Genome-Wide High-Resolution Mapping by Recurrent Intermating Using *Arabidopsis thaliana* as a Model

Sin-Chieh Liu,* Stanley P. Kowalski,* Tien-Hung Lan,* Kenneth A. Feldmann† 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, Deleware 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 F_2 individuals derived from crosses between homozygous parents. Comparison of intermated progenies to F_2 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 F_2 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; PUTTERRILL *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 F_2 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; KWOLEK *et al.* 1986; WELLS and KOFOID 1986; TYAGI 1987; FATMI *et al.* 1992). The general approach involves recurrent intermating among

Corresponding author: Andrew H. Paterson, Department of Soil and Crop Sciences, Texas A&M University, College Station, TX 77843-2474. E-mail: ahp2343@bioch.tamu.edu

individuals within a population, mimicking the random-mating behavior of many natural populations. In a population random 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 random 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 intermating as a strategy for genome-wide high-resolution genetic mapping, an intermated population of *Arabidopsis thaliana* was developed and compared with F_2 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 provide a homozygous population of *ca.* 100 individuals that afford rapid genetic orientation of Arabidopsis YACs or contigs along the chromosomes, contributing to identification of a minimal set of contiguous DNA clones that span the genome.

MATERIALS AND METHODS

Population development: *Arabidopsis thaliana* ecotypes Wassilewskija (WS) and mutant stock M13 (biological ecotype Landsberg, carrying *angustifolia,* Leaf/Silique Phenotype, *distorter-1,* Trichome Phenotype, and *erecta,* Size/Inflorescence Phenotype; KRANZ and KIRCHHEIM 1987), were hybridized by hand-crossing, the F_1 verified by RFLP analysis, and the F_1 selfed to generate F_2 seed. A total of 120 F_2 individuals were the starting point for two additional mating schemes, as follows:

Recombinant inbred (RI-selfed population): Each F_2 individual was selfed to generate F_3 progeny, a single F_3 progeny individual was selfed to generate F_4 progeny, *etc.*, to F_6. At F_6, 95 of the 120 F_2 lineages were still represented, the remainder lost due to sterile plants at intervening generations. Only a subset of RI lines were analyzed.

Recurrent random-intermating (intermated population): Each F_2 individual was used as the pollen parent in a hand-cross to another randomly chosen F_2 individual. Each F_2 individual was also used as a seed parent in such a cross, receiving pollen from another F_2 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 F_2 lineages were still represented, the remainder having been lost due to a sterile plant at one of the intervening generations.

F_1 and F_2 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.

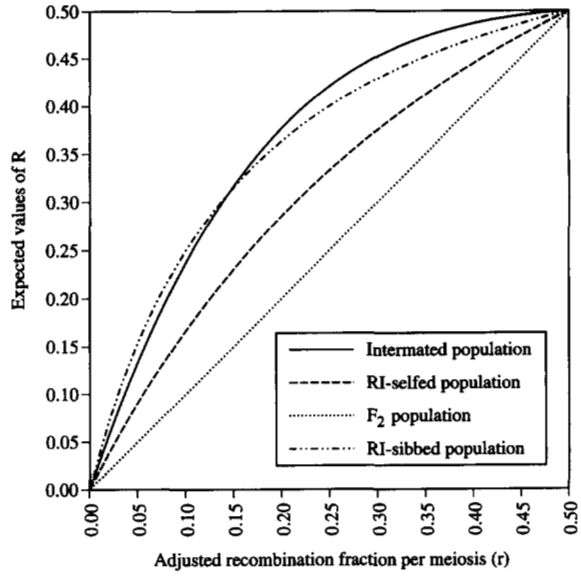

FIGURE 1.—Relationship between the adjusted recombination fraction per meiosis (r) and the expected values of the observed recombination fraction (R) of F_2, RI-selfed, intermated, and RI-sibbed populations. Values for RI-selfed, intermated, and RI-sibbed populations are derived from (1), (3), and (9), respectively.

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 F_2 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 (R_o) in the RI-selfed and intermated maps with the values of the expected R (R_e), which were derived for the respective populations from the R values of corresponding intervals in the F_2 map. The test statistics are

$$2 \ln \left[\frac{R_o^{nR_o}(1 - R_o)^{n(1-R_o)}}{R_e^{nR_o}(1 - R_e)^{n(1-R_o)}} \right]$$

for RI-selfed population and

$$2 \ln \left[\frac{A_o^{nA_o} B_o^{nB_o} C_o^{nC_o} D_o^{nD_o}}{A_e^{nA_e} B_e^{nB_e} C_e^{nC_e} D_e^{nD_e}} \right]$$

for intermated population, where n = number of observations and A_o, B_o, C_o, and D_o, and A_e, B_e, C_e, and D_e are $(1 - R)^2/2$, $R^2/2$, $2R(1 - R)$, $[R^2 + (1 - R)^2]/2$ and with $R = R_o$ and $R = R_e$, respectively. Both test statistics asymptotically distribute as a distribution when n is large (WILKS 1938).

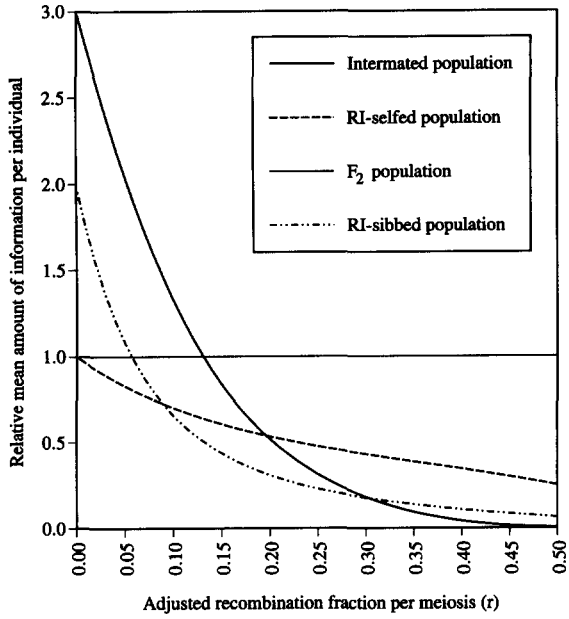

FIGURE 2.—The precision of the estimated r (adjusted recombination fraction per meiosis) of F_2, RI-selfed, intermated, and RI-sibbed populations. Measurement is based on the mean amount of information per individual (i_r) relative to F_2 population: $i_{r\ \text{intermated}} > i_{r\ F2} > i_{r\ \text{RI-selfed}}$ when $0 < r < 0.131$; $i_{r\ F2} > i_{r\ \text{intermated}} > i_{r\ \text{RI-selfed}}$ when $0.131 < r < 0.196$; $i_{r\ F2} > i_{r\ \text{RI-selfed}} > i_{r\ \text{intermated}}$ when $0.196 < r < 0.5$. Values for F_2, RI-selfed, intermated, and RI-sibbed populations are derived from (4), (5), (7), and (10), respectively.

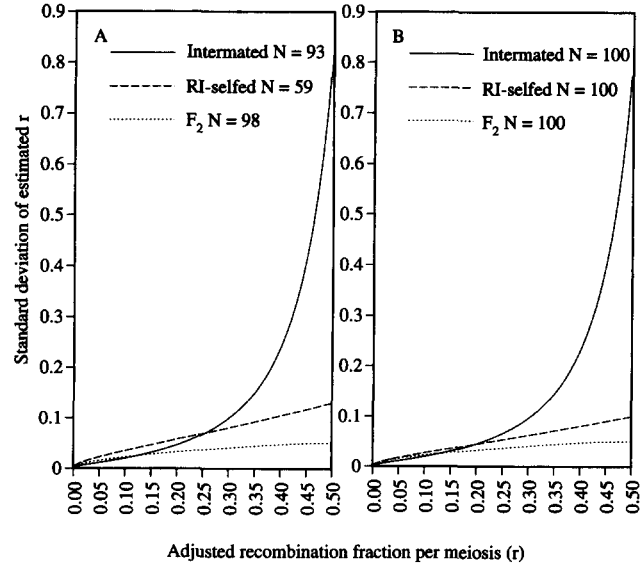

FIGURE 3.—Standard deviation (s) of the estimated r (adjusted recombination fraction per meiosis) calculated from (8) for populations with N individuals. (A) Our F_2 ($N = 98$), RI-selfed ($N = 59$), and intermated ($N = 93$) populations: $s_{\text{intermated}} < s_{F2} < s_{\text{RI-selfed}}$ when $0 < r < 0.131$; $s_{F2} < s_{\text{intermated}} < s_{\text{RI-selfed}}$ when $0.131 < r < 0.255$; $s_{F2} < s_{\text{RI-selfed}} < s_{\text{intermated}}$ when $0.255 < r < 0.5$. (B) F_2, RI-selfed, and intermated populations having equal population sizes ($N = 100$): $s_{\text{intermated}} < s_{F2} < s_{\text{RI-selfed}}$ when $0 < r < 0.131$; $s_{F2} < s_{\text{intermated}} < s_{\text{RI-selfed}}$ when $0.131 < r < 0.196$; $s_{F2} < s_{\text{RI-selfed}} < s_{\text{intermated}}$ when $0.196 < r < 0.5$.

However, the tests performed within a population are not independent, because the intervals are not independent.

RESULTS

Expected recombination under different breeding systems: The observed recombination fraction (R) of an F_2 population is equal to its adjusted recombination fraction (r) per meiosis, because an F_2 population is derived from a single meiosis. Therefore, the F_2 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

$$R = \frac{2r}{1 + 2r}. \qquad (1)$$

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: $R_{t+1} = R_t - rRt + r/2$, where t is the number of generations of intermating following F_2, and $R_0 = r$. Therefore, the relationship between R and r in an intermated population can be derived as

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

and the R and r of our intermated population have a relationship of

$$R = \frac{1}{2}[1 - (1 - r)^4(1 - 2r)]. \qquad (3)$$

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 F_2) 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

$$\frac{d}{dr}\left(\frac{2r}{1 + 2r}\right)\Big|_{r=0} = 2$$

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

respectively.

Efficiency of detecting recombination: The relative

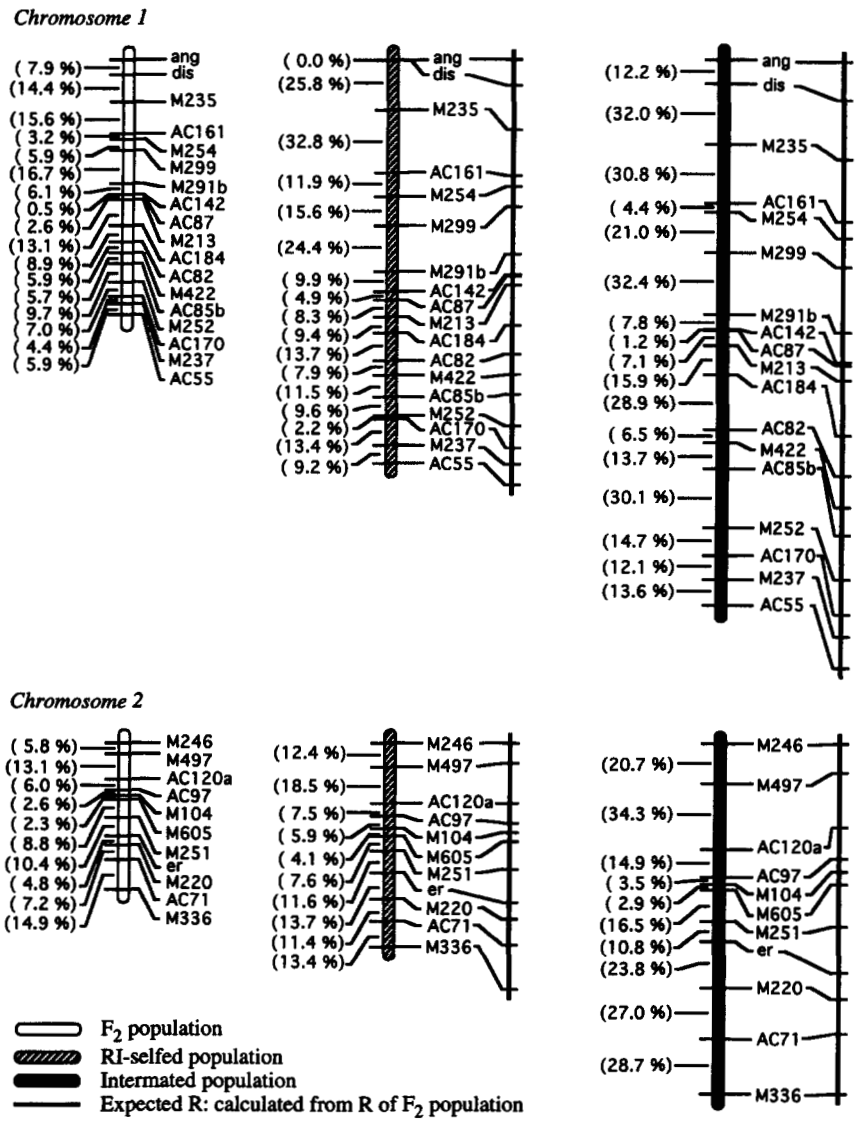

FIGURE 4.—Genetic linkage maps of F_2, RI-selfed, and intermated populations of *Arabidopsis thaliana* WS × M13. Interval lengths are expressed as observed recombination fraction (R) estimated without adjusting for multiple meioses. Expected maps were constructed for RI-selfed and intermated populations by calculating the expected R for each interval from (1) and (3), respectively, using observed R of the F_2 population as the adjusted recombination fraction (r) per meiosis.

precision of estimates of r, in F_2, RI-selfed, and intermated populations can be compared based on the mean amount of information (i_r) provided by a single individual in each population. MATHER (1936) derived i_r of a single F_2 individual as

$$i_r = \frac{2(1 - 3r + 3r^2)}{r(1 - r)(1 - 2r + 2r^2)}. \quad (4)$$

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

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

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

$$i_r = \frac{(1 - r)^{2t-2}[2(1 - r) + t(1 - 2r)]^2 \times [1 + 3(1 - 2r)^2(1 - r)^{2t}]}{[1 - (1 - 2r)^4(1 - r)^{4t}]}, \quad (6)$$

where t is the number of generations of intermating following F_2 (APPENDIX A). Thus, i_r of a single lineage of our intermated population is

$$i_r = \frac{(1 - r)^6[2(1 - r) + 4(1 - 2r)]^2 \times [1 + 3(1 - 2r)^2(1 - r)^8]}{[1 - (1 - 2r)^4(1 - r)^{16}]}, \quad (7)$$

Populations intermated for four generations yield threefold more information per individual than F_2 and RI-selfed populations, for values of r approaching 0 (Figure 2). The values of i_r for intermated population remain the highest among the three populations when $r < 0.131$, 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

$$s = \sqrt{\frac{1}{Ni_r}}, \quad (8)$$

where N is the population size, Ni_r = the inverse of the variance of the estimate of $r = I_r$ is the amount of information concerning the r estimate. In our experi-

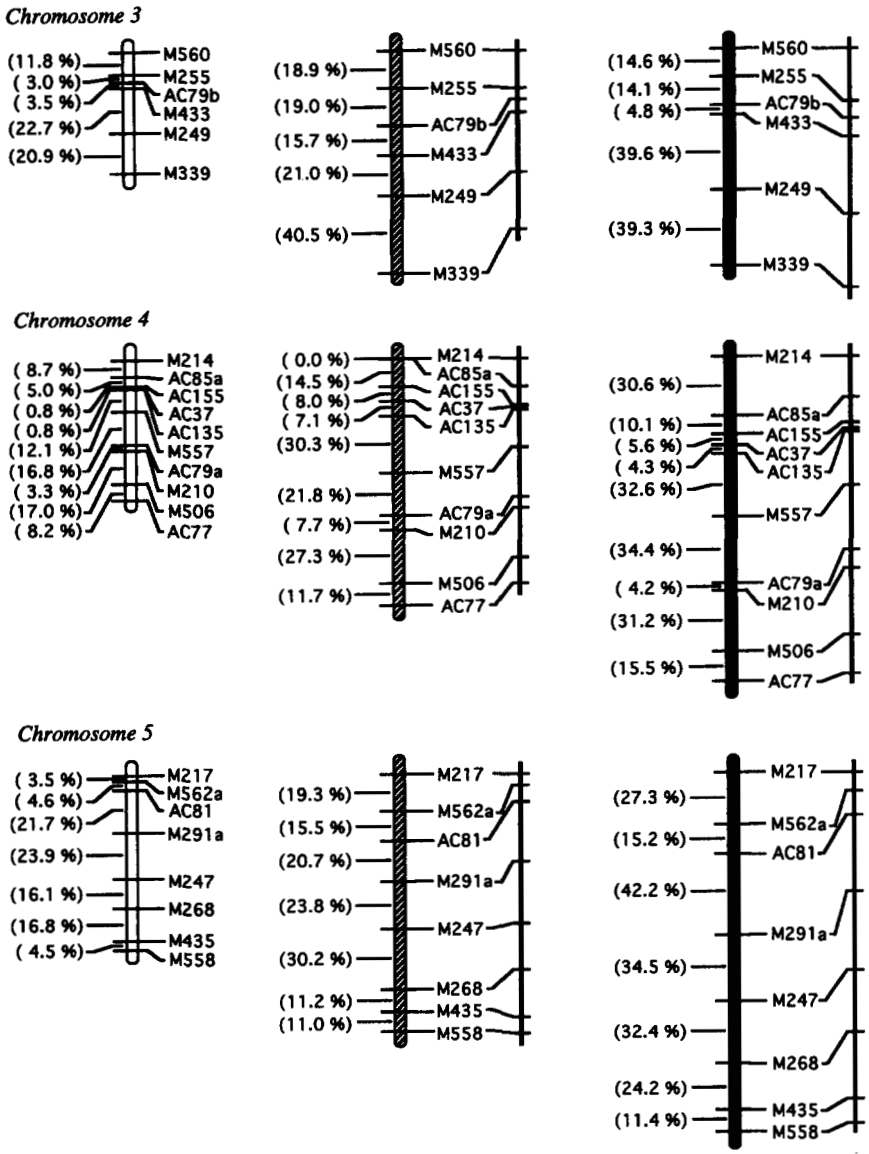

FIGURE 4.—*Continued*

ment, the precision of recombination estimates for F_2, and intermated populations correspond approximately to that of i_r (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 (F_2, 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 F_2 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 concensus order shown (Figure 4) uses the order of closely linked markers in the RI or intermated populations in cases that the F_2 could not resolve, and relies upon F_2 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 F_2 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 F_2, but were significantly different (LOD $\geq$ 2) for the RI-selfed and/or intermated populations. These groups of markers, with the respective LODs in parentheses, were *-AC87-AC142-* (F_2: −0.48, RI-selfed: 0.23, intermated: 2.65) and *-M422-AC85b-* (F_2: 0.79, RI-selfed: 8.89, intermated: 14.12) on chromosome *1*, *M246-M497-* (F_2: 0.20, RI-selfed: 4.37, intermated: 1.52) on chromosome *2*, and

-AC155-AC37-AC135- (F_2: 0.13, RI-selfed: 2.44, intermated: 1.91) on chromosome *4*.

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$_{\text{two-tail}} = 0.0003$; RI: 54.5% WS allele, $t = 2.77$, P value$_{\text{two-tail}} = 0.0080$; intermated: 55.1% WS allele, $t = 5.51$, P value$_{\text{two-tail}} < 0.0001$). 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 intermating 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 intermating among different F_2-derived lineages for four generations, the resulting progeny yield threefold 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" genetic 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-sibbed populations yield a maximum of fourfold expansion of R when r approaches 0 (HALDANE and WADDINGTON 1931; TAYLOR 1978) (Figure 1):

$$R = \frac{4r}{1 + 6r} \tag{9}$$

and

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

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

$$i_r = \frac{4}{r(1 + 2r)(1 + 6r)^2}, \tag{10}$$

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

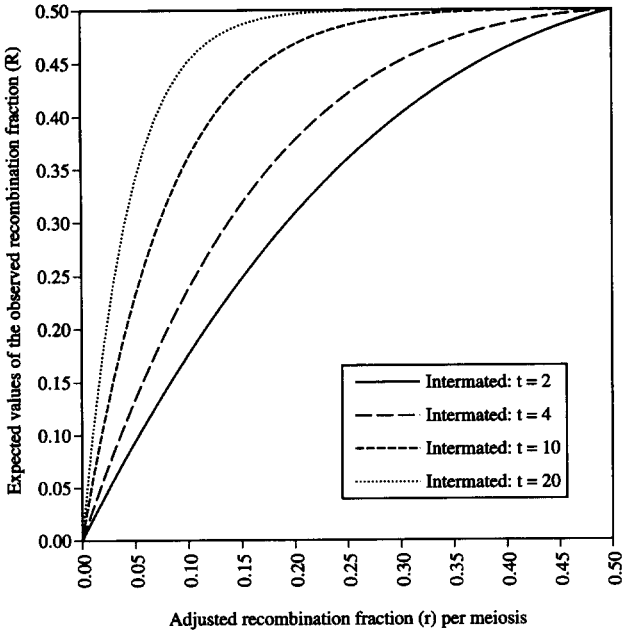

FIGURE 5.—Relationship between the adjusted recombination fraction (r) per meiosis and the expected values of the observed recombination fraction (R) of intermated populations undergo different numbers of generations (t) of intermating. Values are derived from (2).

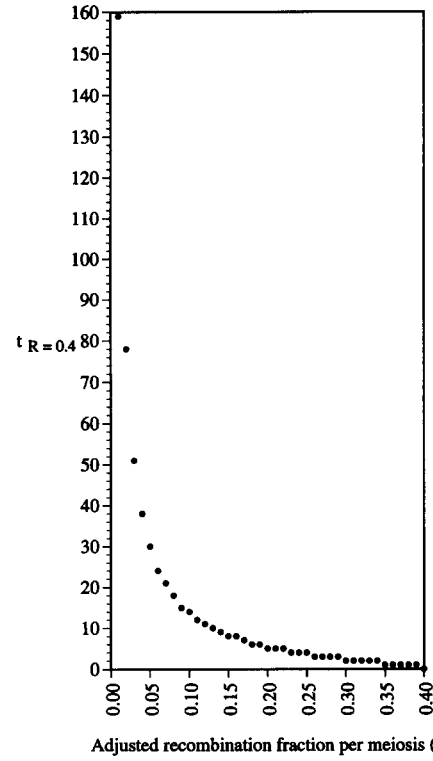

FIGURE 6.—Relationship between (adjusted recombination fraction per meiosis) and $t_{R>0.4}$, the number of generations of intermating expected to render R (observed recombination fraction) > 40%, a value beyond which it is very difficult to detect linkage between markers. Values are derived from (11).

recombinational information and can be subjected to many cycles of recombination to afford further accumulation of information (Figures 5–7).

Many other combinations of intermating 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 intermating has individuals comprised of two different gametes harboring unique recombination sites. Each generation of a recurrent intermating population retains heterozygosity at 50% (theoretically) and accumulates new recombination sites at a constant rate through additional generations of intermating. 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 intermating, and will segregate independently after sufficient generations of intermating. The frequencies of genotypes for such a pair of markers, at the tth generation of intermating are as follows:

$$\text{frequency } (AABB) = \text{frequency } (aabb) = \frac{1}{16}[(1-2r)(1-r)^t + 1]^2,$$

$$\text{frequency } (AAbb) = \text{frequency } (aaBB) = \frac{1}{16}[(1-2r)(1-r)^t - 1]^2,$$

$$\text{frequency } (AABb) = \text{frequency } (aaBb) = \text{frequency } (AaBB) = \text{frequency } (Aabb) = \frac{1}{8}[1-(1-2r)^2(1-r)^{2t}],$$

and

$$\text{frequency } (AaBb) = \text{frequency } (AB/ab) + \text{frequency } (Ab/aB) = \frac{1}{8}[(1-2r)(1-r)^t + 1]^2 + \frac{1}{8}[(1-2r)(1-r)^t - 1]^2,$$

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 intermating. For a pair of markers linked at r recombination fraction (adjusted recombination fraction per meiosis), the number of generations ($tR_{>0.4}$) of intermating 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:

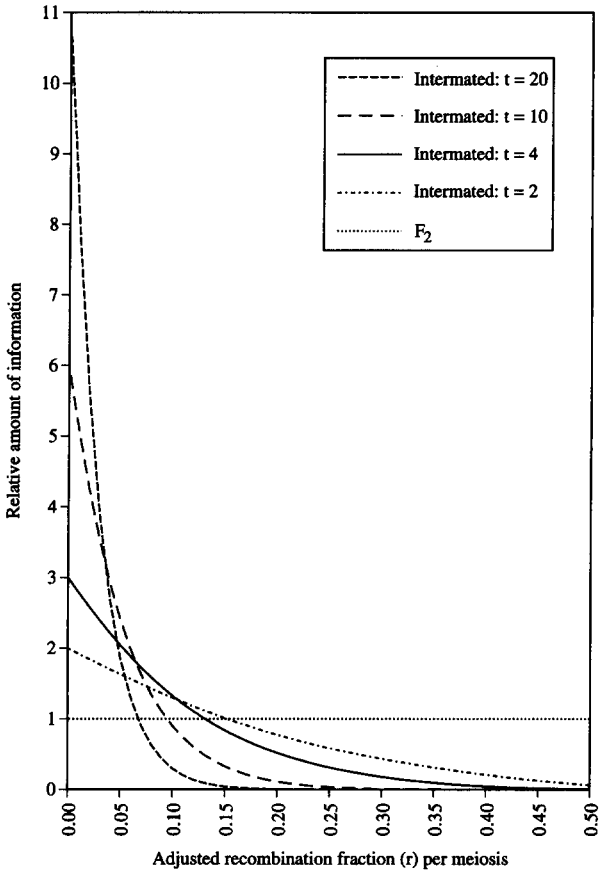

FIGURE 7.—The precision of the estimation of r (adjusted recombination fraction per meiosis) of intermated populations undergo different numbers of generations (t) of intermating. Measurement is based on the amount of information relative to F_2 population. Values for F_2 and intermated populations are derived from (4) and (6), respectively.

$$t_{R>0.4} = \min\left(\text{integer} > \frac{\ln(1-2R) - \ln(1-2r)}{\ln(1-r)}\bigg|_{R=0.4}\right). \quad (11)$$

Figure 7 shows the relationship between $t_{R>0.4}$ 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 the F_2 population (not significant), but LODs of 8.89 and 14.12 greater than the best alternative order in RI-selfed and intermated populations, respectively (both highly significant). Likewise, the order between *AC87* and *AC142* on chromosome *1* yielded a LOD of 0.48 *less* than the best alternative order, but LODs of 0.23 and 2.65 (significant) greater than the best alternative order in RI-selfed and intermated populations, respectively. In this case, use of the best order in F_2 population would have incorrectly ordered the markers. In the other two cases of closely linked markers, *M246-M497-* on chromosome *2* and *-AC155-AC37-AC135-* on chromosome *4*, uncertain orders in F_2 were also better resolved in RI-selfed and intermated populations; however, RI-selfed population showed higher LODs for the best order. Because both chromosomal regions involved markers with severe segregation distortion in the intermated population, it is likely that the expected better resolution from the intermated population was affected by selection at these regions, through generations of intermating. This may also account for the significant difference observed in the likelihood ratio tests comparing the observed R and the expected R of the intermated population. The likelihood ratio tests were used to compare the observed R with the expected R derived from the R observed in the F_2 population. Because the F_2 population does not provide more precise estimation of the r for a small interval than an intermated population, the test results reflect the differences in the estimation of recombination fraction among populations, rather than the precision of the estimation of recombination fraction in the populations. The five significant intervals showing similar deviation from the expected R in both RI-selfed and intermated populations are likely to result from inaccurate estimation of the recombination fraction in the F_2 population. The problems involving selection causing recombinational variation in the intermated population can be minimized by increased population size as discussed below.

The greater recombination afforded by intermated populations was also reflected in the observed segregation distortion in these populations. Specifically, the intermated population showed 16 distinct regions of segregation distortion, while the F_2 population showed only 12. The greater number of distorted regions in the intermated population is expected, due to the greater recombination in this 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 intermating 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 intermating. In an intermated population, population size (N) and the number of generations of intermating (t) determine the precision of the r estimate, which can be measured by the amount of information (I_r) = Ni_r [i_r 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 intermating, should 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; WEISSENBACH *et al.* 1992).

Development of homozygous high-resolution mapping populations: As a high-resolution alternative to RI-selfed populations, intermating 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 (*Zea 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 intermating 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 intermating, 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; PUTTERRILL *et al.* 1993).

The improved genetic map resolution afforded by intermating 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 oleracea, Lycopersicon esculentum, Z. mays* and *Solanum tuberosum* have 290, 400, 500, 540, 750, 1400, and 2500 kb/cM, respectively (CHANG *et al.* 1988; GEBHARDT *et al.* 1989; ARUMUNGANATHAN and EARLE 1991; LANDRY *et al.* 1992; TANKSLEY *et al.* 1992; COE and NEUFFER 1993; CHITTENDEN *et al.* 1994; REINISCH *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; WHITKUS *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. FORSTHOEFEL and L. CHITTENDEN for technical assistance and K. SCHERTZ, 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

AHN, S., and S. D. TANKSLEY, 1993 Comparative linkage maps of rice and maize genomes. Proc. Natl. Acad. Sci. **90:** 7980–7984.

AHN, S., J. A. ANDERSON, M. E. SORRELLS and S. D. TANKSLEY, 1993 Homoeologous relationships of rice, wheat, and maize chromosomes. Mol. Gen. Genet. **241:** 483–490.

ALEXANDER, D. E., 1988 Breeding special nutritional and industrial types, pp. 869–880 in *Corn and Corn Improvement,* Ed. 3, No. 18, edited by G. F. SPRAGUE and J. W. DUDLEY. American Society of Agronomy, Inc., Madison.

ARUMUGANATHAN, K., and E. D. EARLE, 1991 Nuclear DNA content of some important plant species. Plant Mol. Biol. Rptr. **9:** 208–218.

BAILEY, D. W., 1981 Recombinant inbred strains and bilineal congenic strains, pp. 223–239 in *The Mouse in Biomedical Research,* Vol. 1, edited by H. L. FOSTER, J. D. SMALL and J. G. FOX. Academic Press, New York.

BONIERBALE, M. D., R. L. PLAISTED and S. D. TANKSLEY, 1988 RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics **120:** 1095–1103.

BRIM, C. A., 1966 A modified pedigree method of selection in soybeans. Crop Sci. **6:** 220.

BRUCE, A. B., 1910 The Mendelian theory of heredity and the augmentation of vigor. Science **32:** 627–628.

BURR, B., and F. A. BURR, 1991 Recombinant inbreds for molecular mapping in maize: theoretical and practical considerations. Trends Genet. **7:** 55–60.

BURR, B., F. A. BURR, K. H. THOMPSON, M. C. ALBERTSON and C. W. STUBER, 1988 Gene mapping with recombinant inbreds in maize. Genetics **118:** 519–526.

BURR, B., F. A. BURR and E. C. MATZ, 1993 Mapping genes with recombinant inbreds, pp. 249–254 in *The Maize Handbook,* edited by M. FREELING and V. WALBOT. Springer-Verlag, New York.

CHANG, C., J. L. BOWMAN, A. W. DEJOHN, E. S. LANDER and E. M. MEYEROWITZ, 1988 Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana.* Proc. Nat. Acad. Sci. USA **85:** 6856–6860.

CHITTENDEN, L. M., K. F. SCHERTZ, Y. R. LIN, R. A. WING and A. H. PATERSON, 1994 A detailed RFLP map of *sorghum bicolor* × *S. propinquum* suitable for high-density mapping suggests ancestral duplication of chromosomes and chromosomal segments. Theor. Appl. Genet. **87:** 925–933.

CHURCHILL, G. A., J. J. GIOVANNONI and S. D. TANKSLEY, 1993 Pooled-sampling makes high-resolution mapping practical with DNA markers. Proc. Nat. Acad. Sci. USA **90:** 16–20.

COE, E. H., and M. G. NEUFFER, 1993 Gene loci and linkage map of corn (maize) (*Zea mays* L.) (2N=20), pp. 6.157–6.189 in *Genetic Maps: Locus Maps of Complex Genomes,* Ed. 6, edited by S. J. O'BRIEN. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

COHEN, D., I. CHUMAKOV and J. WEISSENBACH, 1993 A first-generation physical map of the human genome. Nature **366:** 698–701.

COLLINS, F. S., 1992 Positional cloning: Let's not call it reverse anymore. Nat. Genet. **1:** 3–6.

COULSON, A., R. WATERSTON, J. KIFF, J. SULSTON and Y. KOHARA, 1988 Genome linking with yeast artificial chromosomes. Nature **335:** 184–186.

EAST, E. M., 1908 Inbreeding in corn. Rep. Ct. Agri. Exp. Sta. **1907:** 419–428.

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.

FREDERICKSEN, L. J., and W. E. KRONSTAD, 1985 A comparison of intermating and selfing following selection for heading date in two diverse winter wheat crosses. Crop Sci. **25:** 555–560.

GEBHARDT, C., E. RITTER, T. DEBENER, U. SCHACHTSCHABEL, B. WALKMEIER *et al.,* 1989 RFLP analysis and linkage mapping in *Solanum tuberosum.* Theor. Appl. Genet. **78:** 65–75.

GREEN, E., L., 1981 Genetics and probability in animal breeding experiments. Oxford University Press, New York.

GRILL, E. and C. SOMERVILLE, 1991 Construction and characterization of a yeast artificial chromosome library of *Arabidopsis* which is suitable for chromosome walking. Mol. Gen. Genet. **226:** 484–490.

HALDANE, J. B. S., and C. H. WADDINGTON, 1931 Inbreeding and linkage. Genetics **16:** 357–360.

HANSON, W. D., 1959a 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., 1959b The breakup of initial linkage blocks under selected mating systems. Genetics **44:** 857–868.

HAUGE, B. M., S. M. HANLEY, S. CARTINHOUR, J. M. CHERRY, H. M. GOODMAN *et al.,* 1993 An integrated genetic/RFLP map of the *Arabidopsis thaliana* genome. Plant J. **3:** 745–754.

HULBERT, S. H., T. E. RICHTER, J. D. AXTELL and J. L. BENNETZEN, 1990 Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. Proc. Natl. Acad. Sci. USA **87:** 4251–4255.

HWANG, I., T. KOHCHI, B. M. HAUGE, H. M. GOODMAN, R. SCHMIDT *et al.,* 1991 Identification and map position of YAC clones comprising one-third of the *Arabidopsis* genome. Plant J. **1:** 367–374.

KOWALSKI, S. P., T. H. LAN, K. A. FELDMANN and A. H. PATERSON, 1994 QTL mapping of naturally-occurring variation in flowering time of *Arabidopsis thaliana.* Mol. Gen. Genet. **245:** 548–555.

KRANZ, A. R., and B. KIRCHHEIM, 1987 Genetic resources, pp. 3.2.52–3.2.107 in *Arabidopsis. Arabidopsis* Information Service, J. W. Goethe-University Frankfurt, Germany.

KWOLEK, T. F., R. E. ATKINS and O. S. SMITH, 1986 Comparisons of agronomic characteristics in C0 and C4 of IAP3BR(M) random-mating grain sorghum population. Crop Sci. **26:** 1127–1131.

LANDE, R., 1975 The maintenance of genetic variability by mutation in a quantitative character with linked loci. Genet. Res. **26:** 221–235.

LANDER, E. S., P. GREEN, J. ABRAHAMSON, 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.

LANDRY, B. S., N. HUBERT, R. CRETE, M. S. CHIANG, 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 *Plasmodiophora brassicae* (Woronin). Genome **35:** 409–420.

LANGLEY, C. H., E. MONTGOMERY and W. F. QUATTLEBAUM, 1982 Restriction map variation in the *Adh* region of *Drosophila*. Proc. Nat. Acad. Sci. USA **79:** 5631–5635.

LEIGH BROWN, A. J., 1983 Variation at the 87A heat shock locus in *Drosophila melanogaster*. Proc. Nat. Acad. Sci. USA **80:** 5350–5354.

LISTER, C. and C. DEAN. 1993 Recombinant inbred lines for mapping RFLP and phenotypic markers in Arabidopsis thaliana. Plant J. **4**: 745–750.

MACPHERSON, J. N., B. S. WEIR and A. J. LEIGH BROWN, 1990 Extensive linkage disequilibrium in the achaete-scute complex of *Drosophila melanogaster*. Genetics **126:** 121–129.

MATHER, K., 1936 Types of linkage data and their value. Ann. Eugenics **7:** 251–264.

MILLER, P. A., and J. O. RAWLINGS, 1967 Breakup of initial linkage blocks through intermating in a cotton breeding population. Crop Sci. **7:** 199–204.

O'BRIEN, S. J., J. E. WOMACK, L. A. LYONS, K. J. MOORE, N. A. JENKINS *et al.*, 1993 Anchored reference loci for comparative genome mapping in mammals. Nat. Genet. **3:** 103–112.

ORKIN, S.H., 1986 Reverse genetics in human disease. Cell **47:** 845–850.

PATERSON, A. H., J. W. DEVERNA, B. LANINI and S. D. TANKSLEY, 1990 Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. Genetics **124:** 735–742.

PATERSON, A. H., M. E. SORRELLS, S. D. TANKSLEY. 1991 DNA markers in crop improvement. Adv. Agron. **46:** 39–90.

PUTTERRILL, J., F. ROBSON, K. LEE and G. COUPLAND, 1993 Chromosome walking with YAC clones in Arabidopsis: isolation of 1700 kb of contiguous DNA on chromosome 5, including a 300 kb region containing the flowering-time gene *CO*. Mol. Gen. Genet. **239:** 145–157.

REINISCH, A. J., J. DONG, C. L. BRUBAKER, D. M. STELLY, J. F. WENDEL *et al.*, 1994 A detailed map of cotton, *Gossypium hirsutum* × *G. barbadense*: Chromosome organization and evolution in a disomic polyploid genome. Genetics **138:** 829–847.

REITER, R. S., J. G. K. WILLIAMS, K. A. FELDMANN, J. A. RAFALSKI, S. V. TINGEY *et al.*, 1992 Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. Proc. Natl. Acad. Sci. USA **89:** 1477–1481.

SCHMIDT, R., G. CNOPS, I. BANCROFT and C. DEAN, 1992 Construction of an overlapping YAC library of the *Arabidopsis thaliana* genome. Aust. J. Plant Physiol. **19:** 341–351.

SHULL, G. H., 1911 The genotypes of maize. Amer. Nat. **45:** 234–252.

SORRELLS, M. E., and S. E. FRITZ, 1982 Application of a dominant male-sterile allele to the improvement of self-pollinated crops. Crop Sci. **22:** 1033–1035.

ST. MARTIN, S. K., and N. E. EHOUNOU, 1989 Randomness of intermating in soybean populations containing male-sterile plants. Crop Sci. **29:** 69–71.

STRICKBERGER, M. W., 1968 *Genetics*. Macmillan, New York.

TANKSLEY, S. D., M. W. GANAL, J. P. PRINCE, M. C. DEVICENTE, M. W. BONIERBALE *et al.*, 1992 High density molecular linkage maps of the tomato and potato genomes. Genetics **132:** 1141–1160.

TAUTZ, D., 1989 Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. **17:** 6463–6471.

TAYLOR, B., 1978 Recombinant inbred strains: use in gene mapping, pp. 423–438 in *Origins of Inbred Mice*, edited by H. MORSE. Academic Press, New York.

TYAGI, A. P., 1987 Correlation studies on yield and fiber traits in upland cotton (*Gossypium hirsutum* L.). Theor. Appl. Genet. **74:** 280–283.

WANG, G. L., D. J. MACKILL, J. M. BONMAN, S. R. MCCOUCH and R. J. NELSON, 1993 RFLP mapping of genes conferring complete and partial resistance in a rice cultivar with durable resistance to blast, pp. 209–215 in *Durability of Disease Resistance*, edited by T. JACOBS and J. E. PARLEVLIET. Kluwer Academic Publishers, Netherlands.

WARD, E., and G. C. JEN, 1990 Isolation of single-copy-sequence clones from a yeast artificial chromosome library of randomly-sheared *Arabidopsis thaliana* DNA. Plant Mol. Biol. **14:** 561–568.

WEISSENBACH, J., G. GYAPAY, C. DIB, A. VIGNAL, J. MORISSETTE *et al.*, 1992 A second-generation linkage map of the human genome. Nature **359:** 794–801.

WELLS, W. C., and K. D. KOFOID, 1986 Selection indices to improve an intermating population of spring wheat. Crop Sci. **26:** 1104–1109.

WHITKUS, R., J. DOEBLEY and M. LEE, 1992 Comparative genome mapping of sorghum and maize. Genetics **132:** 1119–1130.

WILKS, S., S., 1938 The large sample distribution of the likelihood ratio for testing composite hypotheses. Ann. Math. Stat. 9: 60–62.

WRIGHT, S., 1969 *The Theory of Gene Frequencies*, Vol. 2. University of Chicago Press, Chicago.

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_r) is given by the formula

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

(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, AAbb, aaBB, AABb, aaBa, AaBB, Aabb,* and *AaBb* are m_1, m_2, m_3, m_4, m_5, m_6 m_7, m_8, and m_9, respectively, where

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

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

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

and

$$\begin{aligned} m_9 &= \text{frequency } (AB/ab) + \text{frequency } (Ab/aB) \\ &= \frac{1}{8}[(1 - 2r)(1 + r)^t + 1]^2 \\ &\quad + \frac{1}{8}[(1 - 2r)(1 - r)^t - 1]^2 \\ &= \frac{1}{4}[(1 - 2r)^2(1 - r)^{2t} + 1]. \end{aligned}$$

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

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

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

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

$$\frac{dm_5}{dr}=\frac{dm_6}{dr}=\frac{dm_7}{dr}=\frac{dm_8}{dr}=\frac{1}{2}(1-2r)(1-r)^{2t}+\frac{1}{4}\frac{(1-2r)^2(1-r)^{2t}t}{1-r},$$

and

$$\frac{dm_9}{dr}=-(1-2r)(1-r)^{2t}-\frac{1}{2}\frac{(1-2r)^2(1-r)^{2t}t}{1-r}.$$

The $1/m_j/(dm_j/dr)^2$ term for each genotype class can be derived as

$$\frac{1}{m_1}\left(\frac{dm_1}{dr}\right)^2=\frac{1}{m_2}\left(\frac{dm_2}{dr}\right)^2=\frac{1}{4}\left[-2(1-r)^{t}-\frac{(1-2r)(1-r)^{t}t}{1-r}\right]^2$$

$$\frac{1}{m_3}\left(\frac{dm_3}{dr}\right)^2=\frac{1}{m_4}\left(\frac{dm_4}{dr}\right)^2=\frac{1}{4}\left[-2(1-r)^{t}-\frac{(1-2r)(1-r)^{t}t}{1-r}\right]^2$$

$$\frac{1}{m_5}\left(\frac{dm_5}{dr}\right)^2=\frac{1}{m_6}\left(\frac{dm_6}{dr}\right)^2=\frac{1}{m_7}\left(\frac{dm_7}{dr}\right)^2=\frac{1}{m_8}\left(\frac{dm_8}{dr}\right)^2=\frac{1}{2}\frac{\left[2(1-2r)(1-r)^{2t}+\frac{(1-2r)^2(1-r)^{2t}t}{1-r}\right]^2}{1-(1-2r)^2(1-r)^{2t}}$$

and

$$\frac{1}{m_9}\left(\frac{dm_9}{dr}\right)^2=\frac{\left[2(1-2r)(1-r)^{2t}+\frac{(1-2r)^2(1-r)^{2t}t}{1-r}\right]^2}{1+(1-2r)^2(1-r)^{2t}}.$$

The mean amount of information (i_r) is the sum of the term from each genotype class.

$$i_r=\sum_{j=1}^{9}\left[\frac{1}{m_j}\left(\frac{dm_j}{dr}\right)^2\right]$$

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

$$+\frac{\left[2(1-2r)(1-r)^{2t}+\frac{(1-2r)^2(1-r)^{2t}t}{1-r}\right]^2}{1+(1-2r)^2(1-r)^{2t}}. \quad \text{(A1)}$$

Let a = $(1-2r)$ and b = $(1-r)^t$. Equation (A1) gives

$$i_r=\left(-2b-\frac{abt}{1-r}\right)^2+2\frac{\left(2ab^2+\frac{a^2b^2t}{1-r}\right)^2}{1-a^2b^2}+\frac{\left(2ab^2+\frac{a^2b^2t}{1-r}\right)^2}{1+a^2b^2}=b^2\left(2+\frac{at}{1-r}\right)^2+2a^2b^4\frac{\left(2+\frac{at}{1-r}\right)^2}{1-a^2b^4}+a^2b^4\frac{\left(2+\frac{at}{1-r}\right)^2}{1+a^2b^2}$$

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

$$=\frac{\left(2+\frac{at}{1-r}\right)^2\times(b^2-a^4b^6+2a^2b^4+2a^4b^6+a^2b^4-a^4b^6)}{(1-a^4b^4)}=\frac{\left[\frac{2(1-r)+at}{1-r}\right]^2(b^2+3a^2b^4)}{(1-a^4b^4)}$$

$$=\frac{(1-r)^{-2}[2(1-r)+at]^2b^2(1+3a^2b^2)}{(1-a^4b^4)}. \quad \text{(A2)}$$

Because a = $(1-2r)$ and b = $1-r)\,t$, equation (A2) gives

$$i_r=\frac{(1-r)^{-2}[2(1-r)+t(1-2r)]^2(1-r)^{2t}\times[1+3(1-2r)^2(1-r)^{2t}]}{[1-(1-2r)^4(1-r)^{4t}]}$$

$$=\frac{(1-r)^{2t-2}[2(1-r)+t(1-2r)]^2\times[1+3(1-2r)^2(1-r)^{2t}]}{[1-(1-2r)^4(1-r)^{4t}]}.$$