

Experimental design and power

Sound experimental design is essential to good science. Scientific review committees reviewing research proposals usually look for evidence of careful experimental design. They may also desire, and it may be in the researcher's self-interest, that experiments be economical. Much of good experimental design is a mixture of common sense, pragmatism, and careful forethought, which are difficult to codify. Nonetheless, some general principles for experimental design can be laid out. In this chapter, we discuss issues special to QTL experiments with the help of the R package, *R/qtlDesign*.

In the context of QTL experiments this would include adjusting for covariates potentially influencing the phenotype, performing reciprocal crosses (if maternal effects are suspected or if the QTL is X-linked), deciding whether to consider one sex or both, adjusting for litter or foster mom effects, etc.

Before a QTL experiment can be conducted, the experimenter has to make many choices. What strains should be crossed, and what type of cross should be used? What phenotypes should be measured? Should they be replicated? What covariates should be collected? What markers should be used, and how dense should the genotyping be? Can selective genotyping be used to save money? How many progeny should be collected?

The calculations performed with *R/qtlDesign* provide quick answers to the above questions but rely on a number of approximations that may not always be accurate. A more cumbersome but potentially more accurate approach is to use computer simulation. We conclude the chapter with a brief discussion of the use of computer simulation to estimate the power to detect a QTL and the precision of localization of QTL.

6.1 Phenotypes and covariates

The most important choice is the phenotype of interest to the experimenter. Sometimes, the choice is natural. Someone interested in hypertension may measure blood pressure, while someone interested in cancer may count tumors.

Researchers studying human diseases in model systems, such as mice or rats, have to consider what phenotype best approximates the human analog. For example, a researcher studying anxiety in mice may use the open field test.

In addition to the primary phenotype of interest, it is useful to record all covariates that may affect the phenotype. Sex and cross direction should always be recorded. Whenever possible date/time of experiment, experiment batch, technician name, cage number, litter number, and parent IDs should be recorded. These are useful for diagnosing and correcting oddities in the data. In addition, factors that may affect the primary phenotype of interest should be recorded. For example, an obesity study might want to record average room temperatures because lower temperatures might lead to greater energy expenditure and lower body weight.

6.2 Strains and strain surveys

After the primary phenotype of interest has been chosen, the experimenter has to decide what strains to cross. We would generally want to cross two strains that exhibit a consistent difference in the phenotype of interest (although a cross between strains of similar phenotype may segregate QTL). Preliminary data in the experimenter's laboratory might suggest natural choices for strains to cross. In the absence of such data, an experimenter may perform a strain survey by comparing the phenotype of interest in a number of available strains. This may be done in one of two ways. The strain survey may be done *in silico*, that is, in the comfort of one's office with a computer, by simply surveying published data on the phenotypes of many strains (e.g., the Mouse Phenome Database, <http://www.jax.org/phenome>). Alternatively, it may be done the hard way, by actually phenotyping all of the strains in one's own laboratory.

The advantage of the first approach is convenience, and the ability to survey a large number of strains relatively easily. The disadvantage is that some phenotypes vary from lab to lab, and with time (as equipment and the strains may slowly drift with time). A recent study showed that the stability of phenotyping over time and space varies by phenotype, and there were no discernible differences in stability between strains (Wahlsten *et al.*, 2003). The advantage of the second approach is that the experimenter has complete control over the phenotyping protocols, and therefore more reliable data is obtained. The disadvantage is cost, and time.

Once there is reliable data on the phenotypes of strains, if we see differences in a phenotype between two strains, then we can be confident that the difference is genetic. Our next step would be to cross the two strains creating genetic variation, so that we can study the association between genetic markers and the phenotype.

6.3 Theory

Having decided which strains to cross, an investigator will have to decide what type of cross to use (e.g., backcross, intercross, or recombinant inbred lines), how many progeny to raise and phenotype, and the genotyping and phenotyping strategies. The ability to detect QTL and the precision of estimated QTL effects depend on design choices through three key quantities: the variance attributable to a given locus, the residual error variance, and the information content of the cross. We will now review the theory behind each of these quantities. This will help us understand how we can leverage cross type, number of progeny, and genotyping strategies to our advantage.

For simplicity we assume that we wish to detect a single locus contributing to variation in the phenotype. The phenotypic variance is composed of genetic and nongenetic components. A part of the genetic variance may be attributable to a locus, the rest being background genetic variance which may be due to multiple loci.

The variance attributable to a locus depends on the cross type and the mode of inheritance. Specifically, a locus with an additive mode of action (i.e., with the average phenotype for heterozygotes at the midpoint between the phenotype averages for the two homozygote groups) explains twice as much variance in an intercross, and four times as much variance in recombinant inbred lines (RILs), compared to a backcross population. On the other hand, the genetic similarity of a backcross population is greater than an intercross population, which, in turn, is more similar than a RIL population. Thus the background genetic variance is greatest in RILs, followed by intercross populations, and least in backcross populations.

Sample size calculations for QTL experiments are usually presented in terms of the proportion of variance explained by the locus of interest we wish to detect (i.e., the heritability due to the QTL). This approach provides a partial picture for comparing different cross types because the variance attributable to a locus and the background variance depend on the cross (for an illustration see Sec. 6.4).

Therefore, we present a complementary approach which directly considers the effect of a locus on the phenotype, as well as the background genetic variance in the population. To develop this approach, we begin by considering the variance attributable to a locus, followed by a discussion of the residual error variance which includes the background genetic variance. Finally, we consider manipulation of information content (or the “effective sample size”) using selective genotyping and variable marker spacing.

6.3.1 Variance attributable to a locus

Let the two strains under consideration be A and B, and let AA and BB be the parental genotypes at a locus of interest. The possible genotypes at each autosomal locus are AA, AB, and BB. Let μ be the overall mean, α be the

Table 6.1. Genotype probabilities, genotype means, and the variance attributable to a segregating locus, as a function of genetic model and cross type. BC_A denotes backcross to the A strain, and BC_B denotes backcross to the B strain.

Genotype	Mean	Probability			
		Intercross	BC_A	BC_B	RILs
AA	$\mu + \alpha - \delta/2$	1/4	1/2	0	1/2
AB	$\mu + \delta/2$	1/2	1/2	1/2	0
BB	$\mu - \alpha - \delta/2$	1/4	0	1/2	1/2

Model	Parameter	Variance attributable to locus			
		Intercross	BC_A	BC_B	RILs
General		$\frac{1}{4}\delta^2 + \frac{1}{2}\alpha^2$	$\frac{1}{4}(\alpha - \delta)^2$	$\frac{1}{4}(\alpha + \delta)^2$	α^2
No dominance	$\delta=0$	$\frac{1}{2}\alpha^2$	$\frac{1}{4}\alpha^2$	$\frac{1}{4}\alpha^2$	α^2
A dominant	$\delta=\alpha$	$\frac{3}{4}\alpha^2$	0	α^2	α^2
B dominant	$\delta=-\alpha$	$\frac{3}{4}\alpha^2$	α^2	0	α^2
No additive	$\alpha=0$	$\frac{1}{4}\delta^2$	$\frac{1}{4}\delta^2$	$\frac{1}{4}\delta^2$	0

additive effect of the locus (half of the difference between the means of the homozygotes), and δ be the dominance effect (the difference between the mean of the heterozygotes and the average of the homozygote means). The variance attributable to a segregating locus depends on the type of cross as well as the genetic model (see Table 6.1).

A purely additive locus (no dominance, $\delta=0$) segregating in a set of RILs accounts for twice the variance compared to an intercross, and four times compared to the two possible backcrosses. A dominant locus ($\delta=\pm\alpha$) segregating in an intercross explains 75% of the variance compared to RILs. If we happen to perform a backcross to the correct parental strain, the locus explains as much variance as in RILs, otherwise it does not contribute to the phenotypic variation. Thus, for a suspected dominant locus, it is safer to perform an intercross, barring specific knowledge about the dominant allele. A segregating locus would explain the most variance in RILs unless the locus is overdominant ($|\delta| > \alpha$). When a locus has no additive effect, it will explain no variance in RILs, but both an intercross and a backcross would explain the same amount of variance. Thus, an intercross, which segregates all three genotypes, offers the opportunity to detect the widest variety of genetic models.

6.3.2 Residual error variance

The residual error variance is composed of background genetic variance due to all QTL not linked to the locus of interest and the nongenetic residual variance. If measurement error is negligible, then the nongenetic residual variance is due to environmental effects specific to each individual. By using biological replication (more than one individual with the same genotype), we can reduce nongenetic residual variance. This is usually only possible for RILs, where we can create multiple individuals with the same genotype. On a similar note, we can reduce the measurement error contribution to the nongenetic residual variance by replicating measurements.

Nongenetic error variance

Let the measurement error variance be σ_M^2 (the variance of phenotype measurements in the same individual), and the environmental variance be σ_E^2 (the variance of phenotypes in individuals with the same genotype, assuming no measurement error). Assume we have m individuals per unique genotype. For backcross and intercross populations, $m=1$. For RILs, we can choose m , subject to cost constraints. Assume that we have k replicate measurements per individual (e.g., the number of times blood pressure is measured on the same mouse). Then the nongenetic residual error variance is $(\sigma_E^2 + \sigma_M^2/k)/m$. It is in the interest of the investigator to choose an instrument with negligible measurement error (σ_M^2), or to replicate the measurement enough times so that σ_M^2/k is small. In that case, the nongenetic residual error variance is approximately σ_E^2/m . Estimates of the measurement error variance (σ_M^2) and the environmental variance (σ_E^2) may be obtained from pilot studies.

Genetic variance

As mentioned earlier, the background genetic variance depends on the cross type. For simplicity, assume that the variance attributable to any single locus is small, so that the background genetic variance is approximately equal to the genetic variance. Let c be a constant that depends on the cross type, equal to 4 for backcrosses, 2 for intercrosses, and 1 for RILs, and let $\sigma_G^2(c)$ be the corresponding genetic variance. Then, the variance attributable to an additive locus is α^2/c (see Table 6.1). If we assume that all loci are additive, then the genetic variance in a cross is $\sigma_G^2(c) = \sigma_G^2/c$, where σ_G^2 is the genetic variance in RILs. Then the residual error variance would be $\sigma_G^2/c + (\sigma_E^2 + \sigma_M^2/k)/m$. Thus the ratio of the variance attributable to an additive locus to the residual error variance (the “signal-to-noise ratio”), would be

$$\begin{aligned} \frac{\alpha^2/c}{\sigma_G^2/c + (\sigma_E^2 + \sigma_M^2/k)/m} &= \frac{\alpha^2}{\sigma_G^2} \left[1 + \frac{c}{m} \left(\frac{\sigma_E^2}{\sigma_G^2} + \frac{\sigma_M^2}{k\sigma_G^2} \right) \right]^{-1} \\ &\simeq \frac{\alpha^2}{\sigma_G^2} \left(1 + \frac{c}{m} \frac{\sigma_E^2}{\sigma_G^2} \right)^{-1}. \end{aligned}$$

Table 6.2. Effect size multiplier by cross type, number of environmental replicates per genotype (m), and the ratio of environmental relative to genetic variance, measured by σ_E^2/σ_G^2 , where σ_E^2 is the within-genotype variance and σ_G^2 is the between-genotype variance in RILs. We assume that all loci are additive, so that the between-genotype variance (the genetic variance) is $\sigma_G^2/2$ and $\sigma_G^2/4$ in intercross and backcross populations, respectively.

σ_E^2/σ_G^2	Backcross	Intercross	RILs	
	$m=1$	$m=1$	$m=1$	$m=4$
1/16	0.80	0.89	0.94	0.98
1/4	0.50	0.67	0.80	0.94
1/2	0.33	0.50	0.67	0.89
1	0.20	0.33	0.50	0.80
2	0.11	0.20	0.33	0.67
4	0.06	0.11	0.20	0.50
16	0.015	0.030	0.059	0.200

The approximation holds when σ_M^2/k is negligible (i.e., when technical measurement error is small or we have sufficient technical replicates). It is easier to detect a QTL when the signal-to-noise ratio is higher.

The signal-to-noise ratio depends on α^2 , σ_G^2 , σ_E^2 , c , and m . Of these, the first three are determined by nature, over which the experimenter has no control. However, by choosing the cross type (which determines c), and the number of environmental replicates per genotype (m), the experimenter can manipulate the signal-to-noise ratio. We focus on the effect size multiplier, $[1+c\sigma_E^2/(m\sigma_G^2)]^{-1}$, displayed in Table 6.2 as a function of cross, number of replicates per genotype, and the ratio of the environmental variance to the genetic variance. As one would expect, the signal-to-noise ratio is highest when the environmental variance is small; in this setting, there is little difference between the different crosses. However, when the environmental variance is high, the signal-to-noise ratio is low for all crossing designs. In this setting, RILs are most advantageous. This advantage can be magnified by replication.

6.3.3 Information content

We have greater control over the information content of a cross compared to our control over the effect size and the error variance. The information content of an experiment may be interpreted as the effective sample size. It is a fundamental quantity which affects the power to detect a QTL, the expected LOD score, and the precision of the estimated QTL effects. We define the information content in the sense of Fisher information (see Cox and Hinkley,

1974): the reciprocal of the expected variance of the genetic model parameters for unit residual variance.

The information content depends on the number of progeny, genotyping strategies, and phenotyping strategies. Genotyping strategies that can affect information content include marker spacing and selective genotyping (where a fraction of the individuals with extreme phenotypes are genotyped). Phenotyping strategies include choosing a subset of individuals based on their genotype for expensive phenotyping, and replication of noisy measurements. By making choices that maximize information content, based on the cost structure of the experiment, the experimenter can most efficiently allocate resources.

6.4 Examples with R/qtlDesign

R/qtlDesign is an R package for facilitating experimental design choices for QTL mapping. It may be obtained from the Comprehensive R Archive Network (CRAN, <http://cran.r-project.org>) and installed in the same way that the R/qtl package is installed (see App. A).

We assume that the phenotypes follow a normal distribution, that the effect size of a QTL is small relative to the residual variance, and that the sample size is large. (A warning is printed if the effective sample size is smaller than 30.) The normality assumption facilitates analytical tractability, and the sample size assumption is needed to use χ^2 approximations in power calculations. The small effect size assumption simplifies information content calculations; the resulting approximation is accurate for most practical situations. We also assume that measurement error is negligible. If that is not the case, it may be advisable to consider replicating measurements.

The package assumes that cost functions are linear; this ignores economies of scale, but provides a useful guide. The optimal marker spacing and the selection fraction (the proportion of extreme phenotypic individuals that are genotyped) should therefore be seen as approximations; they are not necessarily optimal. The function approximating the residual error variance assumes that all loci are additive. This is intended as an approximation; in practice one would consider a range of possibilities (see the examples below).

6.4.1 Functions

The main functions in the R/qtlDesign package are the following:

powercalc	Calculates the power to detect a QTL given the effect size, the residual error variance, the genome-wide LOD threshold, the width of the marker interval containing the QTL, the sample size, and the proportion of extreme individuals genotyped (the selection fraction).
------------------	---

detectable	Calculates the minimum detectable QTL effect as a function of the target power and other powercalc inputs.
samplesize	Calculates the minimum sample size needed to detect a QTL effect given the desired power and other powercalc inputs.
info	Calculates the approximate expected information as a function of the selection fraction and the marker interval width.
optspacing	Calculates the optimal marker spacing and selection fraction given the genotyping cost and the genome size.
optselection	Calculates the optimal selection fraction given the genotyping cost, marker spacing, and genome size.
error.var	Calculates the residual error variance given the cross type, environmental variance, background genetic variance, and the number of environmental replicates per unique genotype.
thresh	Calculates the genome-wide LOD threshold required for QTL detection given the cross type, genome length, and marker density, using the approximations of Dupuis and Siegmund (1999).

6.4.2 Choosing a cross

Barring specific knowledge about the mode of action of the QTL, the intercross is often the best cross choice. It segregates all possible genotypes, and therefore permits detection of QTL with any mode of action (dominant, recessive, additive, or overdominant). If the phenotype is noisy, with a lot of environmental variation, then RILs (provided that they are available) are the best choice, as we can use replicate individuals to decrease noise. If we are confident of the nature of the effect, or suspect substantial genetic variance due to epistasis, then performing a backcross would be a good choice. Sometimes investigators will perform more than one type of cross, and then combine the evidence from multiple populations.

Power calculations in research grant applications traditionally are presented in terms of the proportion of variance detectable with high power given a population of a certain size. Let us explore what effects we can detect using a backcross or intercross population with 100 individuals. We assume that we desire 80% power in a mouse cross.

We must first load the R/*qtlDesign* package.

```
> library(qtlDesign)
```

We first estimate the 5% genome-wide LOD threshold that we will use for the mouse genome (of size 1440 cM), assuming infinitely dense markers. We use the `thresh` function.

```
> thresh(G=1440, cross="bc", p=0.05)
```

```
[1] 3.190
```

We get the analogous threshold for an intercross as follows.

```
> thresh(G=1440, cross="f2", p=0.05)
```

```
[1] 4.183
```

Note that the LOD threshold for an intercross population is a bit higher. This is because we have two degrees of freedom in an intercross (versus one in a backcross), and because the recombination density in an intercross is twice that of a backcross.

We can now calculate the minimum detectable effect sizes using the function `detectable`.

```
> detectable(cross="bc", n=100, sigma2=1, thresh=3.2)
```

```
      effect percent.var.explained
[1,]    0.936                17.97
```

```
> detectable(cross="f2", n=100, sigma2=1, thresh=4.2)
```

```
      additive.effect dominance.effect percent.var.explained
[1,]           0.726                0                20.86
```

Thus, we can detect loci explaining a similar percentage of variance in backcrosses and intercrosses. However, the effect size that can be detected in an intercross is smaller if the environmental variance is high. We will see this in examples below.

Table 6.3 displays the approximate detectable effects (measured as the percent variance explained by a QTL) for a range of sample sizes, in each of a backcross and an intercross.

Suppose we are planning to map blood pressure QTL in mice by crossing the A and B strains, whose blood pressure means are 85 mm of Hg and 105 mm of Hg, respectively. The within-strain standard deviations are 8 mm of Hg. Let us also assume that we are interested in detecting a locus with an additive effect of 5 mm of Hg. We will compare crossing designs assuming that we want to detect the locus with 80% power. (For brevity, we suppress measurement units below.) How do the choices of backcross, intercross, and recombinant inbred lines shape up?

We estimate σ_E^2 , the environmental variance, by the within-strain variance, $8^2=64$. We estimate the background genetic variance with some assumptions, and therefore consider a range of possibilities, guiding our choices using the

Table 6.3. The minimum percent variance attributable to a QTL for it to be detectable with 80% power, as a function of sample size, marker spacing (in cM), and the selection fraction (the proportion of extreme phenotypic individuals genotyped). We use a significance threshold of 3.2 for a backcross and 4.2 for an intercross. These are the estimated thresholds corresponding to dense markers for the mouse genome (1440 cM). The power is calculated for a locus at the center of a marker interval with the given spacing.

Sample size (n)	Selection fraction							
	100%				50%			
	Marker spacing in cM				Marker spacing in cM			
	0	5	10	20	0	5	10	20
Backcross								
100	17.9	18.7	19.5	21.4	19.1	19.9	20.7	22.7
200	9.9	10.3	10.8	12.0	10.5	11.0	11.6	12.8
400	5.2	5.4	5.7	6.4	5.6	5.8	6.1	6.8
Intercross								
100	20.8	21.7	22.6	24.7	22.1	23.0	23.9	26.1
200	11.6	12.2	12.7	14.1	12.4	13.0	13.6	15.0
400	6.2	6.5	6.8	7.6	6.6	6.9	7.3	8.1

between-strain variance, $(105 - 85)^2/4 = 100$. This would be the genetic variance in RILs if a single QTL accounted for the strain difference. If there are two or more QTL, the genetic variance would be smaller, assuming no epistasis and that all QTL have effects of the same sign. Thus we consider three values of σ_G^2 , 25, 50, and 100 (these correspond to percent variance attributable to the QTL equal to 14.0%, 12.3% and 9.9%, respectively). We may use the `samplesize` function to get an approximate sample size. For $\sigma_G^2=25$, we use:

```
> samplesize(cross="f2", effect=c(5,0), env.var=64, gen.var=25,
+           thresh=4.2)

      sample.size percent.var.explained
[1,]          162              14.04
```

This gives us a sample size of 162. The additive and dominance effects we wish to detect are given in the `effect` argument. The residual error variance is approximated using our estimates of the environmental and genetic variances (alternatively one can specify the residual error variance directly). If the genetic variance is 50 or 100, we get sample size estimates of 188 and 241, respectively.

```
> samplesize(cross="f2", effect=c(5,0), env.var=64, gen.var=50,
+           thresh=4.2)
```

```
      sample.size percent.var.explained
[1,]          188             12.32
```

```
> samplesize(cross="f2", effect=c(5,0), env.var=64,
+           gen.var=100, thresh=4.2)
```

```
      sample.size percent.var.explained
[1,]          241             9.881
```

By default, the software assumes that our target LOD threshold is 3, that our desired power is 80%, and that all individuals are typed densely.

For a backcross population, we would need more individuals, as the following results show.

```
> samplesize(cross="bc", effect=5, env.var=64, gen.var=25,
+           thresh=3.2)
```

```
      sample.size percent.var.explained
[1,]          247             8.17
```

```
> samplesize(cross="bc", effect=5, env.var=64, gen.var=50,
+           thresh=3.2)
```

```
      sample.size percent.var.explained
[1,]          269             7.553
```

```
> samplesize(cross="bc", effect=5, env.var=64, gen.var=100,
+           thresh=3.2)
```

```
      sample.size percent.var.explained
[1,]          312             6.562
```

To estimate how many RILs we would need (if such a resource existed), we will first have to estimate the LOD threshold required. Assuming that the RILs were created by sibling mating, we can get the threshold by using the `thresh` function for a backcross population, multiplying the genome length by 4 (because of the four-fold map expansion in RILs by sibling mating).

```
> thresh(G=1440*4, cross="bc", p=0.05)
```

```
[1] 3.834
```

RILs are most advantageous when the genetic variance is small relative to the environmental variance. We therefore consider the setting when σ_G^2 is 25.

```
> samplesize(cross="ri", effect=5, env.var=64, gen.var=25,
+           thresh=3.8)
```

```

      sample.size percent.var.explained
[1,]          90          21.93

```

We would need about 90 animals, which is favorable in terms of the number of animals needed, but might be infeasible because RIL populations of such size are expensive to create and maintain. This assumed that we used just one animal per line. With replication, the number of unique RILs decreases, but to a point. We see this by using the `bio.rep` argument.

```

> samplesize(cross="ri", effect=5, env.var=64, gen.var=25,
+           thresh=3.8, bio.rep=2)

```

```

      sample.size percent.var.explained
[1,]          58          30.49

```

```

> samplesize(cross="ri", effect=5, env.var=64, gen.var=25,
+           thresh=3.8, bio.rep=4)

```

```

      sample.size percent.var.explained
[1,]          42          37.88

```

```

> samplesize(cross="ri", effect=5, env.var=64, gen.var=25,
+           thresh=3.8, bio.rep=16)

```

```

      sample.size percent.var.explained
[1,]          30          46.3

```

```

> samplesize(cross="ri", effect=5, env.var=64, gen.var=25,
+           thresh=3.8, bio.rep=100)

```

```

      sample.size percent.var.explained
[1,]          26          49.37

```

This indicates that, with 4–20 replicate animals per line, we can detect the desired effects in a RIL population of modest size (30–40 lines). If we used 4 replicate animals, we would use about 168 animals. The number of intercross animals needed is about 162, and the number of backcross animals needed is about 247. Since the cost of breeding replicate animals is smaller for RILs, one would conclude that using a RIL population with about 40 lines would be a good choice if the genetic variance is small. However, the intercross is quite competitive.

6.4.3 Genotyping strategies

Once a cross has been performed, genotyping strategies can be used to reduce experimental cost. Because of linkage between adjacent markers on the same chromosome, there are diminishing returns as genotyping density increases. An investigator might want to know what genotyping density provides a good

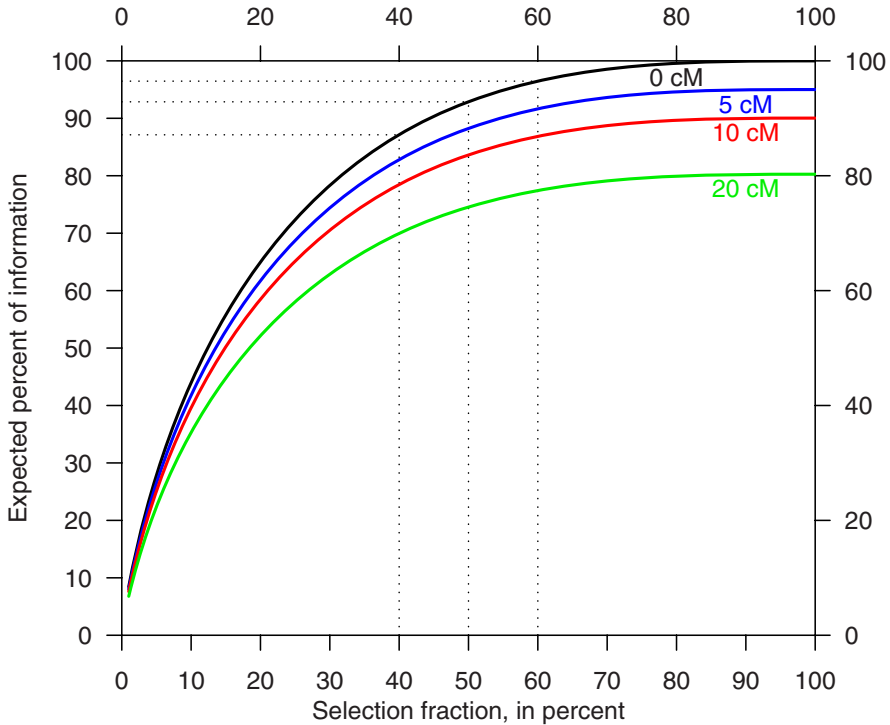


Figure 6.1. Expected information from a selectively genotyped backcross as a function of the selection fraction (the proportion of extreme phenotypic individuals genotyped), in the middle of a marker interval of length 0, 5, 10, and 20 cM. The information is plotted relative to a fully genotyped cross, where all individuals are genotyped at a dense set of markers.

return on investment. Further, if there is a single phenotype of primary interest, selective genotyping, where only extreme phenotypic individuals are genotyped, may be used. This strategy is most effective when the cost of phenotyping and raising an individual is low relative to the cost of genotyping. We will consider these issues with our design tools. First, let us take a look at how information content (“effective sample size”) varies with marker density and the selection fraction (the proportion of extreme phenotypic individuals genotyped).

The expected information from a selectively genotyped backcross is displayed in Fig 6.1. The information is plotted relative to a fully genotyped cross where all individuals are genotyped at a dense set of markers. We can see that if the cross is densely genotyped, genotyping between 40% to 60% of the cross gives us 85% to 95% of the information in the cross. The gains diminish with wider marker spacings. The figure was created using the function `info`.

Suppose we are performing an intercross, and suppose a genotyping facility charges 10 cents per genotype, and that the animal facility per diem rate for mice is \$1.20. The total cost of housing a mouse for 25 weeks is about \$30; therefore the genotyping cost in the units of raising the mouse is about 1/300. To find the genotyping density that gives the best information-to-cost ratio, we use the function `optspacing`.

```
> optspacing(cost=1/300, G=1440, sel.frac=1, cross="f2")
```

Marker spacing (cM)	Selection fraction
17.93	1.00

This suggests that relatively sparse genotyping (18 cM density) would be adequate for detecting QTL.

If we had a single phenotype and could perform selective genotyping, we could find the selection fraction and genotyping density combination that give the best information to cost ratio. We do this by setting the `sel.frac` argument to `NULL`.

```
> optspacing(cost=10/3000, G=1440, sel.frac=NULL, cross="f2")
```

Marker spacing (cM)	Selection fraction
14.5573	0.6063

This suggests that the most economical option would be to genotype approximately 60% of the cross at a 15 cM marker density.

6.4.4 Phenotyping strategies

For many investigators, phenotyping costs, rather than genotyping costs, are high relative to the cost of raising an individual. Examples include behavioral phenotyping or microarrays. In these settings it may be more useful to perform *selective phenotyping*. Selective phenotyping involves phenotyping a subset of individuals, chosen based on their genotype or by based on another (inexpensive, but related) phenotype.

The idea of selective phenotyping based on observed genotypes is to select the most genotypically diverse subset of given size. For example, we may want to select 40 out of 200 intercross individuals for microarray phenotyping. One can select a set of dissimilar individuals using the MMA (minimum moment aberration) method. This method is implemented in the `mma` function.

We illustrate the use of the former strategy by simulating data on five chromosomes of length 100 cM. Then we use the `mma` function to select 40 individuals based on chromosome 1 genotypes (where the QTL is). We compare the LOD scores obtained by selective phenotyping to that obtained from a random subset of 40 individuals.

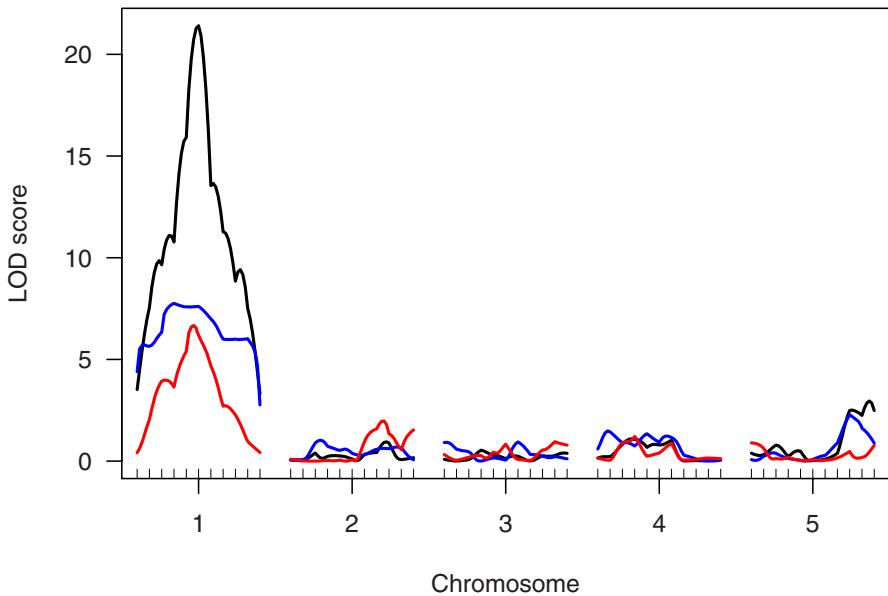


Figure 6.2. Illustration of selective phenotyping using simulated data. We compare the LOD scores obtained using three different strategies: the full cross of 200 individuals (black), 40 individuals selectively phenotyped based on chromosome 1 genotypes (blue), and a random set of 40 individuals (red).

```
> library(qtl)
> mp <- sim.map(len=rep(100,5), n.mar=11, include.x=FALSE,
+             eq.spacing=TRUE)
> cr <- sim.cross(mp, model=c(1,50,1,0), n.ind=200, type="f2")
> idx40 <- mma(pull.geno(cr, chr=1), p=40)
> cr <- calc.genoprob(cr, step=2)
> out1 <- scanone(cr)
> out2 <- scanone(subset(cr, ind=idx40$cList))
> out3 <- scanone(subset(cr, ind=sample(1:200, 40)))
> plot(out1, out2, out3, ylab="LOD score")
```

This type of selective phenotyping is most effective when we have credible knowledge about the region where the putative QTL might be. The better the knowledge, the more effective it is. Otherwise, selectively phenotyping is about as good as phenotyping a random subset of individuals.

6.4.5 Fine mapping

After a QTL has been detected, one will usually want to narrow the location of the QTL. Planning for fine mapping involves different considerations than in the detection stage discussed earlier in the chapter. It is helpful to

have access to genotyping resources to effectively narrow the location of every crossover in every individual. With dense genotyping, the average width of confidence intervals is proportional to the *inverse of the sample size*. By contrast, for most statistical problems, the lengths of confidence intervals are approximately proportional to the inverse of the *square root* of the sample size. This happy circumstance is tempered by the fact that the width of confidence intervals varies from cross to cross depending on the configuration of marker genotypes in the neighborhood of the true QTL. For this reason, the `R/qtlDesign` function, `ci.length`, for calculating confidence interval widths, reports the *median* confidence interval width.

The following commands give us the median width of the 95% confidence interval for QTL location for a backcross or intercross with 250 individuals, assuming dense genotyping. In practice, confidence intervals are likely to be slightly wider than these calculations indicate.

```
> ci.length(cross="bc", n=250, effect=5, p=0.95,
+          gen.var=25, env.var=64)

[1] 13.47

> ci.length(cross="f2", n=250, effect=c(5,0), p=0.95,
+          gen.var=25, env.var=64)

[1] 7.334
```

We see that with an intercross we will get confidence intervals that are expected to be about half as wide as those obtained from a backcross. As with the power for detecting QTL, the expected widths also depend on the background genetic variance, as well as the environmental variance.

6.5 Other experimental populations

Although backcrosses, intercrosses, and recombinant inbred lines are the staples of experimental geneticists, a number of other populations are in widespread use for QTL mapping. All of them are based on the same fundamental idea. We create (or assemble) a genetically diverse set of individuals, genotype and phenotype them, and look for associations between genotype and phenotype. Statistical methods to examine these associations depend on the population. In the following, we briefly describe a number of other populations that might be considered for QTL mapping.

Advanced intercross lines (AIL) are constructed by intercrossing an intercross population for multiple generations. At any given locus, they have approximately the same diversity as an intercross; however the span of linkage disequilibrium (LD) is much shorter due to the multiple generations of recombination. They are useful for fine-mapping, but require a greater breeding effort and very dense genotyping.

Heterogeneous stock (HS) are similar in spirit to advanced intercross lines, but are derived from multiple strains. Typically, an outbreeding mating scheme is used to maintain heterozygosity. The genotypic diversity of HS is greater than AIL, and the span of LD is smaller as well. In the analysis of both HS and AIL, one generally will need to take account of familial relations among individuals.

Introgression lines are a set of congenic strains spanning a locus, a chromosome, or the whole genome. They consist of a series of introgressions from a donor strain onto a recipient strain (i.e., a genomic segment from the donor strain is inserted into the recipient strain by a series of crosses). Thus, they are a collection of single-factor perturbations of a genetic system. They may be used for either genome scanning or fine mapping. Consomic strains (or chromosome substitution strains, CSS) are a special case in which the introgressed segment consists of a whole chromosome.

Recombinant congenic strains (RCS) are like introgression lines, but each strain may contain multiple small introgressed segments, rather than just one.

Inbred-outbred crosses between an inbred strain and an outbred populations may offer benefits of both linkage and association mapping. By tracking identity by descent (IBD) from the parental strains, one can perform linkage mapping (as in a backcross or intercross). By examining association between alleles at any given locus and the phenotype, one can use historical recombinations within the outbred stock to narrow down the location of a QTL.

Strain collections (association mapping) The prospect of using historical recombination for QTL detection or fine mapping also underlies association mapping using a collection of strains. The advantage of this method is that one can tap the great genetic diversity of strain collections. A disadvantage is that the complex relationships among strains may lead to spurious associations.

Natural populations also offer some of the same advantages of strain collections, especially the prospect of using historical recombination for fine mapping. However, one may have to contend with hidden population structure in the collection, and the trait of interest may not be segregating in the population.

Selection experiments apply a selective pressure on a population, and let the population evolve for a small number of generations before genotyping. By observing which genotypes survive the selective pressure, one can identify loci that underly fitness to survive selection.

An affecteds-only design may be seen as a variant of the above where one only genotypes the affected individuals (or individuals exhibiting a particular extreme of a continuous trait). By observing which genotypes are over-represented, one can map loci for the trait of interest.

The Collaborative Cross (CC) is a proposed collection of a large number of recombinant inbred lines derived from eight parental mouse strains. It seeks to establish an immortal, genetically diverse set of experimental genetic factors for mouse biology. Because it is genetically more diverse than RILs derived from two strains and has a smaller span of linkage disequilibrium, it is expected

to be more efficient for mapping loci than RILs. The CC lines can be used as reference strains for complex phenotyping that generate data comparable across laboratories or years.

6.6 Estimating power and precision by simulation

The `R/qtlDesign` package provides quick answers to numerous design questions. However, the calculations in `R/qtlDesign` rely on a number of approximations that may not always be appropriate, and so the results are not always accurate.

An alternate approach to assessing the power to detect QTL and the precision of localization of QTL is to use computer simulation. Computer simulations can be cumbersome and time consuming, but they are extremely flexible and so allow one to address more complex questions.

In this section, we illustrate the use of computer simulation to assess the power to detect a QTL and the precision of localization of a QTL. We focus on the simple case of an intercross with a single QTL.

The simulation of QTL mapping data was described in Sec. 2.5. We must start with a genetic map of marker locations. It is convenient to use the `map10` object that is distributed with `R/qtl`. This is a genetic map modeled after the mouse genome, with evenly spaced markers at approximately a 10 cM spacing. (The marker spacing is slightly different on the different chromosomes so that the markers will be evenly spaced on each chromosome but with chromosome lengths matching those of the mouse genome.)

We first load `R/qtl` and get access to the `map10` object.

```
> library(qtl)
> data(map10)
```

To assess the power to map a QTL, we must first obtain a significance threshold. While one might use a permutation test for each simulated QTL mapping data set, that would be extremely time consuming. Instead, we perform some initial simulations under the null model (of no QTL) to estimate the null distribution of the genome-wide maximum LOD score.

We consider the case of an intercross with 250 individuals and assume no crossover interference and complete marker genotype data with no genotyping errors. We focus solely on the autosomes (chromosomes 1–19) and perform 10,000 simulation replicates.

The code to accomplish this is not too complicated, but requires a bit of knowledge of R.

```
> n.sim <- 10000
> res0 <- rep(NA, n.sim)
> for(i in 1:n.sim) {
+   x <- sim.cross(map10[1:19], n.ind=250, type="f2")
```

```
+ x <- calc.genoprob(x, step=1)
+ out <- scanone(x, method="hk")
+ res0[i] <- max(out[,3])
+ }
```

We first create an empty vector to contain the genome-wide maximum LOD scores. We use a `for` loop to do the 10,000 simulations. We call `sim.cross` to simulate the data (under the null hypothesis of no QTL). We use `calc.genoprob` to calculate the QTL genotype probabilities and then `scanone` to perform a genome scan. (We use Haley–Knott regression, for the sake of speed.) We pull out the maximum LOD score with `max(out[,3])`, since the LOD scores form the third column in the output.

The 95th percentile of the results serves as our estimate of the 5% genome-wide LOD threshold. We use `print` in order to simultaneously assign the 95th percentile to `thr` and print the value.

```
> print(thr <- quantile(res0, 0.95))

95%
3.582
```

It is interesting to compare this result to that obtained with the function `thresh` in `R/qtlDesign`. To use `thresh`, we first need to calculate the length of the genome. This may done by a call to `summary.map`. We add up the values in the first 19 rows (corresponding to the autosomes) in the column labeled “length.”

```
> print(G <- sum(summary(map10)[1:19, "length"]))

[1] 1568
```

We now use `thresh` to estimate the significance threshold. We use `d=10` (the marker spacing) and `p=0.05` (the significance level). (Note that we would need to use `library(qtlDesign)` to load the `R/qtlDesign` package, if it had not already been loaded.)

```
> thresh(G, "f2", d=10, p=0.05)

[1] 3.424
```

This is slightly smaller than that estimated by our simulations.

We can also make a histogram of the genome-wide maximum LOD scores. See Fig. 6.3. We use the function `rug` to create, underneath the histogram, line segments at the individual data points.

```
> hist(res0, breaks=100, xlab="Genome-wide maximum LOD score")
> rug(res0)
```

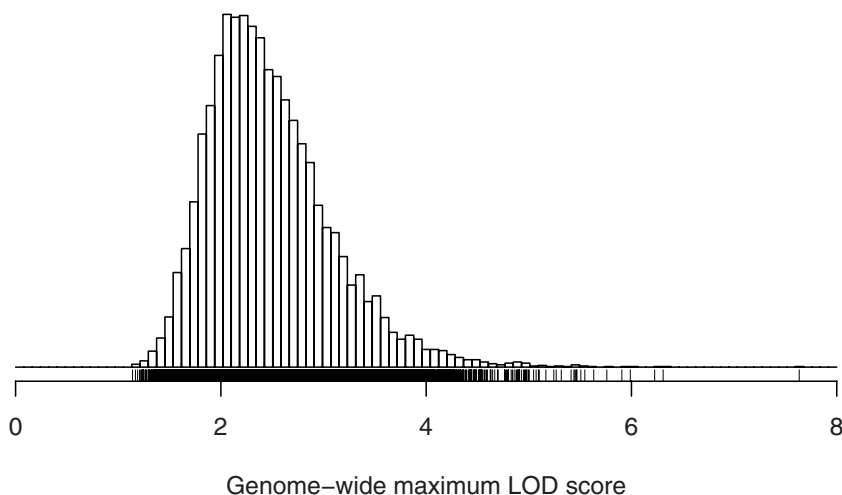


Figure 6.3. Distribution of the genome-wide maximum LOD scores under the null hypothesis of no QTL, for the case of an intercross with 250 individuals and with a genome modeled after that of the mouse and with equally spaced markers at a 10 cM spacing.

With our LOD threshold in hand, we may now turn to simulations in the presence of QTL. We will consider the simplest possible case, of a single QTL. We will assume that the alleles act additively (that is, the average phenotype for the heterozygote is halfway between the averages for the two homozygotes). We will simulate a QTL responsible for 8% of the phenotypic variance. If the average phenotypes for the two homozygotes are $-\alpha$ and α and the residual variance is 1 (as assumed in `sim.cross`), then the heritability due to the QTL is $\alpha^2/2/(\alpha^2/2 + 1)$. (See Table 6.1 on page 156.) Thus, we need $\alpha = \sqrt{2 \times 0.08/(1 - 0.08)} \approx 0.417$.

We will place the QTL at 54 cM on chromosome 1 (halfway between two markers). We may assess power and precision at the same time, and we will also study the width and coverage of the 1.5-LOD support interval (see Sec. 4.5). We need only simulate data for the chromosome containing the QTL. This will save a great deal of computation time.

The code to perform the simulations is a bit more complicated than before.

```
> alpha <- sqrt(2*0.08/(1-0.08))
> n.sim <- 10000
> loda <- est <- lo <- hi <- rep(NA, n.sim)
> for(i in 1:n.sim) {
+   x <- sim.cross(map10[1], n.ind=250, type="f2",
+                 model=c(1, 54, alpha, 0))
+   x <- calc.genoprob(x, step=1)
```

```

+   out <- scanone(x, method="hk")
+   loda[i] <- max(out[,3])
+   temp <- out[out[,3]==loda[i],2]
+   if(length(temp) > 1) temp <- sample(temp, 1)
+   est[i] <- temp
+   li <- lodint(out)
+   lo[i] <- li[1,2]
+   hi[i] <- li[nrow(li),2]
+ }

```

We first calculate the effect of the QTL that corresponds to heritability of 8%. We create empty vectors that will contain the results: the maximum LOD scores, the estimated QTL positions, and the lower and upper endpoints of the 1.5-LOD support intervals. We must be careful about the case that multiple positions give exactly the same LOD score; in such cases, we pick a random location (among those with the maximum LOD) as the estimated location of the QTL. We use the `lodint` function to calculate the 1.5-LOD support interval, and we again must be careful of the case that multiple positions share the maximum LOD score. The first and last elements in the second column of the output from `lodint` are the endpoints of the interval; while there are generally three rows in the output, there can be more.

The estimated power is the proportion of the simulation replicates with LOD score exceeding our threshold.

```

> mean(loda >= thr)

[1] 0.7383

```

This is slightly larger than the estimate provided by `powercalc` in R/qtl-Design. Note that we use `theta=0.09`, the approximate recombination fraction between two markers (from the Haldane map function, for a genetic distance of 10 cM).

```

> powercalc("f2", 250, sigma2=1, effect=c(alpha,0), thresh=thr,
+           theta=0.09)

      power percent.var.explained
[1,] 0.6855                      8

```

We might also wish to look at the distribution of the maximum LOD scores. In 10,000 simulation replicates, we obtained LOD scores as large as 14.7 (see Fig. 6.4).

```

> hist(loda, breaks=100, xlab="Maximum LOD score")

```

Turning to the precision of localization of the QTL, we first consider a histogram of the estimated QTL locations.

```

> hist(est, breaks=100, xlab="Estimated QTL location (cM)")
> rug(map10[[1]])

```

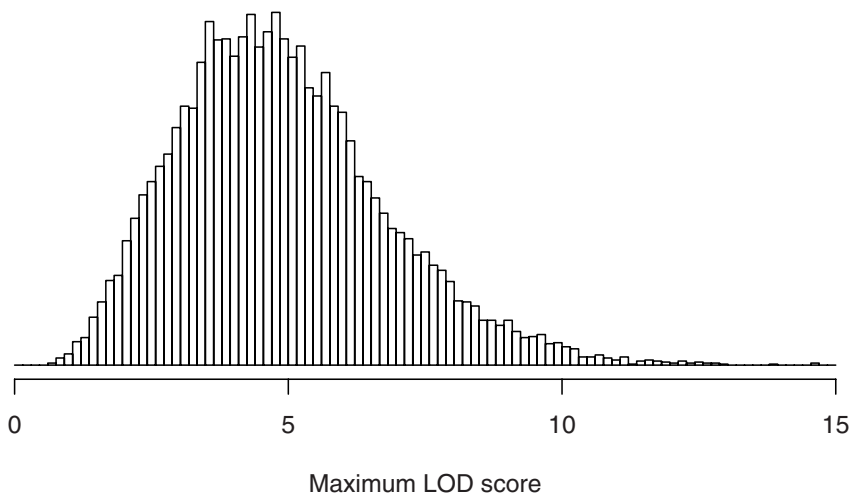


Figure 6.4. Distribution of the chromosome-wide maximum LOD scores in the presence of a single QTL responsible for 8% of the phenotypic variance, for the case of an intercross with 250 individuals and with equally spaced markers at a 10 cM spacing.

We use `rug(map10[[1]])` to place tick marks at the marker locations. (In the results, in Fig. 6.5, we used a slightly fancier method for defining the breakpoints in the histogram; see the detailed code for this and all figures in the book in the online complements at <http://www.rqtl.org/book>.)

Note the large spike in the distribution at the two markers flanking the QTL. The QTL is estimated to be at one or the other of these markers approximately 12% of the time.

The estimate of QTL location is approximately unbiased. (Recall that the QTL was located at 54 cM.)

```
> mean(est)
```

```
[1] 54.37
```

The estimated standard error of the estimated QTL location is approximately 11.6.

```
> sd(est)
```

```
[1] 11.62
```

It is interesting to consider the precision of the estimated QTL location among those cases in which there was significant evidence for a QTL (i.e., for which the LOD score exceeded our threshold, 3.58). We first create an indicator of the cases in which the LOD score exceeded the threshold, and then calculate the SD of the estimated QTL location among those cases.

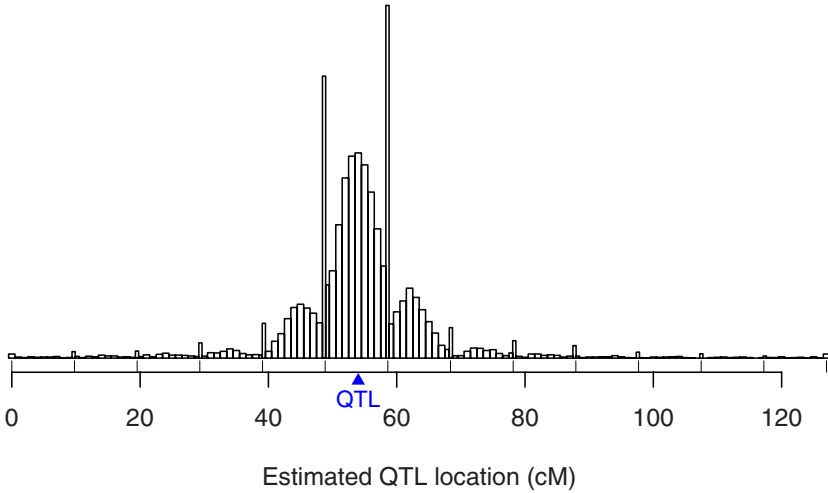


Figure 6.5. Estimated QTL location in 10,000 simulation replicates of an intercross with 250 individuals, with a QTL located at 54 cM (indicated by the blue triangle) and responsible for 8% of the phenotypic variance. The tick marks at the bottom indicate the marker locations (~ 10 cM spacing).

```
> sig <- (loda >= thr)
> sd(est[sig])

[1] 9.07
```

We see that the QTL location is more precisely estimated in the cases in which we had significant evidence for a QTL.

Let us turn to the 1.5-LOD support intervals. We are particularly interested in the estimated coverage: the proportion of simulation replicates in which the left endpoint was ≤ 54 and the right endpoint was ≥ 54 .

```
> mean(lo <= 54 & hi >= 54)

[1] 0.9795

Also interesting is coverage conditional on having significant evidence for a QTL.

> mean(lo[sig] <= 54 & hi[sig] >= 54)

[1] 0.9743
```

In either case, coverage is a bit higher than 95%.

Finally, let us look at the distribution of the width of the 1.5-LOD support interval. We will focus on the cases in which there was significant evidence for a QTL.

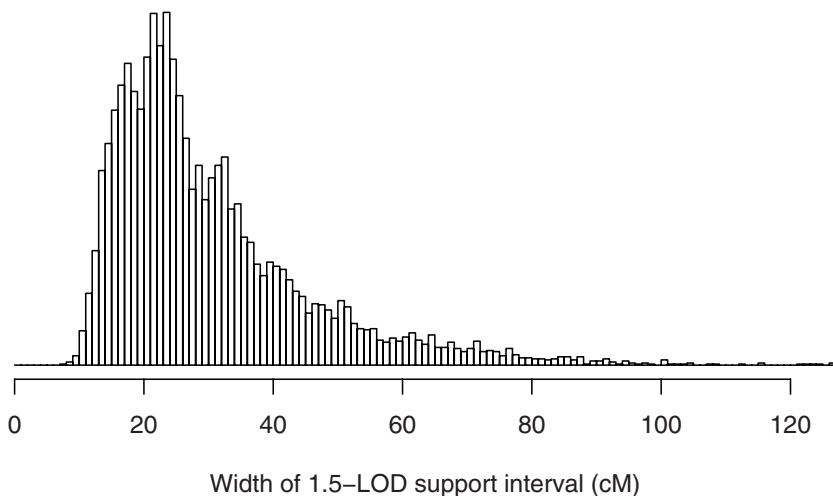


Figure 6.6. Distribution of the width of the 1.5-LOD support interval, conditional on having significant evidence for a QTL, for the case of a single QTL responsible for 8% of the phenotypic variance, in an intercross with 250 individuals and markers at a 10 cM spacing.

```
> hist(hi[sig]-lo[sig], breaks=100,
+       xlab="Width of 1.5-LOD support interval (cM)")
```

The median width of the 1.5-LOD support interval was 26 cM, but it typically varied from 14 to 63 cM long (see Fig. 6.6).

In summary, while the calculations in R/qtlDesign are often accurate, they do rely on a number of approximations. An alternative is to use computer simulations. Simulations can be time consuming and require more detailed knowledge of R, but are quite flexible and so may be used to address more complex design questions.

6.7 Summary

Sound experimental design is the bedrock of good science. A QTL experimenter should follow the general principles of good experimental design. The special structure of QTL experiments offer some additional choices, including the type of cross, the number of progeny to raise, and genotyping and phenotyping strategies. Using the package R/qtlDesign, the experimenter can make choices based on the cost structure of the experiment and the nature of the QTL effects one seeks to identify. The calculations in R/qtlDesign are efficient and convenient but rely on a number of approximations and are not always accurate. Computer simulations, while more cumbersome, can give

more accurate estimates of the power to detect a QTL and the precision of localization of QTL.

6.8 Further reading

The planning of experimental crosses is discussed in Silver (1995) and Lynch and Walsh (1998). Belknap (1998) compared the sample size requirements for recombinant inbred lines (RILs) relative to intercrosses and backcrosses by quantifying the error variance. The advantages of selective genotyping were analyzed by Lander and Botstein (1989) and Darvasi and Soller (1992). Darvasi (1998) gave a comprehensive account of design options for model organisms, including selective genotyping. Selective phenotyping was proposed and analyzed by Jin *et al.* (2004). For a review reflecting recent developments, see Flint *et al.* (2005). Dupuis and Siegmund (1999) discussed genome-wide thresholds for QTL detection and confidence interval construction. Sen *et al.* (2005) framed experimental design through its information content. Also see Sen *et al.* (2009). The web-based program of Purcell *et al.* (2003) performs power calculations for complex trait analysis, although it is not designed for inbred line crosses. Sen *et al.* (2007) is the paper introducing R/qtlDesign.