

Genetic mapping of mutations using phenotypic pools and mapped RAPD markers

John G.K. Williams, Robert S. Reiter, Roslyn M. Young and Pablo A. Scolnik*

Central Research and Development, E. I. Du Pont de Nemours & Co., Wilmington, DE 19880, USA

Received January 7, 1993; Revised and Accepted April 20, 1993

ABSTRACT

Genetic markers facilitate the study of inheritance and the cloning of genes by genetic approaches. Molecular markers detect differences in DNA sequence, and are thus less ambiguous than phenotypic markers, which require gene expression. We have demonstrated a molecular approach to the mapping of mutant genes using RAPD markers and pooling of individuals based on phenotype. To map genes by phenotypic pooling a strain carrying a mutation is crossed to a strain that is homozygous for the wild-type allele of the corresponding gene. A set of primers corresponding to mapped RAPDs distributed throughout the genome and in coupling phase with respect to the wild type parent is then used to amplify DNA from wild type and mutant pools of F_2 individuals. Linkage between the mutant gene and the RAPD markers is visualized by the absence of the corresponding RAPD DNA bands in the mutant pool. We developed a mathematical model for calculating the probability of linkage between RAPDs and target genes and we successfully tested this approach with the model plant *Arabidopsis thaliana*.

INTRODUCTION

The study of genetics is based on the use of markers to study patterns of inheritance. Phenotypic markers such as flower color or leaf shape are easy to identify, but in some cases are difficult to score accurately because of modulation by environment or by modifier genes. Moreover, the utility of phenotypic markers is restricted because only a few markers are available in any single cross and consequently linkage can be detected in only a small fraction of the genome. DNA markers circumvent these difficulties by directly detecting inheritance of allelic differences in DNA sequence. These differences, called DNA polymorphisms, can be detected as changes in the size of restriction fragments (RFLP) or by amplification of DNA by procedures based on the polymerase chain reaction (PCR).

Phenotypic markers have been used in plant genetics to map mutations. However, because of the drawbacks of phenotypic markers it would be desirable to develop new mapping methods that rely exclusively on molecular markers.

The efficiency of linkage detection with molecular markers can be increased by scoring polymorphisms in DNA samples formed from pools instead of individuals (1–4). We have now developed a computer simulation for this process and we have demonstrated the utility of the method by mapping several phenotypic traits in the model plant system *Arabidopsis thaliana*.

MATERIALS AND METHODS

Computer modeling

A computer program was written in the C programming language under UNIX from Sun Microsystems (SunOS 4.1.1, Sun C 1.1). This program used the random number generator function 'addrans' (5, 6). The two parental genotypes in the model are represented as 0 or 1. An individual chromosome is represented as an array of 0's and 1's. Each element of the array represents the genotype at intervals of 1% recombination unit. To simulate meiosis one of the two source arrays (homologous chromosomes) comprising a diploid genome is selected at random and the first element (locus) of the array is copied to an empty destination array, which will become the gamete. Before copying the next element from source to destination array, a random number between 1 and 100 is obtained. Only if the number is 1 (occurs with a probability of 1%) is the opposite source array selected (i.e., a recombination event has occurred). This process is continued through to the end of the array, to create a recombinant gamete. Gametes are combined in pairs to create new individual genomes.

Window sizes in F_2 pools were simulated using this algorithm. The genome model was a single linkage group, 200 recombination units in length. Different pool sizes were tested, ranging from 5 to 100 individuals per pool (i.e. 10 to 200 linkage group homologs per pool). We tested marker amplification potentials of 0.01, 0.025, 0.05. An amplification potential of 0.01, for example, means that the marker is detectably amplified from a pool if the positive:negative allele ratio is 0.01 or greater (the 'positive' allele supports amplification, while the 'negative' allele does not). To compute window size in our model, a pool is first assembled from the required number of random F_2 individuals. These individuals are selected as homozygous for the negative allele at the central locus of the model linkage group

* To whom correspondence should be addressed at: Du Pont Experimental Station, PO Box 80402, Wilmington, DE 19880-0402, USA

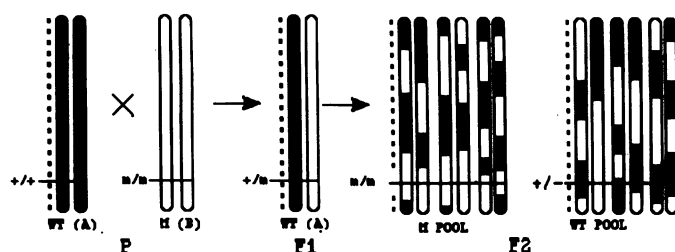


Figure 1. Diagrammatic representation of mapping by phenotypic pooling in a hypothetical diploid organism containing a single linkage group. Dots represent RAPD markers. m and +, mutant and wild-type alleles respectively. Note the absence in the mutant pool of 2 RAPD alleles linked to the subject locus.

(the subject locus is at position 100 in the linkage group of 200 recombination units). Initially, all homologs in the pool are examined at the subject locus and are found, by definition, to be homozygous for the negative allele. The program then steps to the right by 1 recombination unit and at this locus again examines the allele composition of the pool. The program continues stepping to the right until the ratio of positive:negative alleles exceeds the modeled amplification potential for the marker, at which point amplification of the positive allele is detected and the right end of the window is declared. The same stepwise process is then repeated to the left of the subject locus to find the left end of the window. The total window size is computed as the distance, in recombination units, between the left and right ends of the window. To accumulate statistics on window size, 5000 independent pools were generated and scored for window size. This process was repeated for each tested combination of pool size and marker amplification potential. The model does not consider recombination interference.

RAPD mapping

Procedures for RAPD mapping were as previously described (4). Segregation was analyzed by the UNIX version of the Mapmaker program (7).

RESULTS

Genome mapping by phenotypic pooling

In the pooling approach to linkage analysis (bulk segregant analysis, ref. 1–4), individuals of a segregating population are sorted into two categories, or pools, based on either genotype or phenotype. In the simplest case of phenotypic pooling the phenotype is controlled by a single genetic locus. For a recessive mutation in an F_2 population, plants with a mutant phenotype are homozygous for the recessive allele, whereas plants with a wild-type phenotype are either heterozygous or homozygous for the dominant allele. For additive gene action, each of the extreme phenotypes would be homozygous for the respective allele. Since the two pools are distinguished by phenotype, each pool will have a different genotype at the subject locus.

In the approach presented here we define a set of RAPD markers which map at even distances throughout the genome and are amplified from a wild-type strain but are absent from the mutant strain (Fig. 1). A cross between these strains and subsequent selfing of the F_1 generation will lead to genetic segregation of all alleles tested. Most RAPD markers are dominant, thus they will segregate in the F_2 population in a 3:1

ratio for presence and absence of the corresponding DNA band. If the mutation to be mapped is recessive, at the subject locus the mutant pool will have only DNA from the mutant parent (homozygous), whereas the wild-type pool will have DNA from both the wild type and mutant parents in a 2:1 ratio (heterozygous). Thus, DNA markers linked in *trans* to the subject locus will detect polymorphisms between the two pools, while the unlinked markers will not reveal polymorphisms because both marker alleles will be represented in both pools (Fig. 1). Thus, the chromosomal location of the mutant gene is determined by linkage to previously mapped RAPD markers, which is detected by the absence of specific RAPD bands in the mutant pool (Fig. 1). These RAPD markers can then be used to analyze individual F_2 plants to obtain an accurate map position.

A computer simulation of pool analysis

Occasionally polymorphism may be falsely detected at RAPD loci not linked to the subject locus. This can happen if, by random chance, only the mutant allele occurs within the mutant pool at any unlinked locus. Michelmore *et al.*(1) calculated the probability of falsely detecting polymorphism as a function of the number of individuals in a pool. The probability of false polymorphism decreases as pool size increases. However, there are no reported theoretical studies on the probability of detecting polymorphisms with DNA markers as a function of recombination distance between the subject locus and the marker. In this section we describe the results of computer simulations to address this issue.

Description of terms. The probability of encountering an alternate allele in the homozygous pool increases with the recombination distance. The presence of an alternate allele makes the pooled genotype heterozygous and consequently blocks detection of polymorphism between pools. We will define the interval bracketing the subject locus in which no alternate alleles are detected as the 'genetic window'. Window size and its standard error are expected to depend on pool size. As the size of the pool increases, the probability of finding alternate alleles near the subject locus also increases, and the window size is expected to decrease until it reaches a minimum that is dependent on the sensitivity of the technique to detect rare alleles. Moreover, for RAPD markers window size will also depend on the relative amplification potential of a particular marker. Amplification potential is characterized by the minimum ratio of positive:negative RAPD alleles at which a given RAPD marker is amplified to a detectable level. As the amplification potential of a RAPD increases, fewer positive alleles are required to detect amplification and window size should therefore decrease. According to this definition, amplification potential depends not only on the nature of the respective RAPD allele but also on the sensitivity of the detection method used.

Simulation results. The distribution of window sizes for a marker amplification potential of 0.025 is shown in Figure 2A for pool sizes of 5, 10 and 40 individuals. These results illustrate the variance in expected window size as a function of pool size. A pool size of 5 individuals, for example, has a very broad distribution of window sizes. While the most likely window size is 15 recombination units ($p = 182/5000$), one is almost as likely to obtain window sizes of 4 or 26 recombination units ($p = 112/5000$ for each). Figure 2A also shows that as pool size

progresses from 5 to 10 to 40 individuals, the most likely window size (the statistical mode of the distribution) decreases from 15 to 8 to 5 recombination units, respectively, and that there is a concomitant decrease in variance.

The data of Fig. 2A were numerically integrated to display the probability of obtaining a window at, or smaller than, the value indicated on the X-axis (Fig. 2B). For example, the probability of obtaining a window between 0% and 20% recombination units is about 0.80 for a pool size of 5 (Fig. 2B, hatched line).

In Fig. 2C we show a plot of the median window size as a function of pool size for three different amplification potentials.

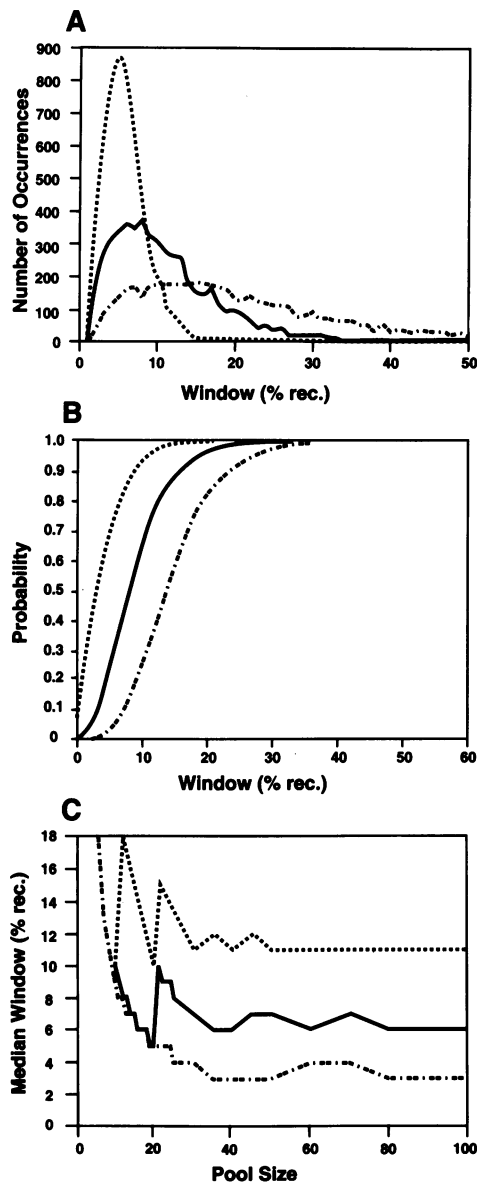


Figure 2. A computer simulation of pool analysis. A, number of occurrences of a certain window size for pools of 5 (hatched), 10 (solid) and 40 (dotted) individuals for a marker with amplification potential of 0.025. B, Numerical integration of data from panel A, showing the probability of obtaining a window size between 0 and the indicated value (lines as in panel A). C, median window as a function of pool size for RAPD markers of amplification potential 0.01 (hatched), 0.025 (solid), and 0.05 (dotted).

Results indicate that the larger the pool, the smaller the expected window size until it reaches a minimum characteristic of the amplification potential. Thus, for any given amplification potential, there is a maximum pool size above which no further reductions in window size are obtained.

We were surprised at the periodic nature of the data presented in Fig. 2C. For example, at an amplification potential of 0.025, median window size decreases smoothly as pool size increases from 5 to 20 individuals. However, when pool size increases by 1 more, to 21 individuals, median window size jumps abruptly from 5 to 10 recombination units. There are similar jumps of progressively smaller amplitude evident at increments of 20 individuals (jumps occur at 20, 40 and 60 individuals). These jumps in window size can be explained by the discontinuous character of the definition of window size. For example, at an amplification potential of 0.025 and pool size of 20 individuals (= 40 homologs), the presence of just 1 positive allele will by definition declare an end to a window ($1/40 = 0.025$). When 1 individual (assuming it is homozygous for the negative allele) is added to this pool increasing it to 21 individuals, the window will open up because the amplification potential is now too weak to detect the positive allele ($1/42 < 0.025$). As still more individuals are added to the pool, the probability of obtaining additional positive alleles increases and the window size drops smoothly until the next multiple of 20 individuals (40 homologs) is encountered. In a real experiment with RAPD markers, it would be difficult for an investigator to distinguish between 1 positive allele in 20 or 21 individuals, as judged by the intensity of a RAPD band in a gel. This practical limitation in data analysis would tend to smooth the periodicity seen in Fig. 2C, dependent on the individual investigator's ability to discriminate band intensities. DNA markers are typically scored as present or absent, black or white with no shades of gray, so a decision must be made whether a weakly-stained band should be scored as present or absent. If the data were perfect and could be scored by computer with programmed thresholds, then the simulation presented in Fig. 2C would apply. In reality, however, data analysis of DNA band patterns on gels is subjective and imperfect.

Mapping by phenotypic pooling

We have previously reported a map of the *Arabidopsis* genome constructed with RAPD markers and recombinant inbred lines derived from a cross between the *Arabidopsis* strain WS and the multiple marker line W100, which carries nine visible markers (3,4). To test this pooling approach to mapping we selected a mapping set of RAPD markers positive for WS alleles and we identified F_2 mutant individuals from a $W100 \times WS$ cross to form pools (Table 1).

By definition, the DNA bands corresponding to the RAPD set will always be present in WS and absent in W100. We tested the amplification potential of 15 of these RAPD markers in a reconstruction dilution experiment. Results indicate that 6 markers (269.1, 269.2, 315.1, 906.1, 906.2, and 1036) are relatively 'weak', or unable to detect WS alleles in a proportion of 0.1. Two markers (269.3 and 929.2) are 'strong', that is they detect WS alleles in a proportion of 0.01 (Table 2). The remaining markers have intermediate amplification potentials.

By scoring a population of 200 F_2 individuals from the $W100 \times WS$ cross we were able to identify at least 20 mutant individuals for the visible markers listed in Table 3, and for *miss broccoli* (*msb*), an extreme apetala phenotype. According to the

Table 1. RAPD mapping set

Chromosome I								Chromosome II								Chromosome IV											
WS	W100	Co	C24	Ler	Nz	NO	FLD	WS	W100	Co	C24	Ler	Nz	NO	FLD	WS	W100	Co	C24	Ler	Nz	NO	FLD				
11	+	-	-	-	+	-	+	+	1212	+	-	+	+	-	+	-	+	+	751.1	+	-	+	-	-	+	+	+
1356	+	-	-	-	+	-	+	+	1213	+	-	+	-	-	+	+	-	-	338	+	-	+	+	-	+	-	-
315.1	+	-	-	-	-	-	-	-	340	+	-	+	-	-	-	-	-	-	774	-	-	-	-	-	-	-	-
315.2	+	-	-	+	-	-	+	+	ap8m1.1	+	-	+	-	-	-	-	-	-	929.1	+	-	+	-	-	-	+	-
1036.1	+	-	-	-	-	-	-	-	870	+	-	+	+	-	-	-	-	-	936.1	+	-	+	-	-	+	+	-
769.1	+	-	+	+	+	+	+	+	Ap11e	+	-	+	+	-	-	-	-	-	901.3	+	-	+	-	+	-	-	-
1389	+	-	-	-	+	+	+	+	1107.3	+	-	+	-	-	+	+	-	-	rAp8m:	+	-	+	-	-	-	-	+
849.1	+	-	-	+	-	-	+	-	516.1	+	-	-	+	-	+	+	-	-	800.2	+	-	+	-	-	-	+	+
1106	+	-	-	-	+	-	+	-	1083.1	+	-	-	-	-	-	+	+	-	901.2	+	-	+	-	+	-	-	-
1506.1	+	-	+	-	+	+	+	+	1386.3	+	-	+	+	-	+	-	+	-	293	+	-	+	-	+	+	+	+
1594.1	+	-	-	-	-	+	+	-	906	+	-	+	-	-	+	+	-	-	800.1	+	-	+	-	-	+	+	-
929.2	+	-	+	+	-	-	-	-	252	+	-	-	-	-	-	-	-	-	739.2	+	-	+	-	+	-	+	-
1350.2	+	-	-	-	-	-	+	-	1107.2	+	-	+	-	-	+	+	-	-	808	+	-	+	-	-	+	+	-
1599	+	-	+	-	-	+	-	-	782.1	+	-	+	-	-	+	+	-	-	1595.2	+	-	+	-	-	+	+	+
ap6.	+	-	+	-	+	-	-	-	734	+	-	+	-	-	-	-	-	-	692	+	-	+	-	-	+	+	+
916.3	+	-	-	-	-	-	+	+											916.2	+	-	+	-	+	+	+	+
1036.4	+	-	-	-	-	-	-	-											901.2	+	-	+	-	+	-	-	-
756.1	+	-	+	+	-	-	-	-	Chromosome III								Chromosome V										
234	+	-	-	+	+	+	+	+	WS	W100	Co	C24	Ler	Nz	NO	FLD	WS	W100	Co	C24	Ler	Nz	NO	FLD			
269.3	+	-	+	+	+	+	+	+	257.2	+	-	+	+	+	+	+	+	+	1440.1	+	-	+	-	+	+	+	+
269.2	+	-	-	+	-	+	+	+	857.2	+	-	+	-	-	-	+	+	-	47	+	-	-	-	-	-	-	+
363.2	+	-	-	+	-	+	+	+	278.1	+	-	+	-	-	-	-	-	-	739.1	+	-	+	-	+	-	-	+
1111.2	+	-	-	-	-	-	+	+	1363.3	+	-	+	-	-	-	+	+	-	936.2	+	-	+	-	+	+	-	-
1177.1	+	-	+	+	-	-	+	+	317.2	+	-	+	+	+	+	+	+	-	257.1	+	-	+	-	+	+	+	+
1595.1	+	-	+	+	-	+	+	+	771	+	-	+	-	+	+	-	-	-	1111.1	+	-	-	-	-	-	+	+
916.1	+	-	+	+	-	+	+	+	1203	+	-	+	-	-	+	-	-	-	727	+	-	-	-	-	+	+	-
731	+	-	+	+	+	+	+	+	1106	+	-	+	+	-	-	-	-	-	1083.2	+	-	-	-	-	-	+	-
756.1	+	-	+	+	-	-	+	+	857.1	+	-	+	-	-	-	+	-	-	78	+	-	+	+	+	+	+	-
916.4	+	-	-	-	-	-	+	+	278.1	+	-	+	-	-	-	-	-	-	1107.1	+	-	+	+	+	-	+	+
									751.3	+	-	+	-	-	+	+	-	-	250	+	-	+	+	+	+	+	+
									807.1	+	-	-	-	-	-	+	-	-	556	+	-	+	-	+	-	+	-
									431.2	+	-	-	+	-	+	-	-	-									
									1488	+	-	-	-	-	-	-	-	-									
									807.1	+	-	-	-	-	-	+	-	-									

Positive (+) and negative (-) alleles for the ecotypes Columbia (Co), Wassileskija (WS), Niederzans (Nz), NO, RLD, C24, Landsberg erecta (Ler), and the marker line W100

Table 2. Detection of WS alleles by RAPD primers

RAPD primer	WS	W100	DNA source ^a			
			1:10	1:15	1:20	1:100
11	+	-	+	+	+	M
78	+	-	+	+	+	M
269.1	+	-	-	-	-	-
.2	+	-	-	-	-	-
.3	+	-	+	+	+	+
315.1	+	-	-	-	-	-
.2	+	-	+	+	+	-
692	+	-	+	+	-	-
734	+	-	+	+	+	M
906.1	+	-	-	-	-	-
.2	+	-	-	-	-	-
929.1	+	-	+	+	+	-
.2	+	-	+	+	+	+
1036	+	-	-	-	-	-
1203	+	-	+	+	-	-

^a Presence (+) or absence (-) of RAPD bands amplified from WS, W100, or increasing dilutions of WS DNA with W100 DNA. M, missing data.

predictions of the model, pools of 20 individuals will result in windows of 5–10 cM according to the amplification potential of the primers used (Fig. 2 C). Based on the density of markers assumed in the model (1/cM) these windows will be recognized by 1–10 RAPDs, depending on the amplification potential. However, with the density of markers used in this work (88 markers, or 6.8 markers/cM), windows in pools of 20 individuals will be recognized on average by a single RAPD.

DNA was extracted from the 6 phenotypic pools and used in amplification reactions with the RAPD primers of the mapping

set. The examples shown in Fig. 3 demonstrate that RAPD markers of the mapping set detect all of the marker genes shown in Table 3 plus *msb*. The same RAPD marker (731) detects *ap1* and *msb*. This suggests that *msb*, which has an extreme apetal phenotype, arose by the action of a modifier gene on *ap1*. A less likely possibility is that *msb* represents a mutation in a second gene linked to *AP1*.

After detecting linkage, we mapped the corresponding RAPD markers in 24 mutant F₂ individuals. Results of this mapping, as well as the genetic distances from the RI map, are shown in Table 3.

In addition to the mutations shown in Table 3, we tested the method with the W100 *cer2* locus. However, we experienced reproducibility problems with 782, the RAPD marker that in initial experiments appeared linked to *cer2*, and the remaining markers of the set did not show linkage.

The results shown in Table 3 for *an*, *ap1*, and *gl1* are consistent with our previous mapping with recombinant inbred lines. We were also able to map *bp* more accurately. In our previous study this marker was placed to a general region in chromosome IV but, unlike the remaining eight markers, it was not possible to provide a reliable (LOD score >3) placement. The pooling results indicate that *bp* is linked to the RAPD marker 929.1.

To investigate the reasons for the previously ambiguous mapping of *bp*, we used the MAPMAKER 'Genotypes' command to determine the number of individuals in the RI population that show double recombination events in the regions of the genome surrounding the markers of the tester line. For example, if an individual recombinant inbred carries the mutant allele (B) of a marker gene and wild-type alleles (A) for flanking molecular markers, the 'Genotypes' command will detect the double recombination event in the sequence ABA from all

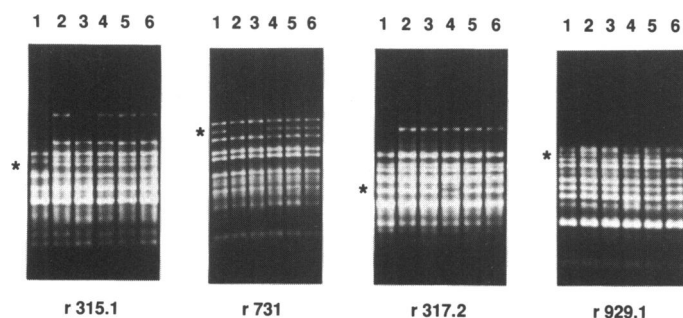


Figure 3. Mapping of *Arabidopsis thaliana* mutations by phenotypic pooling. Phenotypic pools for *an* (1), *ap1* (2), *msb* (3), *gl1* (4), *cer2* (5), and *bp* (6) detected by the RAPD primers of the mapping set (Table 1) indicated at the bottom of each panel. Corresponding DNA fragments are indicated by asterisks.

nonrecombinant (AAA or BBB) and single-recombinant (ABB and BAA) individuals. Only one double recombination event was detected among 115 recombinant inbred lines for the markers *an*, *ap1*, *er*, *py*, *gl1*, *cer2*, and *tt3*. However, 7 double recombinants were found at the *bp* locus. This is likely to have arisen from miscoring of the *bp* phenotype.

Analysis of six ecotypes with the RAPD mapping set

We have shown that the mapping set can be used to map mutations using a WS \times W100 cross. To determine whether these markers can also be used with other *Arabidopsis* crosses, we used the same mapping set to amplify DNA from 6 other commonly used *Arabidopsis* ecotypes. Results indicate (Table 1) that of the 88 markers in the mapping set, 83 can be used in crosses between WS and some ecotypes other than W100 (Table 1). Also, mutations segregating in populations derived from lines other than WS and W100 can also be mapped with at least some of the members of this marker set. We have also mapped W100 RAPD alleles (3) which can also be used as a complementary mapping set.

During the scoring of genomic lambda clones as RFLP markers (3), we surveyed 91 clones for polymorphisms using 5 restriction enzymes and DNA from the *Arabidopsis* ecotypes Co, WS, and W100. We now examined these data for the number of alleles detected by these RFLPs in all 3 ecotypes, and determined that <27% of these clones detect 2 alleles, and only 5.7% detect 3 alleles. Thus, if RAPD markers also have this allele frequency, most negative scores in Table 1 will correspond to the same allele.

DISCUSSION

We have demonstrated the use of an ordered set of RAPD markers to map mutants by phenotypic pooling. Linkage analysis using RAPDs and phenotypic pooling is an attractive alternative to mapping with phenotypic marker lines because the markers used are not subjected to the environmental, epistatic, and pleiotropic effects that often influence the expression of genes used as classical markers.

In classical genetic mapping the position of a subject locus with respect to a marker is determined by analyzing individuals of a segregating population. In the pooling approach to mapping, the position of the subject locus is assigned to a genetic interval (window). Thus, when mapping by pooling the burden of accuracy in linkage determinations is placed on scoring the RAPD markers and, unlike in classical Mendelian genetics, the

Table 3. Detection of W100 phenotypic marker genes by phenotypic pooling

Chromosome	Genetic distance (cM) ^c				
	Locus location ^a	A:B ^b	RAPD marker ^a	RI	F2
<i>an</i>	I-12.1	1:1	315.1	5.5 +/- 2.7	4.3 +/- 3.0
<i>ap1</i>	I-125.3	5:1	731	6.5 +/- 3.0	5.0 +/- 3.6
<i>gl1</i>	III-47.6	2:1	771	8.0 +/- 3.4	10.5 +/- 5.5
			317.2	4.0 +/- 2.2	4.4 +/- 3.1
<i>bp</i>	IV- ?	4:1	1203	6.5 +/- 3.0	13.0 +/- 5.6
			929.1	13.2 +/- 4.3	5.2 +/- 3.8

^a from ref. 3; ^b from scoring WS (A) and W100 (B) alleles in the recombinant inbred (RI) population, ^c as measured by 2-point analysis in a recombinant inbred population of 48 individuals (3) or in 24 mutant individual F₂ plants after detection by pooling.

proportion of individuals in the population that show the corresponding phenotype is no longer relevant, as long as 10–20 mutant individuals can be identified to form a pool. This is particularly important when expression of a phenotype is altered or masked by the action of modifier genes or by the presence of unrelated mutations.

We have developed a model for mapping by phenotypic pooling that predicts the window and optimal pool sizes. With the high density of markers (1 RAPD/cM) assumed in this model, a 10 cM window would be detected by about 10 tested RAPDs. Although useful for modeling purposes, this density is exceedingly high for routine mapping. For instance, an *Arabidopsis* (genome size 600 cM, ref. 3) mapping set would comprise 600 RAPD markers. A more practical number might be 96 RAPD primers, which would allow mapping of a mutant with one microtiter plate of primers. For *Arabidopsis* the resulting number of polymorphisms will be 153 (1.6 polymorphisms per primer, ref. 3), thus resulting in an average density of 1 RAPD marker/3.9 cM. The mapping set used in this work (Table 1) contains 88 markers, with markers averaging every 5.2 cM (chromosome I), 6.5 cM (chromosome II), 4.7 cM (chromosome III), 3.6 cM (chromosome IV), and 10.7 cM (chromosome V). Although this density is reasonably high, the distribution of these markers is not uniform throughout the genome. For instance, the top and bottom of chromosome IV are not covered.

Three of the five mutations mapped by pooling were identified by a single marker. Three markers, however, showed linkage to *gl1*. Markers flanking this region define a 8 cM window which is consistent with the predictions of the model. Of the five mutant loci targeted for mapping, we only failed to map *cer2*. Additional analysis of the recombinant inbred data indicated that this marker should be moved down in our published map to a location near marker c17340, which is consistent with the RFLP map of Nam et al. (8). This region is not covered by our RAPD mapping set. To improve this set we will have to add RAPD markers in regions that are not currently covered and, if necessary, delete some markers in highly saturated regions. Pooling by genotype can be used to add markers in selected regions of the genome (2–4).

We had previously observed that some of the W100 marker genes segregate in the recombinant inbred population in a ratio that indicates a deficiency for the recessive mutant alleles. Five examples are shown in Table 3. The *angustifolia* (*an*) gene, which determines a phenotype of narrow leaves and twisted siliques, segregates in the expected 1:1 ratio for wild-type and mutant phenotypes. However, the remaining three markers show recessive deficiencies that range from 2:1 for *gl1* to 5:1 for *ap1*.

We were able to map reliably (LOD score >3) three of these loci to chromosomal locations using recombinant inbreds, indicating that recessive deficiencies *per se* do not result in ambiguous mapping results. However, the location of *bp* was ambiguous (3). Using phenotypic pooling we have mapped *bp* accurately. Examination of the genotype around this locus suggests that the initial ambiguity was due to miscoring of the *bp* phenotype.

We have determined that RAPD markers have quite different amplification potentials. Thus, the ability of a RAPD marker to detect linkage to a subject locus depends upon both genetic distance and amplification potential. A possible strategy for scanning the entire genome with relatively few markers is to select weak RAPDs, which give large window sizes. For instance, a mapping set formed of markers with amplification potential of 0.05, when used with pools of 20 individuals, will detect mutations within a window of average size 15 cM (Fig. 2).

The consequences of miscoring phenotypes will depend on the amplification potential of the RAPD markers. For instance, if 1 individual in a pool of 20 is miscored, markers with amplification potentials >0.05 will still be of diagnostic value. It is advisable to pool only individuals in which the mutant phenotype has been unambiguously scored, and furthermore to confirm the phenotype by progeny testing.

For many applications it will be sufficient to map the subject locus to a 15 cM interval. For some applications such as chromosome walking, however, it will be necessary to narrow the window further. This can be achieved by using other RAPD markers in the region with higher amplification potentials and by employing detection methods more sensitive than ethidium bromide staining. Churchill *et al.* (9) have recently described the theoretical basis for ordering markers within a window. Classical segregation analysis of individuals may be used to determine the genetic distance between the subject locus and additional markers in the region. However, the classical approach has the disadvantage that one must extract and analyze DNA from many individuals.

Using a set of mapped RAPDs has the additional advantage that fewer amplification reactions are needed, thus saving costs in labor and supplies. For example, we derived this set of 88 markers from screening 1,200 RAPD primers and mapping RAPD markers with 46 individual recombinant inbred lines (3). Thus, to map with this set, 2 DNA pools and 166 amplification reactions are needed. In contrast, to map mutations relative to unmapped RAPDs would have required 2,676 reactions (1,200 primers with 2 pools, plus mapping 6 markers with 46 individuals).

We describe the method here only for recessive mutations. However, one can extend this idea to both co-dominant and dominant mutations. For co-dominant mutations the wild type phenotype pool will contain only homozygous wild-type individuals, whereas the mutant pool will consist of homozygous mutant individuals. In the case of dominant mutations, one pool will contain wild-type individuals and the other mutant individuals, the latter containing both homozygous mutant and heterozygous individuals. Alternatively, homozygous and heterozygous individuals can be distinguished by progeny testing. For recessive mutations the primer set will derive from the wild-type parent, whereas for dominant mutations the RAPD set will be generated by the mutant parent. For co-dominant mutations the RAPD set could derive from either parent.

We determined the allele distribution of markers of the RAPD mapping set in 7 *Arabidopsis* ecotypes. We used this information to extend the applicability of this set to the mapping of mutants isolated from ecotypes other than W100. Mutants genes isolated in other backgrounds and mapped with this RAPD set become automatically integrated into the recombinant inbred map.

An unexpected conclusion of this survey is that W100 and *Ler* contain significant differences in all chromosomes except II (Table 1). Of the 19 positive RAPD alleles detected in *Ler*, 12 occur in clusters of 2 or more markers (Table 1). This indicates that at least some of the lines carrying the individual mutations introgressed into *Ler* to create the multiple marker line were not isogenic with this background. This demonstrates the usefulness of this approach in determining intra-specific relationships in *Arabidopsis*.

Although our mapping set uses exclusively RAPD markers, it is possible to use either RFLPs or other PCR-based markers, such as microsatellite repeats (10). Given the many advantages of the PCR, it is likely that most markers developed in the future will be based on this assay. However, exponential amplification will lead to the detection of mutant alleles regardless of their proportion in a mutant pool. For example, the probability of occurrence of at least 1 wild-type allele in a 20-individual mutant pool is 0.33 for a marker located at 1% recombination and 0.87 for a marker at 5% recombination. Thus, using PCR markers for mapping by phenotypic pooling will result in very small windows. Possible solutions to this problem are either to limit the number of cycles in the PCR reaction to limit the amplification of rare alleles, or to use multiple pools with fewer individuals.

Although we have only applied this strategy to mapping in *Arabidopsis* this approach can be used with other diploid, sexually reproducing organisms, for which a molecular map has been developed and mutants can be isolated from one of the parental strains or from other strains with suitable allele compositions.

REFERENCES

1. Michelmore, R. W., Paran, I., and Kessell, R. V. (1991) *Proc. Natl. Acad. Sci.* **88**, 9828–9832.
2. Giovannoni, J. J., Wing, R. A., Ganai, M. W., and Tanksley, S. D., *Nucleic Acid Research* **19**, 6553–6558.
3. Reiter, R. S., Williams, J. G. K., Feldmann, K. A., Rafalski, J. A., Tingey, S. V., and Scolnik, P. A. (1992) *Proc. Natl. Acad. Sci. USA*. **89**, 1477–1481.
4. Reiter, R. S., Young, R. M., and Scolnik, P. A. (1992) In Koncz, C., Chua, N.-H., and J. Schell, (eds.) 'Methods in Arabidopsis Research', World Scientific, pp. 170–190.
5. Knuth, D. E. (1981) In Knuth, D. E. (ed.) 'The Art of Computer Programming', Addison-Wesley, New York, vol. 2, pp 456–463.
6. Park, S. K., and Miller, K. W. (1988) 'Computer Practices', Communications of the Association for Computing Machinery, Addison-Wesley, New York, pp 1192–1201.
7. Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E., and Newburg, L. (1987) *Genomics* **1**, 174–181.
8. Nam, H.-G., Giraudat, J., den Boer, B., Moonan, F., Loos, W. D. B., Hauge, B. M., and Goodman, H. M. (1989) *Plant Cell* **1**, 699–705.
9. Churchill, G. A., Giovannoni, J. J., and Tanksley, S. D. (1993) *Proc. Natl. Acad. Sci.* **90**, 16–20.
10. Armour, J. A. L., and Jeffreys, A. J. (1992) *FEBS LETT.* **307**, 113–115.