closely linked genes prior to mapping. In plants, the traditional approach has been to develop recombinant inbred lines from an F_2 population derived from a single biparental cross. It has been noted, however, that genetic linkage is broken down more effectively by randomly mating an F_2 population for a small number of generations (Liu *et al.* 1996), and that more precise map locations of quantitative trait loci can be found in such populations (Darvasi and Soller 1995).

Recently, mapping populations that incorporate important aspects of both randomly mated (RM) and recombinant inbred (RI) populations have been constructed in maize [Lee *et al.* 2002; Sharopova *et al.* 2002; details on this population are available at the Maize Mapping Project (http://www.maizemap.org/)]. These intermated recombinant inbred (IRI) populations are randomly mated for some number of generations after the F_2, followed by cycles of self-fertilization to produce a population of homozygous lines (*e.g.*, Figure 1). The development of an IRI mapping population has advantages over a RM mapping population in that a permanent population of fixed lines is produced. A mapping population developed by selfing from the F_2 also produces a population of fixed lines, but mapping resolution is limited relative to an IRI population.

Although IRI populations are being developed, no theory has been established to relate the frequency of recombination events in IRIs to the crossover frequency per meiosis (that is, the frequency expected in an F_2 population). Expectations for RI populations are well known and date to the work of Haldane and Waddington (1931). More recently, a theoretical expectation has been published for RM populations (Liu *et al.* 1996). Mapping studies using IRI populations typically, and erroneously, apply equations for RI and RM populations to adjust recombination rates observed in IRI mapping populations back to an F_2 basis (Coe *et al.* 2002; Lee *et al.* 2002). In this study we derive a relationship between the observed recombination fraction, R, and the F_2-adjusted recombination rate, r, in an IRI mapping population. The derived relationship may be used to estimate R given some knowledge of r, and vice versa.

[1]*Corresponding author:* 7250 NW 62nd Ave., P.O. Box 552, Johnston, IA 50131-0552. E-mail: chris.winkler@pioneer.com

TABLE 1

List of symbols

Symbol	Meaning
R	Recombination fraction observed in a mapping population
r	Recombination fraction per meiosis (*i.e.*, adjusted to an F_2 basis)
t	Number of generations of random mating
n	Number of generations of inbreeding, either via selfing or full-sib mating
C_n, D_n, E_n, F_n, G_n	Proportion of genotypes in genotypic classes *C*, *D*, *E*, *F*, and *G* after n generations of inbreeding (see text for definitions)
C'_n, D'_n, E'_n, F'_n, G'_n	Proportion of genotypes in each class after t generations of random mating and n generations of inbreeding
c_n, f_n	Differences between proportions of genotypic classes after n generations of inbreeding (see text for definitions)
c'_n, f'_n	Differences between proportions of genotypic classes after t generations of random mating and n generations of inbreeding
RI	Recombinant inbred
RM	Randomly mated
IRI	Intermated recombinant inbred

THEORY

An extension of Haldane and Waddington's (1931) method was used to derive the results for IRI populations. The method is first illustrated for self-fertilized populations where the relationship between R and r is well known. A list of the symbols used in this section and their meaning is provided in Table 1.

Selfed populations: Consider two loci on a chromosome with two alleles possible at each locus. Two parents, with genotypes *AABB* and *aabb*, are crossed to produce an F_1 population. Following n generations of self-fertilization, the population contains a distribution of genotypes. We separate this distribution into the classes

$$\begin{aligned} C_n &= AABB + aabb \\ D_n &= AAbb + aaBB \\ E_n &= AABb + AaBB + Aabb + aaBb \\ F_n &= AB.ab \\ G_n &= Ab.aB, \end{aligned}$$

where F_n and G_n represent coupled and repulsion double heterozygotic genotypes, respectively. Moreover, define the relationships

$$\begin{aligned} c_n &\equiv C_n - D_n \\ f_n &\equiv F_n - G_n. \end{aligned}$$

Thus, c_n is the difference between the proportions of homozygous individuals having parental and nonparental gametes (and similarly for f_n but with regard to double heterozygous genotypes). Haldane and Waddington (1931) show that c and f are related between any two generations i and j by

$$c_i + \lambda f_i = c_j + \lambda f_j, \tag{1}$$

where

$$\lambda = \frac{1 - 2r}{2 + 4r}. \tag{2}$$

Equation 2 is appropriate only for self-fertilized populations, however. Different values of λ are appropriate for other types of mating, such as full-sib mating.

As the number of generations of selfing becomes large, only the homozygous genotypes remain in the population. That is, in the limit $n \to \infty$,

$$\begin{aligned} E_\infty = F_\infty = G_\infty = 0 \\ C_\infty + D_\infty = 1 \end{aligned} \tag{3}$$

and thus $f_\infty = 0$. So

$$c_\infty + \lambda f_\infty = c_\infty = c_1 + \lambda f_1, \tag{4}$$

where the subscript 1 indicates the F_2 generation (*i.e.*, after one generation of selfing). Using Haldane and Waddington's Equation 1.1 to show that $2c_1 = f_1 = 1 - 2r$ and substituting into Equation 4 yields

$$c_\infty = c_1(1 + 2\lambda) = \frac{1 - 2r}{2}(1 + 2\lambda). \tag{5}$$

Since $c_\infty = C_\infty - D_\infty$, Equations 2 and 3 can be used to solve for D_∞ in Equation 5. This gives

$$D_\infty = \frac{2r}{1 + 2r}, \tag{6}$$

which is the standard formula relating the fraction of crossover events observed in a self-fertilized inbred population to the recombination frequency per meiosis since $D_\infty = R$.

Again, there is nothing novel about the derivation just described; it is simply presented to illustrate the method used for IRI populations.

IRI populations: Equation 6 does not hold, however,

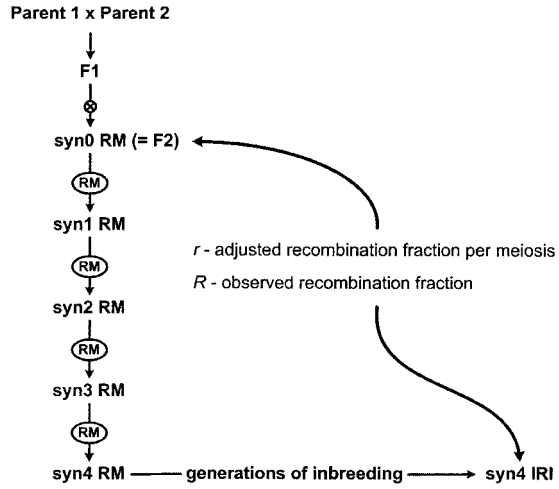

FIGURE 1.—Schematic illustration of the development of an IRI population. Four generations of random mating followed by multiple generations of inbreeding are shown. The syn4 RM designation refers to populations that have been randomly mated for four generations after the F_2. The term syn4 IRI refers to a syn4 RM population that has subsequently undergone self-fertilization or full-sib mating to create a population of inbred lines.

if t generations of random mating have taken place prior to the inbreeding process (*e.g.*, Figure 1). Likewise, equations for RM populations (DARVASI and SOLLER 1995; LIU *et al.* 1996) are inappropriate for IRI populations due to the generations of inbreeding present in IRIs. The correct relationship between R and r for IRIs is outlined below.

The amount of linkage disequilibrium between any two loci is expected to decrease by $1 - r$ for each generation of random mating (FALCONER and MACKAY 1996; OTT 1999), and thus the value of c decreases by $1 - r$ for each generation of random mating after the F_2 generation. Defining c' to be the value of c after t generations of random mating, we have the following relationship:

$$c_1' = c_1(1 - r)^t. \tag{7}$$

Because of the random mating, Equation 4 becomes

$$c_\infty' = c_1' + \lambda f_1' \tag{8}$$

and, likewise,

$$\begin{aligned} 2c_1' &= f_1' \\ c_\infty' &= C_\infty' - D_\infty' \\ C_\infty' + D_\infty' &= 1 \\ D_\infty' &= R. \end{aligned}$$

Substitution of the above relations into Equation 8 and reduction of terms gives the equation

$$R = \frac{1}{2}\left(1 - \frac{1 - 2r}{1 + 2r}(1 - r)^t\right). \tag{9}$$

This relationship associates the observed fraction of crossover events in an IRI population to the crossover frequency seen at the F_2 generation. Note that if $t = 0$, Equation 9 reduces to the equation used for purely self-fertilized populations (*i.e.*, Equation 6) so the recombination fractions in RI *and* IRI populations developed via selfing can be determined with this theory. A population that was randomly mated after the F_2 but had no generations of self-fertilization follows the relationship

$$R = \frac{1}{2}(1 - (1 - 2r)(1 - r)^t)$$

as was previously shown by LIU *et al.* (1996) and in somewhat different form by DARVASI and SOLLER (1995).

Additionally, IRI populations can be developed by full-sib mating, rather than by selfing, after undertaking generations of random mating. Such populations also follow an analytic relationship between R and r. Recognizing that $\lambda = (1 - 6r)/(2 + 12r)$ for full-sib mating populations (HALDANE and WADDINGTON 1931), the derivation outlined above can be repeated with this new value of λ. Doing so yields the equation

$$R = \frac{1}{2}\left(1 - \frac{1 - 2r}{1 + 6r}(1 - r)^t\right). \tag{10}$$

Here again, note that if $t = 0$, Equation 10 reduces to the correct relationship for full-sib-mated recombinant inbred populations, $R = 4r/(1 + 6r)$.

Frequently, however, the quantity of interest is r, rather than R. By knowing the recombination rate on an F_2 basis it is possible to relate marker information between mapping studies and hence produce dense consensus genetic maps. If R has been measured in an IRI population, then Equation 9 can be solved for r as follows (and similarly for Equation 10, but not shown).

Set the dummy variable $z = 1 - r$ and substitute into Equation 9. A little algebra gives

$$1 - 2R = \frac{2z - 1}{3 - 2z} \cdot z^t.$$

Rearranging terms yields the expression

$$2z^{t+1} - z^t + (2 - 4R)z + 3(2R - 1) = 0. \tag{11}$$

Although this equation is complex, it can be solved numerically with standard methods.

RESULTS

Expansion of genetic maps: The degree of map expansion in RI, RM, and IRI populations depends on the value of r. Maximum expansion occurs as $r \to 0$ and decreases as r increases (HALDANE and WADDINGTON 1931; LIU *et al.* 1996). By taking the derivative of the equations showing the relationship between R and r and solving for $r = 0$, it can be shown that the maximum expansion factor is 2 for self-fertilized RI populations and 4 for full-sib-mated RI populations. In

TABLE 2

Genetic map expansion as a function of population type and marker density

No. of marker pairs	Spacing (cM)	syn4 IRI[a]	syn4 RM	RI[a]
5	20.0	206.11	188.56	142.86
10	10.0	281.30	237.56	166.67
100	1.0	385.37	293.08	196.08
200	0.5	392.59	296.52	198.02

Expansion of a genetic map with a total length of 100 cM in the F_2. Different marker densities considered are shown in the first two columns, and the last three columns contain the total map length (in centimorgans) observed in three types of mapping population. Morgan's mapping function (Morgan 1928; Ott 1999) was used to convert map distances to recombination frequencies.

[a]Reported values are for IRI and RI populations developed by self-fertilization.

RM and IRI populations, however, the degree of map expansion varies as a function of the number of generations of random mating undertaken. For example, in an RM population,

$$\frac{d}{dr}\left(\frac{1}{2}[1 - (1 - 2r)(1 - r)^t]\right)\Big|_{r=0} = \frac{t + 2}{2}.$$

Likewise, in IRI populations that are inbred by selfing the map expansion factor is $(t + 4)/2$ and in IRI populations that are inbred by full-sib mating this factor is $(t + 8)/2$. Sib-mated populations have greater map expansion than selfed populations, thus making them potentially more useful for fine mapping. This greater expansion, however, is a consequence of the relatively slow decay of heterozygosity in the population; sib-mated RI populations are thus not preferable to IRI populations developed by selfing for organisms where self-fertilization is possible. For example, an F_2 population that is randomly mated for 4 generations and then self-fertilized for 6 generations will have a maximum map expansion equal to that of a full-sib-mated RI population, but with an inbreeding coefficient greater than that of a population that had been full-sib mated for 10 generations (Fehr 1987).

We stress, however, that the *maximum* map expansion factors apply *only* when $r = 0$; marker pairs that have larger recombination frequencies will have less map expansion. This point appears to have been misunderstood in the literature (Coe *et al.* 2002; Lee *et al.* 2002). To illustrate the fact that map expansion varies as a function of r, as well as the type of mapping population used, consider Table 2. Data relating to four fictitious 100-cM genetic maps that differ in marker density are presented. The theoretically expected genetic map expansions for IRI, RM, and RI populations are shown, where only self-fertilization mating is considered for the inbreeding generations of the RI and IRI populations. The RM and IRI populations have $t = 4$. Such populations are referred to as syn4 RM or syn4 IRI populations, respectively, in Table 2, a notation that is commonly used in the plant-breeding literature (*e.g.*, Lee *et al.* 2002). The total length of the genetic maps observed in the final mapping population varies markedly depending on the marker coverage (*e.g.*, a sparse syn4 IRI map is almost half the size of a dense syn4 IRI map). Moreover, a densely mapped RI population can have nearly as much map expansion as a less dense syn4 IRI or RM population. Clearly, the amount of expansion expected in a genetic map depends on *both* the marker coverage and the type of mapping population considered.

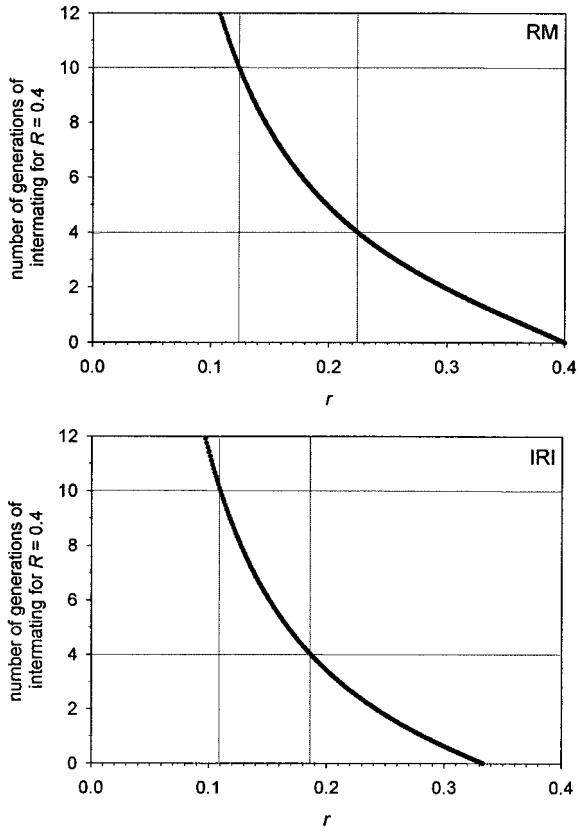

Figure 2.—The relationship between r and $t_{R\geq 0.4}$ for RM (top) and IRI (bottom) populations that have been inbred by self-fertilization.

Limits for detecting linkage between markers: The extent of map expansion is dependent upon the number of opportunities for crossover events during the creation of a mapping population. Additionally, it is difficult to detect linkage relationships when R is large. Liu *et al.* (1996) thus presented the equation

$$t_{R\geq 0.4} = \min\left(\text{integer} > \frac{\ln(1 - 2R) - \ln(1 - 2r)}{\ln(1 - r)}\Big|_{R=0.4}\right) \quad (12)$$

to show the number of generations of random mating required for R to reach 0.4 in a RM mapping popula-

tion. Here $t_{R\geq 0.4}$ is the number of generations of random mating required for R to reach 0.4 and it is assumed that no linkage can be detected between markers when $R \geq 0.4$. Equation 12 can also be used to determine the maximum value of r for which linkage can be detected after t generations of random mating. Figure 2, top, shows this relationship and indicates that in a syn4 RM population, for example, linkage would not be detected between markers with $r \geq 0.224$. Similarly, linkage would not be detected between markers with $r \geq 0.124$ in a syn10 RM population. It has therefore been suggested that a first approximation of marker order be determined using an F_2 mapping population (Liu *et al.* 1996). Clearly the main purpose of developing random mating populations is to provide increased resolution for closely linked marker pairs (*e.g.*, $r < 0.1$).

In an IRI population where the inbreeding is accomplished via selfing, the maximum value of r at which linkage can be detected between two markers for a given generation of intermating is

$$t_{R\geq 0.4} = \min\left(\text{integer} > \frac{\ln(1 - 2R) - \ln(1 - 2r) + \ln(1 + 2r)}{\ln(1 - r)}\bigg|_{R=0.4}\right). \quad (13)$$

Equation 13 is plotted in Figure 2, bottom, and shows that linkage disequilibrium is broken down more rapidly in such IRI populations relative to RM populations. In a syn4 (syn10) IRI population, for example, linkage cannot be detected between markers where $r \geq 0.186$ ($r \geq 0.109$). Significantly, estimates of r from an F_2 population in maize (and that has subsequently been developed into an IRI; Lee *et al.* 2002) indicate that at least 8 of 10 chromosomes contain regions with recombination rates too large for detecting linkage in a syn4 IRI population.

DISCUSSION

We have derived a relationship between R, the observed recombination fraction, and r, the equivalent recombination fraction in an F_2 population, for a mapping population that has undergone some number of generations of intermating followed by generations of either self-fertilization or full-sib mating. Our results suggest that recombination estimates previously reported for IRI maize populations (*e.g.*, Lee *et al.* 2002; Sharopova *et al.* 2002) need to be revised and that the degree of map expansion is in general larger in IRI populations than in RM or RI populations. The relatively larger genetic map afforded by an IRI population is a direct result of the additional meioses that occur in the construction of these populations.

The theoretical expectations presented here, however, are not without limitations. Notably, we have assumed that crossover interference does not take place, that multiple crossover events between two markers do not occur during a single meiotic event, and that segregation distortion is not a factor. Strictly speaking, the theory holds only when the number of generations of selfing or sib mating in the IRI population is sufficient to create inbreds (that is, in the limit $n \rightarrow \infty$). In practice, however, the equations presented are still valid for smaller, realistic values of n. Finally, the theory is technically correct only for "large" population sizes. The results of Equations 9 and 10 will hold on average, of course, but individual realizations of the breeding process can, and usually will, produce discrepancies from expectation when population sizes are small.

LITERATURE CITED

Coe, E., K. Cone, M. D. McMullen, S.-S. Chen, G. Davis *et al.*, 2002 Access to the maize genome: an integrated physical and genetic map. Plant Physiol. **128:** 9–12.

Darvasi, A., and M. Soller, 1995 Advanced intercross lines, an experimental population for fine genetic mapping. Genetics **141:** 1199–1207.

Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics*, Ed. 4. Prentice Hall, Harlow, UK.

Fehr, W. R., 1987 *Principles of Cultivar Development, Volume 1: Theory and Technique.* MacMillan, New York.

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

Lee, M., N. Sharopova, W. D. Beavis, D. Grant, M. Katt *et al.*, 2002 Expanding the genetic map of maize with the intermated B73 × Mo17 (*IBM*) population. Plant Mol. Biol. **48:** 453–461.

Liu, S.-C., S. P. Kowalski, T.-H. Lan, K. A. Feldmann and A. H. Paterson, 1996 Genome-wide high-resolution mapping by recurrent intermating using *Arabidopsis thaliana* as a model. Genetics **142:** 247–258.

Morgan, T. H., 1928 *The Theory of Genes.* Yale University Press, New Haven, CT.

Ott, J., 1999 *Analysis of Human Genetic Linkage*, Ed. 3. Johns Hopkins University Press, Baltimore.

Sharopova, N., M. D. McMullen, L. Schultz, S. Schroeder, H. Sanchez-Villeda *et al.*, 2002 Development and mapping of SSR markers for maize. Plant Mol. Biol. **48:** 463–481.

Communicating editor: J. B. Walsh