Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 16–20, January 1993
Genetics

Pooled-sampling makes high-resolution mapping practical with DNA markers

GARY A. CHURCHILL, JAMES J. GIOVANNONI, AND STEVEN D. TANKSLEY

Department of Plant Breeding and Biometry, Cornell University, Ithaca, NY 14853

Communicated by Michael T. Clegg, September 17, 1992 (received for review March 9, 1992)

ABSTRACT **A pooled-sample approach to the construction of high-resolution genetic maps is described. The strategy depends on the existence of an easily selectable target locus and the ability to produce large segregating populations. If these requirements are met, the pooled-sample mapping approach allows tightly linked markers (e.g., restriction fragment length polymorphisms) to be mapped relative to the target with a great economy of effort. The recombination fractions among loci can be estimated by the maximum likelihood method and a simple approximate estimator is derived. The order of loci is deduced using a Bayesian statistical framework to yield posterior probabilities for all possible orderings of a marker set. Optimal pooling strategies and the effects of misclassification of selected individuals are discussed and studied by computer simulation. The feasibility of this method is demonstrated by the high-resolution mapping of a region on chromosome 5 of tomato that contains a gene regulating fruit ripening.**

The idea of using restriction fragment length polymorphisms (RFLPs) for genetic mapping was introduced in 1980 (1). Since that time there has been rapid progress in the development of genetic maps in a variety of organisms, including humans, mice, and many crop species (2–5). One of the reasons for constructing DNA-based genetic maps is for use in chromosome walking. Currently, many interesting and important genes are known only by their phenotype. Lack of knowledge of their gene products inhibits traditional methods of gene cloning; however, knowing the position of such genes on a DNA-based map opens the opportunity for walking to the gene from adjacent marker(s). In this manner, a number of genes have been cloned from humans and other higher eukaryotes in recent years (6).

Two requirements for chromosome walking are (*i*) availability of tightly linked DNA marker(s) in the vicinity of the gene of interest and (*ii*) knowledge of the position of these markers relative to the targeted gene. In the past, finding markers near a gene of interest proved to be very time consuming. However, increased efforts in genome mapping have led to the generation of RFLP-based maps for many organisms (2–5). In addition, there are now methods in place for rapidly identifying DNA markers specific to any region of a genome (7–11).

Once a number of markers tightly linked to a gene of interest have been identified, the two markers that most closely flank the targeted gene must be identified since it is these two markers that provide the most efficient starting points for a walk. In cases where the markers are very tightly linked to the targeted gene [e.g., <1 centimorgan (cM)], analysis of hundreds or even thousands of segregating progeny may be required to determine the order of markers in the vicinity of the target gene (12). This can be costly and time consuming since it requires isolation and analysis of DNA from each individual in a segregating population. As the relationship between genetic and physical distances may vary across different regions of the genome, additional experiments to verify tight physical linkage may be warranted.

In an effort to overcome the problem encountered with mapping large populations, we have devised and tested a pooled-sample method for high-resolution mapping around genes targeted for cloning. The method exploits the fact that when mapping many markers in a small segment of a chromosome, very few individuals (from a segregating population) contain chromosomes with a crossover in the region of interest and thus most individuals provide little useful information. By pooling individuals for analysis, the effort required to construct a high-resolution map can be reduced manyfold. The steps for using this pooled-mapping technique in a segregating population (e.g., F_2) are as follows: (*i*) identify (by phenotype) those individuals that are homozygous (usually homozygous recessive) for the target gene and (*ii*) divide these individuals into pools (see *Results* for optimum pool sizes). Extract DNA *en masse* from each pool using approximately equal amounts of tissue from each individual. This bulked DNA is then probed with clones known to be located in the vicinity of the gene. The proportion of pools containing at least one crossover event is recorded and the resulting data are used to construct a high-resolution map. By using this method, the number of samples from which DNA must be isolated and analyzed can be reduced by a factor of 10 or more. There are two practical requirements: the trait being selected must be known to be controlled by a single gene and it must be possible to identify single recombination events between the target and marker genes in the pooled DNA sample.

We present here the basic theoretical considerations necessary for utilizing pooled mapping and demonstrate the technique by mapping a gene regulating fruit ripening in tomato. This method leads to the identification of the two markers most closely flanking the targeted gene. It can also be used to determine the map order of other markers and an estimate of map distance from the targeted gene. Although the method has been demonstrated in a plant, it can be used in any diploid sexually reproducing organism for which large segregating populations can be obtained.

MATERIALS AND METHODS

Pooling. Individuals from an F_2 population segregating for a target locus (T) are determined to be homozygous for the target locus (either T/T or t/t) and are divided into n pools of k individuals each. Since each individual contains two chromosomes, the number of independent meiotic events represented by each pool equals $2k$. If a backcross population is used, the number of meiotic events equals k, and the results below should be modified accordingly. DNA is isolated *en masse* from each pool and analyzed with markers thought to be linked to T but not necessarily ordered with respect to T or each other. It is assumed that a single recombinant

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: RFLP, restriction fragment length polymorphism; cM, centimorgan.

chromosome can be detected in a pool of otherwise nonrecombinant chromosomes but that the exact number of recombinant chromosomes cannot be determined. Therefore, the results from probing a pooled DNA sample with a linked marker classify that pool as recombinant (i.e., contains at least one chromosome that is recombinant between the marker and the target gene) or nonrecombinant. The combined result of probing a pool with a set of m markers, $\mathbf{M} = \{M_1, M_2, \ldots, M_m\}$, is a classification of the pool into one of 2^m possible pool types. A pool type is defined by the set S of markers in M that are recombinant with respect to T for at least one chromosome in the pool. The observed number of pools of type S will be denoted by Y_s and the 2^m vector of observed counts will be denoted by Y.

Ordering of Markers. The goal of a pooled sample mapping experiment is to infer the map order among a set of markers relative to the target locus. In particular, we wish to identify the two markers that most closely flank the target locus. One widely used approach to the problem of inferring map order among a set of markers is to choose that order which has the highest maximized likelihood (13, 14). However, the maximum likelihood approach has several limitations. (*i*) When markers are known to be tightly linked to the target locus, as will generally be the case when a pooled-sample approach to mapping is being considered, this prior knowledge should be incorporated into the linkage analysis. (*ii*) The different order hypotheses are not nested and thus one cannot construct a likelihood ratio to formally test the best order. For these reasons we have employed a Bayesian approach to the marker ordering problem.

Let H denote the map order of markers in the set $\mathbf{M} \cup \mathbf{T}$ and let $\mathbf{R} = (r_1, r_2, \ldots, r_m)$ denote the recombination probabilities between adjacent markers in the ordered set. Inference of the map order will be based on the posterior probability (Pr) of H, which can be computed as

$$\Pr(H|\mathbf{Y}) \propto \Pr(H) \int \Pr(\mathbf{Y}|\mathbf{R}, H) \Pr(\mathbf{R}|H) dr, \qquad [1]$$

where $\Pr(H)$ is the prior distribution on map orders, $\Pr(\mathbf{R}|H)$ is the prior distribution on recombination probabilities, and $\Pr(\mathbf{Y}|\mathbf{R}, H)$ is the likelihood of the observed counts. The integral can be evaluated to desired numerical precision by the Monte Carlo method of composition (15).

The likelihood is multinomial on 2^m pool types,

$$\Pr(\mathbf{Y} = \mathbf{y}|\mathbf{R}, H) \propto \prod_{\mathbf{S} \subseteq \mathbf{M}} p_{\mathbf{S}}^{y_{\mathbf{S}}} \qquad [2]$$

where the pool type probabilities $p_{\mathbf{S}}$ are to be defined in *Results* and $y_{\mathbf{S}}$ is the observed count of pools of type S. The multinomial distribution follows from assumptions that individual pools contain identical numbers of chromosomes generated by independent meioses and that each pool is probed independently of others.

Prior Distributions. The prior knowledge available about map order and recombination probabilities will depend on the source of the markers. For example, we may consider markers selected at random from a fixed interval of known size (e.g., markers identified using nearly isogenic lines; ref. 9), markers known to segregate closely with the target locus in previous crosses, or markers selected in the vicinity of the target locus from a previously constructed high density genetic map. For purposes of this analysis, the markers are assumed to be placed uniformly at random in a region around T with known density D markers per cM.

In the cases of unmapped or cosegregating markers, it is natural to assume that all possible orders are equally likely. For a set of m markers plus the target T, there are $(m + 1)!/2$ distinct orders and the equally likely prior distribution is $\Pr(H = h) = 2/(m + 1)!$, for all orders h. In situations where order information is available, the posterior probabilities from previous experiments can be used as the prior distribution for the new experiment.

The prior distribution on recombination probabilities is taken to be a product of $m - 1$ independent Beta distributions with parameters $a = 1$ and $b = 100 \cdot D - 1$. This is an analytically convenient choice and, when the genetic distances involved are small, approximates the distribution of spacings between randomly placed markers. Because recombination probabilities can vary significantly in different crosses, it is not clear what form of prior information should be used when previous experimental data are available.

Decision Rules. The posterior probabilities (Eq. 1) are used to make decisions about the map order. The decision rule that chooses the order with highest posterior probability as an estimate of the true map order is optimal in the sense of being the Bayes rule for a $0 - 1$ loss function (ref. 16, p. 163). However, if the markers are to be used to attempt a chromosome walk, an incorrect decision could be very costly. We wish to ensure that the two nearest markers that span the target are correctly ordered with high probability. A more stringent criterion is to decide that an order is correct only if its posterior probability exceeds a specified critical value (e.g., 0.90). Otherwise, no decision is made and probing of additional pools will be required to resolve the map order.

RESULTS

Estimating the Genetic Map Distance Between a Marker and the Target Gene. Let T denote the target locus, A denote a marker locus near T, and r denote the probability of recombination between A and T within a single chromosome. The probability that a pool contains at least one recombinant is $p_A = 1 - (1 - r)^{2k}$. When r is small, this probability is approximated by $1 - e^{-2kr}$. An approximate maximum likelihood estimator for r is

$$\hat{r} = (1/2k)\ln[1 - (y_A/n)], \qquad [3]$$

where y_A is the total number of recombinant pools. Numerical comparisons of the exact and approximate maximum likelihood estimators show very close agreement when the true value of the recombination probability is small (<0.1). The large sample variance of this estimator,

$$\mathrm{Var}(\hat{r}) = (1/4k^2n)[(1 - e^{-2kr})/e^{-2kr}], \qquad [4]$$

is minimized when the pool size is $k_{\mathrm{opt}} \approx 1.594/2r$ (17). This provides one criterion for selecting an optimal pool size. Note that when the pool size is near k_{opt}, the expected number of recombinations per pool is near 1.

Ordering Markers Relative to the Target Gene. Two-point data can be used to estimate map distances from the target to each member of a set of markers but will not be sufficient to order the markers relative to the target. For this purpose, three-point data (i.e., a target T and two segregating markers A and B) are required. It is well known that there is a significant information gain in standard (unpooled) genetic mapping when one considers triplets of markers rather than pairwise data (14). Thus we expect that three-point data will also be useful in resolving the order of markers that may be unresolved by two-point data.

For the three-point analysis, we will focus on the ordering problem. There are three distinct orders, each with its own set of recombination parameters $\mathbf{R} = (r_1, r_2)$. The probability of a recombination between the "left" pair of markers is r_1 and the recombination probability for the "right" pair is r_2. By assuming a no-interference model, recombination be-

tween the outer pair of markers will occur with a probability of $(1 - r_1)r_2 + r_1(1 - r_2)$.

The individual chromosomes in a pool can be classified into four types (ϕ, A, B, and AB), indicating which markers are recombinant with respect to T. The chromosome type probabilities, denoted by q, depend on the order of markers as shown in Table 1.

The pool type is determined by the collection of chromosome types contained in the pool. For example, a pool of type A will contain at least one chromosome of type A and none of type B or AB. A pool of type AB may contain one or more AB chromosomes or at least one each of the types A and B. The pool type probabilities

$$p_\phi = q_\phi^{2k},$$

$$p_A = (q_\phi + q_A)^{2k} - q_\phi^{2k},$$

$$p_B = (q_\phi + q_B)^{2k} - q_\phi^{2k},$$

$$\text{and } p_{AB} = 1 - (q_\phi + q_A)^{2k} - (q_\phi + q_B)^{2k} + q_\phi^{2k} \quad [5]$$

are substituted into the likelihood (Eq. 2) and together with the prior distributions define the posterior distribution over map orders (Eq. 1).

The results for three-point mapping are readily generalized to the problem of ordering a set of m markers relative to a target. However, the multipoint analysis presents a number of computational problems and will be described elsewhere (18).

Optimum Pool Size. The optimum number of individuals to include in a pool is determined primarily by the local density of markers around the target locus. Consider the nearest adjacent markers to T. The probability that a chromosome is recombinant for either of these markers will decrease as the local density of markers increases. If the pool size is small, most pools will contain no recombinant chromosomes and the amount of mapping information per pool will be very low. If the pool size is very large, most pools are likely to contain multiple recombinant chromosomes and mapping also becomes very inefficient. An optimum pool size is found between these two extremes and is larger for higher marker densities.

To study the effect of pool size on the inference of map order, we considered sets of three adjacent loci (one of which is the target locus) selected from a region with known marker density D markers per cM. Simulations were carried out to estimate the probability of a correct decision, averaged over all realizations of Y and all orders M. The quantity computed is one minus the Bayes risk of the decision rule (ref. 16, p. 159). This probability is shown (Fig. 1) as a function of pool size for different marker densities, numbers of pools, critical values for the decision rule, and misclassification rates (see below). An optimum pool size maximizes the probability of correct ordering.

The local density of markers has the greatest effect on optimal pool size (Fig. 1*A*). At low marker densities smaller optimal pool sizes are obtained and overpooling can result in an inefficient experiment. Higher marker densities give larger optimal pool sizes and are more robust to overpooling. For the lowest marker density studied ($D = 0.16$ marker per cM), the maximum probability of correct ordering is attained with a pool size of three individuals. At this density, an experiment with more than about eight pooled individuals becomes less efficient than probing single individuals. For higher densities the maximum ordering probabilities are attained with larger pool sizes and the range of efficient experiments is much broader.

Table 1. Chromosome type probabilities

	A–T–B	A–B–T	T–A–B
q_ϕ	$(1 - r_1)(1 - r_2)$	$(1 - r_1)(1 - r_2)$	$(1 - r_1)(1 - r_2)$
q_A	$r_1(1 - r_2)$	$r_1(1 - r_2)$	r_1r_2
q_B	$(1 - r_1)r_2$	r_1r_2	$(1 - r_1)r_2$
q_{AB}	r_1r_2	$(1 - r_1)r_2$	$r_1(1 - r_2)$

Order of markers A, T, and B is shown.

As more pools are sampled, the probability of selecting the correct order increases but there is little effect on the optimum pool size (Fig. 1*B*). The critical value of the decision rule also affects the probability of choosing the correct order but has little effect on the optimum pool size (Fig. 1*C*). Although the true order may have the highest posterior probability, if this probability does not exceed c, no decision is made. If the decision rule is very stringent ($c = 0.99$), large sample sizes may be needed to assign order unambiguously but confidence in the chosen order is very high.

Misclassification. When forming pools of individuals of type t/t, it is possible that one or more individuals of type T/t or T/T may be misclassified and included in a pool. Pools that contain a misclassified individual will typically be recombinant for all markers and will inflate the estimated recombination fractions in the intervals immediately flanking the target locus. Pools that are recombinant for markers flanking the target but are nonrecombinant for more distant markers contain an obligate double crossover or a misclassification. The phenotypes of individuals in such pools should be rechecked if possible (see the example below).

We have studied the effects of misclassification using simulations (Fig. 1*D*). If the misclassification rate is low (1–2%), correct ordering inferences are made with high probability. As the misclassification increases, fewer correct inferences are made, the optimum pool size is slightly reduced and the loss of efficiency due to overpooling is more serious. The effect is more pronounced for large pool sizes ($2k > 20$) and for high marker densities. The presence of a misclassified individual in a pool will generally reduce the posterior probability for ordering the two markers flanking the target and thus the procedure is conservative.

Technical Limits in Detecting Recombinants in a Pooled Sample. One of the assumptions of pooled mapping is that a single recombinant chromosome can be detected in a pool of otherwise nonrecombinants. The degree to which this is technically possible depends on the organism being studied and the molecular detection techniques being employed. RFLPs represent the common type of molecular marker now being used in higher organisms and they are normally detected on Southern blots with single-copy probes (1). Tomato (*Lycopersicon esculentum*), like most eukaryotes, has a genome that is relatively large and complex (haploid DNA content = 900 megabases). To test the limits of pooling, DNA from tomato plants of two genotypes was mixed and subjected to Southern blot analysis with a single-copy DNA probe. The results indicate that in a mixture as great as 40:1, one can still detect the rare allele. This is comparable to detecting a single recombinant chromosome in a pool of 20 plants. However, since the pooling is done before DNA extraction and is based on utilizing approximately equal amounts of tissue from individuals within the pool, there is room for additional error. For this reason, for the purposes of testing pooled-mapping, we decided to use pools of five plants (see next section).

Pooled Mapping of the *rin* Locus. To test the pooled mapping strategy, a large F_2 population (1840 plants), segregating for the *rin* (ripening inhibitor) gene, was planted in the field and grown to maturity. Fruit from plants homozygous for the recessive *rin* allele do not ripen and it is believed that *rin* represents an upstream regulatory switch for the ripening process (19). *rin* is on chromosome 5 and a number of DNA

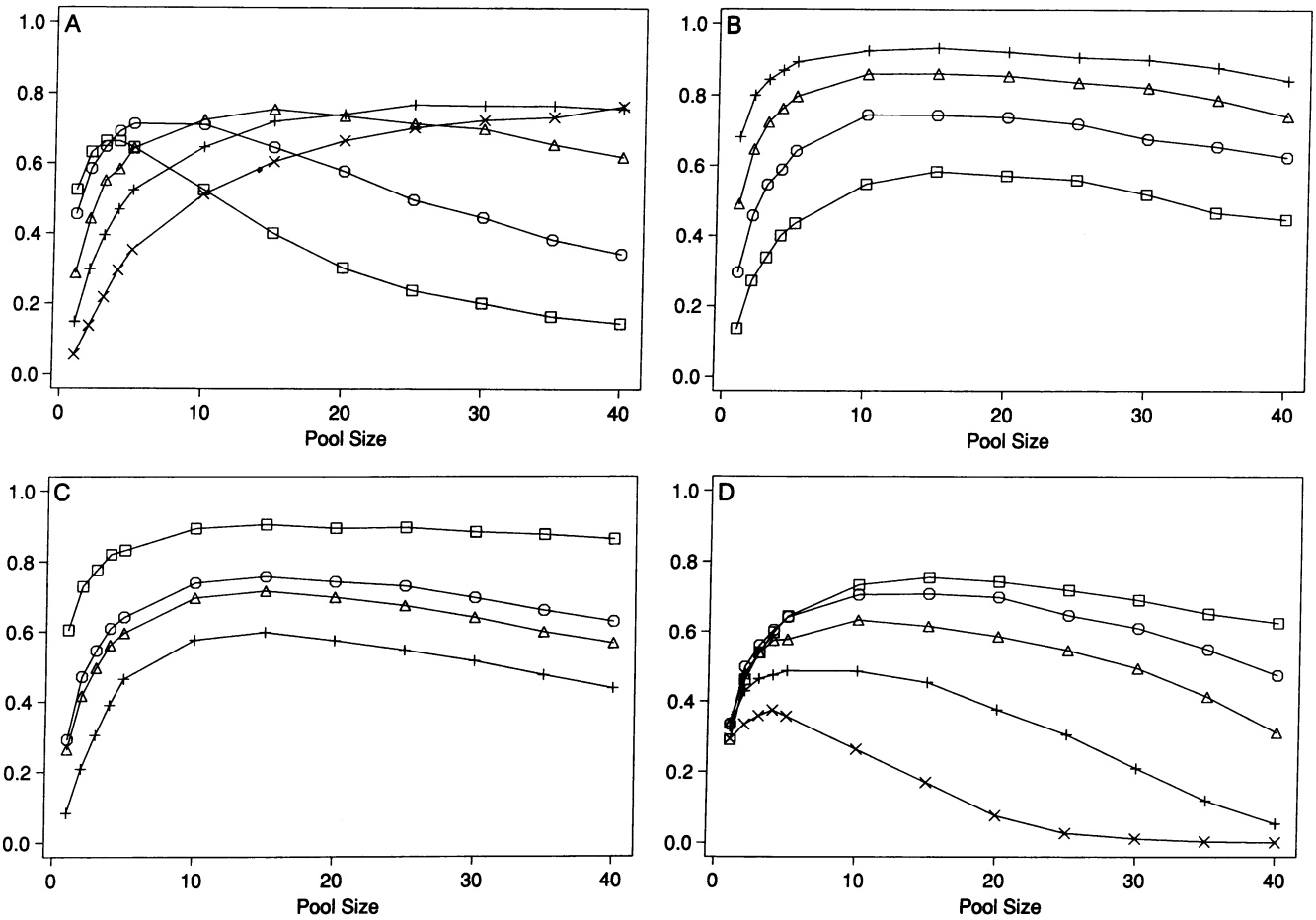

FIG. 1. Probability of correct order. The probability of determining the correct order of three adjacent loci (two markers and the target) is shown as a function of pool size. The effects of marker density D, sample size N, critical value for the decision rule c, and misclassification rate α were studied. The curves shown are based on constant values of $D = 0.64$ marker per cM, $N = 40$ pools, $c = 0.90$, and $\alpha = 0.0$, except that one factor is varied as follows: marker densities $D = 0.16$ (□), 0.32 (○), 0.64 (△), 1.28 (+), and 2.56 (×) markers per cM (*A*); sample sizes $N = 20$ (□), 40 (○), 80 (△), and 160 (+) pools (*B*); critical values $c = 0.90$ (○), 0.95 (△), 0.99 (+), and a decision rule that chooses the most likely order (□) (*C*); proportion of misclassified chromosomes $\alpha = 0\%$ (□), 1% (○), 2% (△), 5% (+), and 10% (×) (*D*). The sample size and critical value were chosen to give curves for which the correct order probability is well below 1 as such curves are most informative regarding optimal pool sizes.

markers have been identified that are in the vicinity of the gene (J.J.G. and S.D.T., unpublished data). Two hundred twenty plants were unambiguously determined to be homozygous (*rin/rin*) and tissue from these plants was pooled into groups of 5 to form 44 pools. DNA was extracted from each pool and scored for seven RFLP markers known from previous experiments to be linked to the *rin* locus. After the pools were formed and scored, a misclassified plant was discovered in one pool. Thus, the analysis presented is based on the remaining 43 pools.

Previous analyses suggested that the eight loci (*rin* plus seven RFLP markers) are located in an ≈20-cM interval (J.J.G. and S.D.T., unpublished data). When markers are considered three at a time, the average density (D) is 0.15 marker per cM. Thus the prior distribution for recombination probabilities in three-point mapping was taken to be Beta ($a = 1$, $b = 14$). All orders were assumed to be *a priori* equally likely.

Two-point analyses were carried out to estimate distances between the marker and the target (Table 2). The left and right groupings were readily established by three-point analysis. The inferred map is shown in Fig. 2. Posterior probabilities for ordering adjacent markers are shown in Table 3. The critical ordering inference involves markers *CT93* and *CT63* that appear to span the *rin* locus. The posterior probability of the three-point ordering *CT93–rin–CT63* is 0.971. When the data of 44 pools, including the misclassified pool, are analyzed (see Table 3), the posterior probability of this order is reduced to 0.904.

DISCUSSION

We have presented the basic theory necessary to create high-resolution genetic maps using pooled DNA samples. Our results suggest that this is a practical and highly efficient approach to high-resolution mapping of DNA markers. A number of factors were shown to affect the probability of choosing the correct order for a set of markers. These include pool size, local density of markers, number of pools sampled,

Table 2. Estimated distances from target locus *rin*

Marker	$\hat{r}$	Standard error
CD64	0.1122	0.0219
CT93	0.0295	0.0089
CT63	0.0024	0.0024
TG503	0.0124	0.0055
TG96	0.0360	0.0100
TG448	0.0871	0.0180
ACC4	0.1122	0.0219

Approximate recombination probabilities (Eq. 3) between markers and the target gene and their standard errors (Eq. 4) were estimated from two-point data.

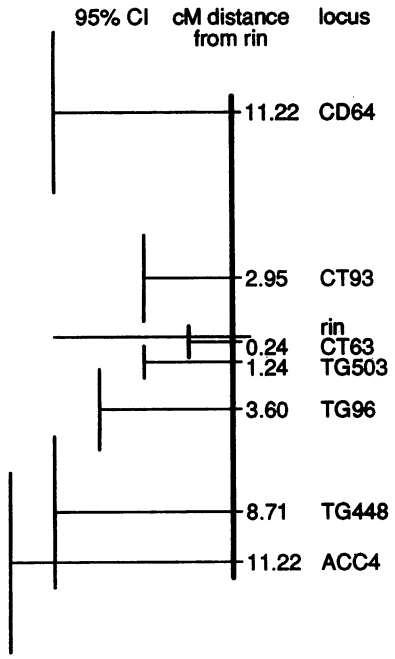

FIG. 2. Map of the *rin* region. The positions of seven markers relative to the target locus *rin* are shown. The distances indicated on the figure are estimated percent recombination (in cM) between each marker and the target. Precisions of the estimates are indicated by 95% confidence intervals $\{\hat{r} \pm 2[\mathrm{Var}(\hat{r})]^{1/2}\}$, shown as vertical bars to the left. Note that the individual markers are not mapped independently and thus overlapping confidence intervals do not necessarily indicate uncertainty in the map order (see Table 3).

the critical value for accepting an order, and the probability of misclassifying an individual. There is an optimum pool size that maximizes the probability of choosing the correct marker order. It is determined primarily by the local density of markers. For a low marker density, small to moderate pool sizes are most likely to yield the correct order, and overpooling can result in an inefficient experiment. For a higher marker density, larger pool sizes are optimal and the analysis is more robust to overpooling.

Map-based cloning represents one of the most promising strategies for isolating genes known only by the phenotype they impart (20). High-resolution mapping is a prerequisite for map-based cloning. The pooled method described in this paper can facilitate this process and may, therefore, aid in the isolation of genes from both plants and animals. In tomato alone, there are >1000 genes identified by the phenotype they impart to the plant (21). Included in the list are genes for resistance to a broad spectrum of plant pathogens and genes controlling differentiation and plant architecture. Currently, none of these genes have been cloned (to our knowledge), but all are prime candidates for map-based cloning. The same situation exists for most crop species as well as model species for genetic and molecular research including *Arabidopsis*, *Drosophila*, and mouse. Pooled mapping could be utilized in these species in the same manner as demonstrated here for tomato. Moreover, in species with a smaller genome (less DNA), larger numbers of individuals could be pooled making this strategy even more effective.

Finally, it should be noted that pooled-mapping results in the ordering of all markers in the vicinity of a scorable locus. Therefore, easily scorable loci can be used to develop region-specific high-resolution maps, even if the scorable loci themselves are not the target of map-based cloning. Thus high-resolution maps might be constructed for genes that are not readily assayed but are linked to genes that are easily scored, including quantitative trait loci.

Table 3. Three-point ordering probabilities

Ordered markers	Posterior probability $N = 43$	Posterior probability $N = 44$
CD64–CT93–rin	0.995	0.996
CT93–rin–CT63	0.971	0.904
rin–CT63–TG503	0.908	0.987
rin–TG503–TG96	0.827	0.938
rin–TG96–TG448	1.000	1.000
rin–TG448–ACC4	1.000	1.000

Posterior probabilities (Eq. 1) were computed for ordering all 21 pairs of markers relative to the target. The Monte Carlo integration used 10,000 random samples for each order H. Only 6 of the 21 ordered triplets are needed to confirm the map intervals in Fig. 2. The second column shows order probabilities using the data from 43 pools. The third column shows results when all 44 pools, including the misclassification, are analyzed.

We thank Walter Federer for helpful discussions and Colleen Bushnell for preparing the manuscript. This work was supported in part by grants from the National Research Initiative Competitive Grants Program, U.S. Department of Agriculture (91-37300-6565 and 91-37300-6418) and by the Binational Agricultural Research and Development Fund (IS-1822-90C). Computer programs used to analyze the mapping data and to generate the curves in Fig. 1 will be made available upon request.

1. Botstein, D., White, R. L. & Skolnick, M. H. (1980) *Am. J. Hum. Genet.* **32**, 314–331.
2. Donis-Keller, H., Green, P., Helms, C., Cartinhour, S., Weiffenbach, B., Stephens, K., Keith, T. P., Bowden, D. W., Smith, D. R., Lander, E. S., Botstein, D., Akots, G., Rediker, K. S., Gravius, T., Brown, V. A., Rising, M. B., Parker, C., Powers, J. A., Watt, D. E., Kauffman, E. R., Bricker, A., Phipps, P., Muller-Kahle, H., Fulton, T. R., Ng, S., Schumm, J. W., Braman, J. C., Knowlton, R. G., Barker, D. F., Crooks, S. M., Lincoln, S. E., Daly, M. & Abrahamson, J. (1987) *Cell* **51**, 319–337.
3. Copeland, N. G. & Jenkins, N. A. (1991) *Trends Genet.* **7**, 113–118.
4. Chang, C., Bowman, J. C., DeJohn, A. W., Lander, E. S. & Meyerowitz, E. S. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6856–6860.
5. Helentjaris, T. (1987) *Trends Genet.* **3**, 217–221.
6. Davies, K. (1991) *Nature (London)* **353**, 798–799.
7. Van Dilla, M. A., Deavan, L. L., Albright, K. L., Allen, N. A., Aubuchon, M. R., Bartholdi, M. F., Brown, N. C., Campbell, E. W., Carrano, A. V., Clark, L. M., Cram, L. S., Crawford, B. D., Fuscoe, J. C., Gray, J. W., Hildebrand, C. E., Jackson, P. J., Jett, J. H., Longmuire, J. L., Lozes, C. R., Luedemann, M. L., Martin, J. C., McNinch, J. S., Meincke, L. J., Mendelsohn, M. L., Meyne, J., Moyzis, R. K., Munk, A. C., Perlman, J., Peters, D. C., Silva, A. J. & Task, B. J. (1986) *BioTechnology* **4**, 537–552.
8. Saunders, R. D. C., Glover, D. M., Ashburner, M., Siden-Kiamos, I., Louis, C., Monastirioti, M., Savakis, C. & Kafatos, F. (1989) *Nucleic Acids Res.* **17**, 9027–9037.
9. Martin, G. B., Williams, J. G. K. & Tanksley, S. D. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2336–2340.
10. Michelmore, R. W., Paran, I. & Kesseli, R. V. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9828–9832.
11. Giovannoni, J. J., Wing, R. A., Ganal, M. W. & Tanksley, S. D. (1991) *Nucleic Acids Res.* **19**, 6553–6558.
12. Messeguer, R., Ganal, M., de Vicente, M. C., Young, N. D., Bolkan, H. & Tanksley, S. D. (1991) *Theor. Appl. Genet.* **82**, 529–536.
13. Lander, E. S. & Green, P. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 2363–2367.
14. Thompson, E. A. (1984) *IMA J. Math. Appl. Med. Biol.* **1**, 31–49.
15. Tanner, M. A. (1991) *Tools for Statistical Inference, Observed Data and Data Augmentation Methods* (Springer, New York).
16. Berger, J. O. (1986) *Statistical Decision Theory and Bayesian Analysis* (Springer, New York), 2nd Ed.
17. Thompson, K. H. (1962) *Biometrics* **18**, 568–578.
18. Churchill, G. A. & Tanksley, S. D. (1992) *Biometrics Technical Report* (Cornell Univ. Press, Ithaca, NY), BU-1150-MA.
19. Grierson, D. (1985) *CRC Crit. Rev. Plant Sci.* **3**, 113–132.
20. Wicking, C. & Williamson, B. (1991) *Trends Genet.* **7**, 288–293.
21. Rick, C. M. (1975) in *Handbook of Genetics*, ed. King, R. C. (Plenum, New York), pp. 247–280.