1-1-1996

Genome-wide high-resolution mapping by recurrent intermating using Arabidopsis thaliana as a model

Sin-Chieh Liu
Stanley Kowalski

University of New Hampshire School of Law, stanley.kowalski@law.unh.edu

Tien-Hung Lan
Kenneth A. Feldmann
Andrew H. Patterson

Follow this and additional works at: http://scholars.unh.edu/law_facpub

Part of the Genetics and Genomics Commons, and the Plant Sciences Commons

Recommended Citation


This Article is brought to you for free and open access by the University of New Hampshire – School of Law at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Legal Scholarship by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact ellen.phillips@law.unh.edu.
Genome-wide high-resolution mapping by recurrent intermating using Arabidopsis thaliana as a model

This article is available at University of New Hampshire Scholars' Repository: http://scholars.unh.edu/law_facpub/156
Genome-Wide High-Resolution Mapping by Recurrent Intermating Using *Arabidopsis thaliana* as a Model

and Andrew H. Paterson*†

*Department of Soil and Crop Sciences, Texas A&M University, College Station, Texas 77843-2474; †Department of Plant Sciences, University of Arizona, Tucson, Arizona 85721 and ‡Department of Plant and Soil Sciences, University of Delaware, Newark, Delaware 19711

Manuscript received May 8, 1995
Accepted for publication October 7, 1995

ABSTRACT

We demonstrate a method for developing populations suitable for genome-wide high-resolution genetic linkage mapping, by recurrent intermating among F2 individuals derived from crosses between homozygous parents. Comparison of intermated progenies to F2 and “recombinant inbred” (RI) populations from the same pedigree corroborate theoretical expectations that progenies intermated for four generations harbor about threefold more information for estimating recombination fraction between closely linked markers than either RI-selfed or F2 individuals (which are, in fact, equivalent in this regard). Although intermated populations are heterozygous, homozygous “intermated recombinant inbred” (IRI) populations can readily be generated, combining additional information afforded by intermating with the permanence of RI populations. Intermated populations permit fine-mapping of genetic markers throughout a genome, helping to bridge the gap between genetic map resolution and the DNA-carrying capacity of modern cloning vectors, thus facilitating merger of genetic and physical maps. Intermating can also facilitate high-resolution mapping of genes and QTLs, accelerating map-based cloning. Finally, intermated populations will facilitate investigation of other fundamental genetic questions requiring a genome-wide high-resolution analysis, such as comparative mapping of distantly related species, and the genetic basis of heterosis.

HIGH-DENSITY genetic linkage maps of many plants and animals are finding utility in a wide range of basic and applied endeavors (cf. Paterson et al. 1991) and are now being used to assemble “contig maps”, contiguous sets of DNA clones that span the genomes of several organisms (Coulson et al. 1988; Hwang et al. 1991; Schmidt et al. 1992; Cohen et al. 1993; Putterill et al. 1993).

A factor of growing importance in genome analysis is the “resolution” of genetic maps, that is the differing power afforded by various experimental designs to detect recombination events between closely linked loci. High-resolution maps that accurately order closely linked markers are crucial in “positional cloning” (Collins 1992), wherein one seeks genetic map resolution compatible with the DNA “carrying capacity” of “artificial chromosomes”. Assembly of “contig maps” is facilitated by use of closely linked DNA markers to quickly identify sets of large DNA clones corresponding to different genetic loci. Determination of the comparative organization of chromosomes in disparate taxa having gene orders conserved over only short distances is facilitated if closely linked markers can be ordered accurately (Kowalski et al. 1994; see especially Figure 2). Finally, resolution of individual quantitative trait loci associated with complex phenotypes is delimited in part by the amount of recombinational information in a mapping population (Paterson et al. 1990).

Many genetic linkage maps are based on segregating backcross or F2 populations, in which gametes have undergone only a single cycle of recombination, and are rarely recombinant between closely linked loci. At the initiation of genetic mapping, such strong “linkage disequilibrium” facilitates establishment of linkage groups among widely dispersed loci. However, as a genetic map becomes densely populated with marker loci, more information is necessary to resolve the linear order of closely linked markers. An efficient method for high-resolution mapping in specific targeted regions of the genome has recently been described (Churchill et al. 1993) — however such methods are less amenable to applications in which information is sought for all genomic regions simultaneously.

Classical plant breeders, faced with the need to overcome “correlations among traits” (unfavorable genetic linkage), long ago devised a technique suitable for genome-wide high-resolution mapping (cf. Hanson 1959a,b; Miller and Rawlings 1967; Fredericksen and Kronstad 1985; Koyle et al. 1986; Wells and Kofoed 1986; Tyagi 1987; Fatmi et al. 1992). The general approach involves recurrent intermating among...
individuals within a population, mimicking the random-mating behavior of many natural populations. In a population randomly mated for many generations, homogenization of the ancestral chromosomes becomes so complete (WRIGHT 1969) that genetic linkage between alleles at nearby loci can only rarely be detected (cf. LANGLEY et al. 1982; LEIGH BROWN 1983; MACPHERSON et al. 1990). However populations randomly mated for only a few generations retain sufficient "disequilibrium" to detect genetic linkage—and harbor more information for ordering closely linked markers, as a result of multiple meiotic cycles (HANSON 1959a,b).

To evaluate the efficacy of recurrent intermingling as a strategy for genome-wide high-resolution genetic mapping, an intermated population of Arabidopsis thaliana was developed and compared with F2 and recombinant inbred-selfed (RI) populations of the same pedigree. The results corroborate theoretical expectations, demonstrate an experimental design that is suitable for a wide range of applications, and impel development of intermated populations in crop plants and other organisms. The theoretical expectations for intermated populations were based on calculations assuming large population sizes that have no selfing and were not influenced by the effect of genetic drift. The intermated progenies described herein are being selfed down to clones that span the genome.

MATERIALS AND METHODS

Population development: Arabidopsis thaliana ecotypes Wassilewskija (WS) and mutant stock was selfed to generate progenies described herein are being selfed down to influenced by the effect of genetic drift. The intermated population sizes that have no selfing and were not in- populations were based on calculations assuming large Newark, DE) with 16 hr photoperiod, 22° night temperature, due to sterile plants at intervening generations. Only a subset as follows:

Recombinant inbred (RI-selfed population): Each F1 individual was selfed to generate F2 progeny, a single F2 progeny individual was selfed to generate F3 progeny, etc., to F6. At F6, 95 of the 120 F2 individuals were the starting point for two additional mating schemes, as follows:

Recurrent random-interminating (intermated population): Each F2 individual was used as the pollen parent in a hand-cross to another randomly chosen F2 individual. Each F2 individual was also used as a seed parent in such a cross, receiving pollen from another F2 individual. Cross-combinations were selected using a simple random-number generator (in Microsoft Excel), with the restriction that no repeats or selfs were allowed. From the seed produced by each cross, one plant was grown to flowering, and the procedure repeated. After four generations of random mating, 99 of the 120 F2 lines were still represented, the remainder having been lost due to a sterile plant at one of the intervening generations.

F1 and F2 generations were grown in the greenhouse (in Newark, DE) with 16 hr photoperiod, 22° night temperature, and ca. 27° day temperature; subsequent generations were grown at 16 hr photoperiod and constant 22°, in a growth chamber.

Genetic mapping: All laboratory procedures were as described previously (KOWALSKI et al. 1994). DNA probes prefixed "M" were provided by E. MEYERowitz, while those prefixed "AC" are anonymous cDNAs from a library provided by Clontech, Inc (Palo Alto, CA).

Data analysis: Determination of recombination fractions utilized MapMaker (LANDER et al. 1987), (generously provided by S. TINGEY, duPont), on a Macintosh Quadra 650. The "observed recombination fraction" (R) is an estimate of the probability of observing a recombinant in a population, and was used to measure the genetic map expansion under the respective breeding systems. Observed recombination fractions and order of loci for RI-selfed and intermated populations were determined by analyzing data as F2 populations, except that residual heterozygotes in the RI-selfed population were scored as missing data, and a LOD score 6.0 was used as linkage threshold in the RI-selfed population (REITER et al. 1992).

A likelihood ratio test was used to compare the values of the observed R (RF) in the RI-selfed and intermated maps with the values of the expected R (Re), which were derived for the respective populations from the R values of corresponding intervals in the F2 map. The test statistics are

\[
2 \ln \left[ \frac{R_{RF}^{n} (1 - R_{F})^{n - R_{F}}}{R_{RF}^{n} (1 - R_{F})^{n - R_{F}}} \right]
\]

for RI-selfed population

\[
2 \ln \left[ \frac{A_{RF}^{n} B_{RF}^{n} C_{RF}^{n} D_{RF}^{n}}{A_{RF}^{n} B_{RF}^{n} C_{RF}^{n} D_{RF}^{n}} \right]
\]

for intermated population, where n = number of observations and A, B, C, and D, and A, B, C, and D are (1 - R)² / 2, R²/2, 2R(1 - R), [R² + (1 - R)²]/2 and with R = R, and R = R, respectively. Both test statistics asymptotically distribute as a distribution when n is large (WILKS 1938).
intermating following F2, and combinations in the previous generation: the contribution of recombinant gametes from all mating generations of intermating is obtained by summing the intermated population, the expected \( R \) for different breeding systems: the observed recombination fraction (\( R \)) of an F2 population is equal to its adjusted recombination fraction (\( r \)) per meiosis, because an F2 population is derived from a single meiosis. Therefore, the F2 population, in which \( R + r \), is used as the base line for comparing map expansion in RI-selfed and intermated populations. HALDANE and WADDINGTON (1931) derived the relationship between \( R \) and \( r \) for a population of recombinant inbred strains (derived by selfing) as

\[
R = \frac{2r}{1 + 2r}.
\]

The \( R \) and \( r \) of our RI-selfed population is also defined by (1), because the RI-selfed population contained 97.3% homozygotes on average (96.9% is expected based on the mating scheme). The small proportion of heterozygotes were analyzed as missing data. For an intermated population, the expected \( R \) for different generations of intermating is obtained by summing the contribution of recombinant gametes from all mating combinations in the previous generation: \( R_{t+1} = R_t - rR_t + r/2 \), where \( t \) is the number of generations of intermating following F2, and \( R_0 = r \). Therefore, the relationship between \( R \) and \( r \) in an intermated population can be derived as

\[
R = \frac{1}{2} [1 - (1 - r)^4 (1 - 2r)],
\]

and the \( R \) and \( r \) of our intermated population have a relationship of

\[
R = \frac{1}{2} [1 - (1 - r)^3 (1 - 2r)].
\]

The expected \( R \) of our intermated population is larger than that of a RI-selfed population for \( 0 < r < 0.5 \) (Figure 1). However, the degree of map expansion in both intermated and RI-selfed populations (relative to the F2) depends on the values of \( r \), i.e., the larger the value of \( r \), the less the expansion. Maximum expansion occurs at values of \( r \) approaching 0: at which point a twofold expansion in RI-selfed population (HALDANE and WADDINGTON 1931) and a threefold expansion in intermated population is expected, because

\[
\frac{d}{dr} \left( \frac{2r}{1 + 2r} \right) \bigg|_{r=0} = 2
\]

and

\[
\frac{d}{dr} \left( \frac{1}{2} \left[ 1 - (1 - r)^4 (1 - 2r) \right] \right) \bigg|_{r=0} = 3,
\]

respectively.

**Efficiency of detecting recombination:** The relative
Fzpopulation amZc0 RI-selfed population - Intermated population - Expected R calculated fmm R of F2 population chromosome 2

FIGURE 4.—Genetic linkage maps of F2, RI-selfed, and intermated populations of Arabidopsis thaliana WS × M13. Interval lengths are expressed as observed recombination fraction (R) estimated without adjusting for multiple meioses. Expected maps were constructed for RI-selfed and intermated populations by calculating the expected R for each interval from (1) and (3), respectively, using observed R of the F2 population as the adjusted recombination fraction (r) per meiosis.

precision of estimates of r, in F2, RI-selfed, and intermated populations can be compared based on the mean amount of information (i,) provided by a single individual in each population. MATHER (1936) derived i, of a single F2 individual as

$$i_2 = \frac{2(1 - 3r + 3r^2)}{r(1-r)(1-2r+2r^2)}.$$  \hspace{1cm} (4)

By following similar derivation, i, of a single lineage in a RI-selfed population was derived as

$$i_r = \frac{2}{r(1+2r)^2},$$  \hspace{1cm} (5)

and i, of a single lineage in an intermated population was derived as

$$i_i = \frac{(1-r)^2[2(1-r) + t(1-2r)]^2}{[1 + 3(1-2r)^2(1-r)^4]},$$  \hspace{1cm} (6)

where t is the number of generations of intermating following F2 (APPENDIX A). Thus, i, of a single lineage of our intermated population is

$$i_i = \frac{(1-r)^6[2(1-r) + 4(1-2r)]^2}{[1 - (1-2r)^4(1-r)^{16}]}.$$  \hspace{1cm} (7)

Populations intermated for four generations yield threefold more information per individual than F2 and RI-selfed populations, for values of r approaching 0 (Figure 2). The values of i, for intermated population remain the highest among the three populations when r < 0.191, but decrease to the lowest when r > 0.196 (Figure 2).

The standard deviation (s) of the estimate of r can be derived for each population as

$$s = \frac{1}{\sqrt{N_i}},$$  \hspace{1cm} (8)

where N is the population size, N, = the inverse of the variance of the estimate of r, is the amount of information concerning the r estimate. In our experi-
ment, the precision of recombination estimates for F2, and intermated populations correspond approximately to that of i, (Figure 3), and RI-selfed populations are somewhat less (i.e., higher value of s), as we obtained linkage information for an average of 98, 93, and 59 individuals (respectively) at each locus.

**Genetic maps:** For each population (F2, RI-selfed, intermated), the 50 DNA and three morphological (an, dis1, er) markers fell into five linkage groups, corresponding to the five haploid chromosomes of A. thaliana (Figure 4). Although there were no overt conflicts in order of loci along the maps of the three populations, there were several cases in which closely linked markers could not be resolved in the F2 population (see below), and several cases in which large gaps between markers in the intermated population precluded orientation of the groups separated by the gaps. The consensus order shown (Figure 4) uses the order of closely linked markers in the RI or intermated populations in cases that the F2 could not resolve, and relies upon F2 information to span large gaps in the RI or intermated maps. The order of markers along the chromosomes was consistent with the published order of markers previously mapped (Chang et al. 1988; Hauge et al. 1993; Lister and Dean 1993), except that markers M217 and M562a on chromosome 5 are inverted.

In several cases, the orders of closely linked markers could not be resolved with confidence in the F2 population, but could in the RI-selfed and/or intermated populations, validating the underlying rationale for our experiment. Specifically, alternate LODs for “ripple” (using MapMaker; Lander et al. 1987) of several groups of markers were not significantly different for the F2, but were significantly different (LOD ≥ 2) for the RI-selfed and/or intermated populations. These groups of markers, with the respective LODs in parentheses, were -AC87-AC142- (F2: -0.48, RI-selfed: 0.23, intermated: 2.65) and -M242-AC85b- (F2: 0.79, RI-selfed: 8.89, intermated: 14.12) on chromosome 1, M246-M497- (F2: 0.20, RI-selfed: 4.37, intermated: 1.52) on chromosome 2, and
The linkage maps included intervals ranging from $R = 0.5\%$ to $23.9\%$ ($F_2$), $2.2\%$ to $40.5\%$ (RI-selfed), and $1.2\%$ to $42.2\%$ (intermated). Two intervals in the RI-selfed population, which showed $R = 0$ were excluded from consideration, because adjacent markers segregated as “dominant” (present or absent), and the possibility of heterozygous recombinants could not be discerned.

Nine intervals in six chromosomal regions and 19 intervals in 12 chromosomal regions of the RI-selfed and intermated populations, respectively, showed significant difference between the observed $R$ and the expected $R$ with $P$ values <0.05 for the likelihood ratio tests. Five of these significant intervals in different chromosomal regions showed similar deviation from the expected $R$ in both RI-selfed and intermated populations (see DISCUSSION).

Average heterozygosity across the genome for the $F_2$ ($50.6\%$) and intermated ($48.2\%$) populations agreed closely with the Mendelian expectation of $50\%$. However, the average percentages of the WS allele in all three populations were higher than the expectation of $50\%$ for the 47 RFLP loci showing codominant segregation ($F_2$: $52.7\%$ WS allele, $t = 3.89$, $P$ value$_{ttwo-tail} = 0.0003$; RI: $54.5\%$ WS allele, $t = 2.77$, $P$ value$_{ttwo-tail} = 0.0080$; intermated: $55.1\%$ WS allele, $t = 5.51$, $P$ value$_{ttwo-tail} < 0.0001$). Twelve regions on five chromosomes of the $F_2$ population showed significant deviations from the Mendelian expectation of monogenic segregation ratios; eight regions on four chromosomes of the RI-selfed population showed significant segregation deviations, and 16 regions on five chromosomes of the intermated population showed significant segregation deviations. We found no evidence of differences in order of DNA marker loci associated with regions of segregation distortion.

**DISCUSSION**

Recurrent intermuting is an experimental design that offers marked efficiencies for many genetic linkage mapping applications, including comparative mapping, QTL mapping, and map-based gene cloning. Intermated populations are ideal for making “second-generation” genetic linkage maps, which simultaneously resolve local orders in many genomic regions densely populated with DNA markers. Previously, recombinant inbred populations derived by self-pollination of plant lineages (“RI-selfed populations”) have been suggested to “permit higher mapping resolution for short linkage distances” than $F_2$ populations (BURR et al. 1993; see also BURR et al. 1988 and BURR and BURR 1991). RI-selfed populations do afford approximate doubling of nominal recombination fraction ($R$), at values approaching 0. However, for accurately ordering closely linked markers, the information content of a single RI-selfed individual is equal to that of a single $F_2$ individual only at a recombination distance of 0, and becomes progressively less than that of a single $F_2$ individual at larger distances. In contrast, by intermuting among different $F_2$-derived lineages for four generations, the resulting progeny yield three-fold more information per individual than $F_2$ or RI-selfed progeny at values of $r$ approaching 0, and remain more informative than $F_2$ or RI-selfed individuals at all values of $r < 0.131$ (Figure 2). Once the “first-generation” genotype map of an organism has reached a molecular marker density such that there are few intervals of $r > 0.131$, an intermated population provides the means to resolve local marker orders on a genome-wide scale, by analysis of a minimal number of individuals.

Mammalian RI populations, derived by sib-mating, resemble intermated populations in that new recombinational information is accumulated during the relatively slow loss of heterozygosity (cf. TAYLOR 1978; BAILEY 1981). RI-selfed populations yield a maximum of four-fold expansion of $R$ when $r$ approaches 0 (HALDANE and WADDINGTON 1931; TAYLOR 1978) (Figure 1):

$$R = \frac{4r}{1 + 6r} \quad (9)$$

and

$$\frac{d}{dr} \left(\frac{4r}{1 + 6r}\right)_{r=0} = 4.$$

The $i$, of a single lineage of a RI-sibbed population can be derived from the standard deviation of $r$ (GREEN, 1981):

$$i = \frac{4}{r(1 + 2r)(1 + 6r)^2} \quad (10)$$

which is smaller than that of the intermated population when $r < 0.310$ (Figure 2). In contrast, plant RI populations, to date all generated by recurrent self-pollination (i.e., single seed descent: BRIM 1966), lose 50% of remaining heterozygosity each generation, yielding a maximum of twofold expansion of $R$ when $r$ approaches 0 (see RESULTS). RI-selfed individuals are thus comprised of two identical gametes which have been through the equivalent of two cycles of recombination, while $F_2$ individuals are comprised of two different gametes that have each been through only one cycle of recombination—providing equivalent information for resolving close linkages (Figure 2). The loss of heterozygosity during selfing is so rapid that gains of information from new recombination are exactly canceled out. By contrast, mammalian RI individuals are comprised of two identical gametes that have been through the equivalent of four cycles of recombination (see below, and HALDANE and WADDINGTON 1931, TAYLOR 1978), with the slower loss of heterozygosity affording twice as much information as RI-selfed individuals for resolving close linkages. Finally, intermated individuals are comprised of two different gametes each carrying unique
recombinational information and can be subjected to many cycles of recombination to afford further accumulation of information (Figures 5–7).

Many other combinations of interating and inbreeding are possible. While map expansion is a disadvantage in initial assembly of genetic maps, it affords maximal exploitation of high density maps by facilitating resolution of close linkages.

Design and development of intermated mapping populations: A population subjected to recurrent interating has individuals comprised of two different gametes harboring unique recombination sites. Each generation of a recurrent interating population retains heterozygosity at 50% (theoretically) and accumulates new recombination sites at a constant rate through additional generations of interating. Consequently, a pair of markers that are linked by a recombination fraction of \( r \) will yield progressively higher values of \( R \) after more generations of interating, and will segregate independently after sufficient generations of interating. The frequencies of genotypes for such a pair of markers, at the \( t \)th generation of interating are as follows:

\[
\begin{align*}
\text{frequency (AABB)} &= \text{frequency (aabb)} \\
&= \frac{1}{16} \left[ (1 - 2r) (1 - r)^t + 1 \right]^2, \\
\text{frequency (AAbb)} &= \text{frequency (aaBB)} \\
&= \frac{1}{16} \left[ (1 - 2r) (1 - r)^t - 1 \right]^2,
\end{align*}
\]

which approach \( 1/16, 1/16, 2/16 \) and \( 4/16 \), respectively, when \( t \) increases. This can also be observed from (2) with an increase of \( t \) (Figure 5). Clearly, the degree of map expansion increases with additional generations of random interating. For a pair of markers linked at \( r \) recombination fraction (adjusted recombination fraction per meiosis), the number of generations \( (tR_{r=0.4}) \) of interating needed before the value of \( R \) will reach 40\%, i.e., no linkage can be detected between the two markers using the common linkage criterion of \( R \leq 0.4 \), was derived from (2) as:
FIGURE 7.—The precision of the estimation of \( r \) (adjusted recombination fraction per meiosis) of intermated populations undergo different numbers of generations \( (t) \) of intermating. Measurement is based on the amount of information relative to \( F_2 \) population. Values for \( F_2 \) and intermated populations are derived from (4) and (6), respectively.

\[
I_{r_{0.1}} = \min \left( \text{integer} \quad \frac{\ln(1 - 2R) - \ln(1 - 2r)}{\ln(1 - r)} \right)_{R=0.1}. \quad (11)
\]

Figure 7 shows the relationship between \( I_{r_{0.1}} \) and \( r \). Our intermated population had been through four generations of intermating and is inadequate for detecting linkage between markers linked at \( r > 0.22 \) (Figure 6). Therefore, several intervals of our map with large \( R \) (= \( r \)) in \( F_2 \), had become so large in the RI-selfed and intermated populations as to preclude detecting linkage. The \( F_2 \) orders had to be used to ascertain orientations of groups of markers flanking these gaps. However, the \( r \) estimated from an intermated population is much more precise for closely linked markers than that from \( F_2 \) or RI populations of the same size (Figure 2).

The better resolution of closely linked markers afforded by intermated populations was evident in our data. In three-point comparisons (MapMaker “ripple” command), the order between \( M422 \) and \( AC85b \) on chromosome 1 yielded a LOD of 0.79 greater than the best alternative order in \( F_2 \) population; as factors closely linked in early generations become uncoupled, natural selection at a larger number of discrete points across the genome can be seen. However, if segregation distortion is due to natural selection, it must be noted that there would be greater opportunity for selection to act during the course of intermating, a factor that is completely confounded with recombinational differences. The absence of a heterozygous class makes it difficult to interpret differences between RI and the other two populations, regarding segregation distortion.

Population size and number of generations of intermating: Intermated populations should include as
many individuals as possible, even if only a subset are used for genetic mapping. Heterozygosity is gradually lost in an intermated population of finite size, but the loss is slower in a large population (STRICKBERGER 1968). Furthermore, repeated intermitting in a small population can magnify the effect of genetic drift, i.e., rare gametes that are recombinant between closely linked markers might be either "propagated" or lost. However, this limitation is easily overcome in crop plants by using large populations segregating for genetic male sterility (cf. SORRELLS and FRITZ 1982) or self-incompatibility (cf. ST. MARTIN and EHOUNOU 1989) to enforce intermitting. In an intermated population, population size \( (N) \) and the number of generations of intermitting \( (t) \) determine the precision of the \( r \) estimate, which can be measured by the amount of information \( (I_r) = N_i \) \( [i \text{ is defined in (6)]} \). Increase of \( N \) improves the precision of the \( r \) estimate under all circumstances. However, increase of \( t \) only improves the precision of the \( r \) estimate when \( r \) is small, but greatly decreases the precision of the \( r \) estimate when \( r \) is large (Figure 7). Therefore, decisions regarding population size, and number of generations of intermitting, should be based on the required precision of the \( r \) estimate, which can also be measured by the standard deviation \( (s) \) \( (8) \).

An intermated population fosters even more rapid accumulation of new recombination if the initial pedigree is complex (HANSON 1959b). For example, one could intermate among many different \( F_1 \) hybrids to create an \( F_2 \) population carrying greater allelic diversity. This complicates genetic mapping because it would be necessary to find RFLP alleles unique to each of the genotypes contributing to the pedigree. However, such an approach might be useful in highly polymorphic species such as maize, or using highly variable DNA markers (cf. TAUTZ 1989; WEISENBACK et al. 1992).

**Development of homozygous high-resolution mapping populations:** As a high-resolution alternative to RI-selfed populations, intermitting for four generations followed by self-pollination (single seed descent) for six to eight generations, would produce a homozygous "intermated recombinant inbred population" (IRI). The IRI will embody both the high resolution of the intermated population and the permanence of the RI population (Burr et al. 1988, 1991, 1993; Reiter et al. 1992; Wang et al. 1993). We are now selfing our intermated population, and the resulting IRI population will be deposited in the Arabidopsis Biological Resources Center (at Ohio State University). Moreover, we are developing an IRI population of sorghum (\( S. \) bicolor L.: A. H. PATERSON and K. F. SCHERTZ, unpublished results), and our colleagues are doing likewise in maize (\( Z. \) mays L.: W. BEAVIS, personal communication).

During the process of selfing to homozygosity, bulk populations of seed derived from one selfing of individual intermated plants can be used as an interim resource (such as we used herein).

**Applications of intermated populations:** Intermated populations are potentially useful for investigating a wide range of questions that require genome-wide high resolution mapping. The facility of plant genetics is likely to make intermated populations a particularly useful tool in studying the genomes of major crops. Intermated populations can expedite the integration of genetic and physical maps. From theoretical expectations (HANSON 1959a,b), a population of 99 individuals derived by intermitting for four generations should harbor about three recombination sites per chromosome in each of their two informative gametes; for a total of ca. 600 recombination sites along an average chromosome. Across the five chromosomes of \( A. \) thaliana, this represents ca. 3000 recombination sites. Because each individual was used as a parent twice in each generation to derive the intermated population, some recombination sites may be identical by descent—by iterating the binomial probability of gain/loss of recombinant gametes over four generations of intermitting, we estimate that 2176 of the 3000 recombination sites are unique. Based on an Arabidopsis genome size of 145 Mb (ARUMUNGANATHAN and EARLE 1991), novel recombination events occur at average spacing of 67 kb. Thus, most Arabidopsis YACs (ca. 150 kb; GRILL and SOMERVILLE, 1991; WARD and JEN 1990; more recent libraries have larger inserts) might be oriented along the genetic map simply by mapping of their respective ends on the intermated population. This may help in orienting new YACs along the chromosomes, and closing gaps between existing contigs (Hwang et al. 1991; Schmidt et al. 1992; PutteRill et al. 1993).

The improved genetic map resolution afforded by intermitting has an even greater potential impact on analysis of crop genomes, in which the physical size of a centiMorgan is much larger than in Arabidopsis, e.g., the genomes of \( A. \) thaliana, Gossypium hirsutum, Sorghum bicolor, Brassica oleraceae, Lycopersicon esculentum, \( Z. \) mays and \( S. \) tumorous have 290, 400, 500, 540, 750, 1400, and 2500 kb/cm, respectively (Chang et al. 1988; Gerhardt et al. 1989; Arumunganathan and Earle 1991; Landry et al. 1992; Tanksley et al. 1992; Coe and Neuffer 1993; Chittenden et al. 1994; Reinsch et al. 1994). In particular, "fine mapping" of quantitative trait loci (QTL) depends upon the level of resolution of pre-existing genetic maps (PATERSON et al. 1990), which can be improved substantially using intermated populations. In principle, QTL mapping directly in intermated populations permits one to improve the resolution of QTLs at the first stage of mapping—however, the time needed to develop intermated populations may constrain the usefulness of this particular application.

Comparative mapping of chromosome organization in disparate taxa (Bonierbale et al. 1988; Hulbert et al. 1990; Tanksley et al. 1992; Whittkus et al. 1992; Ahn and Tanksley 1993; Ahn et al. 1993; O'Brien et al. 1993; Kowalski et al. 1994), or genomes within taxa (Rei-
Nisch et al. 1994), can be facilitated by use of intermated populations, which better resolve gene order in small chromosomal regions remaining homosequential in distantly related species. Improved map resolution will become increasingly important as comparative mapping efforts reach across greater taxonomic distances, and seek conservation across smaller chromosome segments.

Finally, intermated populations may help to resolve classical questions in population biology. For example, alternative explanations of heterosis (hybrid vigor) propose close linkage between dominant and recessive alleles at different loci ("dominance" theory; cf. Bruce 1910), vs. true "heterozygote advantage" at a single locus ("overdominance" theory; cf. East 1908; Shull 1911). Similarly, the persistence of phenotypic variation in populations subjected to intense directional selection (cf. Alexander 1988) has been postulated to be a result of new mutations, or "release" of cryptic variation in the form of closely linked "+" and "-" alleles (cf. Lande 1975). Such questions can be addressed in unprecedented detail, by combining high-density genetic maps of DNA markers with recurrently intermated mapping populations.

The first three authors contributed equally to this work. We thank E. Meyerowitz and the Arabidopsis Biological Resources Center (at Ohio State University) for DNA clones, K. Manly for valuable comments, Clontech Inc for an Arabidopsis cDNA library, N. Forshoeffel and L. Chittenden for technical assistance and K. Schertl, B. McDonald, R. Wing, G. Wang, V. Rastogi and several anonymous reviewers for valuable comments. Novel DNA probes will be deposited at the Arabidopsis Biological Resources Center. This research was funded by the Texas and Delaware Agricultural Experiment Stations (A.H.P.), and National Science Foundation grant DMB-9108442 (K.A.F.).

LITERATURE CITED


Fatmi, A., D. B. Wagner and T. W. Pfeiffer, 1992 Intermating schemes used to synthesize a population are equal in genetic consequences. Crop Sci. 32: 89–94.


Hanson, W. D., 1993a Theoretical distribution of the initial linkage block lengths intact in the gametes of a population intermated for n generations. Genetics 44: 839–846.

Hanson, W. D., 1993b The breakup of initial linkage blocks under selected mating systems. Genetics 44: 857–868.


Landry, B. S., N. Hulbert, R. Grete, M. S. Chang, S. E. Lincoln et al., 1992 A genetic map for Brassica oleracea based on RFLP
markers detected with expressed DNA sequences and mapping of resistance genes to race 2 of *Plasmiodiothora brassicaceae* (Wortman). *Genome* **35**: 409–420.


Communicating editor: B. S. Weir

**APPENDIX A**

If $g$ classes of genotype are expected in the frequencies $m_1, m_2, \ldots, m_g$ being given in terms of $r$, the recombination fraction, the mean amount of information (i) is given by the formula

$$i = \sum_{j=1}^{g} \left[ \frac{1}{m_j} \left( \frac{dm_j}{dr} \right)^2 \right]$$

(*Mather 1936*). In the intermated population, nine genotype classes can be distinguished by RFLP markers showing codominant segregation, with the AaBb class including the two double heterozygous genotypes AB/ab and Ab/Ab. The frequencies of AABB, aabb, aBb, AABb, aabB, AaBb, AAbb, and AbBb arc $m_1$, $m_2$, $m_3$, $m_4$, $m_5$, $m_6$, $m_7$, and $m_8$, respectively, where

$$m_1 = m_2 = \frac{1}{16} \left[ (1 - 2r)(1 - r)^i + 1 \right]^2,$$

$$m_3 = m_4 = \frac{1}{16} \left[ (1 - 2r)(1 - r)^i - 1 \right]^2,$$

$$m_5 = m_6 = m_7 = m_8 = \frac{1}{8} \left[ 1 - (1 - 2r)^2 (1 - r) \right]^2,$$

and

$$m_9 = \text{frequency (AB/ab) + frequency (Ab/aB)} = \frac{1}{8} \left[ (1 - 2r)(1 + r)^i + 1 \right]^2$$

$$+ \frac{1}{8} \left[ (1 - 2r)(1 - r)^i - 1 \right]^2 = \frac{1}{4} \left[ (1 - 2r)^2 (1 - r)^2i + 1 \right].$$

The $dm_j/ dr$ term for each genotype class can be derived as

$$\frac{dm_1}{dr} = \frac{dm_2}{dr} = \frac{1}{8} \left[ (1 - 2r)(1 - r)^i + 1 \right] \times \left[ -2(1 - r) - (1 - 2r)(1 - r)^i \right],$$

$$\frac{dm_3}{dr} = \frac{dm_4}{dr} = \frac{1}{8} \left[ (1 - 2r)(1 - r)^i - 1 \right]$$

\[
\frac{dm_0}{dr} = \frac{dm_{e}}{dr} = \frac{dm_{c}}{dr} = \frac{1}{2}(1-2r)(1-r)^{2t} + \frac{1}{4}\frac{(1-2r)^2(1-r)^{2t}}{1-r},
\]

and
\[
\frac{dm_0}{dr} = -(1-2r)(1-r)^{2t} \frac{(1-2r)^2(1-r)^{2t}}{1-r}.
\]

The \(1/m_i/(dm_i/dr)^2\) term for each genotype class can be derived as
\[
\frac{1}{m_i} \left( \frac{dm_i}{dr} \right)^2 = \frac{1}{m_2} \left( \frac{dm_2}{dr} \right)^2 = \frac{1}{m_3} \left( \frac{dm_3}{dr} \right)^2
\]

\[
= \frac{1}{4} \left[ -2(1-r)^t - \frac{(1-2r)(1-r)^t}{1-r} \right]^2
\]

\[
= \frac{1}{4} \left[ -2(1-r)^t - \frac{(1-2r)(1-r)^t}{1-r} \right]^2
\]

\[
= \frac{1}{4} \left[ -2(1-r)^t - \frac{(1-2r)(1-r)^t}{1-r} \right]^2
\]

\[
+ \frac{1}{4} \frac{(1-2r)^2(1-r)^{2t}}{1-r}
\]

and
\[
\frac{1}{m_0} \left( \frac{dm_0}{dr} \right)^2 = \frac{1}{m_5} \left( \frac{dm_5}{dr} \right)^2 = \frac{1}{m_6} \left( \frac{dm_6}{dr} \right)^2
\]

\[
= \frac{1}{4} \left[ -2(1-r)^t - \frac{(1-2r)(1-r)^t}{1-r} \right]^2
\]

\[
= \frac{1}{4} \left[ -2(1-r)^t - \frac{(1-2r)(1-r)^t}{1-r} \right]^2
\]

\[
= \frac{1}{4} \left[ -2(1-r)^t - \frac{(1-2r)(1-r)^t}{1-r} \right]^2
\]

\[
= \frac{1}{4} \left[ -2(1-r)^t - \frac{(1-2r)(1-r)^t}{1-r} \right]^2
\]

\[
= \frac{1}{4} \left[ -2(1-r)^t - \frac{(1-2r)(1-r)^t}{1-r} \right]^2
\]

The mean amount of information \(i_i\) is the sum of the term from each genotype class.
\[
i_i = \sum_{j=1}^{9} \left( \frac{1}{m_j} \left( \frac{dm_j}{dr} \right)^2 \right)
\]

\[
= 4 \times \frac{1}{4} \left[ -2(1-r)^t - \frac{(1-2r)(1-r)^t}{1-r} \right]^2
\]

\[
= 4 \times \frac{1}{4} \left[ -2(1-r)^t - \frac{(1-2r)(1-r)^t}{1-r} \right]^2
\]

\[
= 4 \times \frac{1}{4} \left[ -2(1-r)^t - \frac{(1-2r)(1-r)^t}{1-r} \right]^2
\]

\[
= \left[ 2(1-2r)(1-r)^{2t} \right] \frac{1}{1+(1-2r)^2(1-r)^{2t}}
\]

\[
= 2(1-2r)(1-r)^{2t} \left[ \frac{1}{1+(1-2r)^2(1-r)^{2t}} \right]
\]

Let \(a = (1-2r)\) and \(b = (1-r)^t\). Equation (A1) gives
\[
i_i = \left( -2b - \frac{abt}{1-r} \right)^2 + 2 \left( \frac{2ab^2 + a^2b^2t}{1-a^4b^4} \right)^2
\]

\[
= \left( \frac{2ab^2 + a^2b^2t}{1-a^4b^4} \right)^2
\]

\[
+ \frac{1}{1-a^4b^4} = b^2 \left( 2 + \frac{at}{1-r} \right)^2
\]

\[
+ 2a^2b^4 \left( \frac{2 + \frac{at}{1-r}}{1-a^4b^4} \right)^2 + a^2b^4 \left( 2 + \frac{at}{1-r} \right)^2
\]

\[
= \frac{(1-a^2b^2)(1+a^2b^2)}{(1-a^2b^2)(1+a^2b^2)}
\]

\[
= \frac{(1+a^2b^2)(2a^2b^4)}{(1-a^2b^2)(1+a^2b^2)}
\]

\[
= \frac{(1+a^2b^2)(a^2b^4)(2 + \frac{at}{1-r})^2}{(1-a^2b^2)(1+a^2b^2)}
\]

\[
= \frac{(1-a^2b^2)}{(1-a^2b^2)}
\]

\[
= \frac{(1-r)^{-2}[2(1-r) + at]^2(1 + 3a^2b^2)}{(1-a^2b^2)^4}.
\]

Because \(a = (1-2r)\) and \(b = (1-r)^t\), equation (A2) gives
\[
(1-r)^{-2}[2(1-r) + t(1-2r)]^2(1-r)^{2t}
\]

\[
i_i = \frac{2(1-r) + at}{1-r} \left( b^2 + 3a^2b^4 \right)
\]

\[
= \frac{2(1-r) + at}{1-r} \left( b^2 + 3a^2b^4 \right)
\]

\[
= \frac{(1-a^2b^2)}{(1-a^2b^2)^4}.
\]