Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 1477–1481, February 1992
Genetics

Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs

(restriction fragment length polymorphisms/chromosome walking/genetic map)

ROBERT S. REITER*, JOHN G. K. WILLIAMS*, KENNETH A. FELDMANN*†, J. ANTONI RAFALSKI‡, SCOTT V. TINGEY‡, AND PABLO A. SCOLNIK*§

*Central Research and Development and ‡Agricultural Products, E. I. Du Pont de Nemours & Company, Experimental Station, Wilmington, DE 19880-0402

Communicated by Nina V. Fedoroff, November 15, 1991 (received for review August 5, 1991)

ABSTRACT **A population of *Arabidopsis thaliana* recombinant inbred lines was constructed and used to develop a high-density genetic linkage map containing 252 random amplified polymorphic DNA markers and 60 previously mapped restriction fragment length polymorphisms. Linkage groups were correlated to the classical genetic map by inclusion of nine phenotypic markers in the mapping cross. We also applied a technique for local mapping that allows targeting of markers to a selected genome region by pooling DNA from recombinant inbred lines based on their genotype. We conclude that random amplified polymorphic DNAs, used in conjunction with a recombinant inbred population, can facilitate the genetic and physical characterization of the *Arabidopsis* genome and that this method is generally applicable to other organisms for which appropriate populations either are available or can be developed.**

The crucifer *Arabidopsis thaliana* is a useful system for basic studies in plant molecular genetics due to its relatively small genome size, small amounts of dispersed repetitive DNA, and rapid generation time (1). These attributes have made *Arabidopsis* an attractive model system for the analysis of genome organization and the development and use of technology to clone genes known only through their genetic map position.

High-density genetic maps based upon DNA markers can provide starting points for chromosome walking experiments. Markers closely linked to a mutation of interest can reduce the amounts of DNA to be cloned and help establish the direction of the chromosome walk. Restriction fragment length polymorphisms (RFLPs) have been used as markers to construct genetic maps (2) and as starting points for chromosome walking (3). To date, two different RFLP maps have been reported in *Arabidopsis* (4, 5).

Recently another class of genetic markers [random amplified polymorphic DNAs (RAPDs)] has been described (6, 7), which relies on the observation that a single oligonucleotide primer, of arbitrary nucleotide sequence, will direct the amplification of discrete loci (for a more detailed description of this method, see ref. 8). We report here the use of RAPD markers to construct a genetic map of *A. thaliana*. This map has been constructed with unprecedented speed by using RAPD markers and a recombinant inbred (RI) population. For many mapping purposes RI populations are superior to F_2 or backcross populations because they constitute a permanent population in which segregation is fixed (9). Additional markers scored on the same RI population are automatically integrated with the existing map, making map information cumulative (9).

Near-isogenic lines have been used to target RFLP (10) or RAPD (11) markers to specific segments of a genome. However, construction of near-isogenic lines is time consuming, and unlinked portions of the donor genome remain even after several crosses to the recurrent parent (12). Pooling DNA based on phenotype has been used as a means of either identifying additional RFLP loci (13) or mapping existing RFLP loci (14) near genes of interest. Recently, Michelmore *et al.* (15) have shown that DNA pools based on phenotype can be used to target markers to a locus responsible for the phenotype. This technique allowed them to quickly identify three additional RAPD markers linked to a disease-resistance locus in lettuce. Pooling DNA from segregants is an attractive alternative to constructing near-isogenic lines. We demonstrate a strategy for targeting markers to discrete regions of the genome, which involves pooling segregants based on genotype. Using RI lines that constitute a permanent set of fixed genotypes in combination with RAPD markers, one can quickly target markers to virtually any portion of a genome from a single locus up to an entire chromosome. This method is generally applicable to any organism in which a set of genotyped individuals is available.

MATERIALS AND METHODS

Population Development. *Arabidopsis* marker line W100, which carries nine phenotypic markers (*an, ap-1, er, py, hy-2, gl-1, bp, cer-2*, and *tt-3*; ref. 16) was crossed with pollen from wild-type *A. thaliana* ecotype Wassileskija (WS). After self-fertilization of the F_1 hybrid, we generated 150 RI lines to F_8 from individual F_2 plants using single-seed descent (17).

DNA Isolation. DNA was isolated from 3- to 4-week-old plants grown in liquid culture. Five to 20 surface-sterilized *Arabidopsis* seeds were inoculated into 50 ml of sterile Gamborg's B5 liquid medium (GIBCO–BRL) and grown under continuous illumination on a rotary platform (50 rpm). Harvested tissue was frozen at −70°C, lyophilized, and ground before DNA isolation. DNA was isolated by the method of Murray and Thompson (18), but in which an additional phenol/chloroform extraction and ethanol precipitation were included as a final step.

Primer Synthesis. Oligodeoxynucleotide primers were synthesized in a Du Pont Coder 300 automated DNA synthesizer or were obtained from either National Biosciences (Hamel, MN) or Operon Technologies (Alameda, CA). Primers were purified by gel filtration on Sephadex G25 (NAP-5 columns,

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: RAPD, random amplified polymorphic DNA; RFLP, restriction fragment length polymorphism; cM, centimorgan(s); RI, recombinant inbred; WS, *Arabidopsis thaliana* ecotype Wassileskija; LOD, logarithm of odds.

†Present address: Department of Plant Science, Forbes Hall, University of Arizona, Tucson, AZ 85721.
§To whom reprint requests should be addressed.

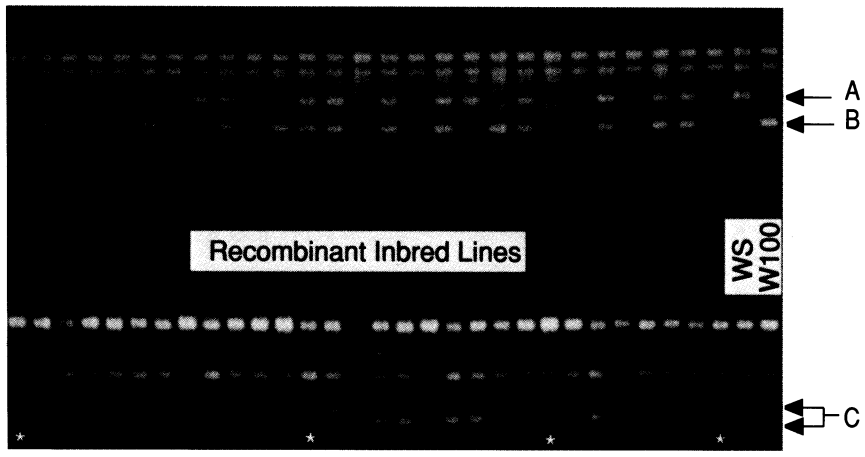

FIG. 1. Example of RAPD band segregation in RI populations. Ethidium bromide-stained electrophoretic pattern of DNA amplified from individual RI or parental lines. Arrows indicate two independent dominant markers amplified by primer r239 (A and B) and a single codominant marker amplified by primer rap14e (C). Asterisks indicate heterozygotes identified by primer rap14e.

Pharmacia). Primer length was 10 nucleotides, except for 96 primers that were nine nucleotides long.

RAPD Markers. Amplification was done in a Biocycler oven (Bios, New Haven, CT). Amplification parameters were 92°C for 3 min followed by 45 cycles with 3°C tolerance at 95°C for 30 sec, 34°C for 15 sec, and 74°C for 1 min. A 72°C incubation for 8 min was included as a final step. Twenty-five-microliter amplification reactions contained 80 mM Tris·HCl (pH 9.0), 20 mM $(NH_4)_2SO_4$, 3.5 mM $MgCl_2$, 100 μM of each dNTP (Boehringer Mannheim, pH 7.0), 0.4 μM primer, 1 unit of AmpliTaq DNA polymerase (Perkin–Elmer/Cetus), and 25 ng of genomic DNA. Reactions were held in 96-well polycarbonate microtiter plates (Techne, Princeton, NJ) and overlaid with 20 μl of mineral oil. With a mapping population of 46 RIs, two RAPD primers can be examined per 96-well microtiter plate. Amplification products were analyzed by gel electrophoresis in 1.2% agarose gels and visualized by ethidium bromide staining. Markers were scored for the presence or absence of the corresponding DNA band among the segregating RI populations.

RFLP Markers. Genomic DNA (1 μg) was digested with one of five restriction enzymes (*Bam*HI, *Eco*RI, *Eco*RV, *Hin*dIII, and *Xba* I) and transferred to Immobilon-N (Millipore), according to the manufacturer's instructions. Prehybridizations and hybridizations were done at 65°C in 5× standard saline phosphate/EDTA (SSPE) (1× SSPE is 0.18 M NaCl/10 mM phosphate, pH 7.4/1 mM EDTA)/0.5% SDS/5× Denhardt's solution/5% dextran sulfate/denatured salmon sperm DNA at 100 μg/ml. DNA isolated either from λ phage (19) or from cosmids (20) was labeled by the random hexamer method (21).

Linkage Analysis. Segregating markers were scored as either A (WS homozygote) or B (W100 homozygote). Residual heterozygotes and ambiguous data were scored as M (missing). Linkage analysis was done by using a Macintosh implementation of the MapMaker program (22) provided by S. V. Tingey and J. A. Rafalski. The data were analyzed as an F_2 population because MapMaker does not handle RI data; this results in a doubling of the logarithm of odds (LOD) scores; thus a LOD score 6.0 was set for linkage threshold. The recombination frequency (r) in a single meiosis was calculated from the fraction of recombinants (R) using the relationship $r = R/2(1-R)$ (23). Map distances in centimorgans (cM) were calculated from recombination frequencies (r) by using Kosambi's mapping function (24).

Targeting Markers to Chromosome 1. Two DNA pools, each containing equal amounts of DNA from six different RI populations, were used to identify additional polymorphisms residing on chromosome 1. These pools were constructed based upon their genotype at all RFLP loci mapped on chromosome 1. The WS pool contained DNA from RI lines that were homozygous for WS alleles at each locus tested on chromosome 1. The W100 pool was analogous, except RI populations were homozygous for W100 alleles. Amplification reactions were done as described using 25 ng of pooled DNA.

RESULTS

Genetic Mapping with RAPD Markers. In an initial screen for polymorphisms between WS and W100, we tested ≈1200 RAPD primers. Twenty percent (245) of these primers showed polymorphisms between the two lines. The total number of polymorphic DNA bands was 392. These polymorphisms were mapped in a population of 46 RI lines. Of the 392 polymorphic DNA bands, 225 polymorphisms segregated in a Mendelian fashion in the RI subpopulation (Fig. 1). The remaining polymorphisms either segregated in a non-

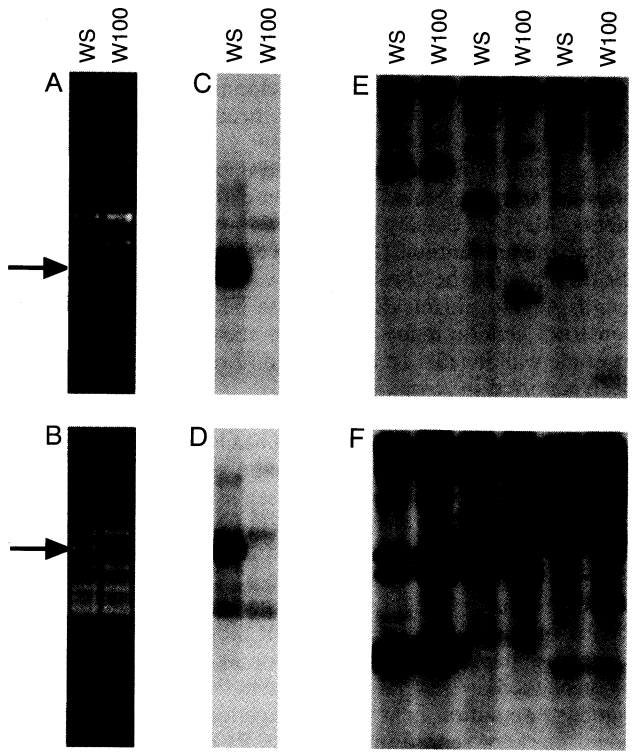

FIG. 2. Blot analyses of RAPD bands. Amplifications used either primer r11 (*A*) or primer r315 (*B*). After agarose gel electrophoresis, the appropriate polymorphic bands (arrows) were excised and used as hybridization probes to blots of duplicate amplification reactions (*C* and *D*) and genomic DNA digested with *Bam*HI, *Eco*RI, and *Eco*RV (*E* and *F*, left to right).

Mendelian fashion or were not reproducible. Ninety-eight percent of the polymorphisms identified by a single primer were >10 cM apart.

The genomic sequence complexity of amplified loci is important if RAPDs are to be used as genetic anchors for physical mapping or as starting points for chromosome walking. To determine the copy number of our RAPD markers, we hybridized DNA from 18 polymorphic RAPD bands to DNA blots (Fig. 2). Results indicate that 9 of the RAPD fragments hybridized to sequences present at 3 copies or less in the genome, whereas the remaining RAPD fragments hybridized to sequences present at 3–10 copies. Thus, one-half of the RAPDs identified may be useful as hybridization probes representing a specific genomic target. Interestingly, 16 of 18 RAPDs (89%), which were used as hybridization probes, also identified an RFLP between WS and W100. This frequency is greater than that seen with genomic λ clones (50%) or cosmids (65%). RAPDs may, thus, be useful in identifying traditional RFLPs.

Genetic Mapping with RFLP Markers. To provide integration with the two existing *Arabidopsis* RFLP maps, we remapped 44 genomic λ clones (4) and 16 cosmid clones (5) by using a collection of 115 RI lines that included the same 46 lines used for RAPD mapping.

Genetic Map Construction. A genetic map was constructed based on the segregation of all three marker types (RAPD, RFLP, and phenotypic; Fig. 3). Five linkage groups were identified and assigned to the five *Arabidopsis* chromosomes by their linkage to phenotypic markers (16, 25). A subset of 126 markers that could be ordered with a LOD score >3 provided a core map with an average distance of 5 cM between markers. The remaining markers could not be ordered unambiguously with the same LOD score, and they are shown in relation to the core map. Ordering of these regions can probably be improved by increasing the mapping population size to provide more recombination information.

Local Mapping. A large number of RAPD markers could provide additional genetic anchor points and fingerprinting information. To increase the marker density on chromosome 1, we have created two DNA pools from RI lines, as described, to collaborate on a project with J. Ecker (University of Pennsylvania) to physically map *Arabidopsis* chromosome 1 by using RAPDs and RFLPs as hybridization probes to yeast artificial chromosomes (26). By testing 384 additional primers we identified 32 more polymorphisms between each pool (Fig. 4). Of these polymorphisms, 23 mapped to chromosome 1 (Fig. 3), 4 mapped to other regions of the genome, and the remaining 5 did not segregate in a Mendelian fashion.

DISCUSSION

We have used an *A. thaliana* RI population to demonstrate the use of RAPD markers for constructing genetic maps. The

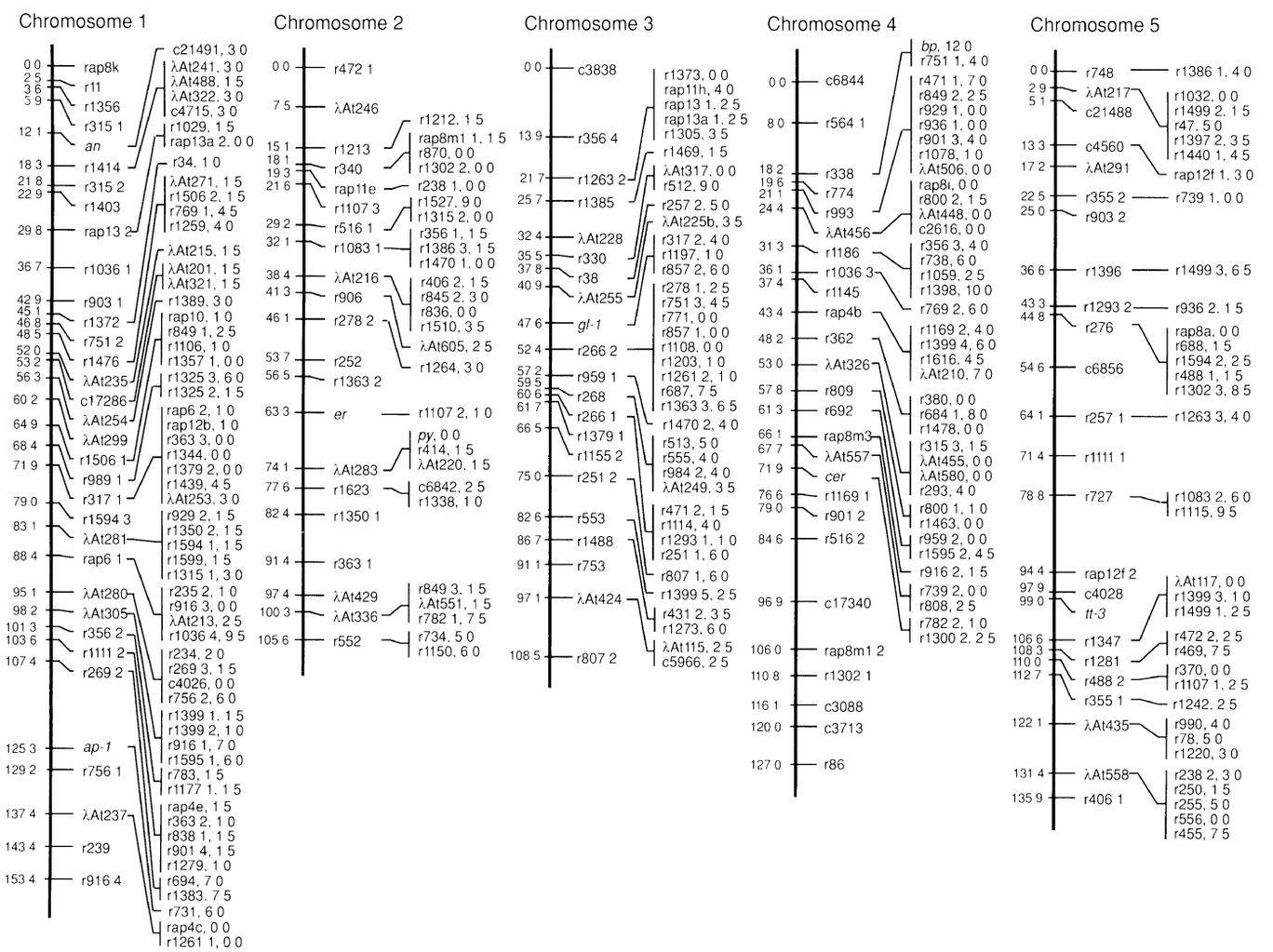

FIG. 3. Linkage map of *A. thaliana*. The 126 markers along the five vertical bars were ordered with LOD score differences (>3.0; see ref. 22). Marker loci listed to the right of each chromosome bar could not be ordered with equal confidence (LOD score differences <3.0). All loci were linked with a LOD score >6.0. The first marker at the top of each chromosome was assigned position 0.0, and approximate marker positions are shown in cM to the left of each chromosome bar. Markers to the right are listed along with an approximate distance in cM from the markers placed on the LOD 3.0 linkage map. Chromosome 1 RAPD markers with numbers >r1300 were placed by local mapping. Phenotypic markers are italicized; RFLP markers are designated either λAt (4) or c (5). The remaining markers are RAPDs.

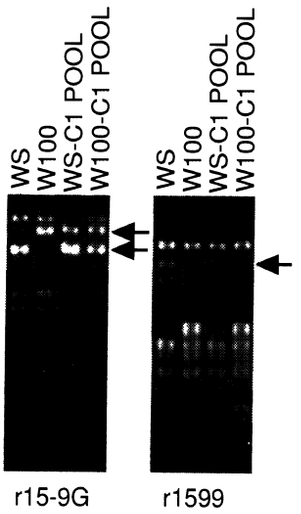

FIG. 4. Example of local mapping for detection of polymorphisms with chromosome 1 DNA pools. RAPD primers r15-9G and r1599 identify polymorphic bands in parental lines. Only primer r1599 identifies a polymorphism between WS-C1 and W100-C1 pools. Primer r1599 was subsequently mapped to chromosome 1 near position 83.1 (see Fig. 3). Arrows indicate polymorphic bands.

final map contains 320 marker loci with an average distance of 2 cM between markers. This map has a total length of 630 cM, slightly greater than previous estimates (4, 5, 25). We have also integrated our RAPD map with the existing classical and RFLP maps (4, 5, 25).

We were able to construct this map in a reasonable time and with a relatively small effort due to the speed with which RAPD markers can be generated. By screening 1200 different primers, 225 marker loci were identified and mapped in *Arabidopsis* by two full-time workers in a 4-mo period. By comparison, it took five full-time workers 2 yr to construct a 600-marker soybean RFLP map (J.A.R. and S.V.T., unpublished work).

We have used a different technique, which we call "local mapping," to target a specific region of the *Arabidopsis* genome for saturation with RAPD markers. This technique is based upon pooling DNA from RI lines selected on the basis of their genotype. The targeted regions can range from an entire chromosome to a small region surrounding a single locus. The technique can be used with RFLP markers but is especially useful with RAPD markers because large numbers of primers can be analyzed very quickly. Using this approach, we targeted markers to chromosome 1, where we added 23 additional markers to the 47 RAPDs obtained through our global mapping efforts. In this experiment, four additional RAPD markers did not map to chromosome 1. This finding is probably due to the fact that only six individuals in our genotyped population were homozygous for all parental alleles tested with chromosome 1 RFLPs. These six individuals were also fixed for alleles from only one parent in other regions of the genome. For experiments in which the mapping target is smaller than an entire chromosome, more individuals should be available for pooling, thus decreasing the probability of having only alleles from one parent in regions outside the target (15).

Our collection of genetically fixed and genotyped RI lines allows targeting of markers to any part of the *Arabidopsis* genome by pooling the appropriate RI lines by genotype. The approach has use for filling gaps in genetic maps and providing either additional markers near single loci to support chromosome walking or entire chromosomes to support physical mapping.

Although we have used *Arabidopsis* as a model system, the approach described here can be applied to any organism for which RI lines exist or can be developed (9, 27). RAPD markers can also be mapped efficiently in any population where segregation can be scored in single recombinant gametes. Several examples include organisms where a haploid generation can be maintained, doubled haploid and backcross populations, and F_2 populations where only markers linked in coupling are considered.

A disadvantage of RAPD markers is that they typically are dominant, whereas RFLP markers are generally codominant. Therefore, the type of mapping population used is important to maximize the amount of recombination information obtained. The relative two-point mapping efficiency of F_2 and backcross populations has been calculated by Allard (28). F_2 populations efficiently map codominant markers, such as RFLPs, but map dominant markers less efficiently (Fig. 5). Backcross populations map dominant and codominant markers with equal efficiency (Fig. 5). We have calculated (Fig. 5 legend) the mapping efficiency of RI populations for both marker types and found that dominant markers can be mapped as efficiently as codominant markers when RI populations are used (Fig. 5). In addition to mapping dominant and codominant markers with the same efficiency, RI lines are advantageous because they constitute permanent mapping populations in which all marker information is cumulative (29). For example, when cloned genes are mapped, a single hybridization provides linkage information to all previously mapped loci in the population (27).

RAPD bands observed in either WS or W100 lines may not be detectable in other *Arabidopsis* ecotypes. Crossing plants of these ecotypes to either WS or W100 would allow use of the corresponding RAPDs from our map. Additional RAPD markers specific for the other ecotype can be added by local mapping. Alternatively, many of our mapped RAPD markers could be used as hybridization probes and, thus, converted to RFLPs.

We intend to use RAPDs as anchors for physical mapping and starting points for chromosome walking; however, RAPDs corresponding to interdispersed multicopy DNA would not be immediately useful to identify, by hybridization, particular cloned DNA corresponding to single genomic locations. According to our results, this procedure could constitute a problem with approximately half of the markers reported here. At present, we foresee two possible approaches to this problem. (*i*) Other RAPD markers identified by local mapping could be preselected for single copy by hybridization to genomic DNA. (*ii*) An alternative approach is the use of sequence information corresponding to the ends

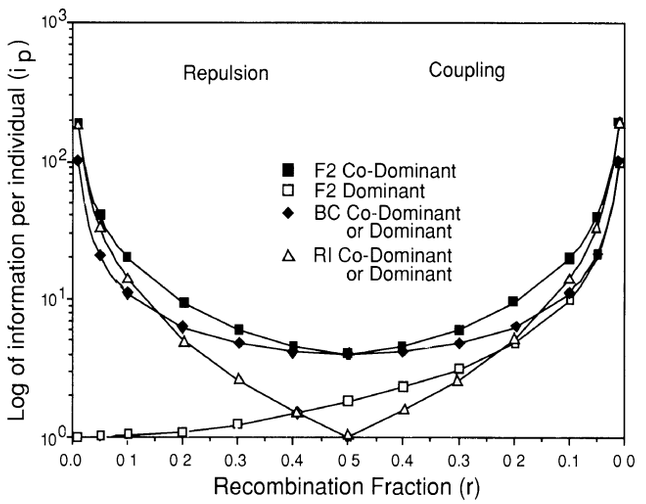

FIG. 5. Theoretical efficiency with which RI, F_2, and backcross (BC) populations can detect recombinants by using either codominant or dominant markers. The amount of information per individual (i_p) in a mapping population is the inverse of the variance divided by population size (28). For an RI population i_p is approximately equal to $2/r(1 + 2r)^2$. Allard (28) previously derived i_p for the other populations shown. The amount of information per individual is represented by the logarithm of i_p and is plotted against the recombination fraction (r) for repulsion- and coupling-phase linkage. An RI population is equally efficient with either codominant or dominant markers and is very efficient for closely linked markers.

of multicopy RAPD fragments. Because these RAPD bands correspond to discrete loci the DNA sequences used as priming sites for amplification are probably specific. Sequence information would permit the conversion of RAPDs into sequence-tagged sites (29). This procedure would allow the use of the PCR to screen genomic libraries (30).

We acknowledge Roslyn Young for excellent technical assistance; Andrew Paterson for valuable comments; Michael Hanafey for assistance with mapping efficiency calculations; Elliot Meyerowitz, Brian Hauge, and Howard Goodman for providing RFLP probes; and Timothy Caspar, Forrest Chumley, Barbara Mazur, David Patton, and Barbara Valent for critical reading of the manuscript.

1. Meyerowitz, E. M. & Pruitt, R. E. (1985) *Science* **299,** 1214–1218.
2. Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. (1980) *Am. J. Hum. Genet.* **32,** 314–331.
3. Rommens, J. M., Iannuzzi, M. C., Kerem, B.-S., Drumm, M. L., Melmer, G., Dean, M., Rozmahel, R., Cole, J. L., Kennedy, D., Hidaka, N., Zsiga, M., Buchwald, M., Riordan, J. R., Tsui, L.-C. & Collins, F. S. (1989) *Science* **245,** 1059–1065.
4. Chang, C., Bowman, J. L., DeJohn, A. W., Lander, E. S. & Meyerowitz, E. M. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 6856–6860.
5. Nam, H.-G., Giraudat, J., den Boer, B., Moonan, F., Loos, W. D. B., Hauge, B. M. & Goodman, H. M. (1989) *Plant Cell* **1,** 699–705.
6. Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A. & Tingey, S. V. (1990) *Nucleic Acids Res.* **18,** 6531–6535.
7. Welsh, J. & McClelland, M. (1990) *Nucleic Acids Res.* **19,** 303–306.
8. Rafalski, J. A., Tingey, S. V. & Williams, J. G. K. (1991) *AgBiotech News Inf.* **3,** 645–648.
9. Bailey, D. W. (1971) *Transplantation* **11,** 325–327.
10. Young, N. D., Zamir, D., Ganal, M. W. & Tanksley, S. D. (1988) *Genetics* **120,** 579–585.
11. Martin, G. B., Williams, J. G. K. & Tanksley, S. D. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 2336–2340.
12. Stam, P. & Zeven, C. (1981) *Euphytica* **30,** 227–238.
13. Arnheim, N., Strange, C. & Erlich, H. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 6970–6974.
14. Patton, D. A., Franzmann, L. H. & Meinke, D. W. (1991) *Mol. Gen. Genet.* **227,** 337–347.
15. Michelmore, R. W., Paran, I. & Kessell, R. V. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 9828–9832.
16. Koornneef, M., Hanhart, C. J., Van Loenen Martinet, E. P. & Van der Veen, J. H. (1987) *Arabid. Inf. Serv. (Frankfurt)* **23,** 46–50.
17. Brim, C. A. (1966) *Crop Sci.* **6,** 220–221.
18. Murray, M. G. & Thompson, W. F. (1980) *Nucleic Acids Res.* **8,** 4321–4325.
19. Chisholm, D. (1989) *BioTechniques* **7,** 21–23.
20. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7,** 1513–1523.
21. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
22. Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E. & Newburg, L. (1987) *Genomics* **1,** 174–181.
23. Haldane, J. B. S. & Waddington, C. H. (1931) *Genetics* **16,** 357–374.
24. Kosambi, D. D. (1944) *Ann. Eugen.* **12,** 172–175.
25. Koornneef, M., van Eden, J., Hanhart, C. J., Stam, P., Braaksma, F. J. & Feenstra, W. J. (1983) *J. Hered.* **74,** 265–272.
26. Ewens, W. J., Bell, C. J., Donnelly, P. J., Dunn, P., Matallana, E. & Ecker, J. R. (1991) *Genomics* **11,** 799–805.
27. Burr, B., Burr, F. A., Thompson, K. H., Albertson, M. C. & Stuber, C. W. (1988) *Genetics* **118,** 519–526.
28. Allard, R. W. (1956) *Hilgardia* **24,** 235–278.
29. Olson, M., Hood, L., Cantor, C. & Botstein, D. (1989) *Science* **245,** 1434–1435.
30. Green, E. D. & Olson, M. V. (1990) *Science* **250,** 94–98.