

MAPPING GENES WITH RECOMBINANT INBREDS

Benjamin Burr, Frances A. Burr and Eileen C. Matz
Biology Department, Brookhaven National Laboratory, Upton, NY 11973
Telephone: 631-344-3396, Fax: 631-344-3407, e-mail: burr@bnl.gov

Recombinant inbreds (RIs) are derivatives of an F₂ population in which linked blocks of parental alleles are now essentially fixed. Like other replicated populations - doubled haploids, for example - these populations can be continuously propagated. This means that different investigators can work with the same materials and that all data generated from the same recombinant inbred population(s) contribute to a common database. Because RIs have undergone multiple rounds of recombination before attaining homozygosity, they also permit higher mapping resolution for short linkage distances. RI families were first employed for gene mapping in mice where they continue to be extensively used. Taylor¹ and Bailey² have published detailed, helpful discussions on the theory of their use.

We have previously written about the use of RIs for mapping in maize^{3,4}. The present paper is meant to be an informal discussion on mapping with RIs in maize drawn from empirical lessons we have learned.

How are recombinant inbreds derived?

The starting materials for a recombinant inbred family are two distinct inbred lines. This assures that there will be only two alleles for any locus segregating in the subsequent population. An F₂ population is obtained from these two original inbreds. The F₂ individuals and their subsequent progeny are self-pollinated, giving rise to new inbred lines. Effective homozygosity is obtained in less than ten generations of inbreeding beyond the F₂. During the inbreeding process it is important to avoid selection in order to prevent skewing allele distributions or maintaining heterozygosity. Since inbreeding leads to the lack of vigor in some genotypes, selection cannot be entirely avoided. Propagation is by the ear-to-row method. When RI seed is distributed, only limited amounts are sent to any one laboratory to ensure that all groups will be working from the same individuals in a given generation.

What can be mapped?

Any trait for which there is polymorphism between the two original inbred parents of the RI family can be mapped. This can be a morphological trait (e.g. plant color), a physiological reaction (e.g. disease resistance), an isozyme or protein mobility polymorphism, or a DNA polymorphism detected by Southern blotting or by PCR. The trait does not have to exhibit simple inheritance but it cannot be cytoplasmically inherited, because in an RI family all the progeny receive only one cytoplasm. Multigenic traits can be mapped. In fact, these replicated families, if they are large enough, are ideal for mapping segregating factors controlling quantitative traits.

How is mapping done?

Mapping is straightforward and simple. After a polymorphism that distinguishes the original inbred parents has been identified, each recombinant inbred line in the family is typed to determine which parental allele it carries. This information comprises a strain distribution pattern for the newly mapped gene. A database search locates previously mapped loci having similar strain distribution patterns. The more tightly linked two genes are, the greater the resemblance in their strain distribution patterns. A computer program reports linked loci and also computes the map distance between the new locus and the previously mapped loci.

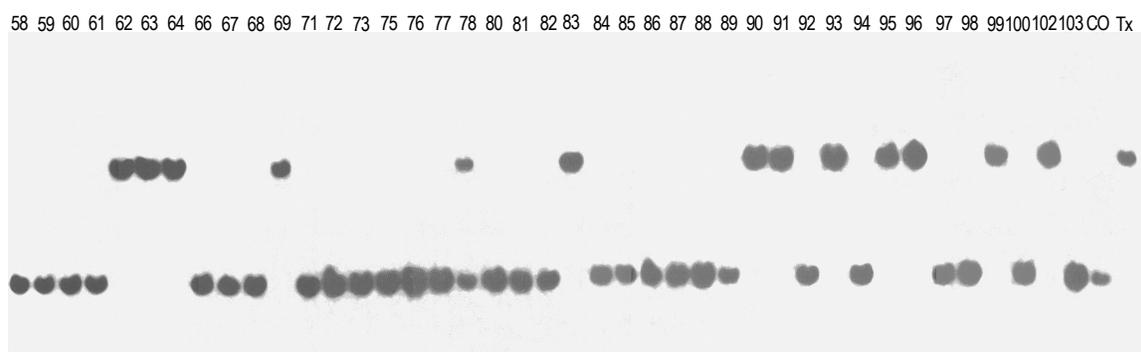
Seed of the most recent expansion of the T×CM and CO×Tx RI families is available by writing to B. Burr, Biology Department, Brookhaven National Laboratory, Upton, NY 11973. Since these plants do not do well in some environments and because we can send only a limited amount of seed to any one investigator, care should be exercised in planting and keeping track of the seed and preparing DNA from the resulting plants so that a long term supply can be maintained. Use of both families serves three functions: 1) Results from one family check the results from a second. 2) As long as the same locus is scored in both families, it is possible to combine the data for a more accurate estimate of map position. 3) When polymorphism cannot be found in one family, it is usually revealed in the other.

A common use of the recombinant inbreds is for mapping newly cloned genes. DNA is first prepared from the original parental inbreds and from all the recombinant inbreds of one or both families. As already mentioned, the inclusion of both families, if possible, will provide greater accuracy in map position. For best results, efforts should be made to prepare DNAs from all the members of the families. We generally prepare DNA from leaves taken from young plants that are ca. 30-35 cm high. The leaves of several plants of a recombinant inbred are pooled. This minimizes sampling error. (If the plants are to be grown to maturity for propagation, DNA may be made from second ears picked when the ears are immature and only about 5 cm long. Young ears are an exceptional source of DNA.)

DNA of the parental inbreds is commonly digested with three restriction enzymes that have six-base pair recognition sites and are not sensitive to cytosine methylation. We generally begin with *Bam*HI, *Bgl*III, and *Eco*RI, but other enzymes have also been useful. Because maize is such a polymorphic species, three enzymes are usually sufficient to reveal a polymorphism between any two inbreds. After probing with the new clone, the enzyme that best shows polymorphism between the parental inbreds is then used to digest the DNAs of the RIs.

The autoradiograms of the resultant Southern blots are then scored for the alternative parental bands (Fig. 1A). We have assigned arbitrary numerical values of 1 or 2 to designate the alleles of each parent in a pair. Thus in the T×CM population, CM37 alleles receive a value of 1 and T232 alleles acquire a value of 2 regardless of the locus being scored. In the other family, Tx303 alleles receive a value of 1 and CO159 alleles a value of 2. In both populations heterozygotes are assigned a value of 3. (The computer actually reads this as 1.5, but we use the integer because it is easier to type.) Zeros are put in when there is no information for that recombinant inbred. The proper format for the data is to put on one line the name of the gene followed after a few spaces with the allele distribution (Fig. 1B). All of the recombinant inbreds for a given family must be accounted for. If data from both families has been obtained, then the first line is used for T×CM and the second line for CO×Tx.

We also keep track of the name of the clone, its origin, the restriction enzyme used to detect the polymorphism, the sizes of the allelic bands, and the name of the investigator who contributed the mapping information. The numerical array is compared with the existing database

A**B**

np1340A

22221112221222222322212222221121211221212

Figure 1. A) Southern blot showing segregation of the RFLP locus *np1340A* in the CO×Tx RI family whose DNA was digested with *Bgl*III. The RI numbers appear above. The Tx303 allele has a band at 9.4 kb and the CO159 allele has a band at 6.6 kb. B) Allele distribution pattern for *np1340A*: 1 = Tx303; 2 = CO159.

using INBRED, a program developed by Keith Thompson at Brookhaven. The program reports the most closely linked loci in the database along with their assigned map positions and calculates the distances between these loci and the new gene (Table 1)

LOCUS ^a	MAP ^b	M ^c	DIFF ^d	R ^e	UNIT ^f	LOD ^g
<i>umc159</i>	6S 15.8	40	0.0	0.0000	0.0000	12.04
<i>nor</i>	6S 20.1	40	3.0	0.0750	0.0405	7.41
<i>np1235</i>	6L 23.1	39	4.0	0.1026	0.0571	6.14

^a Framework marker

^b Map position

^c Number of non-sero comparisons

^d Total of absolute differences between allele distribution

^e Recombination fraction

^f Estimated distance in Morgans

^g LOD score for linkage.

Table 1. The output from INBRED, comparing the allele distribution of *np1340A* shown in Figure 1 with the RI database. The *np1340A* locus was not separated from *umc159* and maps to 6S.

Data obtained with the T×CM or the CO×Tx families can be sent to us preferably by e-mail to burr@bnl.gov. We will run the data through the program INBRED and report the

results back to the investigator expeditiously. As of this writing the present consolidated map of maize has about 2,400 markers most of which have been contributed by other investigators.

Potential problems

Because recombinant inbreds undergo multiple rounds of meiosis while they are heterozygous, there have been more chances for recombination between tightly linked markers. On one hand, this means it is easier to detect non-allelism compared with a similar sized backcross or F_2 population, but it also means that detection of linkage beyond 20 cM can be unreliable.

The most common problems arise from experimental errors and data mishandling. These problems lead to non-linear results in their mildest form or the inability to detect reliable linkage at their worst. The database is arranged so that all mapped loci are aligned from the end of *IS* through the end of *IOL*. As such, the allele values for each RI can be viewed as a linkage map of parental alleles for a given chromosome. When the newly mapped locus is inserted into its optimal position, there should be few, if any, apparent double crossovers. Of course this is contingent on its proximity to adjacent linked markers, but the presence of two or more double crossovers is usually indicative of problems with the data. Obviously it is imperative that the RI numbers and the DNA samples not be accidentally switched. This can be verified, if there is a possibility of confusion, by using a previously mapped clone as a probe. The allele distributions should match. We would like to emphasize that only unambiguous data should be reported. When faint or obscured blots or poorly resolved bands are misinterpreted, the data is useless both for obtaining a map position and as a contribution to the database. There were about 2% heterozygotes at all loci in the RIs after eight generations beyond the F_2 . The presence of a small, but finite, number of heterozygotes is an indication of authentic data. Errors can also creep into data recording. The most frequently encountered mishap is the omission of a data point which produces a frameshift in the allele distribution. Another common mistake is a reversal in the assignment of parental allele values.

The INBRED program is robust enough to permit a few missing data points, but when this number slips to 3/4 or less of an RI family, only approximate map positions can be assigned. The problem with too few data points is that there are not enough opportunities to observe recombination and so the new locus may appear to be allelic with several adjacent loci.

Mapping is conceptually simple. However, it must be approached with sufficient rigor to obtain meaningful results. As is the case with other genetic experiments, care must be taken to keep track of strains and DNA samples, good technique should be used in preparing and interpreting Southern blots, and every effort should be expended to obtain a full data set.

References

1. Taylor, B.A. 1978. In *Origins of Inbred Mice*, ed. Morse, H.C., Academic Press, pp. 423-438.
2. Bailey, D.W. 1981. In *The Mouse in Biomedical Research*, eds. Foster, H.J., J.D. Small, and J.G. Fox, Academic Press, pp. 223-239.
3. Burr, B., F.A. Burr, K.H. Thompson, M.C. Albertsen, and C.W. Stuber. 1988. Gene mapping with recombinant inbreds in maize. *Genetics* 118: 519-526.
4. Burr, B. and F.A. Burr. 1991. Recombinant inbreds for molecular mapping in maize: theoretical and practical considerations. *Trends Genet.* 7: 55-60.

