Chapter 3

Design and Construction of Recombinant Inbred Lines

Daniel A. Pollard

Abstract

Recombinant inbred lines (RILs) are a collection of strains that can be used to map quantitative trait loci. Parent strains are crossed to create recombinants that are then inbred to isogenicity, resulting in a permanent resource for trait mapping and analysis. Here I describe the process of designing and constructing RILs. This consists of the following steps. Parent strains are selected based on phenotype, marker availability, and compatibility, and they may be genetically engineered to remove unwanted variation or to introduce reporters. A construction design scheme is determined, including the target population size, if and how advanced intercrossing will be done, and the number of generations of inbreeding. Parent crosses and F1 crosses are performed to create an F2 population. Depending on design, advanced intercrossing may be implemented to increase mapping resolution through the accumulation of additional meiotic crossover events. Finally, lines are inbred to create genetically stable recombinant lines. I discuss tips and techniques for maximizing mapping power and resolution and minimizing resource investment for each stage of the process.

Key words: Recombinant inbred line, Quantitative trait loci, Advanced intercross, Inbred line, Breeding design, Linkage map, Marker density, Mapping resolution, Mapping power, Drift

1. Introduction

The causative genetic loci underlying phenotypic traits can be mapped and studied using recombinant inbred lines (RILs) (1). RILs are a collection of strains derived from a cross of genetically divergent parent strains (see Fig. 1). Meiotic crossover events create a mosaic of parent genomes in each RIL. Phenotypes that quantitatively vary across the genetically distinct RILs can be mapped to their underlying causal loci, called quantitative trait loci (QTL). The mapping of QTL relies on markers, genotyped in each RIL, falling close enough to the causal loci (i.e., in linkage disequilibrium) to show a nonrandom association with the phenotype. Knowledge of the loci underlying phenotypic variation informs a large range of disciplines including medicine, agriculture, ecology, and evolution.

Scott A. Rifkin (ed.), *Quantitative Trait Loci (QTL): Methods and Protocols*, Methods in Molecular Biology, vol. 871,
DOI 10.1007/978-1-61779-785-9_3,

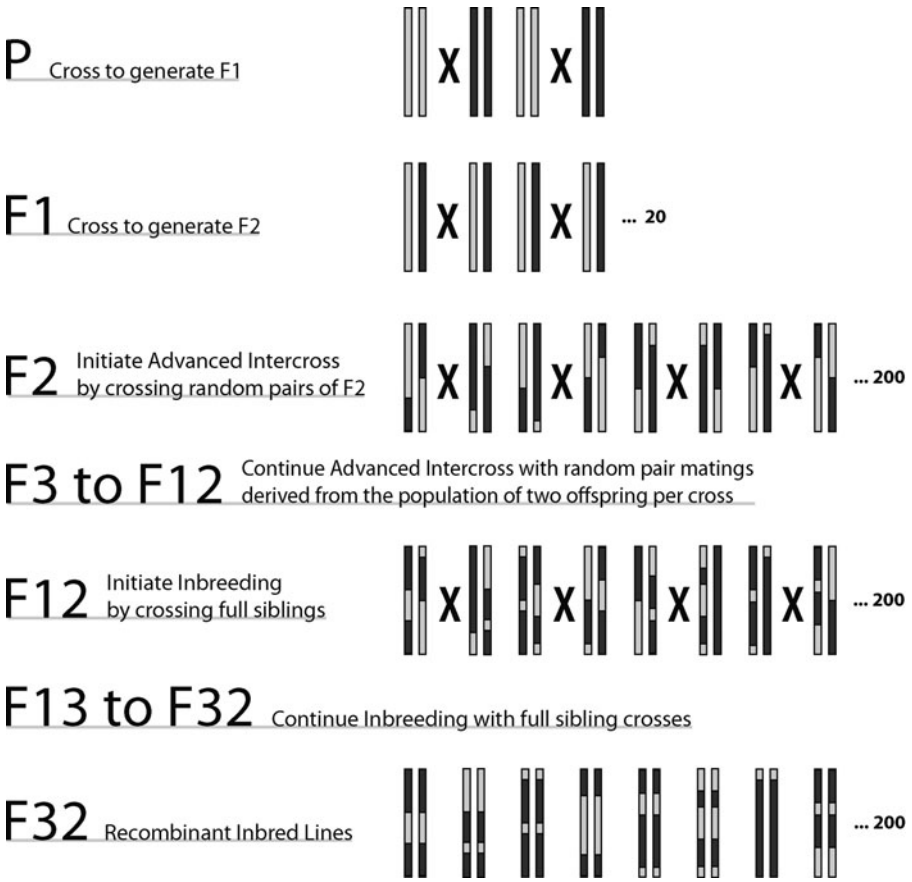

Fig. 1. Example of a RIL construction design. Two replicate parent crosses produce 40 F1. Twenty F1 crosses produce 400 F2. Two hundred random F2 crosses initiate the advanced intercross. Two hundred random pair matings of offspring (two from each cross) in each generation are performed for ten generations of intercrossing. Inbreeding of full siblings in all 200 lines begins at F12 and continues for 20 generations to F32. Individuals are represented by a set of diploid chromosomes. Each parent genotype is represented by either *light* or *dark grey*.

The step-by-step process of designing and constructing RILs for mapping phenotypes of interest is discussed in the following sections of this chapter:

2.1 Select Parent Strains
2.2 Select Construction Design
2.3 Parent Cross & F1 Cross
2.4 Advanced Intercross (Optional)
2.5 Inbreed
2.6 RIL Storage & Maintenance

Due to the enormous variation in species-specific techniques, we focus on the universal aspects of the process.

2. Methods

2.1. Select Parent Strains

The RIL design process begins with selecting parent strains that will be crossed to create the recombinants. While your parent strains may be predetermined, we outline several criteria to consider when options do exist.

1. Select strains with significant phenotypic divergence. Typically, significant divergence in the trait of interest is the initial criterion considered in selecting parent lines (see Note 1 for exceptions). Precisely defining your phenotype measurement assay and establishing the mean and variance of parent strains in your laboratory are prudent to accomplish prior to making a large resource investment in constructing a set of RILs.
2. Select strains with sufficient marker density. As described in detail in Chapter 5 on Map Making and Identifying QTL, mapping QTL depends on markers being in linkage disequilibrium with causal variants. Thus, it is important to select parent strains with a sufficient density of polymorphic markers for your QTL mapping purposes. What is sufficient? There is no universal answer to this question, though one can make some back of the envelope calculations to get a feeling for what to expect. First calculate the expected linkage map length resulting from your RIL construction design (linkage map length is the genetic distance spanned by all the chromosomes—a value that increases with increased recombination). Inbreeding to isogenicity through sibling matings (see Subheading 2.5) expands the F2 linkage map approximately fourfold, while inbreeding through selfing results in approximately twofold expansion (2). Advanced intercrossing for t generations (see Subheading 2.4) adds an additional map expansion of approximately $t/2 + 1$ (3). Given a linkage map of length L in your RILs, the number of randomly placed markers needed (n) to have fraction p loci within m map units of a random marker is:

$$n = \frac{\ln(1-p)}{\ln\left(1-\frac{2m}{L}\right)}$$

(See Note 2 for assumptions behind this derivation) (4). Plotting the number of markers (n) vs. m for different values of p and L can give an intuitive feeling for the relationship of these variables. Once a target number of markers is established, you can confirm that potential parent pairs have sufficient genotypic divergence for this marker density (see Note 3 for a discussion of the downside of extreme genotypic divergence between parents). Prior to RIL construction, the full set of markers should be selected and tested on the parents for accuracy and ease of genotyping.

3. Avoid incompatibility. Parent lines with known incompatibilities should be avoided as they will lead to the loss of some recombinants, resulting in allele frequency distortions, decreasing variation, and QTL detection power in the RILs. For example, RILs were constructed from two *Caenorhabditis elegans* strains with a previously unknown incompatibility that resulted in very strong allele frequency distortion across much of chromosome I (5).
4. Avoid segregation of other important traits. Some traits may be diverged in potential parent lines that would present an inconvenience for phenotyping and characterization of RILs. For example, divergence in a life history trait like developmental growth rate could make synchronizing strains challenging. Divergence in the efficiency of reverse genetic methods such as RNA inference could also present problems when validating causal variants (6). Careful testing of parent strains can produce a vastly more useful set of RILs.
5. Engineer parent strains. Given the resource investment in building RILs, it can be prudent to consider engineering transgenic parent strains. Such engineering could be for the purposes of removing existing variation in traits for which variation is not desirable (see Subheading 2.1, step 4) or for adding useful genes to both parents. For example, if a specific cell type will be the eventual focus of RIL phenotyping, one could engineer parent strains such that both contain transgenes expressing fluorescent markers for this cell type.
6. Inbreed parents. Ideally, parent lines should be inbred for a sufficient number of generations to be effectively isogenic (see Subheading 2.2, step 2). Inbreeding also provides the opportunity to remove recessive alleles that affect the trait of interest or life history traits.

2.2. Select Construction Design

Deciding on a design for the construction of a set of RILs requires some form of cost/benefit analysis. The major factors influencing this decision are the number of RILs produced, how many generations they are inbred, and if, how, and for how many generations they are intercrossed past the F2 generation. Different designs can require vastly different resources and can produce different mapping power and resolution, so each component of the design deserves careful consideration.

1. Determine the target RIL population size. Larger RIL populations have many advantages over smaller populations including reducing the influence of drift on allele frequencies (important for QTL detection power and mapping resolution) and increasing the number of crossing over events (important for mapping resolution). While increasing population size has different effects under different design schemes, it universally benefits

both power and resolution. Populations counted in the hundreds have negligible allele frequency drift and have high mapping resolution, while populations counted in the tens will have low mapping resolution and, as the population size approaches zero, will suffer from increasing allele frequency drift (7). Resource investment, however, scales nearly linearly with population size and must be made both during RIL construction as well as downstream for genotyping, phenotyping, and maintenance.

2. Determine the number of generations of inbreeding. Inbreeding both removes heterozygosity and generates additional independent crossing over events. After t generations of full sibling inbreeding, an initial level of heterozygosity, h_0, is approximately reduced to (8):

$$h_t = h_0 \times 1.17(0.809^t)$$

For selfing species, the expected homozygosity after t generations is $h_0/2^t$. In full sibling inbreeding, h_0 is reduced by 86% in 10 generations and 98.3% after 20 generations. In selfing inbreeding, h_0 is reduced by 99.9% in just 10 generations. With a few exceptions, 10 generations of selfing inbreeding and 20 generations of full sibling inbreeding should be sufficient. The inbreeding phase is highly valuable, straightforward to perform, and an unlikely target of resource economizing.

3. Determine if, how, and for how many generations to advance intercross. The F2 generation can be directly passaged through the inbreeding phase, or, if so desired, subject to a form of advanced intercrossing. Additional generations of intercrossing past the F2 generation has the potential to add many more crossing over events, which expands the linkage map and can improve the mapping resolution of the RILs by several fold (see Subheading 2.1, step 2). There are many schemes for deciding which line mates with which at each generation during the advanced intercross (e.g., circular (9), inbreeding avoidance (10), or random mating), but not all designs equally improve mapping resolution. Rockman and Kruglyak (7) showed through simulations that circular mating schemes and random mating schemes with variance in offspring number are relatively poor at improving mapping resolution compared to inbreeding avoidance and random mating with equal offspring number (11) (see Note 4 for an extreme example). Random mating with equal offspring number has an advantage over inbreeding avoidance in that it requires considerably less effort to implement and is unaffected by loss of lines during the intercross.

   Regardless of breeding scheme, one must determine how many generations to advanced intercross. Each generation of intercrossing expands the genetic map and improves QTL

mapping resolution. However, intercrossing (including the parent cross) also provides opportunities for allele frequencies to drift. While some advanced intercross schemes reduce the effects of drift more than others, Rockman and Kruglyak (7) found that intercross scheme plays a minor role in determining the degree of drift compared to the dominant effect of population size. To avoid problems associated with drift, little to no advanced intercrossing should be used in small populations (counted in the tens), while many generations of advanced intercrossing can be used in large populations (counted in the hundreds). One additional concern for advanced intercrosses with many generations is mutation. Mutation plays a relatively small role in RILs without intercrossing because mutations accumulate independently in lines and will therefore be unlikely to affect QTL mapping. However, mutations that take place during the intercross phase can spread through the population, potentially confounding QTL mapping. So to decide whether and for how many generations to advanced intercross, the advantages of increased mapping resolution (assuming sufficient marker density) must be weighed against potential problems from drift (in small populations), shared mutations, and the resource expense of the additional generations of crossing.

2.3. Parent Cross and F1 Cross

The goal of the parent cross is to generate an F1 population with equal chromosomal contributions from each parent and that is sufficiently large to generate the desired F2 population.

1. Set up sufficient numbers of parent crosses. Replicate parent crosses are often needed to generate the desired RIL population (see Note 5 on reciprocal crosses). Given an average brood size of B, equal sex ratios, and monogamous outcrossing, the construction of a RIL population of size N will require at minimum $4N/B^2$ replicated parent crosses (see Fig. 1). For example, to construct a RIL population of 200 for a species with average brood sizes of 20, a minimum of 2 parent crosses are needed. For practical reasons, it is always recommended to set up more crosses than are needed to guarantee sufficient numbers of F1s.
2. F1 crosses. A minimum of $2N/B$ F1 crosses are required to generate the desired F2 population (see Fig. 1). From the example above ($N = 200$, $B = 20$), 20 F1 crosses are needed to generate an F2 population of 400 from which 200 inbreeding lines can be set up. As with the parent crosses, it is always recommended to set up more crosses than the minimum required to guarantee sufficient numbers of F2s.

2.4. Advanced Intercross (Optional)

The accurate and efficient implementation of an advanced intercrossing design depends on careful planning and organization. Even for a random mating design, you need to keep careful track of the

provenance of every mated strain. See Note 6 for an important discussion of avoiding selection during intercrossing and inbreeding.

1. Initiate the intercross. Set up crosses from the F2 population (see Fig. 1). Use a simple naming scheme to give each mating a unique identifier. For example, use M1F2 for mating 1 in the F2 generation. Using a spreadsheet, record the names of the F2 crosses. As previously suggested, set up more crosses than your desired population size as some crosses might not produce offspring during the intercrossing and inbreeding. Note that many cross designs assume an even population size.
2. Plan the next intercross. Enumerate the next generation of crosses on your spreadsheet (based on your breeding design scheme). Indicate which plate will be crossed with which using the cross names. Assign new names to the new crosses. For example, you might indicate that mating 79 in the F3 generation is a cross of mating 1 and mating 164 from the F2 generation as M1F2 × M164F2 = M79F3. By planning ahead and printing out a spreadsheet that tells you what to do and what to name everything, you will have a clear record and you will minimize errors.
3. Implement the next cross. Follow your planned instructions and set up the next generation of crosses. See Note 7 on lost crosses and lines.
4. Repeat steps 2 and 3. Continue until the planned number of advanced intercross generations has been reached.

2.5. Inbreed

Inbreeding is a straightforward process of sibling pair mating (or singling for selfing species) for the desired number of generations (see Fig. 1). See Note 6 for an important discussion on avoiding selection during intercrossing and inbreeding.

1. Initiate inbreeding. Initiation of inbreeding from an F2 population involves the random pairing of F2 individuals (or singling F2 individuals for selfing species). Nothing needs to be done to initiate inbreeding from an advanced intercrossed population because they are already organized into lines and simply need to be switched to sibling pair mating or singling. Assign a unique name to each inbreeding line. Also, record which cross the inbreeding line was derived from if initiating from an advanced intercross.
2. Repeat. Continue inbreeding through sibling mating or singling until the desired number of generations has been reached.

2.6. RIL Storage and Maintenance

1. Store RILs if you can. Upon completion of the construction of the RILs, it is very important to protect your valuable resource. For organisms where storage is possible (e.g., seeds for plants or freezing microorganisms and worms), it is important to do this quickly to minimize the risks of loss and mutation.

2. Maintain large RIL populations. For the purposes of genotyping and phenotyping as well as for species that cannot be stored, it is very important to maintain large populations of each RIL to minimize the chance of fixing new mutations.

3. Notes

1. Selecting parent strains based on significant phenotypic divergence is typical, though there are exceptions. RIL panels might be constructed as a general resource, not meant for any specific trait (e.g., the Collaborative Cross of *Mus musculus* (12)). Additionally, parents with similar phenotypes may be chosen in cases where transgressive trait variation (offspring trait variation exceeds parental extremes) might be sought or expected (such as cases of compensatory evolution or canalization).
2. The equation listed in Subheading 2.1, step 1 assumes a circular genome, which will result in an underestimation of the required number of markers for linear chromosomes due to loss of information at chromosome ends. See Lynch and Walsh (4) for a correction for linear genomes.
3. High genotypic divergence is important for providing sufficient densities of markers for high-resolution QTL mapping. However, polymorphism levels greatly exceeding that needed for markers can result in inefficiencies in identifying causal variants within QTL harboring a great many polymorphisms, as there may be too many candidates.
4. At small population sizes (<32), Rockman and Kruglyak (7) found that random mating with variance in offspring number can actually decrease mapping resolution relative to RILs with no advanced intercrossing. This is due to the high degree of drift and fixation of recombinant chromosomes during the advanced intercross in this small population.
5. It is sometimes desirable to perform reciprocal crosses of the parent lines to get equal contributions of cytoplasm, mitochondrial DNA, and other sex-specific DNA to the RIL population.
6. Over the course of many generations of intercrossing and inbreeding, it is quite straightforward to inadvertently impose selection on life history traits, like growth rate. Even if parent strains show little divergence in a life history trait, segregating alleles can lead to transgressive phenotypes. Selection on traits during RIL construction causes departures from balanced allele frequencies near any genes associated with the traits, potentially affecting QTL mapping power and resolution for your traits of interest. To avoid imposing selection, attempt to implement a

procedure for picking progeny for subsequent generations in as unbiased and random a way as possible. In nematode worms, researchers have attempted to avoid selection on life history traits by using a bleaching technique on gravid adults to extract a population of fertilized embryos from which to select random individuals for the next generation (13).

7. When crosses fail, your intercross design cannot be implemented as planned. Unless you are depending on your intercross design to be perfectly implemented, it may be more desirable to violate the design and keep more lines going than to strictly maintain the design and lose lines. Therefore, one might set up the wrong crosses for a generation just to keep as many lines going as one can, make note of the change, and then adjust the plan for the next generation.

References

1. Bailey DW (1971) Recombinant-inbred strains. An aid to finding identity, linkage, and function of histocompatibility and other genes. Transplantation 11:325–327
2. Haldane JB, Waddington CH (1931) Inbreeding and linkage. Genetics 16:357–374
3. Teuscher F, Guiard V, Rudolph PE, Brockmann GA (2005) The map expansion obtained with recombinant inbred strains and intermated recombinant inbred populations for finite generation designs. Genetics 170:875–879
4. Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer, Sunderland
5. Seidel HS, Rockman MV, Kruglyak L (2008) Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. Science 319:589–594
6. Felix MA (2008) RNA interference in nematodes and the chance that favored Sydney Brenner. J Biol 7:34
7. Rockman MV, Kruglyak L (2008) Breeding designs for recombinant inbred advanced intercross lines. Genetics 179:1069–1078
8. Nagylaki T (1992) Introduction to theoretical population genetics. Springer, Berlin, New York
9. Kimura M, Crow JF (1963) On maximum avoidance of inbreeding. Genet Res 4:399–415
10. Wright S (1921) Systems of mating. II. The effects of inbreeding on the genetic composition of a population. Genetics 6:124–143
11. Crow JF, Kimura M (1970) An introduction to population genetics theory. Harper and Row, New York
12. Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J et al (2004) The collaborative cross, a community resource for the genetic analysis of complex traits. Nat Genet 36:1133–1137
13. Ebert RH 2nd, Cherkasova VA, Dennis RA, Wu JH, Ruggles S, Perrin TE et al (1993) Longevity-determining genes in *Caenorhabditis elegans*: chromosomal mapping of multiple noninteractive loci. Genetics 135: 1003–1010