18

Linkage Mapping And Characterizing QTLs: Inbred Line Crosses

QTL detection experiments do not map a gene, but rather a genetic effect that might consist of many linked genes within $\sim$ 30 cM of each other. Mapping loci in inbred crosses can, therefore, disguise the genetic architecture of a QTL. Flint and Mott (2001)

Version 8 Dec. 2022

Crosses between completely inbred lines offer an ideal setting for detecting and mapping QTLs by marker-trait associations, as all F_1s are genetically identical and show complete linkage disequilibrium for genes differing between lines. A number of crossing designs have been proposed to exploit these features. While usually involving flies and crop plants, QTL line-cross analysis has also been applied to a number of other animal species, especially mice (e.g., Frankel 1995; Flint and Mackay 2009; Aylor et al. 2011; Atamni et al. 2018). More recently, QTL mapping designs have exploited special biological features of the species being used, such as fungal systems (Liti and Louis 2012). A particularly interesting example is the mapping of QTLs for honey bees foraging behavior using designs that exploited the haploid nature of the males (Hunt et al. 1995; 2007; Rüppel et al. 2004).

We note that the heyday of linkage-based QTL mapping (i.e., using a set of individuals from a cross or pedigree), that was at its height when our first edition was published, has now likely passed. Thanks to modern genomics, association mapping (using a dense set of markers in a population sample; Chapter 20) is now feasible for just about any species. This allows for a much larger, and (usually) easier to gather, sample than one requiring that all individuals come from a cross or pedigree, coupled with much finer mapping resolution (kilobase, rather than megabase, scales in many cases). However, linkage-based QTL mapping is still viable for many questions and introduces some of the statistical issues for association mapping that we examine more deeply in Chapter 20. Cross-based QTL mapping (this chapter) and association mapping (Chapter 20) also ***examine different genetic architectures***. The former examines the genetic basis of *differentiation* between the parental lines generating the cross, the latter examines the nature of *segregating variation* for the focal trait in a given population. These two architectures can be largely independent, as the former may be shaped by selection for trait divergence, while the later may be shaped (in part) by selection against deleterious alleles (Chapter 15; WL Chapter 28).

While the vast QTL mapping literature may seem complex to the uninitiated, it is based on a few simple ideas. We start by reviewing these foundational building blocks, first considering basic line-cross designs useful for QTL mapping. The key element from which the formal theory of QTL mapping is constructed is the ***conditional probability of a particular QTL genotype given an observed marker genotype***. These probabilities allow a full development of the two principal methods for QTL detection and estimation: least-squares (LS) linear models (which use differences in marker mean) and maximum likelihood (ML, and, by extension, Bayesian methods) which use information from the entire marker-trait distribution. Both approaches are examined in some detail. We conclude our discussion of statistical issues by considering adjustments for multiple comparisons, examine methods that accommodate multiple linked QTLs, develop more complex designs, discuss the bias in effect size estimates introduced by the Beavis effect (or the winner's curse), and consider the sample sizes required to achieve a given power for detecting, and locating, a QTL. Finally,

we reward the persevering reader with a review of some interesting studies that have used inbred lines for QTL mapping.

FOUNDATIONS OF LINE-CROSS MAPPING

The idea behind using marker information to map and characterize QTLs is quite simple: by crossing two inbred lines, the excess of parental over recombination gametes at linked loci (Chapter 5) generates linkage disequilibrium in their progeny (Example 5.6). This in turn creates associations between alleles at marker loci and alleles at linked segregating QTLs. A number of diverse experimental designs and statistical methodologies have been developed to exploit this information. Our attempt to make this approach more accessible starts with an overview of some of the basic experimental designs and key tools for the analysis of results from these designs.

Basic Crossing Designs

The large number of possible designs can be categorized by the type of line-cross populations used for generating disequilibrium (e.g., F_2 vs. backcross populations), the population in which individuals are genotyped, the population in which phenotypes are scored, and the unit of marker analysis used (e.g., single markers vs. interval mapping). We consider these in turn.

Starting with two completely inbred parental lines, P_1 and P_2, a number of line-cross populations derived from their F_1 can be used for QTL mapping. The **F_2**, or **intercross**, **design** examines marker-trait associations in the progeny from a cross (or selfing) of F_1s, while the **backcross design** examines marker-trait associations in the progeny formed by backcrossing the F_1 to one (or both) of the parental lines. While these are the most widely used designs, other line-cross populations can offer further advantages (and disadvantages). As discussed in Chapter 17, the F_1 can be used to create **recombinant inbred lines** (**RILs**) and **doubled haploid lines** (**DHLs**). While generation of these lines is a laborious process, they allow marker-trait associations to be scored in a completely homozygous background and offer the ability (via replication) of scoring the line phenotype with arbitrary precision and across multiple environments. The F_2 design has an advantage over designs using backcross, RIL, or DHL populations, because it generates three genotypes at each marker locus, which allows the estimation of the degree of dominance associated with detected QTLs. Designs using an F_t population (formed by randomly mating F_1s for $t - 1$ generations) allow for even higher resolution of QTL map positions than do F_2s, albeit at the expense of decreased power of QTL detection (due to lower effective marker density). The properties of such **advanced intercross lines** (AILs) are discussed below.

More complex designs can be considered wherein individuals are genotyped in one population, while trait values are scored in a future population derived from the genotyped individuals. Fisch et al. (1996) present a general treatment for such designs. One example is the **$F_{2:3}$ design**, wherein F_2 individuals are genotyped and then selfed (Zhang and Xu 2004). The trait value associated with a genotyped individual is estimated by the mean value of the resulting F_3 family. Scoring the phenotype as the mean of several individuals (as opposed to measurement of a single individual) can offer increased power over a standard F_2 design by reducing the sampling variance.

Designs combining information from multiple crosses (Chapter 11) are expected to be more powerful than those involving a single cross, and a number of such designs have been proposed, in particular, for estimating components associated with dominance and heterosis. One can augment these designs with markers to use them for QTL mapping. Examples include diallel designs whose basic structure is examined in Chapters 23 and 25 (Rebaï and Goffinet 1993; Rebaï et al. 1994a), and Comstock and Robinson's (1952) classic Design III wherein the F_2 from two inbred lines is backcrossed to both parental lines (Cockerham and Zeng 1996). We consider such multiple-line designs later in the chapter. Finally, several

workers have considered designs involving crosses between lines that are not completely inbred, such as a cross of an outbred line to a completely inbred tester (Beckmann and Soller 1988; Dudley 1992; Haley et al. 1994).

Experimental designs are also classified by the unit of marker analysis chosen by the investigator. Marker-trait associations can be assessed using one-, two-, or multiple-locus marker genotypes. Under a **single-marker analysis**, the distribution of trait values is examined separately for each marker locus. Each marker-trait association test is performed independent of information from all other markers, so that a chromosome with n markers offers n separate single-marker tests. As discussed below, single-marker analysis is generally a good choice when the goal is simple *detection* of a QTL linked to a marker, rather than *estimation* of its position and effects. Under **interval mapping** (or **flanking-marker analysis**), a separate analysis is performed for each *pair* of adjacent marker loci. Here one computes the probability of a putative QTL at a series of discrete locations between the two markers defining the target interval. The use of such two-locus marker genotypes results in $n - 1$ separate tests of marker-trait associations for a chromosome with n markers (one for each marker interval). Interval mapping offers a further increase in power of detection (albeit usually a slight one) and more precise estimates of QTL effects and position. Both single-marker and interval mapping approaches are biased when multiple QTLs are linked to the marker/interval being considered.

Methods simultaneously using three or more marker loci attempt to reduce or remove such bias. **Composite interval mapping** (Jansen 1993b, 1994b; Zeng 1993, 1994; Jansen and Stam 1994) considers a marker interval plus a few other well-chosen markers in each analysis, so that (as above) $n - 1$ tests for interval-trait associations are performed on a chromosome with n markers. **Multiple interval mapping** extends this approach by simultaneously fitting multiple intervals (Kao et al. 1999; Zeng et al. 1999). Finally, **multipoint mapping** considers all of the linked markers on a chromosome simultaneously, resulting in a single analysis for each chromosome (Kearsey and Hyne 1994; Hyne and Kearsey 1995; Wu and Li 1994, 1996a, 1996b; Charmet et al. 1998).

Conditional Probabilities of QTL Genotypes

The basic element upon which the formal theory of QTL mapping is built is the conditional probability that the QTL genotype is Q_k, given the observed (possibly multilocus) marker genotype is M_j. From the definition of a conditional probability (Equation 3.3a), this is

$$\Pr(Q_k \mid M_j) = \frac{\Pr(Q_k M_j)}{\Pr(M_j)} \tag{18.1}$$

The joint and marginal probabilities, $\Pr(Q_k M_j)$ and $\Pr(M_j)$, are functions of the experimental design and the linkage map (the position of the putative QTLs with respect to the marker loci). Computing these probabilities is a relatively simple matter of bookkeeping (see Example 18.1), but can get rather tedious as the number of markers and/or QTLs under consideration increases.

When computing joint probabilities involving more than two loci, one must also account for recombinational interference between loci (Chapter 17). Consider a single QTL flanked by two markers, M_1 and M_2. The gamete frequencies depend on three parameters: the recombination frequency c_{12} between markers, the recombination frequency c_1 between marker M_1 and the QTL, and the recombination frequency c_2 between the QTL and marker M_2. Under the assumption of no interference, $c_{12} = c_1 + c_2 - 2c_1 c_2$, while $c_{12} = c_1 + c_2$ under complete interference (Chapter 17). When c_{12} is small, gamete frequencies are essentially identical under either interference assumption. Typically, c_{12} is assumed known, leaving two unknown recombination parameters (c_1 and c_2) under general assumptions about interference. In either case, there is only one parameter to estimate, as assuming complete interference $c_2 = c_{12} - c_1$, and assuming no interference $c_2 = (c_{12} - c_1)/(1 - 2c_1)$. Note that $1/(1 - 2c_1)$ has values of 1.11, 1.25, 1.67, and 2 for c_1 values of 0.05, 0.1, 0.2, and 0.25, respectively. Hence, for flanking-marker analysis, we restrict attention to the single recombination parameter c_1, the distance from marker locus M_1 to the QTL. When considering

analysis of single-marker loci, for notational ease we drop the subscript, using c in place of c_1.

Conditional probabilities involving more than three linked loci are generally dealt with by first assuming an appropriate mapping function on which distances are additive (Chapter 17), and then translating these distances into recombination frequencies. When a large number of markers is considered, missing marker information can become a problem. Many individuals can be left with incomplete multilocus marker genotypes, excluding them from further analysis. Martínez and Curnow (1992) show how information from linked markers can be used to estimate the genotype at missing or incomplete (i.e., dominant) markers. Predicting the values of missing marker genotypes, **imputation**, is an important concept that arises in association mapping, and will be considered in detail in Chapter 20.

Example 18.1.   Consider a single-marker analysis using the F_2 formed by crossing two inbred lines, $MMQQ \times mmqq$. If the recombination frequency between the marker locus and the QTL is c, the expected F_1 gamete frequencies (Chapter 5) are

$$\Pr(MQ) = \Pr(mq) = (1-c)/2, \qquad \Pr(Mq) = \Pr(mQ) = c/2$$

The probability that an F_2 individual is $MMQQ$ is $\Pr(MQ) \cdot \Pr(MQ) = [(1-c)/2]^2$. Likewise, $2 \cdot \Pr(MQ) \cdot \Pr(mQ) = 2(c/2)[(1-c)/2]$ is the probability of an $MmQQ$ individual, and so on. Because the probabilities of the marker genotypes MM, Mm, and mm are $1/4$, $1/2$, and $1/4$, Equation 18.1 gives the F_2 conditional probabilities as

$$\Pr(QQ \mid MM) = (1-c)^2, \quad \Pr(Qq \mid MM) = 2c(1-c), \quad \Pr(qq \mid MM) = c^2$$

$$\Pr(QQ \mid Mm) = c(1-c), \ \Pr(Qq \mid Mm) = (1-c)^2 + c^2, \ \Pr(qq \mid Mm) = c(1-c)$$

$$\Pr(QQ \mid mm) = c^2, \quad \Pr(Qq \mid mm) = 2c(1-c), \quad \Pr(qq \mid mm) = (1-c)^2$$

This same logic extends to multiple marker loci. Suppose the QTL is flanked by two scored markers, and consider the F_2 in a cross of lines fixed for M_1QM_2 and m_1qm_2. What are the conditional probabilities of the QTL genotypes when the marker genotype is $M_1M_1M_2M_2$? Because all F_1s are M_1QM_2/m_1qm_2, under the assumptions of no interference, the frequency of F_1 gametes involving M_1M_2 are

$$\Pr(M_1QM_2) = (1-c_1)(1-c_2)/2, \qquad \Pr(M_1qM_2) = c_1 c_2/2$$

giving expected frequencies in the F_2 of $M_1M_1M_2M_2$ offspring as

$$\Pr(M_1QM_2/M_1QM_2) = [(1-c_1)(1-c_2)/2]^2$$
$$\Pr(M_1QM_2/M_1qM_2) = 2[(1-c_1)(1-c_2)/2][c_1 c_2/2]$$
$$\Pr(M_1qM_2/M_1qM_2) = (c_1 c_2/2)^2$$

where $c_2 = (c_{12} - c_1)/(1 - 2c_1)$. The overall frequency of $M_1M_1M_2M_2$ individuals, $\Pr(M_1M_1M_2M_2)$, is the sum of the three above terms, or $(1 - c_{12})^2/4$. Substituting into Equation 18.1 gives

$$\Pr(QQ \mid M_1M_1M_2M_2) = \frac{(1-c_1)^2(1-c_2)^2}{(1-c_{12})^2}$$

$$\Pr(Qq \mid M_1M_1M_2M_2) = \frac{2c_1c_2(1-c_1)(1-c_2)}{(1-c_{12})^2} \qquad (18.2)$$

$$\Pr(qq \mid M_1M_1M_2M_2) = \frac{c_1^2 c_2^2}{(1-c_{12})^2}$$

Conditional probabilities for other marker genotypes are computed in a similar fashion. Because $c_1 c_2 \leq (c_{12}/2)^2$ is usually very small if c_{12} is moderate to small, essentially all $M_1 M_1 M_2 M_2$ individuals are QQ. For example, assuming $c_1 = c_2 = c_{12}/2$ (the worst case), the conditional probabilities of an $M_1 M_1 M_2 M_2$ individual being QQ are 0.96, 0.98, and 0.99 for $c_1 = c_2 = 0.25, 0.2$, and 0.1.

We now consider the conditional probabilities for other single-marker line cross designs, starting with backcrosses. For a B_1 population, where the F_1 is backcrossed to P_1 (with genotype $MMQQ$), one parental gamete is always MQ. Hence, for a single-marker analysis, there are only two marker genotypes, MM and Mm. Using the frequencies for the four possible gametes (Example 18.1) of the F_1 parent gives the following conditional probabilities

$$\Pr(QQ\,|\,MM) = 1 - c, \qquad \Pr(Qq\,|\,MM) = c$$
$$\Pr(QQ\,|\,Mm) = c, \qquad\qquad \Pr(Qq\,|\,Mm) = 1 - c \qquad (18.3a)$$

Likewise, when backcrossing to the P_2 ($mmqq$), the two possible single-locus marker genotypes are Mm and mm, and the conditional probabilities become

$$\Pr(qq\,|\,mm) = 1 - c, \qquad \Pr(Qq\,|\,mm) = c$$
$$\Pr(qq\,|\,Mm) = c, \qquad\qquad \Pr(Qq\,|\,Mm) = 1 - c \qquad (18.3b)$$

Historically, backcross designs were often used when only dominant markers (e. g., RAPDs; Chapter 8) were available. However, the backcross design can be driven by the biology of the problem. A good example is genetics of speciation (Chapter 17), wherein backcrosses of a hybrid (F_1) to one species can be viable, while the reciprocal backcross and the hybrid-hybrid ($F_1 \times F_1$) crosses are not.

For designs involving more than one generation of recombination, the single generation recombination frequency c is simply replaced by a corrected frequency $\widetilde{c}$ that is a function of the particular design. We consider three such designs: advanced intercross lines (AILs), recombinant inbred lines (RILs), and double-haploid lines (DHLs).

Advanced intercross lines (Darvasi and Soller 1995; Liu et al. 1996) are obtained by crossing two inbred lines, but instead of stopping at the F_2, random mating proceeds for t generations, generating an F_t. In this case, unlike the strategy used to create RILs (Chapter 17), inbreeding is avoided by keeping the breeding population size large. As the result of the multiple rounds of recombination, markers in an F_t individual show an expansion of the genetic map relative to an F_2, with the expected frequency of a recombinant gamete in the F_t for a pair of loci at recombination fraction c (following from Equation 5.16c) being

$$\widetilde{c} = \frac{1 - (1-c)^{t-2}(1-2c)}{2} \simeq \frac{t}{2}\,c \qquad (18.4)$$

where the approximation holds for $ct \ll 1$ (Darvasi and Soller 1995; Liu et al. 1996). For example, if the marker-QTL recombination frequency is $c = 0.01$, only 1% of the F_2 gametes are recombinant (Mq, mQ), but this increases to 2.5% in an F_5 and 9.1% in an F_{20}. The conditional genotype probabilities for an F_t AIL are given by the F_2-design expressions in Example 18.1, with $\widetilde{c}$ substituted for c.

Recombinant inbred lines (RILs) also involve several generations of recombination, but here genotypes are eventually fixed by inbreeding (Chapter 12). Starting with a MQ/mq F_1 parent, there are only four possible genotypes in the resulting RILs: $MMQQ$, $MMqq$, $mmQQ$, and $mmqq$. The frequency of recombinant gametes (Mq, mQ) in RILs approaches a limiting value of $\widetilde{c} = 2c/(1 + 2c)$ for selfed lines and $\widetilde{c} = 4c/(1 + 6c)$ for lines formed by brother-sister mating (Haldane and Waddington 1931). Thus, the expected frequencies of genotypes in RILs are

Line genotype	Frequency
$MMQQ$, $mmqq$	$(1 - \widetilde{c})/2$
$MMqq$, $mmQQ$	$\widetilde{c}/2$

While doubled-haploid lines (DHLs) also have only these four genotypes, they are formed by a single generation of meiosis (if based on using F_1s), so that $\widetilde{c} = c$. If based on F_2s, Equation 18.4 gives $\widetilde{c} = 3c/2 + c^2$. Hence, among either RILs or DHLs, the conditional QTL probabilities are

$$\Pr(QQ \,|\, MM) = 1 - \widetilde{c}, \qquad \Pr(qq \,|\, MM) = \widetilde{c}$$
$$\Pr(QQ \,|\, mm) = \widetilde{c}, \qquad \Pr(qq \,|\, mm) = 1 - \widetilde{c} \tag{18.5a}$$

where

$$\widetilde{c} = \begin{cases} c, \; 1.5c + c^2 & \text{for } F_1 \text{ and } F_2 \text{ DHL, respectively} \\ 2c/(1 + 2c) & \text{for RILs formed by selfing} \\ 4c/(1 + 6c) & \text{for RILs formed by brother-sister mating} \end{cases} \tag{18.5b}$$

Values of $\widetilde{c}$ for intermediate generations of inbreeding are given by Teuscher et al. (2006).

Expected Marker-class Means

With these conditional probabilities in hand, the expected trait values for the various marker genotypes follow immediately. Suppose there are N QTL genotypes, $Q_1, \cdots, Q_N$, where the mean of the kth QTL genotype is μ_{Q_k}. The mean value for marker genotype M_j is just

$$\mu_{M_j} = \sum_{k=1}^{N} \mu_{Q_k} \Pr(\, Q_k \,|\, M_j \,) \tag{18.6}$$

The **QTL effects** enter through the μ_{Q_k}, while the **QTL positions** enter through the conditional probabilities $\Pr(\, Q_k \,|\, M_j \,)$. Equation 18.6 is completely general, allowing for multilocus marker genotypes and multiple QTLs.

Example 18.2.   Consider the single-marker F_2 design with a single QTL linked (at recombination frequency c) to the marker. Denote the QTL genotypic values by

$$\mu_{QQ} = \mu + 2a, \quad \mu_{Qq} = \mu + a(1 + k), \quad \text{and} \quad \mu_{qq} = \mu$$

where a measures the additive value and k the degree of dominance ($d = ak$; Chapter 4). Applying the conditional probabilities developed in Example 18.1 to Equation 18.6, the mean values for the marker genotypes become

$$\mu_{MM} = \mu + 2a(1 - c)^2 + 2c(1 - c)(1 + k)a$$
$$\mu_{Mm} = \mu + 2ac(1 - c) + [1 - 2c(1 - c)](1 + k)a$$
$$\mu_{mm} = \mu + 2ac^2 + 2c(1 - c)(1 + k)a$$

If the marker and QTL are unlinked ($c = 1/2$), all marker genotypes have the same mean, $\mu + a[1 + (k/2)]$. Rearranging these equations gives

$$(\mu_{MM} - \mu_{mm})/2 = a(1 - 2c) = a^* \tag{18.7a}$$

$$\frac{\mu_{Mm} - (\mu_{MM} + \mu_{mm})/2}{(\mu_{MM} - \mu_{mm})/2} = k(1 - 2c) = k^* \tag{18.7b}$$

Hence, one strategy for detecting QTLs is to test for ***significant differences between the mean trait values associated with different marker genotypes***. This is the basis for QTL detection via regression or ANOVA, which we generically refer to as **least-squares (LS) linear model** approaches (Chapter 10).

This example shows that while contrasts of single-marker means can be used to estimate both a^* and k^*, these underestimate the magnitude of a and k by the (unknown) fraction $1 - 2c$. If the marker and QTL are tightly linked, this error is small, but it increases rather dramatically as c approaches $1/2$. A small difference between marker-homozygote means is thus compatible with either a tightly linked QTL of small effect or a loosely linked QTL of large effect. As we will show shortly, when multilocus marker genotypes are considered, the use of appropriate combinations of marker means allows for separate estimates of QTL effects (a, k) and position (c).

If there are N QTLs linked to the marker, the ith of which is at recombination frequency c_i from the marker and has associated additive and dominance effects a_i and k_i, then (from Edwards et al. 1987),

$$(\mu_{MM} - \mu_{mm})/2 = \sum_{i=1}^{N} a_i^* \qquad (18.8a)$$

$$\frac{\mu_{Mm} - (\mu_{MM} + \mu_{mm})/2}{(\mu_{MM} - \mu_{mm})/2} = \sum_{i=1}^{N} a_i^* k_i^* \bigg/ \sum_{i=1}^{N} a_i^* \qquad (18.8b)$$

where $a_i^* = a_i(1 - 2c_i)$ and $k_i^* = k_i(1 - 2c_i)$. If some of the linked QTLs have effects of opposite sign, some cancellation occurs, reducing the marker-trait association. Moreover, with multiple linked QTLs, the degrees of dominance (k_i) are confounded with the homozygous effects (a_i).

Marker-class means for other designs follow by applying the appropriate conditional probabilities to Equation 18.6. For example, for the B_1 (= $F_1 \times P_1$) design, from Equation 18.3a,

$$\mu_{MM} - \mu_{Mm} = (\mu_{QQ} - \mu_{Qq})(1 - 2c) = a(1 - k)(1 - 2c) \qquad (18.9a)$$

Thus, under a backcross design the scaled QTL effects are influenced strongly by the (unknown) degree of dominance k. If Q is completely dominant to q, $k = 1$, and there is no marker-QTL effect. Conversely, if q is dominant to Q, $k = -1$ and the scaled effect becomes $2a(1-2c)$, which is the same as under an F_2 design. Recalling Equation 18.3b, the reciprocal backcross ($B_2 = F_1 \times P_2$) yields a similar expression,

$$\mu_{Mm} - \mu_{mm} = (\mu_{Qq} - \mu_{qq})(1 - 2c) = a(1 + k)(1 - 2c) \qquad (18.9b)$$

Note that the ratio of Equation 18.9a to 18.9b gives $(1 - k)/(1 + k)$, so that (provided only a single QTL is linked to the marker) an estimate of k can be obtained if one has access to *both* backcross populations (B_1 and B_2).

The expressions developed in Example 18.2 for F_2 analysis hold for an F_t population, provided $\widetilde{c}$ (given by Equation 18.4) replaces c. For example, $\mu_{MM} - \mu_{mm} = 2a(1 - 2\widetilde{c})$, and so forth. Because $(1 - 2c) > (1 - 2\widetilde{c})$, AILs have smaller differences between marker means, and hence reduced power of QTL detection, relative to the F_2 design. Despite this, Darvasi and Soller (1995) advocate the use of AILs for fine-mapping of QTLs, as the expansion of the genetic map (relative to the F_2) offers a higher precision (improved **mapping resolution**) of estimates of QTL position.

For RILs and DHLs, the recombination parameter $\widetilde{c}$ is given by Equation 18.5b, and from Equations 18.5a and 18.6 it follows that

$$\mu_{MM} = \mu_{QQ}(1 - \widetilde{c}) + \mu_{qq}\widetilde{c} \quad \text{and} \quad \mu_{mm} = \mu_{QQ}\widetilde{c} + \mu_{qq}(1 - \widetilde{c}) \qquad (18.10a)$$

giving

$$\frac{\mu_{MM} - \mu_{mm}}{2} = a\,(1 - 2\,\widetilde{c}) = a^* \tag{18.10b}$$

again providing an estimate of a composite parameter of the QTL effect (a) and position (c). Because $\widetilde{c}$ is smallest in DHLs (see Equation 18.5b), the largest marker effect (and greatest power for QTL detection) occurs in this type of line, followed by selfed RILs, and finally by sib-mated RILs. Conversely, the map expansion from recombination during line formation that reduces power increases the precision of QTL position estimates. Hence, confidence intervals are smallest for sib-mated RILs and largest for DHLs.

Finally, note that by considering two-locus (rather than single-locus) marker means, separate estimates of QTL effect and position can be obtained. Taking the genotype at two adjacent marker loci (M_1/m_1 and M_2/m_2) as the unit of analysis, consider the difference between the contrasting double homozygotes in an F_2. If the markers flank a QTL, then under the assumption of no interference, Equation 18.2 (and its analog for $m_1m_1m_2m_2$ probabilities) implies

$$\frac{\mu_{M_1M_1M_2M_2} - \mu_{m_1m_1m_2m_2}}{2} = a\left(\frac{1 - c_1 - c_2}{1 - c_1 - c_2 + 2c_1\,c_2}\right)$$
$$\simeq a\,(1 - 2c_1\,c_2) \tag{18.11a}$$

where c_1 is the M_1 – QTL recombination frequency. Equation 18.11a is essentially equal to a when the distance between flanking markers $c_{12} \leq 0.20$, as here $(1 - 2c_1\,c_2) \geq 0.98$. Thus, recalling from Equation 18.7a that $\mu_{M_1M_1} - \mu_{m_1m_1} = 2a(1 - 2c_1)$, we can obtain estimates of the recombination frequencies by substituting Equation 18.11a for a and rearranging to give

$$c_1 = \frac{1}{2}\left(1 - \frac{\mu_{M_1M_1} - \mu_{m_1m_1}}{2a}\right)$$

$$\simeq \frac{1}{2}\left(1 - \frac{\mu_{M_1M_1} - \mu_{m_1m_1}}{\mu_{M_1M_1M_2M_2} - \mu_{m_1m_1m_2m_2}}\right) \tag{18.11b}$$

Using the same logic (and bookkeeping), estimates for other flanking-marker designs were developed by Knapp et al. (1990), Knapp and Bridges (1990), and Knapp (1991).

Marker Variances and Higher-order Moments

In addition to differing in trait *means*, marker genotypes can also differ in *higher-order moments* as well, such as the variance or skew. Such differences can arise in segregating populations (e.g., F_2s) simply as a consequence of changes in the mixture proportions of QTL genotypes among different marker genotypes (Example 18.1). Such differences are not uncommon. For example, in a cross of tomato species, Weller et al. (1988) found significantly different variances for 28% (40 of 180) of the possible marker-trait associations, while 17% showed significant differences in skewness.

Several workers have suggested the use of these higher-order moments for detection of a linked QTL and estimation of its effects (Zhuchenko et al. 1978, 1979; Korol et al. 1981, 1983, 1987; Ginzburg 1983; Asins and Carbonell 1988; Zhang et al. 1992). One difficulty with this approach is that variances (and higher moments) are estimated with far less precision than means, reducing both the power of detection and the accuracy of estimates. Another complication is that not all designs are capable of revealing significant changes in higher moments (e.g., Asins and Carbonell 1988).

RILs and DHLs provide one case where functions of higher-order moments (e.g., a correlation coefficient) can be exploited. Here, single-locus marker information can be used to estimate the recombination frequency c (Hu et al. 1995). As shown in Table 18.1, coding the alternative marker homozygotes as $x = \pm 1$, the expected marker-trait correlation becomes

$$\rho = \frac{\sigma(z, x)}{\sigma(x)\,\sigma(z)} = \frac{a(1 - 2\widetilde{c})}{\sqrt{a^2 + \sigma_e^2}} = \frac{1 - 2\widetilde{c}}{\sqrt{1 + C^2}} \tag{18.12a}$$

Table 18.1 The expected correlation between marker genotype (coded as $x = 1$ for *MM*, $x = -1$ for *mm*) and phenotypic value z can be used to estimate c. We assume that the difference in QTL means is $E(z_{QQ}) - E(z_{qq}) = 2a$ and that the phenotypic distributions conditioned on the QTL genotypes have common (within-line) variance σ_e^2. Because $\sigma(x, z)$ and $\sigma^2(z)$ are unchanged by a change in the mean of z (Chapter 3), we can arbitrarily set $E(z_{QQ}) = -E(z_{qq}) = a$. Coded this way, $E(x) = E(z) = 0$, simplifying calculation of $\sigma(x, z)$ and $\sigma^2(z)$.

Genotype	Freq.	x	$E[z]$
MMQQ	$(1 - \widetilde{c})/2$	1	$\mu_{QQ} = a$
MMqq	$\widetilde{c}/2$	1	$\mu_{qq} = -a$
mmQQ	$\widetilde{c}/2$	-1	$\mu_{QQ} = a$
mmqq	$(1 - \widetilde{c})/2$	-1	$\mu_{qq} = -a$

$$\sigma(x, z) = E(x \cdot z) = a(1 - 2\widetilde{c}), \qquad \sigma^2(x) = E(x^2) = 1$$

$$\sigma^2(z) = E(z^2) = (1/2)[E(z_{QQ})^2 + \sigma_e^2] + (1/2)[E(z_{qq})^2 + \sigma_e^2] = a^2 + \sigma_e^2$$

where $C = \sigma_e/a$, with a being the QTL effect and σ_e^2 being the within-line variance (see Table 18.1). (The C term was neglected by Hu et al. 1995.) Equation 18.12a assumes no segregation distortion (both RIL homozygotes have equal frequency). Rearranging and letting r be an estimate of ρ suggests the estimator

$$\widetilde{c} = \frac{1 - r\sqrt{1 + C^2}}{2} \leq \frac{1 - r}{2} \tag{18.12b}$$

While the value of C is unknown, by ignoring it one can obtain an upwardly biased estimate of c by first taking $\widetilde{c} = (1 - r)/2$ and then using Equation 18.5b to translate this value of $\widetilde{c}$ into c. Rearranging Equation 18.10b,

$$a = \frac{\mu_{MM} - \mu_{mm}}{2(1 - 2\widetilde{c})} \tag{18.12c}$$

which, upon using $0 \leq \widetilde{c} \leq (1 - r)/2$, gives

$$\frac{\mu_{MM} - \mu_{mm}}{2} \leq a \leq \frac{\mu_{MM} - \mu_{mm}}{2r} \tag{18.12d}$$

Hence, the use of both the observed correlation r and the difference in marker means ($\overline{z}_{MM} - \overline{z}_{mm}$) allows the estimation of upper bounds for both c and a.

Variance QTLs (vQTLs)

While differences in trait variance across marker genotypes could simply reflects differences on the weights on underlying, but equally variable, distributions, they also could have a much more interesting cause: the QTL genotypes themselves may actually have different variances. This is commonly seen among inbred lines (WL Chapter 17), where some marker genotypes are more variable than others. Denoted as **vQTLs (variance QTLs)** by Rönnegård and Valdar (2011) and **veQTLS (variance in expression-level traits)** by Huang et al. (2015), such QTLs denote sites where the trait variance differs over marker genotypes. In some settngs, vQTLs can be of more interest than QTLs impacting the mean (Weller and Wyler 1992). For example, a reduction in the variance of flowering time shortens the harvesting window, and by reducing costs this may be more significant than changing mean harvesting time per se. While candidate vQTLs could be tagged simply using among-marker differences in variance (e.g., Edwards et al. 1987), a number of more formal approaches to specifically map vQTLs have been proposed (Ordas et al. 2008; Paré et al. 2010; Struchlain et al. 2010;

Visscher and Posthuma 2010; Jimenez-Gomes et al. 2011; Rönnegård and Valdar 2011, 2012; Hothorn et al. 2012; Shen et al. 2012).

Differences in trait variances among QTL genotypes can result from two different, but not mutually exclusive, causes (WL Chapter 17). They could reflect differences in *sensitivity to environmental conditions* ($G \times E$), such that the environmental variance of, say, QQ is 20, while that for qq is 10. Alternatively, they may reflect **sensitivity to genetic backgrounds** (epistasis). Indeed, Paré et al. (2010) and Deng and Paré (2011) suggested that using variance heterogeneity across markers as a preliminary scan for potentially epistatic loci. Distinguishing between genetic vs. environmental sensitivity as causes for vQTLs can be done using panels of RILs (e.g., Fraser and Schadt 2010; see WL Example 17.1). Within a given RIL, one can use replication to estimate the within-line variance very accuracy (sensitivity to environmental conditions). Conversely, one can measure the among-line variance in RILs means for a given marker genotype, which is a direct measure of sensitivity to genetic background.

Genome-wide Significance Level with Multiple Tests

A final statistical issue to address is the problem of the proper significance level for an *entire mapping experiment*. This is the issue of **multiple comparisons**, which is examined in much greater detail in Appendix 6. The basic concern is as follows. Suppose that 100 markers are tested for departure from the null (no linkage to a QTL), each using an α value of 0.05 (the probability of a false positive: incorrectly declaring linkage when the marker is unlinked to a QTL). If all 100 markers were unlinked to QTLs, we would still (on average) declare $0.05 \cdot 100 = 5$ as being significant (linked to a QTL), resulting in (on avenge) 5 **false positives**. The level of error control, α, for any *single* test is referred to as the **comparison-wise error rate** (**CWER**). In contrast, the **family-wide error rate** (**FWER**), γ, is the probability of one (or more) false positives over the *entire collection of tests*. In the mapping literature, the term **genome-wide error rate** (**GWER**) is often used in place of FWER.

If n independent tests, each with significance level α, are conducted, the probability γ that at least one test shows a false positive (the FWER) is

$$\gamma = 1 - (1 - \alpha)^n \tag{18.13a}$$

For our values of $\alpha = 0.05$ and $n = 100$, $\gamma = 0.9940$. Setting $\alpha = 0.01$ for each individual test, the probability of at least one false positive in 25 tests is 0.22, which increases to 0.633 for 100 tests, and is essentially one for 500 tests. The latter number of tests is not uncommon in QTL mapping studies (n is several orders of magnitude larger in association studies; Chapter 20). Hence, unless we use a very stringent significance value for each test, we run a very high risk of declaring false associations. This tighter stringency for each test results in a higher Type II error rate (failing to declare a linked marker as being significant), resulting in a tradeoff between power of detection and control of the experiment-wide error rate (Appendix 6).

Suppose we wish to achieve an overall significance level γ for the *entire* experiment. With n *independent* tests, the standard **Bonferroni correction** for multiple comparisons, derived by rearranging Equation 18.13a, states that an overall significance level γ requires that each individual test be based on a significance level of

$$\alpha = 1 - (1 - \gamma)^{1/n} \simeq \frac{\gamma}{n} \tag{18.13b}$$

This standard Bonferroni correction is too strict for two reasons. First, sequential Bonferroni corrections (Holm's, Simes-Hochberg, Hommel's; Appendix 6), which follow from procedures based on the ordered p values of the tests (Example A6.4), still control the FWER, but with less stringent control on the CWER. The second issue is that Equation 18.13b assumes *independent* tests. This is appropriate for tests using unlinked markers (such as those on different chromosomes), but is problematic for markers on the same chomosome. Linked markers result in tests generally being positively correlated, which has the effect that the

effective number of tests, n_e, is *less* than n. Hence, one correction is to adjust for the effective number of tests (usually done separately for each chromosome). Appendix 6 reviews n_e estimates based on the eigenvalues of the correlation matrix of the tests, such as the methods of Cheverud (2000) and Li and Ji (2005) (Equations A6.5a and A6.5b).

A more robust approach for obtaining overall significance levels utilizes resampling procedures such as **permutation tests**, wherein the original analysis is replicated many times on data sets generated by appropriate reshuffling of the original data to simulate draws from the null (Churchill and Doerge 1994; Doerge and Churchill 1996). Let $\mathbf{m}_i$ and z_i denote the vector of marker genotypes and the trait value(s), respectively, for individual i. The original data is of the form $(z_1, \mathbf{m}_1), (z_2, \mathbf{m}_2), \cdots, (z_k, \mathbf{m}_k)$ and one uses some method to search for associations between z_i and one (or more) of the elements of $\mathbf{m}_i$. Under the null, there is no association between z and $\mathbf{m}$. A resampled data set would be (say) of the form $(z_{15}, \mathbf{m}_1), (z_{105}, \mathbf{m}_2), \cdots, (z_{27}, \mathbf{m}_k)$, randomizing z over different draws of $\mathbf{m}$, as would be expected under the null. The test statistic is then computed on this new sample, and this procedure is repeated many times, generating an empirical distribution of the test under the null hypothesis of no marker-trait associations. For example, if the highest test score over all markers in the randomized samples exceeds 4 in only 2.5% of the samples, any marker with a score above 4 is significant at the a FWER of $\gamma = 0.025$. Note that one can obtain significance levels on either a per chromosome (using the distribution of randomized scores for that chromosome) or an entire genome basis. Churchill and Doerge suggest that 1,000 resamplings is sufficient for a significance level of 5%, but that 10,000 or more resamplings may be required to generate a stable critical value for the 1% level. By keeping the marker information ($\mathbf{m}_i$) for each individual together, this approach nicely accounts for missing markers, differences in marker densities, correlations among markers, and any nonrandom segregation of marker alleles. The latter is not uncommon in crosses between widely divergent lines.

A key assumption of permutation tests is that the *permuted sampling units are exchangeable under the null* (Welch 1990). Consider the case of selective genotyping (Chapter 17), wherein one phenotypes a sample of individuals, but only genotypes a selected subset of these. In this setting, threshold values must be computed using **stratified permutation tests** (Manichaikul et al. 2007), where only those phenotypes with associated genotypes are shuffled over the genotypes. Similarly, with family-based QTL mapping (Chapter 19), care use to taken to find the appropriate unit for shuffling.

An alternative strategy for error control under multiple comparisons is the idea of the **false discovery rate** (**FDR**). As detailed in Appendix 6, instead of attempting to limit the *number* of false positives over the *entire set* of tested markers (FWER), one limits the *fraction* of false positives among *the set of tests that are declared to be significant*. For example, setting an FDR control of five percent implies that one expects only 5% of the *significant tests* to be false positives. Conversely, setting an FWER of 5% says that there is only a 5% chance that the *entire* experiment would have one (or more) false positives. The FWER results in more stringent control (and thus less power) than FDR. The FWER setting assumes that essentially none of the tests are true positives, while the FDR framework assumes that some nontrivial fraction of the tests are indeed true positives. Appendix 6 examines the subtle, but quite important, differences between these approaches in detail.

FDR is often used in association studies (Chapter 20), and has been proposed for use in linkage-based QTL mapping (Weller et al. 1998; Lee et al. 2002; Sabatti et al. 2003; Bernardo 2004; Benjamini and Yekutieli 2005). However, Chen and Storey (2006) noted the standard FDR method is not appropriate when a *set of linked markers* are each tested, as signal from the same underlying QTL generates multiple discoveries that are taken as distinct. They instead suggest an FDR-related approach that builds on the Churchill and Doerge resampling approach, essentially looking at the distribution of the *second* highest genome-wide scores (see Appendix 6 for details). We will return to multiple comparisons adjustments in Chapter 20, as association mapping has major differences relative to linkage-based mapping. It has a far higher number of markers, many more linked markers, but a smaller *fraction* of correlated

markers among these tested (most markers on the same chromosome are uncorrelated in the GWAS setting, as they are essentially in linkage equilibrium; Chapter 5). Further, association mapping can be computationally intensive, limiting the utility of resampling.

Example 18.3   When the marker density on a chromosome is low, Bonferroni corrections are reasonable. Lander and Botstein (1989) called this the **sparse-map** case, where consecutive markers are well-separated. At the other extreme is the **dense-map** case, where the spacing between consecutive markers approaches zero. Such settings arise when each point on a chromsome generates a test statistic. Lander and Botstein (1989; Lander and Schork 1994; Lander and Kruglyak 1995) approximated this process by using an Orenstein-Uhlenbeck diffusion (LW Appendix 1). Here the value of the test statistics under the null as one moves along a chromosome is approximated by Brownian motion (generating correlations among adjacent tests) with a restoring force that constrains the variance of the process. Their model assumes the Haldane mapping function (Chapter 17) and uniform recombination rates throughout the genome to model the covariance function between marker genotypes (and, hence, between tests based on these genotypes). Using this framework, results from random-walk theory give approximate expressions for the probability that the walk exceeds some value T over some defined duration (here, the genome length).

In particular, if X is the linkage statistics (a LR value or t or more general score test), then the FWER, $\gamma_T = \Pr(X > T$ somewhere in the genome), can be approximately relatived to the point-test value (CWER), $\alpha_T = \Pr(X > T$ at a single site), by

$$\gamma_T \simeq [C + 2\rho G h(T)]\alpha_T \qquad (18.13c)$$

where C is the number of chromosomes, G the genome length (in Morgans), ρ is a measure of the crossover rate in the design (values tabulated in Lander and Kruglyak 1995), and $h(T)$ depends on the test (Lander and Schork 1994). Note by comparison to Equation 18.13b that $[C + 2\rho G h(T)]$ is a measure of the effective number of tests. See Feingold et al. (1993), Dupuis and Sigmund (1999), Piepho (2001a), Zou et al. (2004) and Kao and Ho (2012) for further discussions on dense-map approximations.

Selective Genotyping and Phenotyping

As introduced in Chapter 17, one mapping design is **selective genotyping**: a large sample of individuals are phenotyped, with a subset, usually those showing extreme values, chosen for genotyping. This design is partly a relict of the halcyon days when phenotyping was much cheaper than genotyping. Such a design can produce biased estimates unless suitable corrections are performed (Lander and Botstein 1989; Darvasi and Soller 1992; Muranty and Goffinet 1997; Muranty et al. 1997; Ronin et al. 1998; Johnson et al. 1999; Xu and Vogl 2000; Lee et al. 2014). An especially interesting approach to adjust for bias was suggested by Henshall and Goddard (1999). Typically, QTL detection searches for markers showing an association between trait value as a function of the number of copies of a reference marker allele (trait value conditioned on the marker value). Henshall and Goddard suggested that one instead looks at the expected number of copies of a reference allele as a function of trait value (marker value conditioned on the trait value). In effect, this is a logistic regression of allele frequency on trait value. The idea is that the frequency of a focal allele influencing the trait should increase with trait value.

One extension of selective genotyping is to multiple traits. Lin and Ritland (1997) suggested that an individual be genotyped when they display extreme values in at least one of the traits, while Muranty et al. (1997) suggested using a weighted index combining the traits of interest and then selecting individuals with extreme index values. Another variant of selective genotyping is **selective mapping**: the use of selective genotyping to improve map resolution. The idea is to genotype individuals using low marker density (**framework markers**) and then increase the density of genotyping for those individuals

showing recombinations within the region of interest (Vision et al 2000; Ronin et al. 2003; Xu et al. 2005). Finally, one can go full-circle with **selective phenotyping**: individuals are chosen to be phenotyped by some marker genotype criteria, such as maximal diversity within a target regions (Jin et al. 2004; Sen et al. 2005, 2009; Fu and Jansen 2006).

QTL DETECTION AND ESTIMATION USING LS LINEAR MODELS

We now have all of the necessary machinery in place to consider particular methods for QTL detection and estimation in greater detail. In a LS linear model framework, detection is very straightforward, while estimation of effects and positions can be more challenging. Equations 18.7–18.11 show the general estimation strategy for linear models: one constructs some appropriate contrast of marker genotype means for the parameter(s) of interest (the **method-of-moments** approach). As we will shortly demonstrate, likelihood (Appendix 4) provides a much better overall framework for *estimation*, so we will focus solely on using linear models for *QTL detection*.

Linear Model Detection of Continuous Traits

As noted above, the simplest test for a marker-trait association involves the comparison of the trait means of alternate marker genotypes. When only two genotypes are compared (such as with single-marker backcross-, RIL-, or DHL-designs), this can be accomplished with a simple t test. Many designs, however, involve more than two marker genotypes. For example, the single-marker F_2 design has three marker genotypes: *MM, Mm, mm*. In such cases, all marker genotypic means (or some subset of them) can be compared by using standard LS linear-model approaches, such as ANOVA or regression.

The simplest framework is **gene-dosage regression** (also called the **additive model**), which assumes a single marker locus and only additive effects at a single linked QTL. The phenotypic value z_{ik} of the kth individual with marker genotype i is modeled as a mean value μ plus an (additive) marker effect b and a residual error e_{ik},

$$z_{ik} = \mu + bN_i + e_{ik} \tag{18.14a}$$

where N_i is the number of copies (or **gene dosage**) of the focal marker allele (e.g., values of 0, 1, and 2, for mm, Mm, and MM, respectively). A significant value of b implies one (or more) QTLs linked to the marker. Note that this is a one degree-of-freedom test.

A more general approach that accommodates dominance, and easily extends to multiple markers (i.e., multilocus marker genotypes), is given by the **general genotype model**

$$z_{ik} = \mu + b_i + e_{ik} \tag{18.14b}$$

This is a one-way ANOVA model (Chapter 22), with the presence of a linked QTL being indicated by a significant between-marker variance (i.e., one, or more, of the b_i is significantly different from zero). The associated degrees-of-freedom is the total number of marker genotypes minus one. Equivalently, we can express this model given by Equation 18.14b as a multiple regression, with the phenotypic value for individual j given by

$$z_j = \mu + \sum_{i=1}^{n} b_i \, x_{ij} + e_j \tag{18.14c}$$

where the x_{ij} are n indicator variables (one for each marker genotype),

$$x_{ij} = \begin{cases} 1 & \text{if individual } j \text{ has marker genotype } i, \\ 0 & \text{otherwise.} \end{cases}$$

The number of marker genotypes (n) in Equations 18.14b and 18.14c depend on both the number of marker loci and the type of design being used. With a single marker, $n = 2$ for

a backcross, RIL, or DH design, while $n = 3$ for an F_2 design (using codominant markers). When two or more marker loci are simultaneously considered, b_i corresponds to the effect of a *multilocus* marker genotype, and n is the number of such genotypes considered in the analysis. In the regression framework, evidence of a linked QTL is provided by a significant r^2, the fraction of character variance accounted for by the marker genotypes (Chapter 10). Finally, as mentioned above, the presence of a linked QTL can cause different marker genotypes to have different trait variances. If the difference in variance between marker classes is substantial, the standard ANOVA assumption of variance homogeneity is violated and corrections are required for hypothesis testing (Asins and Carbonell 1988; Xu 1995).

Estimation of dominance requires information on all three genotypes at a marker locus, i.e., an F_2, F_t, or other design (such as *both* backcross populations). In these cases, dominance can be estimated using an appropriate weighted combination of the marker means (e.g., Equations 18.7b and 18.8b). For a single marker locus, a simple test for dominance is whether Equation 18.14b gives a significantly better fit than the strictly additive model (Equation 18.14a).

Epistasis between QTLs can be modeled by including **interaction terms** (Chapter 10). Here, an individual with genotype i at one marker locus and genotype k at a second is modeled as $z = \mu + a_i + b_k + I_{ik} + e$, where a and b denote the single-locus marker effects, and I is the interaction term due to epistasis between QTLs linked to those marker loci. In linear regression form this model becomes

$$z_j = \mu + \sum_{i}^{n_1} a_i\, x_{ij} + \sum_{k}^{n_2} b_k\, y_{kj} + \sum_{i}^{n_1} \sum_{k}^{n_2} I_{ik}\, x_{ij}\, y_{kj} + e_j \tag{18.14d}$$

where x_{ij} and y_{kj} are indicator variables for two different marker genotypes (with n_1 and n_2 genotypes, respectively). Significant a_i and/or b_k terms indicate significant effects at the individual marker loci (**main** or **marginal effects**), while significant I_{ik} terms indicate epistasis between the effects of the two markers. More formally, we could express these interaction terms using our general epistatic model (Equation 5.5), e.g., Kao and Zeng (2002) and Zeng et al. (2005).

Essentially the same approach can be used to look for QTL $\times$ environment interactions (Chapter 27) when markers are examined over several environments. The basic model for an individual with marker genotype i measured in the kth environment is $z = \mu + b_i + E_k + I_{ik} + e$, where a significant E_k indicates an environmental effect, while a significant I_{ik} implies a marker $\times$ environment interaction. For example, if the character has significant sex-specific effects, these can be incorporated by using the model $z = \mu + b_i + s_k + I_{ik} + e$ for an individual of marker genotype i and sex k. A significant s_k implies a significant sex effect, while a significant I_{ik} implies a significant marker $\times$ sex interaction. Long et al. (1995) gave an example of the utility of this approach, finding very significant sex-specific effects for bristle number in *Drosophila*. Detecting QTL $\times$ environment interaction is examined more fully in Chapter 27.

Example 18.4.   Edwards et al. (1987) examined two F_2 maize populations. Cross 1 consisted of 1776 individuals scored for 16 markers, while Cross 2 used a different set of parental lines and consisted of 1930 individuals scored for 20 markers. As the frequency distribution (below) shows, the detected marker effects (measured by the fraction r^2 of total F_2 phenotypic variance accounted for by each significant marker-trait association) were generally quite small.

A total of 82 vegetative characters were examined, with 60% (Cross 1) and 64% (Cross 2) of all possible marker-trait combinations showing significant effects (at the $\alpha = 0.05$ level) using single-marker ANOVA. On average, each trait showed 10 (Cross 1) and 14 (Cross 2) significant marker associations. Dominance was common, while pairwise epistasis, as tested by incorporating a marker $\times$ marker interaction term into the linear model (Equation 18.14d), was rare. The same two F_2 populations were used by Stuber et al. (1987) to examine 25 yield-related characters, with similar results. In that study, most marker-trait combinations were significant (66% and 72% at the $\alpha = 0.05$ level), and most marker effects were small (over

half of the significant associations having r^2 values less than two percent). As a group, yield-related traits displayed more dominance than vegetative traits, but many yield traits were still largely additive. Edwards et al. (1992) examined a subset of the vegetative characters in Cross 2, using a much larger number of markers (114 RFLPs). While only 187 F_2 individuals were scored, 15% of marker-trait associations were significant, and the overall results with respect to the distribution of effects were similar to those for the 1987 experiments
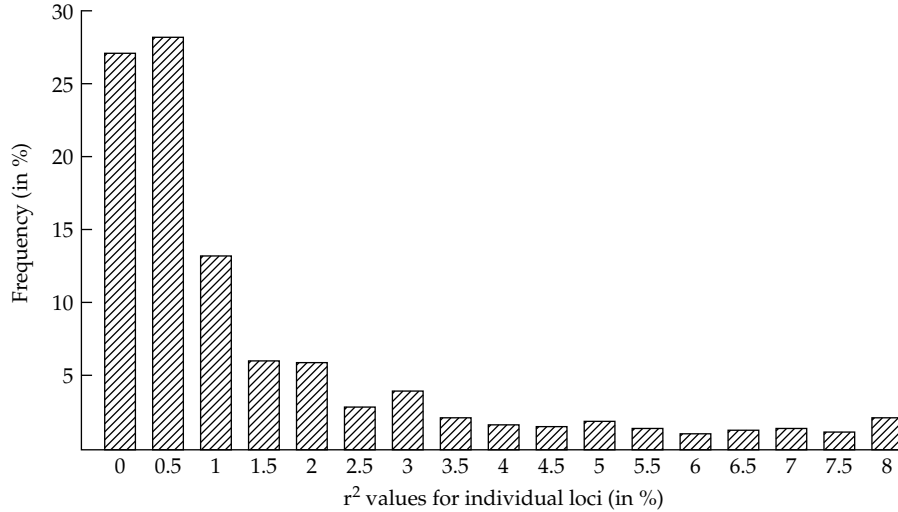
Detecting QTLs for Dichotomous Traits: The Cochran-Armitage Trend Test

Many QTL experiments are concerned with dichotomous (binary) traits, such as disease or pest resistance in crop plants or disease susceptibility in humans. In many cases, one can score quantitative physiological traits contributing to the binary trait, such as blood pressure or number of lesions per leaf, and the above methodology for QTL detection with continuous traits applies. However, such variables are often either unknown or unmeasured, and the data are simply scored as presence/absence values (or **cases** versus **controls** in the medical literature). The simplest procedure to detect marker-trait associations in this setting is to test for independence using standard association tables (such as χ^2 or Fisher's exact tests). As shown in the table below, the n total observations are partitioned into counts for each particular class, e.g., n_{P1} is the sample number of Mm individuals showing the trait.

	Marker Genotype			
	mm	Mm	MM	Totals
Present	n_{P0}	n_{P1}	n_{P2}	n_P
Absent	n_{A0}	n_{A1}	n_{A2}	n_A
Totals	n_0	n_1	n_2	n

This same approach easily extended to polychotomous (ordinal) characters. With three marker genotypes and two trait values, the result χ^2 test has two degrees of freedom, with a significant value indicating linkage to one (or more) QTLs.

Note that the contingency table approach is akin to Equation 18.14b, making no assumptions about the trait values (percentage showing the trait) associated with each genotype, hence the two degrees of freedom for the test. A more powerful approach, akin the gene dosage regression (Equation 18.14a), is the **Cochran-Armitage trend test** (Cochran 1954; Armitage 1955) which uses one degree of freedom to test for a *trend in proportions* as the gene dosage increases. The basic structure of the test is as follows. First one chooses a weight vector $\tau = (\tau_0, \tau_1, \tau_2)$ which is an assumption about the nature of the trend over the three

genotypes. The resulting test statistic becomes

$$T = \frac{n_A n_P}{n} \cdot \frac{\left[\sum_{i=0}^{2} \tau_i \left(\frac{n_{Pi}}{n_P} - \frac{n_{Ai}}{n_A} \right) \right]^2}{\sum_{i=0}^{2} \tau_i^2 \frac{n_i}{n} - \left(\sum_{i=0}^{2} \tau_i \frac{n_i}{n} \right)^2} \tag{18.15}$$

where, for large samples, $T \sim \chi_1^2$. The classical Cochran-Armitage test assumes $\tau = (0, 0.5, 1)$, namely an additive trend. Note that one can easily modify Equation 18.15 under the assumption that M is linked to a completely dominant ($\tau = [0, 1, 1]$), or recessive ($\tau = [0, 0, 1]$), QTL allele. Lee (2015) discusses optimal weighting schemes under other scenarios. Note that if we use new weights $k_i = a\tau_i$ (all the weights are scaled by the same constant), the a values in the numerator and denominator cancel, recovering Equation 18.15.

Example 18.5. Consider the following data from Zhang et al. (2005), who examined the association between genotypes at the DNA repair gene *ADPRT* and lung cancer:

	Genotype			
	mm	*Mm*	*MM*	Totals
Present	307 [0.461]	509 [0.502]	184 [0.573]	1000
Absent	359 [0.539]	504 [0.498]	137 [0.427]	1000
Totals	666	1013	321	2000

The number in the square bracket is the fraction of the marker class with (Present) or without (Absent) lung cancer. There appears to be a trend with lung cancer risk increasing with the number of copies of *M*. Is this statistically significant? Performing a standard χ^2 test on this contingency table returns a test statistic value of 10.9664, with an associated p value of $\Pr(\chi_2^2 \geq 10.9664) = 0.0042$. The Fisher-exact test returns (to four decimal places) the same p value.

To compute the Cochran-Armitage statistic (assuming an additive trend), note that

$$\frac{n_A n_P}{n} = \frac{1000 \cdot 1000}{2000} = 500$$

$$\sum_{i=0}^{2} \tau_i \left(\frac{n_{Pi}}{n_P} - \frac{n_{Ai}}{n_A} \right) = 0.5 \left(\frac{509}{1000} - \frac{504}{1000} \right) + 1 \left(\frac{184}{1000} - \frac{137}{1000} \right) = 0.0495$$

$$\sum_{i=0}^{2} \tau_i^2 \frac{n_i}{n} = \frac{1}{4} \cdot \frac{1013}{2000} + 1 \cdot \frac{321}{2000} = 0.2871$$

$$\sum_{i=0}^{2} \tau_i \frac{n_i}{n} = \frac{1}{2} \cdot \frac{1013}{2000} + 1 \cdot \frac{321}{2000} = 0.4138$$

Substituing these values into Equation 18.15 gives

$$T = 500 \cdot \frac{0.0495^2}{0.2871 - 0.4138^2} = 10.5733$$

The associated p value is $\Pr(\chi_1^2 \geq 10.5733) = 0.0011$. Note the nearly four-fold smaller p value relative to a standard χ^2 test, which results from the Cochran-Armitage test requiring fewer degrees of freedom.

QTL DETECTION AND ESTIMATION VIA MAXIMUM LIKELIHOOD

Maximum likelihood (ML) methods (Appendix 4) are especially popular in the QTL mapping literature. While LS linear models use only marker means, ML uses the full information from the marker-trait distribution and, as such, is expected to be more powerful. The trade-off is that ML is computationally intensive, requiring rather special programs to solve the likelihood equations, while linear model analysis can be performed with almost any standard statistical package. Further, while modifying the basic model (such as adding extra factors) is rather trivial in the linear model framework, with ML new likelihood functions need to be constructed and solved for each variant of the original model. Although writing down a set of likelihood equations for the model of interest is relatively straightforward (e.g., Example 18.6), obtaining the ML estimates is more difficult. One approach outlined in Appendix 4 is to use specialized algorithms, of which **EM (expectation-maximization) methods** have been successfully adapted to many of the mixture-model problems in QTL mapping (e.g., Lander and Botstein 1989; Carbonell and Gerig 1991; Luo and Kearsey 1992; van Ooijen 1992; Carbonell et al. 1992; Luo and Wolliams 1993; Jansen 1992, 1993a, 1994a, 1996; Jansen and Stam 1994; Kao and Zeng 1997; Xu et al 2003; Chen 2005; Xu et al 2005; Xu and Hu 2010). Alternatively, as we discuss later, a creative use of regressions can often provide excellent approximations to ML solutions. For the remainder of this chapter, we assume that the reader has recently read Chapter 16 and Appendix 4, which introduces much of the ML machinery used here.

Assuming that the distribution of phenotypes for an individual with QTL genotype Q_k is normal with genotypic-specific mean μ_{Q_k} and common variance σ^2, and following the logic of Chapter 16, the likelihood for an individual with phenotypic value z and marker genotype M_j becomes

$$\ell(z \mid M_j) = \sum_{k=1}^{N} \varphi(z, \mu_{Q_k}, \sigma^2) \Pr(Q_k \mid M_j) \tag{18.16}$$

where $\varphi(z, \mu_{Q_k}, \sigma^2)$ denotes the density function for a normal distribution with mean μ_{Q_k} and variance σ^2 (Equation 16.5c), and a total of N QTL genotypes is assumed. This likelihood is a mixture-model distribution (Chapter 16). The mixing proportions, $\Pr(Q_k \mid M_j)$, are functions of the genetic map (the assumed position(s) of the QTL(s) with respect to the observed markers) and the experimental design, while the QTL effects enter only though the means μ_{Q_k} and variance σ^2 of the underlying distributions.

Example 18.6. Consider the single-marker F_2 design with a single QTL linked to the marker. Making the standard assumption that phenotypes are normally distributed about each QTL genotype, substitution of the F_2 conditional probabilities (Example 18.1) into Equation 18.16 gives the likelihood functions for the three different marker genotypes as

$$\ell(z \mid MM) = (1-c)^2 \varphi(z, \mu_{QQ}, \sigma^2) + 2c(1-c)\,\varphi(z, \mu_{Qq}, \sigma^2) + c^2 \varphi(z, \mu_{qq}, \sigma^2)$$
$$\ell(z \mid Mm) = c(1-c)\varphi(z, \mu_{QQ}, \sigma^2) + \left[(1-c)^2 + c^2\right]\varphi(z, \mu_{Qq}, \sigma^2)$$
$$+ c(1-c)\varphi(z, \mu_{qq}, \sigma^2)$$
$$\ell(z \mid mm) = c^2 \varphi(z, \mu_{QQ}, \sigma^2) + 2c(1-c)\,\varphi(z, \mu_{Qq}, \sigma^2) + (1-c)^2 \varphi(z, \mu_{qq}, \sigma^2)$$

as obtained by Weller (1986). For example, if individual one has trait value 10 and marker genotype MM, then its contribution to the likelihood function is

$$\ell(z_1 \mid M_1) = (1-c)^2 \varphi(10, \mu_{QQ}, \sigma^2) + 2c(1-c)\,\varphi(10, \mu_{Qq}, \sigma^2) + c^2 \varphi(10, \mu_{qq}, \sigma^2)$$

The total likelihood for n F_2 individuals is the product of the individual likelihoods,

$$\ell(\mathbf{z}) = \prod_{i=1}^{n} \ell(z_i \mid M_i)$$

While rather complex, the total likelihood is a function of just five parameters: the QTL position (c), the three QTL means ($\mu_{QQ}, \mu_{Qq}, \mu_{qq}$), and the common variance (σ^2).

The standard likelihood model assumes continuous trait values and normally disturbed residuals, although likelihood models have been developed to map continuous traits with censored observations (Diao et al. 2004). However, a variety of important traits are discrete. Much of human genetics deals with dichotomous (binary) traits, which have binomially distributed residuals (Chapter 2). Plant breeders often score resistance in an ordinal fashion, such as fully resistant, partly resistant, or fully susceptible. Such polychotomous traits have multinomially distributed residuals (Chapter 2). As we saw with segregation analysis (Chapter 16), the likelihood equations can be modified to handle such traits by using generalized linear models (Chapters 14 and 20). Dichotomous and polychotomous traits can be modeled through the use of logistic regressions and probit scales (Ghosh et al. 1993; Hackett and Weller 1995; Visscher et al. 1996a; Xu and Atchley 1996; Rao and Xu 1998; Yi and Xu 2000; Deng et al. 2006). While the generalized linear model framework is formally appropriate in such cases, often (as a first pass), the discrete structure of the data is ignored by an investigator, treating them as if they were continuous (e. g., coding alternative binary characters as 0/1) and applying standard ML. When flanking markers are used, this approach gives essentially the same power and precision as methods specifically designed for polychotomous traits (Hackett and Weller 1995; Visscher et al. 1996a), but when single markers are used, this approach can give estimates for QTL position that are rather seriously biased (Hackett and Weller 1995).

A final class of discrete data are those that can take on a value of zero and otherwise have a very limited range of possible values, such as number of eggs laid by a wild bird. Again, one could treat these discrete values as continuous and use standard ML, but such data are usually poorly approximated by a normal distribution. For example, they often have inflated zero values (an excess of zero values). Such data are best treated as Poisson-distributed (perhaps with extra point mass at zero; Chapter 14, WL Chapter 29). In the generalized linear model framework, such data are modeled using log-linear models (Cui et al. 2006; Cui and Yang 2009). Chen and Liu (2009) and Xu and Hu (2010) present a general EM treatment for using generalized linear models in QTL mapping. An alternative approach for treating nonnormally distributed characters is given by Kruglyak and Lander (1995c), who develop a nonparametric interval mapping procedure.

Finally, we note that Bayesian analysis (Appendix 7) is essentially a generalization of likelihood methods, incorporating any prior information to yield a posterior distribution for the unknown parameters. There is a growing literature for Bayesian-based QTL mapping (e.g., Hoeschele and VanRaden 1993a, 1993b; Satagopan et al 1996; Thaller and Hoeschele 1996a, 1996b; Sillanpää and Arjas 1998; Yi et al. 2003; Zhang et al. 2005; Xu et al. 2009). Sen and Churchill (2001) present a unified treatment, based on conditioning on the unobserved QTL genotype (akin to Equation 18.6), which allows the issues of QTL position and QTL effect to be separated into two independent and manageable parts. Using this partitioned, they developed an MCMC approach (Appendix 8) to accommodate multiple, interacting QTLs, allowing for non-normally distributed traits, as well as accommodating missing genotype data and genotyping errors. Two powerful uses of Bayesian methods are in estimating the number of QTLs and dealing epistasis, and we will return to these later in the chapter.

Likelihood Maps (Profile Plots)

In the likelihood framework, tests of whether a QTL is linked to the marker(s) under consideration are based on the likelihood-ratio statistic,

$$\text{LR} = -2\ln\left[\frac{\max \ell_r(\mathbf{z})}{\max \ell(\mathbf{z})}\right]$$

where $\max \ell_r(\mathbf{z})$, given by Equation 16.8, is the maximum of the likelihood function under
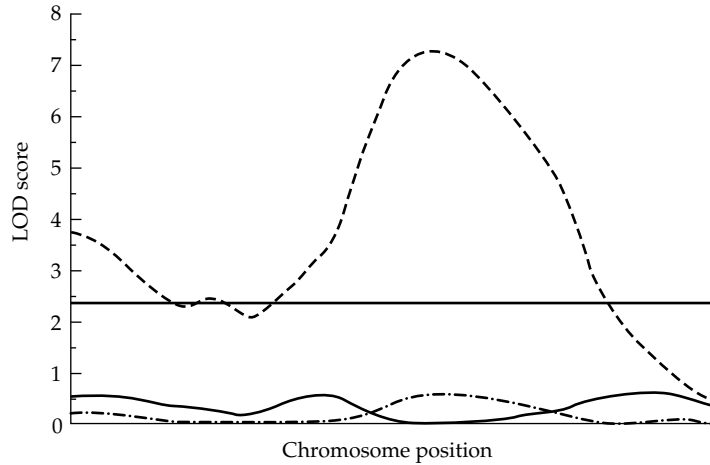
Figure 18.1 Likelihood map (or profile) for QTL positions on chromosome 10 in a cross of two tomato species. Evidence for a QTL is provided when the likelihood function exceeds the significance threshold (indicated by the horizontal line). The upper dashed curve gives the LOD score for fruit pH as a function of map position, showing strong evidence of a QTL near the middle of the chromosome. The lower two curves (solid and broken) are for fruit weight and soluble-solid concentration, neither of which shows a significant QTL effect on this chromosome. (After Paterson et al. 1988.)

the null hypothesis of no segregating QTL (i.e., under the assumption that the phenotypic distribution is a single normal). This test statistic is approximately χ^2-distributed, with the degrees of freedom given by the extra number of fitted parameters in the full model. For a model assuming a single QTL, most designs have five parameters in the full model (the three QTL means, the common variance, and the QTL position), and two in the reduced model (the mean and variance), giving three degrees of freedom. Certain designs (such as a backcross, RIL, or DHL) involve situations where only two QTL means enter (e.g., *QQ* and *qq* for RILs/DHLs, *Qq* and *QQ* or *qq* for a backcross), and here the likelihood ratio test has two degrees of freedom.

The amount of support for a QTL at a particular map position is often displayed graphically through the use of **likelihood maps** (Figures 18.1 and 18.2), also called **profile plots**, which graph the likelihood-ratio statistic (or a closely related quantity) as a function of map position of the putative QTL. For example, the value of the likelihood map at $c = 0.05$ gives the likelihood-ratio statistic that a QTL is at recombination fraction 0.05 from the marker vs. a model assuming no QTL. This approach for displaying the support for a QTL was introduced by Lander and Botstein (1989), who plotted the LOD (**likelihood of odds**) scores (Morton 1955b). The LOD score for a particular value of c is related to the likelihood-ratio test statistic (LR) by

$$\text{LOD}(c) = \log_{10}\left[\frac{\max \ell_r(\mathbf{z})}{\max \ell(\mathbf{z}, c)}\right] = \frac{\text{LR}(c)}{2\ln 10} \simeq \frac{\text{LR}(c)}{4.61} \tag{18.17}$$

showing that the LOD score is simply a constant times the likelihood-ratio statistic. Here $\max \ell(\mathbf{z}, c)$ denotes the maximum of the likelihood function given a QTL at recombination frequency c from the marker. Another variant is simply to plot $\max \ell(\mathbf{z}, c)$ instead of the likelihood-ratio statistic, as the restricted likelihood, $\max \ell_r(\mathbf{z})$, is the same for each value of c. Testing over the entire genome yields a **genomic scan** for QTLs.

The likelihood profile projects the multidimensional likelihood surface (which is a function of the QTL means, the common variance, and the map position) on to a single dimension, that of the map position, c. The ML estimate of c is that which yields the maximum value on the likelihood map, and the values for the QTL means and variance that maximize
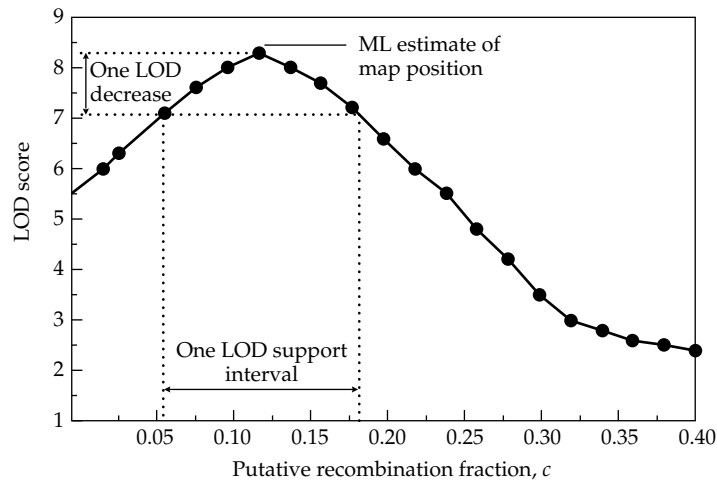
Figure 18.2 Hypothetical likelihood map for the marker-QTL recombination frequency c in a single-marker analysis. Points connected by straight lines are used to remind the reader that likelihood maps are computed by plotting the maximum of the likelihood function for each c value, usually done by considering steps of 0.01 to 0.05. A QTL is indicated if any part of the likelihood map exceeds a critical value. In such cases, the ML estimate for map position is the value of c giving the highest likelihood. Approximate confidence intervals for QTL position (one-LOD support intervals) are often constructed by including the set of all c values giving likelihoods within one LOD score of the maximum value.

the likelihood given this value of c are the ML estimates for the QTL effects. Thus, in the likelihood framework, *detection* of a linked QTL and *estimation* of its position are coupled—if the likelihood ratio exceeds the critical threshold for that chromosome, it provides evidence for a linked QTL, whose position is estimated by the peak of the likelihood map. If the peak does not exceed this threshold, there is no evidence for a linked QTL.

Precision of ML Estimates of QTL Position

Because ML estimates are approximately normally distributed for large sample sizes, confidence intervals for QTL effects and position can be approximated using the asymptotic sampling variances for the ML estimates (Appendix 4). Approximate confidence intervals are often constructed using the **one-LOD rule** (Figure 18.2), with the confidence interval being defined by all those values falling within one LOD score of the maximum value (Conneally et al. 1985; Lander and Botstein 1989). The motivation for such **one-LOD support intervals** follows from the fact that the large-sample distribution of the LR statistic follows a χ^2 distribution. If only one parameter in the likelihood function is allowed to vary, as when testing whether c equals a particular value (say the observed ML estimate), the LR statistic has one degree of freedom. Because a one-LOD change corresponds to an LR change of 4.61 (Equation 18.17), which for a χ^2 with one degree of freedom corresponds to a significance value of 0.04 (e.g., $\Pr(\chi_1^2 \geq 4.61) = 0.04$), it follows that one-LOD support intervals approximate 95% confidence intervals under the appropriate settings.

While widely used, the one-LOD rule is a *large-sample property*, and, as such, is only an approximation. When the true QTL effect is small, this approach often yields intervals that are *too small*. Mangin et al. (1994a, 1994b) showed that one-LOD confidence intervals have between 60% and 95% probability of actually containing small-effects QTLs, and they developed an improved method for such cases. This observation lead to a series of **LOD-drop rules** using intervals based on decreases in LOD scores in excess of one. For example, simulation studies led van Ooijen (1992) to suggest that support intervals should be based on **two-LOD differences** in order to have a high probability of containing the QTL, while Dupuis and Siegmund (1999) suggested using a **1.5-LOD support interval**. Very extensive

simulations (using 450 CPU days!) by Manichaikul et al. (2006) showed that the LOD drop value increased with decreasing power (i.e., smaller QTL effects), ranging from around 2 for low power to close to 1.5 to 1 for high power. Visscher and Goddard (2004) showed that the likelihood surface around the MLE is often rather leptokurtic (Chapter 2), rather than the quadratic required by asymptotic theory for the one-LOD rule to be valid.

A more rigorous approach, although still a large-sample approximation, for obtaining standard errors of both map position and QTL effects is to use the inverse of the Fisher information matrix associated with the likelihood function (Appendix 4). Kao and Zeng (1997) and Chen (2005) showed how to obtain this asymptotic matrix in an EM framework. Bayesian methods (Appendix 7) offer another approach, and tend to have better small-sample properties than likelihood. Example 18.7 outlines a Bayesian credible interval approach, based on areas under the likelihood profile, suggested by Sen and Churchill (2001).

Resampling methods have been suggested as a robust procedure for constructing confidence intervals for QTL position, and Visscher et al. (1996b), Walling et al. (1998, 2002), and Talbot et al. (1999) suggested using a **nonparametric bootstrap** approach (Efron 1979, 1982). Suppose the original data set consists of n individuals. A bootstrap sample is generated by drawing n values, *with replacement*, from the original data set. Such a sample will have some of the original observations present multiple times and others not present at all. A series of N such samples are generated and an estimate (map position in this case) is computed for each, generating a distribution of estimates (the **empirical bootstrap distribution**). The resulting 95% bootstrap confidence interval has as its lower value the estimate corresponding to the 2.5% cumulative frequency point of the empirical bootstrap distribution, while the upper value is that corresponding to the upper 97.5% of the bootstrap distribution. While the bootstrap approach is generally fairly robust, for QTL mapping it appears to be compromised by the bias from ML (and regression) methods tending to place QTLs *exactly* at the markers, as opposed to between them (Walling et al. 2002; Manichaikul et al. 2006). Manichaikul et al. noted that a consequence of this feature is that the accuracy of bootstrap confidence intervals *critically* depends on where the true QTL is relative to the markers, and, as a result, recommend that LOD drop off or Bayesian credible intervals (Example 18.7) be used instead, as these tend to be a bit more robust to true QTL position.

The length of the confidence interval is influenced by the number of individuals sampled, the effect of the QTL in question, and the marker density. Darvasi et al. (1993) showed that precision is not significantly increased by increasing marker density beyond a certain point (around one marker every 5 to 10 cM). Given such a dense map, van Ooijen (1992) found that ML mapping using flanking markers with reasonable sample sizes (200–300 F_2 or backcross individuals) allowed a QTL accounting for 5% of the total variance to be mapped to a 40 cM interval, while one accounting for 10% could be mapped to a 20 cM interval. Unfortunately, these interval sizes are distressingly large for cloning QTLs or even defining their positions to smaller intervals for RIL construction. Equation 18.44b (below) gives a simple approximation for the sample size required for a desired interval length.

One strategy for increasing the precision of mapping is to use lines with expanded genetic maps, such as RILs or AILs. With these designs, estimates of the map position are in terms of the cumulative recombination frequency $\widetilde{c} \simeq \tau c$ (where $\tau > 1$; Equation 18.5), so that the confidence interval for c is reduced by a factor of $1/\tau$. For example, recombinant inbred lines have a two- to four-fold expansion of the map (Equation 18.5b), and hence reduce the length of the confidence interval for c by $1/2$ to $1/4$ relative to an F_2. Even more dramatic reductions are possible using advanced intercross lines. A sample size and marker density that yields a 20 cM confidence interval in an F_2 design gives a 3.4 cM confidence interval for the same QTL in an F_{10} design. (This follows from Equation 17.4, which shows that a Haldane distance of 20 cM, corresponding to $c = 0.165$, translates into $\widetilde{c} = c/5 = 0.033$ and a Haldane distance of 3.4 cM with an F_{10} AIL.) Similarly, an F_{20} design would give a 1.7 cM confidence interval.

Example 18.7.   Sen and Churchill (2001) proposed an **approximate Bayes credible interval** for QTL position. Bayesian analysis is examined in more detail in Appendix 7, but the basic idea (Equation A7.3d) is that one starts with some initial prior distribution $p(c)$ for QTL position, and then updates this information with data $\mathbf{z}$ via the likelihood $\ell(\mathbf{z}, c)$, to yield posterior distribution for the QTL position,

$$p(c \,|\, \mathbf{z}) = \frac{\ell(\mathbf{z}, c)\, p(c)}{\int \ell(\mathbf{z}, c)\, p(c)\, dc} = \frac{\ell(\mathbf{z}, c)}{\int \ell(\mathbf{z}, c)\, dc}$$

Integration is performed to ensure that the posterior is a proper probability distribution (the area under the posterior is one). The right-hand simplification follows when one assumes a **flat prior**, $p(c) = 1/L$ (L being the length of the chromosome in Morgans), namely we assume, a priori, that all map position values are equally likely QTL locations. Sen and Churchill noted that this distribution can be approximated using profile likelihoods. Using Equation 18.17 to express $\ell(\mathbf{z}, c)$ in terms of LOD scores,

$$p(c \,|\, \mathbf{z}) \simeq \frac{10^{LOD(c)}}{\sum_{i=1}^{L} 10^{LOD(c_i)}}$$

where $LOD(c)$ is the LOD score if the QTL is assumed to be at map position c on a chromosome, and the sum is over the possible c values (moving from one end of a chromosome to the other). Using $p(c \,|\, \mathbf{z})$, a 95% credible interval is simply the smallest interval in c (map position) that contains 95% of the probability mass of c. Manichaikul et al. (2006) found that this approach tended to outperform the bootstrap.

ML Interval Mapping

ML mapping with line crosses usually employs the genotypes of a pair of flanking markers as the unit of analysis. The likelihood functions for such **ML interval mapping** (**IM**) follow from Equation 18.16 using the appropriate conditional probabilities for QTL genotypes given the two-locus marker genotypes (Jensen 1989; Lander and Botstein 1989; Knapp et al. 1990, Carbonell et al. 1992; van Ooijen 1992; Korol et al. 1996). Example 18.8 shows the basic structure of the resulting likelihood functions. As with single-marker analysis, support for a QTL is evaluated with a likelihood profile over a region (typically a chromosome; Figures 18.1 and 18.5) with the peak of the likelihood profile corresponding to the ML estimate of QTL position within that region and its significance given by a likelihood-ratio test.

Example 18.8.   Likelihood functions for interval mapping follow by substituting the appropriate conditional probabilities into Equation 18.16. For example, consider the F_2 formed by crossing two inbred lines. Assuming no interference, from Equation 18.2 the likelihood for marker genotype $M_1 M_1 M_2 M_2$ is

$$\ell(z \,|\, M_1 M_1 M_2 M_2) = \left[\frac{(1 - c_1)^2\, (1 - c_2)^2}{(1 - c_{12})^2} \right] \cdot \varphi(z, \mu_{QQ}, \sigma^2)$$

$$+ \left[\frac{2\, c_1\, c_2\, (1 - c_1)\, (1 - c_2)}{(1 - c_{12})^2} \right] \cdot \varphi(z, \mu_{Qq}, \sigma^2)$$

$$+ \left[\frac{c_1^2\, c_2^2}{(1 - c_{12})^2} \right] \cdot \varphi(z, \mu_{qq}, \sigma^2)$$

Likelihoods for the other eight flanking-marker genotypes follow similarly and can be found in Luo and Kearsey (1992), Carbonell et al. (1992), and van Ooijen (1992). Even though these likelihoods involve three recombination parameters (c_{12}, c_1, c_2), the distance between markers (c_{12}) is usually taken as known, and hence $c_2 = (c_{12} - c_1)/(1 - 2\,c_1)$ (assuming no interference) or $c_2 = c_{12} - c_1$ (complete interference). This leaves five parameters to estimate: three QTL means, the common variance σ^2, and the position c_1 of the putative QTL within the interval. Likelihoods for other designs follow using the appropriate conditional probabilities.

One of the first applications of ML interval mapping was performed by Paterson et al. (1988), who examined 237 backcross individuals in a cross between the tomato species *Lycopersicon esculentum* and *L. chmielewskii* for several fruit-related traits (Figure 18.1 gives the chromosome 10 likelihood maps for three traits). By using 68 markers (63 RFLP and 5 isozyme variants), 95% of the genome was within 20 cM of a marker. Six QTLs affecting fruit mass, four affecting concentration of soluble solids, and five affecting fruit pH were detected. A follow-up study (Paterson et al. 1990) using NILs (Chapter 17) detected additional QTLs. However, this finer mapping could not confirm the presence of one putative QTL that showed a highly significant peak on the likelihood map in the 1988 study, suggesting it was a false positive.

With ML interval mapping, the likelihood map for an entire chromosome is constructed by splicing together the likelihood maps for each successive interval. If the order of markers on a particular chromosome is $M_1 - M_2 - M_3 - \cdots - M_n$, the likelihood map for the $M_1 - M_2$ interval is constructed using only marker information from these two loci, the profile plot for the $M_2 - M_3$ interval uses only information from M_2 and M_3, etc.

Given the multiple-test nature of these plots (because each map is actually a set of multiple intervals), the appropriate threshold value for the collection of internal maps that constitutes the likelihood map for a chromosome is debatable. Knott and Haley (1992a) noted that the total number of independent tests is bounded above by the number of intervals examined, but because these intervals are linked, they are not independent tests (Zeng 1993). The lower bound is set by the number of chromosomes examined, as these segregate independently. Hence, we first set a threshold level for each chromosome that ensures a desired genome-wide significance level for the entire collection of chromosomes. If C chromosomes are examined, Equation 18.13b implies that in order to obtain a genome-wide significance level γ, the significance level used to set thresholds for each chromosome is

$$1 - (1 - \gamma)^{1/C} \simeq \gamma/C \qquad (18.18)$$

Rebaï et al. (1994b) suggested an improved approach that takes into account differences in chromosome lengths. Turning now to the significance values for intervals on a given chromosome, suppose the chromosome of interest has m intervals and we have set the chromosome-wide significance as γ_i. Simulation studies by Zeng (1994) suggest that if the number of markers is not too large, then, for large sample sizes, the critical value for each interval is approximately given by a χ^2_k value with significance γ_i/m. Here k is the number of free parameters in the likelihood-ratio test. Approximations assuming a dense marker map have been developed (Example 18.3; Lander and Botstein 1989; Feingold et al. 1993; Lander and Schork 1994; Lander and Kruglyak 1995; Piepho 2001a; Kao and Ho 2012), as have those for a sparse marker map (Zeng 1994; Rebaï et al. 1994b). Simulations by Doerge and Rebaï (1996) show that dense marker approximations (assuming a very large number of markers per chromosome) are conservative, with the probability of a test statistic exceeding the α-level threshold being less than α when no QTL is present. For comparison among the intervals on a *specific chromosome*, one can use sequential Bonferroni tests, as well as adjusting for the effective number of tests, both previously discussed above.

Example 18.9.   Suppose five chromosomes are used for ML-interval mapping in an F_2 design. Chromosomes 1 through 5 have 10, 5, 20, 30, and 40 markers, respectively. In order to achieve a genome-wide level of significance of $\gamma = 0.10$, what are the approximate critical values for each chromosome? Applying Equation 18.18, the overall level of significance for each chromosome is $1 - (1 - 0.1)^{1/5} = 0.021$.

The critical values for each chromosome vary with the number of markers. For chromosome 1, the significance levels for each test become approximately $0.021/10 = 0.0021$. Recall that the degrees of freedom for the test of no QTLs in an F_2 design are $5 - 2 = 3$. Because $\Pr(\chi_3^2 > 14.71) = 0.0021$, this implies that the critical values for the likelihood ratios for chromosome 1 is 14.7. Similarly, the critical values for the remaining four chromosomes are 13.2, 16.2, 17.0, and 17.6.

An alternative approach to obtaining critical values is to use permutation tests to set the threshold levels (Churchill and Doerge 1994; Doerge and Churchill 1996). This resampling procedure has the advantage of being robust to the actual distribution of effects. Further, resampling is superior to analytical approximations for data with missing and incomplete marker information, as the permutation test, by keeping genotypes intact during reshuffling, automatically incorporates the special nature of each data set (Doerge and Rebaï 1996). Piepho (2001a) noted that resampling can be computationally intensive, and suggested an approximate threshold approach that is much less demanding. He used results from Davies (1987) for situations where a nuisance parameter is only present in the alternative hypothesis (here, the QTL position, which is absence under the null). Using this theory, Piepho developed an approximation method that uses information from the geometry of the profile plot (the number of times the derivative of the profile plot changes along a chromosome) to obtain a chromosome-wide error rate.

Finally, it should be mentioned that the null hypothesis usually assumed, that of no QTLs, may be misleading. Crossed lines are often chosen *because* they differ in traits of interest, so that there is certainly segregating genetic variance in the F_2 and other line-cross populations. Visscher and Haley (1996) noted that if such background variance is present, it results in a more frequent rejection of the null hypothesis of no QTL than expected. They argued the more appropriate null hypothesis should be that, taking the strain differences into account, the amount of genetic variance explained by a chromosome segment is that expected by chance, and they propose several tests of this hypothesis. Xu (2013) presented general mixed model to account for nonadditive (dominance and epistasis) background variation.

Approximating ML Interval Mapping by Haley-Knott Regressions

One problem with ML estimators is that they can be rather computationally demanding. Among other things, this limits the applicability of resampling methods, which require thousands of ML estimates to be computed per experiment. Fortunately, a simple regression procedure usually gives an excellent approximation of the likelihood map for ML interval mapping (Haley and Knott 1992; Martínez and Curnow 1992; Xu 1995, 1998; Kao 2000). This procedure greatly facilitates resampling, as regressions are quickly and easily computed. Haley and Knott's (1992) idea is to express the regression coefficients as a function of the unknown QTL parameters. Using the Falconer parameterization for genotypic means,

$$\mu_{QQ} = \mu + a, \qquad \mu_{Qq} = \mu + d, \qquad \mu_{qq} = \mu - a \qquad (18.19a)$$

this is done by considering the regression

$$z_j = \mu + a \cdot x(M_j) + d \cdot y(M_j) + e_j \qquad (18.19b)$$

The variables x and y, which depend on both the flanking-marker genotype of the individual (M) and the assumed map position of the putative QTL, are obtained as follows. Taking the expectation of Equation 18.19b over all individuals with marker genotype M_i gives

$$\mu_{M_i} = \mu + a \cdot x(M_i) + d \cdot y(M_i) \tag{18.20a}$$

From Equation 18.6,

$$\mu_{M_i} = (\mu + a)\Pr(QQ\,|\,M_i) + (\mu + d)\Pr(Qq\,|\,M_i) + (\mu - a)\,\Pr(qq\,|\,M_i)$$
$$= \mu + a \cdot \big[\,\Pr(QQ\,|\,M_i) - \Pr(qq\,|\,M_i)\,\big] + d \cdot \Pr(Qq\,|\,M_i) \tag{18.20b}$$

Equating like terms in Equations 18.20a and 18.20b gives the conditional means of x and y, given marker genotype, as

$$x(M_i) = \Pr(QQ\,|\,M_i\,) - \Pr(qq\,|\,M_i), \qquad y(M_i) = \Pr(Qq\,|\,M_i) \tag{18.21}$$

Thus, the x and y values are functions of the conditional QTL probabilities given the flanking-marker genotypes. For the F_2 design with no interference, Equation 18.2 gives

$$x(M_1 M_1 M_2 M_2) = \frac{(1 - c_1)^2 (1 - c_2)^2 - c_1^2 \, c_2^2}{(1 - c_{12})^2}$$

$$y(M_1 M_1 M_2 M_2) = \frac{2 c_1 c_2 (1 - c_1)(1 - c_2)}{(1 - c_{12})^2}$$

Haley and Knott give expressions for the eight other F_2 marker genotypes, and values for other designs easily follow when the appropriate conditional probabilities are employed. This regression approach was independently suggested by Martínez and Curnow (1992) for the analysis of backcross populations. These authors also detailed how missing marker information can be accommodated (Martínez and Curnow 1994a).

By analogy with likelihood maps, the regression given by Equation 18.19b is computed for each c_1 value within the $M_1 - M_2$ interval, with the value giving the regression with the largest r^2 being taken as the estimate of QTL position. For each c_1 value, Equation 18.21 yields the set of x and y values, allowing μ, a, and d to be estimated by ordinary least-squares regression (Equation 10.9a),

$$\mathbf{b}_{c_1} = \begin{pmatrix} \widehat{\mu} \\ \widehat{a} \\ \widehat{d} \end{pmatrix} = \big(\, \mathbf{X}_{c_1}^T \, \mathbf{X}_{c_1} \,\big)^{-1} \mathbf{X}_{c_1}^T \mathbf{z} \tag{18.22}$$

where the ith row of the design matrix $\mathbf{X}_{c_1}$ is $(1, x(M_i, c_1), y(M_i, c_1)\,)$.

Haley and Knott show that r^2 plots for this regression are related to likelihood plots. Assuming that phenotypes are normally distributed about each QTL genotype, then if the QTL is completely linked to either marker ($c_1 = 0$ or $c_1 = c$), the residuals for the regression given by Equation 18.19b are normally distributed. In this case, the regression estimates are also ML estimates and the likelihood-ratio test can be expressed as

$$\text{LR} = n \, \ln \left(\frac{\text{SS}_T}{\text{SS}_E} \right) = -n \, \ln(1 - r^2) \tag{18.23}$$

where SS_T and SS_E are the total and error (or residual) sums of squares associated with the regression (Equations A3.16a and A3.16c), with the second equality following from Equation A3.15. If the QTL is not completely linked to either marker, the distribution of residuals follows a mixture of normals, as some marker genotype classes will contain different QTL genotypes. ML accommodates this mixture of normals, while LS does not. However, Haley

and Knott (1992) and Rebaï et al. (1995) showed that the function given by Equation 18.23 often gives extremely similar values to the true likelihood ratio. Haley and Knott suggested that the number of degrees of freedom appropriate for this test is the number of estimated QTL parameters plus an additional degree of freedom for map position c_1. Xu (1995) noted that this regression approximation tends to overestimate the residual variance, and presented a correction. More generally, if the linear model has additional factors (accounting for, say, differences due to sex and age), the LR test is modified to become

$$\text{LR} = n \, \ln \left(\frac{\text{SS}_E(reduced)}{\text{SS}_E(full)} \right) \tag{18.24}$$

where the error sums of squares are now for the full model and the reduced model (the latter incorporating all factors but the QTL effects). Improvements building on the Haley-Knott regression approach (such as iteratively reweighted least squares) have been proposed (Xu 1998a, 1998b; Feenstra et al. 2006; Han and Xu 2008).

Example 18.10.   Consider the following hypothetical data set: 10 F_2 individuals scored for flanking marker genotypes M_1/m_1 and M_2/m_2, separated by recombination frequency $c_{12} = 0.30$. The following marker genotypes and their associated character values are observed:

$M_1 m_1 M_2 m_2$	$M_1 M_1 M_2 M_2$	$M_1 m_1 M_2 M_2$	$m_1 m_1 M_2 m_2$	$M_1 M_1 M_2 m_2$
3.9	5.6	3.7	3.9	5.3

$m_1 m_1 m_2 m_2$	$M_1 m_1 M_2 M_2$	$M_1 M_1 M_2 M_2$	$M_1 m_1 M_2 M_2$	$M_1 m_1 M_2 m_2$
1.1	3.6	5.4	3.7	3.3

This yields the observation vector

$$\mathbf{z}^T = (3.9,\ 5.6,\ 3.7,\ 3.9,\ 5.3,\ 1.1,\ 3.6,\ 5.4,\ 3.7,\ 3.3)$$

Assuming no interference, $c_2 = (0.3 - c_1)/(1 - 2c_1)$. For each c_1 value within the M_1 - M_2 interval ($0 \leq c_1 \leq 0.3$), a regression is fitted by first using Equation 18.21 to compute the elements of the design matrix for that value of c_1 and then using Equation 18.22 to obtain the regression coefficients. For example, consider three different QTL positions: $c_1 = 0$ (QTL at marker M_1), $c_1 = 0.15$ (QTL in the middle), and $c_1 = 0.3$ (QTL at marker M_2). The resulting regressions for these three c_1 values are

c_1	$\widehat{\mu}$	$\widehat{a}$	$\widehat{d}$	r^2
0.00	3.97	1.47	−0.33	0.730
0.15	3.70	1.89	−0.26	0.732
0.30	2.75	1.65	1.35	0.597

These regressions are obtained using the design matrices

$$\mathbf{X}_0 = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \\ 1 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \quad \mathbf{X}_{0.15} = \begin{pmatrix} 1 & 0.00 & 0.85 \\ 1 & 0.91 & 0.09 \\ 1 & 0.35 & 0.60 \\ 1 & -0.56 & 0.40 \\ 1 & 0.56 & 0.40 \\ 1 & -0.91 & 0.09 \\ 1 & 0.35 & 0.60 \\ 1 & 0.91 & 0.09 \\ 1 & 0.35 & 0.60 \\ 1 & 0.00 & 0.85 \end{pmatrix}, \quad \mathbf{X}_{0.3} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

To complete the analysis, regressions are computed for the full range of c_1 values over the M_1 - M_2 interval, generating the following plot of regression r^2 as a function of c_1.

Recombination frequency, c_1, between marker and QTL
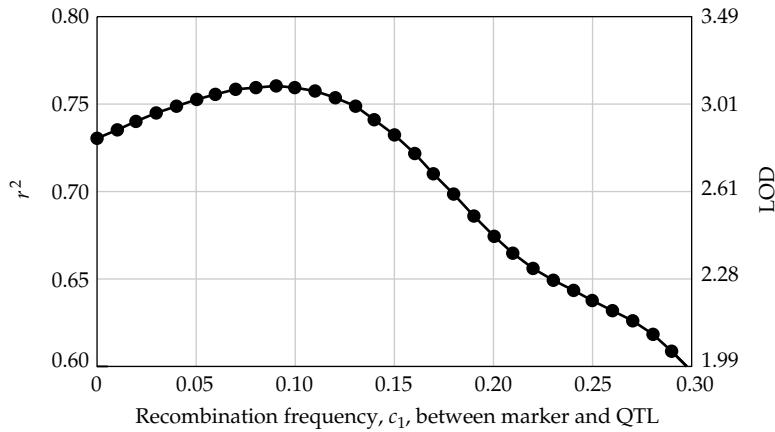
The maximum value of r^2 (0.76) occurs at $c_1 = 0.09$, and the associated regression coefficients are $\widehat{\mu} = 3.90$, $\widehat{a} = 1.76$ and $\widehat{d} = -0.46$. Hence, the data suggest that a QTL lies between these two markers at recombination fraction $c_1 = 0.09$ from marker locus M_1, with estimated genotypic means

$$\widehat{\mu}_{QQ} = \widehat{\mu} + \widehat{a} = 5.66, \quad \widehat{\mu}_{Qq} = \widehat{\mu} + \widehat{d} = 3.44, \quad \widehat{\mu}_{qq} = \widehat{\mu} - \widehat{a} = 2.14$$

Does this example show significant evidence of a QTL? From Equation 18.23, with $n = 10$ and $r^2 = 0.76$, the likelihood ratio (LR) becomes $-10 \cdot \ln(1-0.76) = 14.27$. Note that only two QTL parameters are fitted (a and d) because the reduced model fits a mean μ. Hence, the critical value for the likelihood ratio is a χ^2 with three degrees of freedom (for a, d, c_1),

$$\Pr[\chi_3^2 > -n \ln(1 - r^2)] = \Pr[\chi_3^2 > 14.27] = 0.003$$

showing that the QTL effect is indeed significant.

Approximate confidence intervals can be constructed by using those values giving scores within one LOD of the maximum value. We can translate r^2 values into LOD scores by using LOD = LR/4.61 = $-n \ln(1 - r^2)/4.61$. The MLE has $r^2 = 0.76$ and $n = 10$, for a LOD score of $-10 \ln(1-0.76^2)/4.61 = 3.10$. Hence any c_1 value with a LOD score of 2.10 or greater is in the one-LOD support interval for QTL position. The resulting interval is $c_1 = 0$ to 0.28, so that although there is very strong evidence for a QTL, there is extreme uncertainty as to its position within the interval. This is not surprising given the very small sample size.

DEALING WITH MULTIPLE LINKED QTLs

All of the methods discussed so far are best characterized as **one-at-a-time approaches** for mapping QTLs, as they all assume a single QTL linked to the marker(s) of interest. While such methods can detect the presence of multiple QTLs (e.g., finding marker effects on a number of different chromosomes), they cannot discern whether significant effects at several linked markers/intervals are due to a common QTL or due to several linked QTLs. The presence of multiple linked QTLs also introduces serious biases into estimates of QTL effects and positions derived from the one-at-a-time approach. We have already mentioned that several tightly linked QTL, depending on the sign of their effects, can result in either an enhanced (same signs), or reduced (alternating signs), marker signal (Equation 18.8a).

A much more subtle effect is that while the presence of multiple (significant) peaks on a likelihood profile for a given chromosome is generally taken as an indication of multiple QTLs, *such peaks do not necessarily correspond to the correct QTL positions* (Martínez
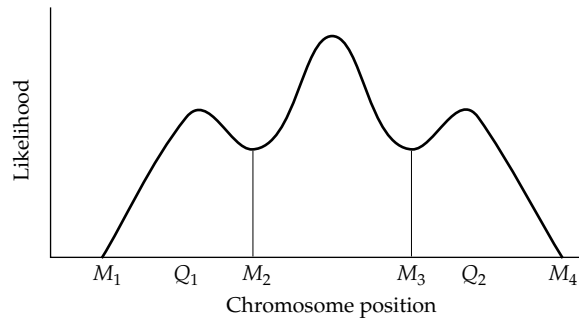
Figure 18.3   A false (or **ghost**) QTL generated by using a single-QTL likelihood function when two linked QTLs are actually present.

and Curnow 1992; Haley and Knott 1992; Wright and Kong 1997). Figure 18.3 gives an example of two linked QTLs embedded within four markers. Using a likelihood function that assumes only a single QTL, interval mapping correctly indicates likelihood peaks in the intervals flanked by $M_1 - M_2$ and $M_3 - M_4$. However, the resulting map also shows a much higher peak between $M_2 - M_3$, incorrectly suggesting the presence of a third QTL in this region. Under a single QTL model, analysis would start with this central peak first.

    One approach to mitigate the impact of a detected QTL on the signal from other QTLs is to include the closest marker (or flanking marker interval) as a cofactor in the analysis. This has the effect of reducing or removing background segregation variance contributed by the detected QTL, reducing the residual variance for the marker/interval under consideration, increasing the power for detection and improving the precision of estimates. Example 18.11 (below) provides a dramatic illustration of this. Adding a cofactor for a detected QTL to Equation 18.14a would yield

$$z_{ik} = \mu + bN_i + hM_i + e_{ik}$$

where M_i is the gene dosage in individuals i from the marker associated with the putative QTL, and N_i the dosage for the current marker being tested. Alternatively, one could first adjust the data to

$$z_{ik}^* = z_{ik} - hM_i$$

and then use these adjusted values to fit additional QTL effects. Dominance effects at the detected QTL can be accommodated by using Equation 18.14c.

    While this cofactor approach works well when the detected QTL is unlinked to the region being probed for additional QTLs, it can introduce serious bias when testing for QTLs linked to the detected location. In particular, Figure 18.3 showed that using markers associated with the highest peak would force the detected QTL in the wrong location. While specific tests for the presence of linked QTLs in adjacent intervals using sets of three overlapping markers have been suggested (Martínez and Curnow 1992, 1994b; Haley and Knott 1992), these are not without problems (Whittaker et al. 1996).

    Most of the single-QTL methods developed above can be extended to multiple QTLs by considering additional marker loci and using conditional probabilities for multilocus genotypes. This approach has been used to develop explicit models for two or three linked QTLs (e.g., Knapp 1991; Haley and Knott 1992; Martínez and Curnow 1992, 1994b; Jansen 1996; Satagopan et al. 1996; Wright and Kong 1997; Goffinet and Mangin 1998). We focus here on three particularly flexible regression-based approaches. The first is **marker-difference regression**, which considers all of the markers on one chromosome in a single analysis by using the regression of differences between the mean values of different genotypes. The second is **composite interval mapping** (**CIM**), which controls for both the effects of linked and unlinked QTLs by using the appropriate marker cofactors. This approach can be extended to **multiple interval mapping** (**MIM**) to simultaneously fit multiple linked intervals (Kao and Zeng 1997; Kao et al. 1999; Kao and Zeng 2002; Chen 2005; Verbyla et

al. 2007). Finally, Wright and Mowers (1994) and Whittaker et al. (1996) have shown how positional information for linked QTLs can be extracted from the regression coefficients of a standard multiple regression incorporating several linked markers.

Example 18.11.   Lin et al. (1995) examined flowering date through ML interval mapping of 370 F_2 individuals from a cross between cultivated and exotic sorghum (*Sorghum bicolor × S. propinquum*). Only a single QTL for flowering date was detected, and this accounted for 85.7% of the total variance. The data were then adjusted to account for the effects of this major gene by using $(z - b_i)$ in place of the trait value z for an individual with genotype i at a marker linked to the major gene. Here, the b_i ($1 \leq i \leq 3$) are the regression coefficients generated by a standard marker-trait regression using this marker locus. While the uncorrected F_2 phenotypic distribution was clearly bimodal, the adjusted data did not appear to deviate from normality. Using the marker-adjusted data, two additional QTLs for flowering time were found (both unlinked to the original QTL), accounting for an additional 8.3% and 4.2% of the total variance. This example illustrates the potential importance of including additional marker information into the analysis when multiple QTLs are present. In this case, removing the effects of a major unlinked QTL reduced the residual variance sufficiently to enable detection of additional, smaller-effect, QTLs.
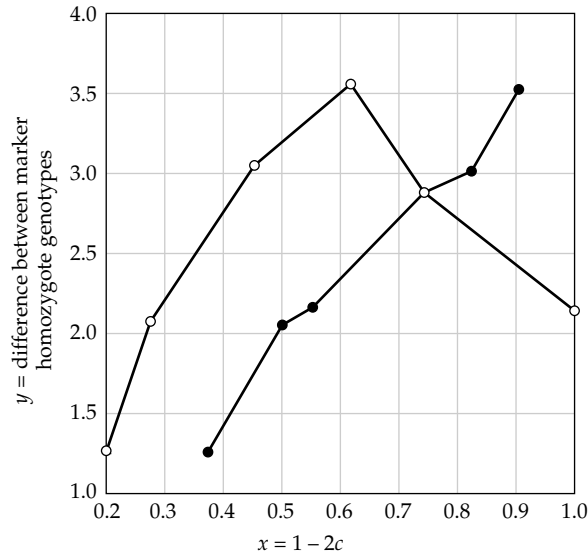

Marker-Difference Regression

Two groups (Kearsey and Hyne 1994; Hyne and Kearsey 1995; Wu and Li 1994, 1996a, 1996b) proposed a very simple, yet powerful, regression method that simultaneously considers all of the markers on a single chromosome. While the authors refer to this method as **marker regression** or **joint mapping**, we will use the more descriptive term **marker-difference regression**, or **MDR**, to emphasize that this approach is rather different from the regressions that we have considered up to this point. With MDR, each data point in the regression corresponds to a marker *mean value*, rather than to values for *single individuals* (as in our previous regressions). While this data structure results in far fewer points in the regression, the use of means allows the inclusion of individuals missing some marker information and also allows the joint incorporation of information from several experiments. Because the method only considers differences between the marker homozygotes, it is best suited for a collection of RILs or DH lines. It can also be formulated to apply to a BC analysis by contrasting the homozygote and heterozygote.

 The motivation for MDR follows from Equation 18.7a. We first present the method under the assumption of a single QTL to illustrate the main points before extending it to multiple QTLs. Suppose there are n linked markers on a chromosome containing a single QTL (with alleles Q and q). If the ith marker is at recombination frequency c_i from the QTL, the expected difference between marker homozygote means is

$$y_i = \mu(M_i M_i) - \mu(m_i m_i) = 2a(1 - 2c_i)$$

Thus, if we plot the differences y_i vs. $(1 - 2c_i)$ for each marker on the chromosome, the resulting n points are expected to fall on a straight line passing through the origin with slope $2a = \mu_{QQ} - \mu_{qq}$. Figure 18.4 illustrates this point, showing two regressions using the same set of marker differences but assuming two different locations for the QTL. The regression computed using the correct position of the QTL is linear, while that assuming the incorrect position is highly nonlinear. As with Haley-Knott regressions, one slides the position of a putative QTL along the chromosome, computing a regression at each point. The regression giving the best fit (i.e., the largest r^2) corresponds to the estimate of QTL position, and the slope of that regression divided by two provides an estimate of the QTL effect, a.

Figure 18.4   Marker-difference regression plot for the data given in Example 18.12. Open circles assume a QTL at map position 90 cM, closed circles a QTL at position 60 cM (the true position). Note that the relationship is linear when the correct position is used, but highly nonlinear under the incorrect position. Further note by comparing the open and solid points that their vertical positions are the same (each corresponding to a difference in marker homozgotes), but their horizontal (location x) values change as we change assumptions as to where the QTL is located (e.g., the solid point at 0.55 corresponds to the open point at 1.0)

To formally develop this approach, suppose that there are n linked markers scored along a single chromosome, and consider the regression

$$y_i = \overline{z}(M_i M_i) - \overline{z}(m_i m_i) = \beta\, x_i + e_i \tag{18.25}$$

with the $x_i = 1 - 2c_i$ values obtained by fixing the QTL position and then computing c_i for each marker (the marker-putative QTL distance). Because the residuals are correlated and potentially heteroscedastic, generalized least-squares regression (Chapter 10) must be used, with

$$\widehat{\beta} = \left(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y} \tag{18.26a}$$

which has sample variance

$$\sigma^2(\widehat{\beta}) = \left(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}\right)^{-1} \tag{18.26b}$$

where

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 - 2c_1 \\ \vdots \\ 1 - 2c_n \end{pmatrix} \tag{18.26c}$$

and

$$\mathbf{V}_{ij} = \begin{cases} \dfrac{\mathrm{Var}(M_i M_i)}{n(M_i M_i)} + \dfrac{\mathrm{Var}(m_i m_i)}{n(m_i m_i)} & i = j \\[4mm] (1 - 2c_{ij}) \sqrt{\mathbf{V}_{ii}\, \mathbf{V}_{jj}} & i \neq j \end{cases} \tag{18.26d}$$

where $\mathrm{Var}(M_x)$ is the sample variance of $\overline{z}(M_x)$, $n(M_x)$ is the sample size for marker class M_x, and c_{ij} is the distance between markers i and j (Wu and Li 1996a).

Assuming normally distributed residuals, from Equation A3.11a the residual sum of squares,

$$\mathrm{SS}_E = \widehat{\mathbf{e}}^T \mathbf{V}^{-1} \widehat{\mathbf{e}} = (\mathbf{y} - \mathbf{X}\widehat{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\widehat{\beta}) \tag{18.27}$$

follows a χ^2 distribution with $n-2$ degrees of freedom (n data points minus two estimated parameters, the QTL effect $\beta = 2a$ and the assumed position). The test for a significant QTL effect compares the SS_E for this regression with that for the regression assuming no marker effect ($y_i = \mu + e_i$). The SS_E for the reduced model is also χ^2-distributed, but with $n-1$ degrees of freedom. Recalling (Equation A5.14c) the additivity property of the chi-square, the difference in residual sums of squares for these two models follows a χ_1^2 distribution under the null hypothesis. Hence, the regression is significant at the α level if SS_E(reduced model) $- SS_E$(QTL model) exceeds $\chi_1^2(\alpha)$, the α-level cutoff for a χ_1^2. Separate regressions are computed for each chromosome, so that to obtain a genome-wide level of significance γ, each chromosomal regression is tested with significance level $\alpha = 1 - (1 - \gamma)^{1/C} \simeq \gamma/C$, where C is the number of chromosomes examined. Piepho (2001b) noted that the χ^2 assumption only approximate, and suggested an improved significance test for MDR.

Example 18.12.   Consider the following hypothetical data (plotted in Figure 18.4) generated by assuming a single QTL with effect $a = 2.0$ at map position 60 cM along a chromosome containing six markers:

Marker Position (cM)	$\overline{z}(M_i M_i) - \overline{z}(m_i m_i)$
10	1.26
25	2.06
50	3.04
65	3.54
75	2.90
90	2.15
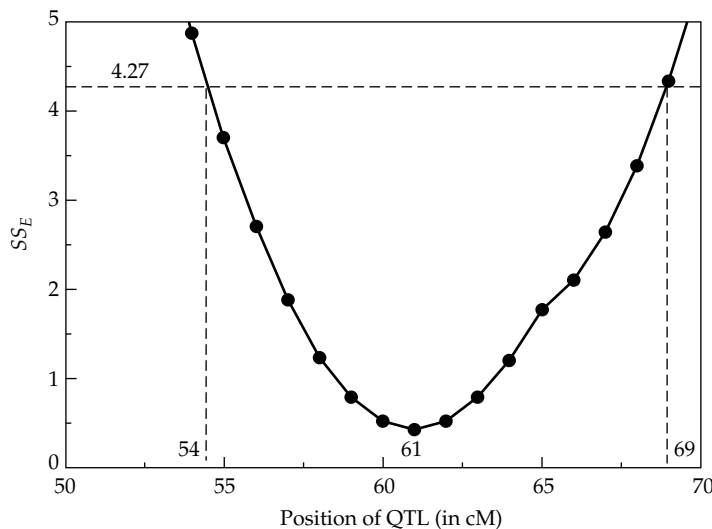
We assume that the variance associated with each marker class is the same with $\text{Var}(M_x)$ = 5, and that 50 individuals of each marker class were scored, giving $V_{ii} = 2 \cdot 5/50 = 0.2$. Using this and the c_{ij} values with Equation 18.26d fills out the rest of $\mathbf{V}$. For a MDR analysis, one computes a separate regression for each possible QTL position. Consider the regression for a QTL assumed to be at map position 50 cM. For the first marker, the QTL-marker map distance is 40 cM, which (assuming a Haldane map distance; Equation 17.3) translates into a recombination frequency of

$$c_1 = [1 - e^{(-2 \cdot 0.4)}]/2 \simeq 0.275$$

giving $x_1 = (1 - 2\,c_1) = 0.45$, and the data point associated this marker becomes (0.45, 1.26). Computing the remaining data points and applying Equations 18.26 and 18.27 gives a regression with $SS_E = 11.03$. After this procedure is repeated for all positions along the chromosome,

the resulting plot of SS_E vs. putative QTL position (shown above) exhibits a minimum value (0.43) at map position 61, and hence r^2 is maximized at this position (see Equation A3.15).

Whether the fit under the single-QTL model is a significant improvement over a model assuming no QTL can be assessed by comparing the error sum of squares of the QTL model ($SS_E = 0.43$) with the error sum of squares of the reduced (no QTL) model $y_i = \mu + e_i$. Because the QTL model fits an extra parameter, the difference in sums of squares follows a χ^2 distribution with one degree of freedom under the hypothesis of no QTL effect. For the reduced model, $SS_E = 19.16$, which is obtained by setting $\mathbf{X}$ equal to a vector of ones and applying Equation 18.26a. Hence, the QTL effect is highly significant as

$$\Pr(\chi_1^2 > 19.16 - 0.43) = \Pr(\chi_1^2 > 18.73) = 0.000015$$

The adequacy of the single-QTL model can be assessed by noting that if this model is correct, SS_E follows a χ_4^2 distribution (there are six data points and two fitted parameters, for four degrees of freedom). Because $\Pr(\chi_4^2 > 0.43) = 0.99$, SS_E is not larger than expected by chance, suggesting that there is no need to consider additional QTLs.

Using the estimated map position, the resulting regression has slope 3.84, giving the estimated QTL effect as $\widehat{a} = 3.84/2 = 1.92$. From Equation 18.26b, $\sigma^2(2\widehat{a}) = \left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}$ $= 0.16$, giving the standard error of $\widehat{a}$ as $\sqrt{0.16}/2 = 0.20$. Because SS_E follows a χ_1^2 distribution, the 95% confidence interval for QTL position contains those values giving regressions with SS_E not exceeding $\chi_1^2(0.05) = 3.84$ of the minimal SS_E value of 0.43 (i.e., SS_E values less than 4.27). This gives the confidence interval for the QTL position as 54 to 69 cM (see figure).

This approach easily extends to multiple QTLs. Recalling Equation 18.8a, if there are N linked QTLs, the jth of which is at recombination frequency c_{ji} from marker i, then (assuming no epistasis),

$$y_i = \mu(M_i M_i) - \mu(m_i m_i) = 2a_1(1 - 2c_{1i}) + \cdots + 2a_N(1 - 2c_{Ni}) \tag{18.28a}$$

This immediately suggests the multiple regression

$$y_i = \beta_1 \cdot x_{1i} + \cdots + \beta_N \cdot x_{Ni} + e_i \tag{18.28b}$$

where $x_{ji} = (1 - 2c_{ji})$ and $\beta_j = 2a_j$. The estimates are still given by Equation 18.26a, with $\mathbf{y}$ and $\mathbf{V}$ being defined in the univariate case, and

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_N \end{pmatrix} \quad \text{and} \quad \mathbf{X} = \begin{pmatrix} 1 - 2c_{11} & \cdots & 1 - 2c_{N1} \\ \vdots & \ddots & \vdots \\ 1 - 2c_{1n} & \cdots & 1 - 2c_{Nn} \end{pmatrix}$$

where N is the number of assumed QTLs, and n is the number of markers. As above, one computes the regression over the set of all possible QTL positions, with the estimates of QTL positions being given by the regression with the smallest SS_E value (or largest r^2). Each additional QTL reduces the degrees of freedom of SS_E by two (one for QTL effect, one for position). The test for whether adding another QTL significantly improves the fit compares the difference in the resulting two error sums of squares (for models assuming N versus $N - 1$ QTLs) with the appropriate critical value for a χ_2^2. Extension of the MDR approach to allow for interacting QTLs was given by Charmet et al. (1998).

Interval Mapping with Marker Cofactors

The careful reader will note that marker-difference regression does not require knowledge of the multilocus marker genotypes of any individual, as all that enters into the analysis are the population means for each separate marker. An alternative approach for dealing
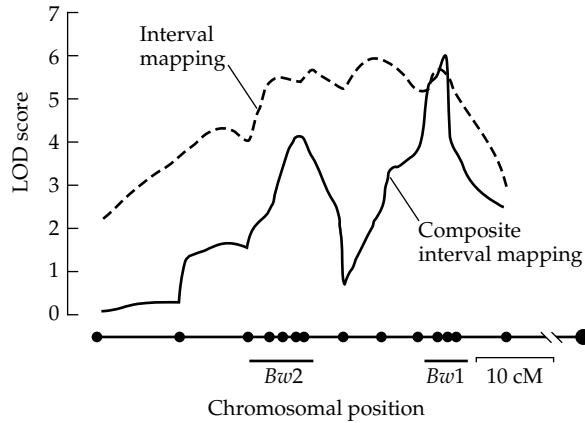
Figure 18.5 Likelihood plots for the X chromosome for QTLs influencing body weight in mice. The likelihood map under standard ML interval mapping (dashed line) shows a single very broad peak. Using the same data, the CIM likelihood map (solid lines) shows two distinct peaks. *Bw1* and *Bw2* denote the two putative body-weight QTLs, and the dots on the chromosome indicate the positions of the marker loci. Note that the ML interval mapping curve shows a ghost QTL, giving the largest peak between the true flanking QTLs (c.f., Figure 18.3). (After Dragani et al. 1995.)

with multiple QTLs that incorporates multilocus marker information from individuals is to modify standard interval mapping to include *additional markers as cofactors* in the analysis. Using appropriate unlinked markers can partly account for the segregation variance generated by unlinked QTLs (Jansen 1992, 1993b; Zeng 1993, 1994), while the effects of linked QTLs can be reduced by including markers linked to the interval of interest (Stam 1991; Zeng 1993, 1994; Rodolphe and Lefort 1993). This general approach of adding marker cofactors to an otherwise standard interval analysis, often referred to as **composite interval mapping (CIM)**, results in substantial increase in the precision of estimates of QTL position (Jansen 1993b, 1994a, 1994b, 1996; Jansen and Stam 1994; Jansen et al. 1995; Zeng 1994; van Ooijen 1994; Utz and Melchinger 1994). Figure 18.5 shows a rather dramatic example of the improvement using CIM over interval analysis.

Suppose the interval of interest is flanked by markers i and $i+1$. One way to incorporate information from additional markers is to consider the sum over some collection of markers outside the interval of interest,

$$\sum_{k \neq i, i+1} b_k \cdot x_{kj} \tag{18.29a}$$

where k denotes a marker locus and j the individual being considered. Letting M_k and m_k denote alternative alleles at the kth marker, the values of the indicator variable x_{kj} depend on the marker genotype of j, with

$$x_{kj} = \begin{cases} 1 & \text{if individual } j \text{ has marker genotype } M_k M_k \\ 0 & \text{if individual } j \text{ has marker genotype } M_k m_k \\ -1 & \text{if individual } j \text{ has marker genotype } m_k m_k \end{cases} \tag{18.29b}$$

This is simply a convenient recoding of the gene-dosage regression (Equation 18.14a) for each marker. Hence, b_k is an estimate of the additive marker effect for locus k. For a backcross or RIL design, each marker has only two genotypes and the indicator variable takes on values 1 and -1. More generally, if there is considerable dominance, the effects of the kth marker locus can be more fully accounted for by adding a dominance term to Equation 18.29a,

$$\sum_{k \neq i, i+1} (b_k \cdot x_{kj} + d_k \cdot w_{kj}) \tag{18.29c}$$
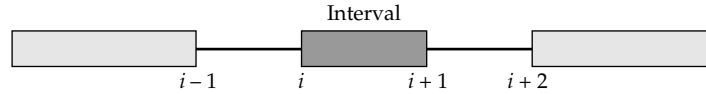
Figure 18.6   Suppose the interval being examined by CIM is between markers i and $i + 1$. Addition of the adjacent markers $i - 1$ and $i + 2$ as cofactors absorbs the effects of any linked QTLs to the left of marker $i - 1$ and to the right of marker $i + 2$. Their inclusion, however, does not remove the effects of QTLs present in the two intervals, $(i - 1, i)$ and $(i + 1, i + 2)$, flanking the interval of interest.

where

$$w_{kj} = \begin{cases} 0 & \text{if individual } j \text{ has marker genotype } M_k M_k \\ 1 & \text{if individual } j \text{ has marker genotype } M_k m_k \\ 0 & \text{if individual } j \text{ has marker genotype } m_k m_k \end{cases} \quad (18.29d)$$

Composite interval mapping proceeds by adding this regression term to the model being considered. For example, upon adding marker cofactors, the Haley-Knott regression focusing on the interval bracketed by markers i and $i + 1$ becomes

$$z_j = \left[\mu + a \cdot x(M_i) + d \cdot y(M_i) \right] + \sum_{k \neq i, i+1} b_k \cdot x_{kj} + e_j \quad (18.30)$$

Estimation of the QTL parameters (μ, a, d, c_i) for the interval proceeds as before, e.g., $x(M_i)$ and $y(M_i)$ are given by Equation 18.21 using marker loci i and $i+1$ as the flanking markers, with c_i being the putative QTL – marker i recombination frequency. For each c_i value in the interval, the regression given by Equation 18.30 is fitted (i.e., a, d, and the b_k), and (as before) the c_i value giving the regression with the largest r^2 is taken as the estimate of the QTL position. The significance of the interval can be tested by using Equation 18.24 to compare the full model (Equation 18.30) with the reduced model,

$$z_j = \mu + \sum_{k \neq i, i+1} b_k \cdot x_{kj} + e_j \quad (18.31)$$

which includes the marker cofactors but ignores the interval.

While most QTL mapping papers use some version of ML (or Bayesian) CIM, making it seem a gold standard, there is some debate on just which markers should be added. While there is no single solution, the two markers *directly flanking* the interval being analyzed should always be included. Suppose the interval of interest is delimited by markers i and $i + 1$ (Figure 18.6). Zeng (1994) showed that adding markers $i - 1$ and $i + 2$ as cofactors accounts for all linked QTLs to the left of marker $i - 1$ and to the right of marker $i + 2$. Thus, while these cofactors **do not account for the effects of linked QTLs in the intervals immediately adjacent to the one of interest** (i.e., the intervals $(i - 1, i)$ and $(i + 1, i + 2)$ in Figure 18.6), they do account for all other linked QTLs.

The number of *unlinked* markers that should be used as cofactors is unclear, as inclusion of too many factors greatly reduces power (Zeng 1994). Jansen and Stam (1994) recommend that the number of cofactors not exceed $2\sqrt{n}$, where n is the number of individuals in the analysis. A first approach would be to include all unlinked markers showing significant marker-trait associations (detected, for example, by standard single-marker regression). If several linked markers from a single chromosome all show significant effects, one might just use the marker having the largest effect. A related strategy, suggested by Jansen (1992, 1993b; Jansen and Stam 1994), is to first perform a multiple regression using all markers and then eliminate those that are not significant. Piepho and Gauch (2001) extend this into a more formal model selection framework.

A multiple-trait extension of composite interval mapping given by Jiang and Zeng (1995) offers improved power for QTL detection and increased precision in estimation

(relative to single-trait analysis) by incorporating the correlated error structure among traits (see also Ronin et al. 1995, 1998; Korol et al. 1995; Weller et al. 1997). Jiang and Zeng also develop likelihood-ratio tests for genotype $\times$ environment interaction and for tests of pleiotropy versus close linkage (one pleiotropic QTL vs. multiple linked QTLs each influencing separate characters).

Hypothesis testing and estimation for CIM follow by simple modifications of the appropriate results for interval mapping. Zeng (1993, 1994) showed that CIM test statistics for linked intervals are only weakly correlated, so that one can approximate each interval as an independent test. Zeng also found that the likelihood ratios within each interval are close to χ^2-distributed, so that an overall significance level of γ for an experiment examining m intervals can be obtained by equating the critical value within each interval to a χ^2 with significance level γ/m.

Resampling methods are easily extended to CIM. Doerge and Churchill (1996) suggested the following permutation test to account for multiple QTLs. A standard permutation test is first used to detect the marker with the greatest marker-trait association. Individuals are then divided (or **stratified**) according to their genotypes at this marker locus, and permutations are performed *within each stratified group* to generate new test statistics to find the next most significant QTL. This procedure is repeated until no significant effects are detected. Although permutation and bootstrap approaches are numerically intense, the rapid computation of solutions using Haley-Knott regressions makes these approaches feasible.

Finally, we note that other mapping approaches besides interval mapping can be improved by considering marker cofactors. For example, we can enhance the power of marker-difference regression by including unlinked markers to reduce the residual variance from unlinked QTLs. Because MDR uses the mean values for each marker, the individual data must be adjusted first to remove the effects from unlinked QTLs. Suppose n markers (unlinked to the chromosome of interest) are chosen because they show significant effects. The marker-adjusted value z_j^* of the original trait value of individual j, z_j, is given by

$$z_j^* = z_j - \left(\sum_{k=1}^{n} b_k \cdot x_{kj} \right) \tag{18.32}$$

and a MDR analysis is then performed using these adjusted values.

Detecting Multiple Linked QTLs Using Standard Marker-Trait Regressions

Consider the standard gene-dosage multiple regression of trait value on the single-locus genotypes at each of n markers,

$$z_j = \mu + \sum_{k=1}^{n} b_k \cdot x_{kj} + e_j \tag{18.33}$$

where j indexes the individual being considered, and the x_{kj} are given by Equation 18.29b (i.e., only considering additive effects). A rather remarkable finding, due to Wright and Mowers (1994) and Whittaker et al. (1996), is that the regression coefficients b_k for adjacent markers provide information on whether these markers flank a QTL. Further, the b_k can be used in many cases to obtain direct estimates of QTL effect and position.

When a QTL is **isolated**—an interval contains a single QTL and both flanking intervals are free of QTLs—the regression coefficients for the two markers immediately flanking the QTL depend only on this QTL and are not influenced by other linked QTLs (Stam 1991; Zeng 1993). A consequence of this finding is that markers flanking a QTL have regression coefficients of the same sign, while markers not adjacent to a QTL (i.e., there is at least one marker in the regression between the marker of interest and the nearest QTL) have expected regression coefficients of zero. Hence, one can simply scan the regression coefficients to see which intervals show support for a QTL (see Example 18.13).

Whittaker et al. (1996) further showed, for an isolated additive QTL, that the regression coefficients for the flanking markers can be directly used to estimate QTL effect and position.

Suppose markers i and $i + 1$ flank an isolated QTL. Whittaker et al. found that for an F_2 population, the estimated distance from marker i to the QTL is

$$c_i = \frac{1}{2} \left[1 - \sqrt{1 - \frac{4\,b_{i+1}\theta_i\,(1 - \theta_i)}{b_{i+1} + b_i\,(1 - 2\theta_i)}} \right] \tag{18.34a}$$

where $\theta_i = c_{i,i+1}$ is the distance between the markers. Likewise, an estimate of the QTL's additive effect a, independent of amount of dominance at this QTL, is given by

$$a^2 = \frac{[\,b_i + (1 - 2\theta_i)\,b_{i+1}\,] \cdot [\,b_{i+1} + (1 - 2\theta_i)\,b_i\,]}{1 - 2\theta_i} \tag{18.34b}$$

where both b_i and b_{i+1} have the same sign as a.

Example 18.13.   Whittaker et al. (1996) simulated 2000 F_2 progeny with three chromosomes, each with five markers evenly spaced at 25 cM (implying $c \simeq 0.2$ under Haldane's mapping function). QTLs were placed in the intervals flanked by markers (1, 2), (4, 5), (7, 8), (13, 14), and (14, 15). The multiple regression involving all 15 markers (Equation 18.33) had associated regression coefficients of:

Marker	1	2	3	4	5
b_i	−0.2996	−0.1422	−0.0221	0.2209	0.1956

Marker	6	7	8	9	10
b_i	−0.0189	−0.1922	−0.2404	0.0100	−0.0108

Marker	11	12	13	14	15
b_i	−0.0254	0.0371	0.3019	0.2644	0.3370

Looking for pairs of adjacent regression coefficients that have the same sign and are both significantly different from zero (as judged using standard regression tests, not shown) suggests evidence for QTLs in the intervals (1, 2), (4, 5), (7, 8), (13, 14), and (14, 15). The regression using just these nine markers had essentially the same SS_E as the full regression using all 15 markers, suggesting that none of the omitted markers are adjacent to QTLs (or they are adjacent to multiple linked QTLs whose effects cancel). However, removal of any one of the nine markers results in a regression with a significantly greater error sum of squares, supporting the hypothesis that all of these markers are adjacent to QTLs. Using these nine markers only, the new regression coefficients become

Marker	1	2	4	5	7	8	13	14	15
b_i	−0.2975	−0.1323	0.2296	0.1962	−0.2407	−0.2377	0.3145	0.2640	0.3355

Because the QTLs in intervals (1, 2), (4, 5), and (7, 8) appear to be isolated (no evidence for QTLs in adjacent intervals), Equations 18.34a and 18.34b can be used to estimate their effects and positions. For the QTL in the interval flanked by markers 1 and 2,

$$c_1 = \frac{1}{2} \left[1 - \sqrt{1 - \frac{4\,(-0.1323) \cdot 0.2\,(1 - 0.2)}{(-0.1323) + (-0.2975)(1 - 2 \cdot 0.2)}} \right] = 0.074$$

and the estimate of the squared effect of the QTL is

$$a_1^2 = \frac{[\,(-0.2975) + (1 - 2 \cdot 0.2)(-0.1323)\,]\,[\,(-0.1323) + (1 - 2 \cdot 0.2)(-0.2975)\,]}{1 - 2 \cdot 0.2}$$
$$= (0.442)^2$$

implying $a_1 = -0.442$ (because the regression coefficients $b_1, b_2 < 0$). Similarly, the estimates for the QTL in the interval (4, 5) are $c_4 = 0.105$ and $a_4 = 0.440$, while for the QTL in (7, 8), we

find $c_7 = 0.112$ and $a_7 = -0.494$. The estimated values were rather close to the true values used in the simulations ($a_4 = -a_7 = -a_1 = 0.447$, $c_1 = 0.07$, $c_4 = 0.11$, and $c_7 = 0.11$).

Whittaker et al. (1996) made a final important point that applies to all multiple-QTL methods. Unless a QTL is isolated—it is the only QTL in a particular interval and the flanking intervals lack QTLs—these methods cannot separate out the effects of multiple linked QTLs. In particular, if an interval contains multiple QTLs, we cannot estimate their effects and positions (or even the correct number of QTLs), a point stressed by McMillan and Robertson (1974; Example 17.2). While one obvious solution is simply to increase the marker density to the point where each QTL is indeed isolated, any increase in the marker density must be accompanied by a sufficient increase in sample size to ensure that a sufficient number of recombination events have occurred between adjacent markers.

MAPPING INTERACTIONS

Complications with Mapping Epistatic QTLs

When multiple QTLs are present, one must entertain the possibility of epistasic interactions between them. Detecting, and mapping, such interacting loci has a number of challenges, as we now detail. At first blush this task seems rather doable, as we have already detailed how to modify linear models to search for interactions between two-locus (or higher-order) marker genotypes (Equation 18.14d). Just as one could expand the gene-dosage single-marker regression (Equation 18.14a) to allow for dominance (Equation 18.14b), Equation 18.14d can be further decomposed into estimates of specific epistatic interactions (e.g., Kao and Zeng 2002; Zeng et al. 2005; Alvarez-Castro and Carlborg 2007). While detecting epistasis is thus seems straightforward, several critical (and delicate) issues remain. As a result, there is still considerable debate about its importance in QTL mapping (Carlborg and Haley 2004; Malmberg and Mauricio 2005). However, it is also becoming apparent that when epistasis is present, models that incorporate searches for pairwise effects often have higher power for detecting QTL than models that ignore possible interactions (Coffman et al. 2005; Evans et al. 2006; Marchini et al. 2006; Verhoeven et al. 2010).

Most of these concerns revolve around the massive expansion in the number of possible tests. For k markers, there are k potential tests for main effects, where single-locus marker genotypes show significant differences in trait means. However, there are $k(k-1)/2$ additional tests for (two-locus) epistatic interactions, for a total of $k(k+1)/2$. If 100 markers are scored (a modest number for most QTL experiments), there 4950 tests for two-marker interactions, for a total of 5050 tests. Hence, when epistasis is considered, the significance threshold to control the genome-wide error rate (GWER) is considerably more strict than when only single-marker tests are considered. With 100 markers, the standard Bonferroni correction when only single-marker tests are considered to obtain a GWER of 5% would be to test each marker using $\alpha = 0.05/100 = 5 \cdot 10^{-4}$. However, when epistatic interactions are included, to control the GWER all tests must now use $\alpha = 0.05/[101 \cdot 100/2] = 1 \cdot 10^{-5}$, resulting in a significant loss in power. There is also the philosophical issue in that if tests for epistasis are done, the number of *all* tests performed *must* be included in the GWER correction, even if none were significant. Thus, the cost of testing epistasis reduces the power to detect main effects if both classes of tests are performed simultaneously. Conversely, including interaction effects in a model can potentially reduce residual error, increasing the power to detect weak margin effects (Carlborg and Andersson 2002). A second issue is that we usually expect $k(k+1)/2 \gg n$, with potentially far more epistatic parameters to estimate than there are degrees of freedom (a **saturated model**).

The final concern is a more subtle issue in that the *genotypic sample size* for detecting two-locus interactions is less than that for testing single-locus markers. For example, with a design involving n F_2s, the expected number of $M_i M_i$ or $m_i m_i$ individuals is $n/4$, while the

expected number of $M_i M_i M_j M_j$ or $m_i m_i m_j m_j$ is $n/16$. Hence, two-locus marker differences are measured with higher standard errors (due to the smaller sample size for each multilocus genotype). This combination of more stringent testing coupled with higher standard errors poses serious power constraints on detection.

Epistatic Mapping Strategies

While little can be done to address the epistatic sample-size issue (short of creating mapping populations to test specific combinations of genotypes), three basic strategies have been proposed to deal with the number of possible tests.

1. **Test only among pairs of markers that both show significant main effects.** If only $s \ll k$ such markers are detected, then the additional number of tests, $s(s-1)/2$ is relatively modest.

2. **Test all interactions between a significant main-effect marker and all other markers.** These first two approaches are called **marginal scans**, requiring at least one marker showing a marginal effect. The rough number of additional tests here is $s(k-1)/2$, which, while far more than the first strategy, is still less than testing all combinations.

3. **Test all combinations, independent of whether they have significant main effects,** a **full scan.**

Most of the early scans for epistasis relied on strategy 1, making the assumption that if a locus has a strong interaction effect with other loci, at least part of that will also be reflected as a main (marginal) effect. This is often true when the loci involved exhibit somewhat extreme allele frequencies (minor allele frequencies are small). However, by the nature of a two-line cross, all segregating alleles have minor allele frequency of $1/2$, the optimal setting for large interaction effects with potentially small main effects (given the right type of epistasic interaction; e.g., Jana 1971).

What do the data say? Unfortunately, significant epistatic effects between loci with weak (i.e. nonsignificant) main effects can occur in QTL studies. Lark et al. (1995) reported several examples of soybean QTLs where one locus had a significant main effect, but its magnitude was strongly influenced by a second locus (which itself showed no significant main effect). In a mouse QTL experiment on lung cancer susceptibility, Fijneman et al. (1996) found that none of the four *Sluc* QTLs had detectable main effects under interval mapping, while *Sluc1* had a strong interaction with the (unlinked) *Sluc2*, and *Sluc3* interacted with the unlinked *Sluc4*. Similarly, Cubillos et al. (2011) found yeast growth loci only detectable by their interaction effects. A larger study involving 20 yeast traits by Bloom et al. (2015) detected 153 pairwise interactions. The vast majority of these had both loci showing significant marginal effects. However, in 36 cases just one locus was significant, and in 16 cases neither locus was significant. Finally, Holland et al. (1997) found 35 significant QTLs interactions in an oat experiment. Only four of these involved situations where both loci had significant main effects, 28 had only one of the loci with significant main effects, and for three interactions neither of the loci had significant main effects.

These observations suggestion that, depending on the nature of the study, strategy 3 must often be employed. Various schemes have been proposed for the challenging task of testing all possible combinations of markers. As summarized by Broman and Speed (2002), this is a **model selection** problem, choosing the best model from a set of candidates. They note that this is a two-step process: (i) choose the model dimensionality (**variable selection**; here number of QTLs and their corresponding associated markers), and then (ii) search within this space for the best (by some criteria) model. A number of approaches have been suggested to accomplish this task.

Kao et al. (1999), Zeng et al. (1999), Siegmund (2004), Storey et al. (2005), Baierl et al. (2006), Stich et al. (2007), Bogdan et al. (2008), and Zou and Zeng (2008) proposed a step-

wise selection process wherein one adds components to a regression (marker candidates for direct and/or interaction effects), and then uses some model selection criteria, such as AIC (Equation 16.9a) or BIC (Equation 16.9c), to compare non-nested models. Carlborg et al. (2000) and Carlborg and Andersson (2002) searched for interacting pairs using **genetic algorithms**, a computer science approach for maximization of a complex structure by essentially evolving solutions and having them compete (Holland 1975). Other search strategies have been proposed by Ljungberg et al. (2004), Pattin et al. (2009), and Wei et al. (2010). Jannink and Jansen (2001) proposed a one-dimensional search for loci with strong background interaction effects (which requires more than a two-way cross). Such loci then could be used to seed strategy 2 above. The basic logic of this idea has been independently suggested by a number of workers (Paré et al. 2010; Struchlain et al. 2010; Deng and Paré 2011; Hothorn et al. 2012).

Many of the methods for dealing with epistasis are set within a Bayesian framework (Appendix 7), with a detailed overview of different approaches given by Yi and Shriner (2008). The simulation methodology of MCMC sampler (Appendix 8) can offer an efficient alternative strategy to a **grid search**, wherein QTLs are placed at regular intervals (e.g., 1 cM) across the genome and models are fit given these assumed positions. Such a two-dimensional search examines points on a grid where location (x, y) corresponds to a QTL at genomic position x and a second at genomic position y. The resulting search is on a lower-triangular space ($x + y \leq L$, L being the total map length) and support could be presented as either two-dimensional profile plots (e.g., in the form of a **heat map** over the entire space, where the color of a region shows the strength of support for a pair of QTLs), or by only highlighting significant regions. See Malmberg et al. (2005) for an example of the latter in their search for fitness-related traits in *Arabidopsis*.

Yi et al. (2003) and Xu (2007) proposed using **stochastic search variable selection** (**SSVS**). Introduced by George and McCulloch (1993, 1997), the SSVS approach starts with some initial vector of weights associated with each of the potential model variables and then uses an MCMC sampler to continually update these weights. The resulting variables selected for final model is given by those that have sufficient posterior weight. Given this set of chosen variables, the optimal model parameters are then estimated, e.g., once the variables for a regression are chosen, the regression is them fit using those variables. Xu (2007) noted that large sample sizes (at least 600) are required for SSVS to be efficient.

A more popular approach is to use a **reversible jump MCMC** (Green 1995). Here, the dimensionality (number of QTL) is treated as a random variable, and one constructs an MCMC with two levels. On one level, the MCMC samples within a fixed number of QTLs. Level two, operating on a longer time scale, allows the sampler to jump up (or down) by one QTL. This generates a posterior distribution for the number of QTL, and the posterior for model parameters (locations and effects) within each QTL number class. Various versions of this method have been proposed for QTL mapping, e.g., Heath (1997), Sillanpää and Arjas (1998), Yi and Xu (2002a), Yi et al. (2003, 2007), and Yi (2004). A final approach is to use the idea of **shrinkage**, where all of the model parameters are estimated, but estimates are shrunk back towards a mean of zero, so that only those with substantial effects survive the shrinkage. The reader might think that is a version of SSVS, but the later uses these posterior weights to include or exclude model components which are then fit, whereas the shrinkage approach fits everything, and then shrink most back to zero. This is done by either treating model effects as random (Chapter 10) or by using penalized regression approaches (Example 20.4). Examples of using these approaches for QTL mapping are given Wang et al. (2005a), Zhang and Xu (2005), Zhang et al. (2005), Xu (2007, 2010), Yi and Xu (2008), Xu et al. (2009), Cai et al. (2011), and Huang et al. (2015).

Example 18.14.   Xu and Jia (2007) examined 7 traits in 145 DH barley lines, scored for 127 markers. Each DH line was replicated over an average of 25 plots, with the line means over all plots used for trait values, resulting in very accurate estimates of line genotypic values. Xu

and Jia used the method of Xu (2007) to fit 127 main effects and 127·126/2 = 8001 interaction effects (all pairs of markers). For each effect, REML (Chapter 32) is used to estimate the variance associated with that effect. These variances were then used to shrink the estimate (Equation 10.24b), with those effects with small variances being essentially shrunk to zero.

By construction, the only (pair-wise) sepistasis in DH lines is additive-by-additive. Xu and Jia found an average of 11.7 main effects and 6 interaction effects over the seven traits. The largest main effect accounted for around 18% of the phenotypic variance, while the largest interaction effect was only 2.6%. Taking the sum of all significant effects (judged by 20 reshuffled samples), main effects account for 35% of the variance, while interaction effects for only 6%. Perhaps the most interesting result was that *none* of the significant epistatic effects had *both* of their loci with significant main effects. Very surprisingly, 90% of all interactions had *neither* of the loci with a significant main effect, with 10% having only one locus with a significant main effect.

QTL x E Interactions

Another important potential interaction involves that between the environment and a detected QTL (**QTL** $\times$ **E**). We offer just a few brief remarks on QTL $\times$ E, which is examined in detail in Chapter 27. Two different approaches have been used to study this interaction: the consistency of marker-trait associations across environments and linear models that incorporate specific terms for marker $\times$ environment interactions. The simplest model would be a modification of Equation 18.14d, where b_k is now the effect for enviroment k and I_{ik} the QTL $\times$ E interaction term. The consistency measure is a very crude metric of QTL $\times$ E, simply asking whether a marker-trait association is detected in all scored environments. If it is, this is generally taken as evidence against G$\times$E, while detection of an association in only some of the environments is often taken as evidence for G$\times$E. However, a QTL can have a significant effect in all environments even in the presence of very significant G$\times$E interaction (Chapter 27). Likewise, low power can result in a QTL being detected in only some of the replicates of an experiment, even when its effects are identical across environments. Consistent with this expectation, Koester et al. (1993) found that QTLs with small effects are less likely to be detected across environments than are QTLs with large effects. By explicitly testing for marker $\times$ environment interactions (some of the I_{ik} are significant), linear model methods provide a more sensitive, and accurate, measure of G $\times$ E effects (Chapter 27).

SAMPLES SIZES REQUIRED FOR QTL EXPERIMENTS

Before investing the time and expense in a QTL mapping experiment, it is critical to have an understanding of the sample sizes required for the detection of QTLs of specified effects. The probability of detecting a marker-trait association is increased by increasing the difference between means and/or by decreasing the within-marker class (or residual) variance. Increasing the sample size reduces the within-class variance, while changing the experimental design can increase the difference between means. Further, the residual variance can often be decreased by adding explanatory factors to the model, such as sex- or age-effects.

This section examines a number of issues related to sample size, power, and accuracy. Sen et al. (2007) presented a general tool for aiding in the design of a QTL experiment that considers many of these issues. The important bottom line is that ***most QTL experiments are dramatically underpowered*** (de Koning and Haley 2005). A further point in terms of resource expenditures is that ***adding more individuals to the study is almost always a better strategy than adding more markers***. This is strictly true from a statistical sense. However, this strategy is also true from a practical standpoint, in that it is usually straightforward to store DNA from individuals, allowing for subsequent more dense mapping within focal regions. However, such denser mapping is only appropriate when there are a sufficient number of

recombinant events between markers, which is a function of nc. Increasing density lowers c, so that for fixed n (original sample size), as some point not enough recombinants are expected between the dense markers and they become completely redundant.

Power and Sample Size

The following discussion of sample size is restricted to single-marker t tests using the F_2 or backcross designs, where a QTL is indicated if the means for two alternative marker genotypes are significantly different. Using the theory of power calculations (reviewed in Appendix 5), simple expressions can be obtained for these designs. The broad utility of the results developed below is that both theoretical (Simpson 1989, 1992; Haley and Knott 1992; Darvasi et al. 1993; Rebaï et al. 1995; Hu and Xu 2008; Kao and Zeng 2010) and empirical (e.g., Stuber at el. 1992; deVicente and Tanksley 1993; Nodari et al. 1993; Damerval et al. 1994; Champux et al. 1995; Kennard and Harvey 1995; Mayer et al. 2004; Mayer 2005) studies show that t tests and more elaborate flanking-marker methods have very similar power for detection, especially when adjacent markers are no farther than 20 cM apart. Hence, the sample size expressions developed below provide a baseline for most designs.

We start by considering the t test for an F_2 design, where the presence of a linked QTL is indicated when $\overline{z}_{MM} - \overline{z}_{mm}$ is significantly different from zero. Suppose that the marker is completely linked to a single QTL with additive value a, in which case $E(\overline{z}_{MM} - \overline{z}_{mm}) = 2a$. Assuming that the distribution of phenotypes about each QTL genotype has constant variance σ_e^2, then if the numbers of MM and mm individuals measured are n_1 and n_2,

$$\sigma^2(\overline{z}_{MM} - \overline{z}_{mm}) = \sigma^2(\overline{z}_{MM}) + \sigma^2(\overline{z}_{mm}) = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\sigma_e^2 \qquad (18.35)$$

If n total F_2 individuals are scored, we expect only one in four to be a particular marker homozygote, giving $n_1 = n_2 = n/4$ and the expected variance $8\sigma_e^2/n$. If $r_{F_2}^2$ denotes the fraction of the total F_2 phenotypic variance $[\sigma_z^2(F_2)]$ due to segregation at the QTL, then $\sigma_e^2 = (1 - r_{F_2}^2)\sigma_z^2(F_2)$. Hence, if n is reasonably large, the observed difference in marker means is approximately normally distributed, with

$$\overline{z}_{MM} - \overline{z}_{mm} \sim N\left[2a, \, 8(1 - r_{F_2}^2)\sigma_z^2(F_2)/n\right] \qquad (18.36a)$$

Under the null hypothesis of no QTL, this difference is distributed as a normal with mean zero and variance $8\sigma_z^2(F_2)/n$.

Using the machinery developed in Appendix 5, the sample size required to have probability $1 - \beta$ of detecting a QTL using a test with an α level of significance becomes

$$n_{F_2} = \frac{8(1 - r_{F_2}^2)}{\delta_{F_2}^2}\left(\frac{z_{(1-[\alpha/2])}}{\sqrt{1 - r_{F_2}^2}} + z_{(1-\beta)}\right)^2 \qquad (18.36b)$$

where $z_{(p)}$ satisfies $\Pr(U \leq z_{(p)}) = p$ with $U \sim N(0, 1)$, and

$$\delta_{F_2} = \frac{\mu_{QQ} - \mu_{qq}}{\sigma_z(F_2)} = \frac{2a}{\sigma_z(F_2)} \qquad (18.37a)$$

is the difference in QTL means in units of F_2 phenotypic standard deviations. The variance contributed by F_2 segregation at this locus is $\sigma_Q^2(F_2) = a^2(2+k^2)/4$, where k is the dominance coefficient, implying

$$r_{F_2}^2 = \frac{\sigma_Q^2(F_2)}{\sigma_z^2(F_2)} = \frac{a^2(2+k^2)/4}{\sigma_z^2(F_2)} = \frac{\delta_{F_2}^2(2+k^2)}{16} \qquad (18.37b)$$

which for a completely additive QTL ($k = 0$) is $r_{F_2}^2 = \delta_{F_2}^2/8$. Using Equation 18.37b, we can alternatively express the required sample size in terms of the fraction of variation accounted for by the QTL,

$$n_{F_2} = \left(\frac{1 - r_{F_2}^2}{r_{F_2}^2}\right) \left(\frac{z_{(1-[\alpha/2])}}{\sqrt{1 - r_{F_2}^2}} + z_{(1-\beta)}\right)^2 [1 + (k^2/2)] \qquad (18.38)$$

Example 18.15.   What sample sizes are required to detect a completely linked QTL using a test with $\alpha = 0.05$ and $\beta = 0.1$ (i.e., a 5% probability of a false positive and a 10% probability of missing a true association)? Note that $\Pr(U < 1.96) = 0.975$ and $\Pr(U < 1.28) = 0.9$, so that $z_{(1-[\alpha/2])} = z_{(0.975)} = 1.96$ and $z_{(1-\beta)} = z_{(0.9)} = 1.28$. Substituting these into Equation 18.38 gives the following sample sizes for a completely additive ($k = 0$) and a completely dominant or completely recessive ($k \pm 1$) QTL whose segregation accounts for r^2 of the total F_2 variance:

r^2	0.5	0.3	0.1	0.05	0.01
Additive QTL	16	31	101	206	1046
Dominant QTL	25	46	151	309	1568

Note that the presence of dominance can significantly inflate the required F_2 sample size.

Turning now to the backcross designs, consider $B_1 = F_1 \times P_1$ (i.e., $MQ/mq \times MQ/MQ$). Here $n_1 = n_2 = n/2$, while $\mu_{QQ} - \mu_{Qq} = a(1 - k)$, giving

$$\overline{z}_{MM} - \overline{z}_{Mm} \sim N\left[a(1 - k),\ 4(1 - r_{B_1}^2)\sigma_z^2(B_1)/n\right] \qquad (18.39a)$$

Using the same logic as above, the required sample size is found to be

$$n_{B_1} = \left(\frac{1 - r_{B_1}^2}{r_{B_1}^2}\right) \left(\frac{z_{(1-[\alpha/2])}}{\sqrt{1 - r_{B_1}^2}} + z_{(1-\beta)}\right)^2 \qquad (18.39b)$$

with

$$r_{B_1}^2 = \frac{\delta_{B_1}^2}{4}, \qquad \text{where} \qquad \delta_{B_1} = \frac{a(1 - k)}{\sigma_z(B_1)} \qquad (18.39c)$$

For the B_2 population, the results are similar, except that the comparison is now $\overline{z}_{Mm} - \overline{z}_{mm}$ and $-k$ replaces k in the above expressions. Comparing the F_2 and the backcross design (for small to modest r^2), the ratio of samples sizes to achieve the same power is approximately

$$\frac{n_{B_1}}{n_{F_2}} \simeq \left[\frac{2}{(1 - k)^2}\right] \left[\frac{\sigma_z(B_1)}{\sigma_z(F_2)}\right]^2 \qquad (18.40)$$

Thus, if the backcross and F_2 phenotypic variances are the same, the backcross design requires twice as many individuals as an F_2 for a completely additive QTL ($k = 0$). When dominance is present, depending on its direction relative to the backcross population used, the backcross design can require more than twice as many individuals as an F_2 ($k > 0$ for B_1, $k < 0$ for B_2) or fewer individuals than the F_2. (If $k = -1$, the required sample size for the B_1 is only half of that for an F_2 design.) A further complication is that the phenotypic variance is generally rather different in the F_2 and backcross populations due to changes in the variance from segregation of background QTLs. In the F_2 population,

all QTL alleles have frequency 1/2, which gives maximum additive variance (provided all QTLs are additive). In a backcross, the minor allele frequency is 1/4, and additive genetic variance is often reduced significantly relative to that in the F_2. Thus, if background QTLs contribute significantly to the character, the backcross can show a reduced variance and more power.

If the QTL is not completely linked to the marker, two corrections are required for the above expressions. First, the difference in means for F_2 homozygous marker genotypes now estimates $2a(1 - 2c)$. A more subtle correction is that the variance about the marker means increases when $c \neq 0$, as the phenotypic distribution for each marker class is now a mixture of distributions with different means (Equation 16.5d). In spite of these complications, to a very good approximation the sample sizes required for a specific power of QTL detection are given by $n_0/(1 - 2c)^2$, where n_0 is the required sample size under complete linkage (Soller et al. 1976; Soller and Genizi 1978). Thus, the power to detect a linked QTL falls off as $(1 - 2c)^2$ decreases, being very weak when $c > 0.2$ (25 cM under the Haldane map).

Example 18.16.   Suppose we wish to have a 90% chance of detecting (using a test with $\alpha =$ 0.05) a QTL whose segregation accounts for 10% the total F_2 variance. Further assume that all of the genetic variation at this locus is additive. From Example 18.15, 101 individuals are required to detect this QTL using a completely linked marker. With a marker at recombination frequency c from the QTL, $n = 101/(1-2c)^2$, giving sample sizes of 281, 158, and 125 for $c =$ 0.2, 0.1, 0.05, respectively.

One can increase the power to detect a linked QTL either by increasing the number of markers (which decreases c and hence increases the difference between marker means) or by increasing the number of individuals genotyped (which decreases the sampling variance). To see the relative importance of each, note from Equation 18.36a that the t statistic has approximate expected value

$$E\left[\frac{\mu_{MM} - \mu_{mm}}{\sigma(\overline{z}_{MM} - \overline{z}_{mm})} \right] \simeq \sqrt{n}\,(1 - 2c)\left[\frac{a}{\sqrt{2}\,(1 - r_{F_2}^2)\,\sigma_z(F_2)} \right] \qquad (18.41)$$

The term in brackets on the right hand side is fixed for a given QTL, so that the test statistic scales with the square root of the sample size. Increasing the number of markers results in an increase in the test statistic, but there is a point of diminishing returns when markers are already closely spaced. For example, for $c = 0.2$ (corresponding to markers spaced 50 cM apart), moving to an infinitely dense map ($c = 0$) requires that only 36% as many individuals be scored to give the same power. However, for markers spaced at $c = 0.1$ and 0.05, these percentages become 81% and 90%.

Darvasi and Soller (1994b) showed, under rather general conditions, that the spacing of markers giving the highest chance of detecting a QTL, given the constraint of scoring a fixed total number of marker genotypes (marker loci $\times$ individuals) is 20 to 30 cM. Here each QTL is no further than 10 to 15 cM (and on average is within 5 to 7.5 cM) from any marker. Thus, for markers spaced 10 cM or closer, there is really little point in further increasing the marker density when the goal is simple detection of a linked QTL. Increasing marker density does become important if the goal is a highly precise estimate of QTL position or the dissection of a cluster of tightly linked QTLs. However, in order to exploit increased density, we also need to have a large enough sample size to realize a sufficient number of recombinational events between the dense markers. Without such recombinations, added markers are completely redundant. For example, Mayer et al. (2004) and Mayer (2005) used simulation studies with 500 F_2 individuals to show that increasing the marker density for 10cM to 5 cM significantly improves the precision of QTL estimations With fewer F_2 individuals, there are not a sufficient number of recombination events to exploit the increased

marker density. Thus, it very rarely pays to simply increase marker density without also increasing sample size. One exception is the use of AIL lines, where extra generations (and hence additional meioses) can exploit a denser marker map.

Example 18.17.   As mentioned in Example 18.4, Edwards et al. (1987, 1992) examined the same cross of two maize strains with two different designs. Their 1987 design used 1,776 F_2 individuals and 17 markers, while their 1992 design used 187 F_2 individuals and 114 markers. The two designs represent a tradeoff between increased marker density (1992 design) and increased sample size (1987 design), as both examined a somewhat similar number of total marker genotypes ($1776 \times 17 = 30{,}200$ vs. $187 \times 114 = 21{,}300$). Comparisons of c values in the two studies is problematic, given that only a fraction of the genome was covered in the 1987 study (about 40% of the genome was within 20 cM of a marker), while under the 1992 design most of the genome was 5 to 10 cM from a marker. Choosing $c = 0.25$ (1987 design) and $c = 0.08$ (1992 design), from Equation 18.41 the expected ratio of t statistics becomes

$$\frac{\sqrt{1776}\,(1 - 2 \cdot 0.25)}{\sqrt{187}\,(1 - 2 \cdot 0.08)} = 1.8$$

showing that (for these c values) the 1987 design had greater power.


ML interval mapping is expected to be somewhat more powerful than the simple single-marker t test, so the above results can be considered as upper bounds for the required sample size, although they are not greatly exaggerated. For example, the power of ML interval mapping to detect QTLs has been examined by several authors (Lander and Botstein 1989; van Ooijen 1992; Carbonell et al. 1993; Darvasi et al. 1993), who concluded that with a reasonable density of markers (one every 20 cM), 250 F_2 individuals are sufficient to detect a QTL whose segregation accounts for at least 5% of the F_2 variation. How does this compare with the required sample size for a t test? Because markers spaced at 20 cM intervals imply that a marker is within 10 cM from the QTL, using the result for $r^2 = 0.05$ from Example 18.15 gives the required sample size for a t test as $206/(1 - 2 \cdot 0.1)^2 = 263$. Hence, the above t test guidelines are also reasonable for ML interval mapping. More power calculations for QTL mapping are presented in Appendix 6.

Power under Selective Genotyping

The idea behind selective genotyping is that, historically, scoring (phenotyping) characters was much less expensive than scoring markers. Under selective genotyping, each choosen individual is genotyped directly, rather than genotyping the bulk mixture of all selected individuals (this latter approach is bulk segregant analysis, discussed in Chapter 17). Hence, if n individuals are scored and genotyped in a normal design, there may be merit in scoring a larger number of individuals $n_z > n$ for the trait value, and then choosing a subset $n_g \leq n$ of these for genotyping. Typically, the uppermost and lowermost fractions (p) of scored individuals are genotyped, giving $n_g = 2p\,n_z \leq n$. Darvasi and Soller (1992) noted that there is typically more linkage information in the tails, rather than the middle, of the trait distribution, so that observations from the tails have more of an impact that observations near the center (Figure 18.7). In particular, they show that selective genotyping by scoring n_z individuals and genotyping $n_g = 2\,p\,n_z$ gives the same power as an analysis genotyping all n individuals, when

$$n_z = \frac{n}{2p + 2\,z_{(1-p)}\,\varphi(z_{(1-p)})} \tag{18.42}$$

Here, $\varphi(z_{(1-p)})$ is the unit normal density function evaluated at $z_{(1-p)}$, where $\Pr(U > z_{(1-p)}) = p$ with $U \sim N(0, 1)$. Figure 18.7 plots the ratio n_z/n as a function of p. For example,
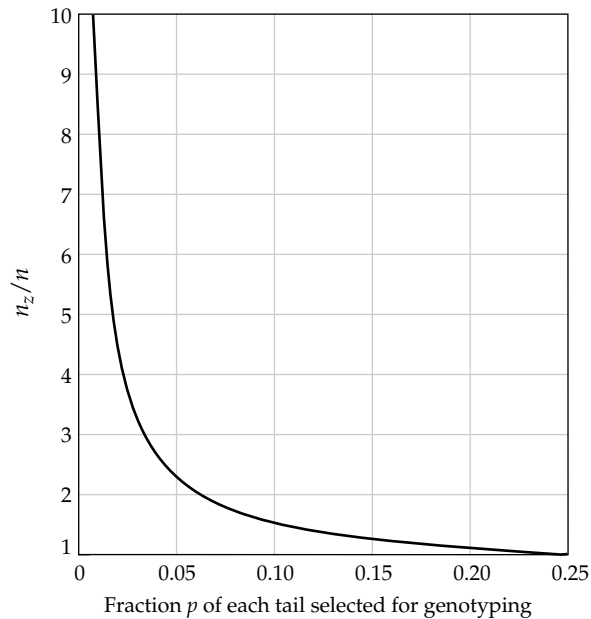
Figure 18.7   Under selective genotyping, n_z individuals are scored for the trait value, with the uppermost and lowermost fraction p of these being genotyped, giving $n_g = 2\,p\,n_z$. Here we plot, as a function of p, the number of individuals scored for the trait (n_z) that yields the same power as scoring and genotyping all individuals in a population of size n.

selective genotyping using the uppermost and lowermost 10% ($p = 0.1$) of the population requires that $n_z = 1.54\,n$ individuals be phenotyped but only $n_g = 2 \cdot 0.1 \cdot n_z = 0.3\,n$ be genotyped. Because a decrease in p reduces the number of individuals that must be genotyped but increases the number that must be scored for the trait, the optimal p value depends on the relative costs of phenotyping and genotyping each individual (Darvasi and Soller 1992). One of the ironies of the genomics revolution that *phenotypes* are now often much more expensive to score than genotypes. In theory, one could think about **selective phenotyping**, genotyping a sample of individuals, and then phenotyping those with more extreme values (by some genotype criteria). In essence, this is what genomic selection (Chapter 31) does.

This raises an extremely important issue that is often neglected in mapping studies. There are a variety of metrics (**quality control**, **QC**, approaches) that allow one to judge the quality of marker genotypes long past when the experiment was performed (Chapter 20). Conversely, they are generally no robust QC metrics for *phenotypes*. Once the trait data has been collected, there is often no way to examine its quality years later. Outliers, which are very informative if true, might well be measurement or recording errors. Poor trait data have the effect of reducing power. One real-world example is lumping different diseases as the same, confounding linkage information (Chapter 20). There is currently a major push, especially in plant breeding and some model systems (such as yeast and mice), to automate **high-throughput phenotyping**, using robotic systems to measure a number of traits with high accuracy and repeatability (Solberg et al. 2006; Liti et al. 2009, Dhondy et al. 2013; Araus and Cairns 2014; Ohya et al. 2015; Shi 2016; Bazakos et al. 2017). It is reflection of the power of modern genomics that phenotyping (**phenomics**) (Houle et al. 2010) is now the bottleneck in many mapping experiments, and that high-quality phenotypes (or lack thereof) can make, or break, a very expensive mapping experiment.

The Beavis Effect: Overestimation of Effects Declared to be Significant

The contributions of *detected* (i.e., those *declared to be significant*) QTLs can be substantially

Declared Significant

High power

μ*

Declared Significant

Low power

μ   μ*
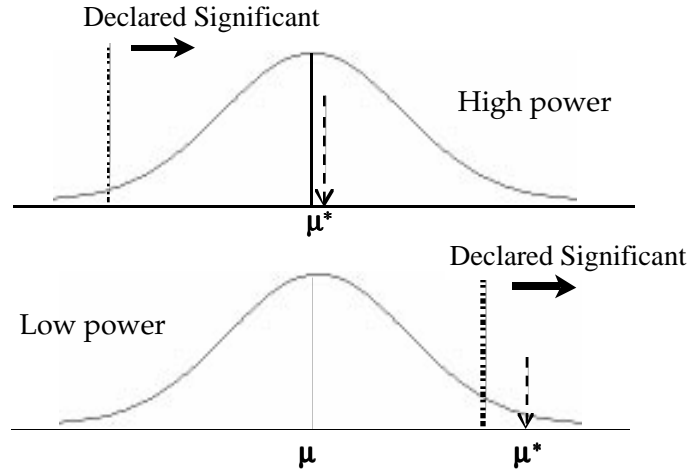
Figure 18.8   The connection between bias and power. Both panels represent the distribution of realizations of a test statistic for a marker effect, with μ the true value. The consequence of choosing only *significant* results is left-truncation of this distribution (Chapter 2). In the top panel, the power is very high, as most of the mass of the distribution (i.e., realizations of tests for a marker effect) is to the right of the significance threshold (dotted line). The resulting mean of this left-truncated distribution, μ^*, is close to the true value, and there is little bias. In the bottom panel, the power is low, as most of the distribution is to the left of the threshold, and hence most tests are not declared significant. The resulting mean of the left-truncated distribution (the tiny fraction of realizations declared significant) is substantially greater than the true value, resulting in significant bias. It is immediately apparent that the bias increases as power (area to the right of the threshold) decreases.

(often *very* substantially) overestimated (Lande and Thompson 1990; Beavis 1994; Utz and Melchinger 1994; Melchinger et al. 1998; Göring et al. 2001; Xu 2003). In QTL mapping, this upward bias in estimated effect size is called the **Beavis effect**, based on an important set of simulations by Beavis (1994) that will be discussed shortly (Figures 18.9 and 18.10). This phenomena is typically called the **winner's curse** in the GWAS literature (e.g., Zöllner and Pritchard 2007).

While the Beavis effect may at first seem a bit perplexing, it is simply a *consequence of sampling from a truncated distribution* (Göring et al. 2001; Xu 2003). As illustrated in Figure 18.8, there is a distribution associated with the realized value of a marker effect, with the true mean effect size, μ, influenced by sampling in any particular experiment. If that realized value exceeds some threshold, it is declared to be significant. The power of the test is the area of the distribution to the right of the significance threshold (Figure A5.1). Under a high power setting (top panel of Figure 18.8), most of the realizations are declared to be significant (most of the probability of the sampling distribution is to the right of the threshold). Given that we *condition* on an effect being significant, the mean value of a *significant result* is the mean, μ^*, of the left-truncated distribution. In the high power setting, this is very close to μ. Conversely, in a lower power setting (bottom panel), only a small fraction of the realizations exceed the threshold (are declared to be significant), so that the mean of the truncated distribution is *much* greater than the true value, $\mu^* \gg \mu$. Further note that this ***bias is expected to increase as power decreases*** (the significance threshold moves further to the right). Xu (2003) formalized this relationship between bias ($\mu^* - \mu$) and the power $(1 - \beta)$ of a test (β being the probability of a Type II error, a false negative) by using results for the mean of a truncated normal distribution (Equation 2.14), which yields

$$\mu^* - \mu = \sigma \cdot \left(\frac{\varphi(x_{[\beta]})}{1 - \beta} \right) \qquad (18.43)$$
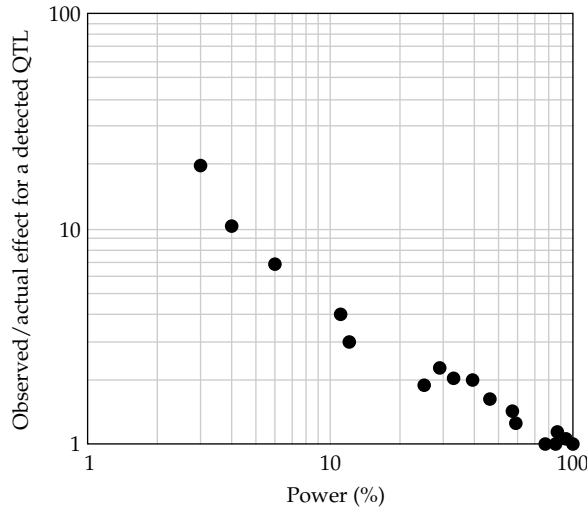
Figure 18.9 Relationship between the probability (power) of detecting a QTL and the amount by which the estimated effect of a *detected* QTL overestimates it actual value. (Based on results from a simulation study of Beavis 1994.)
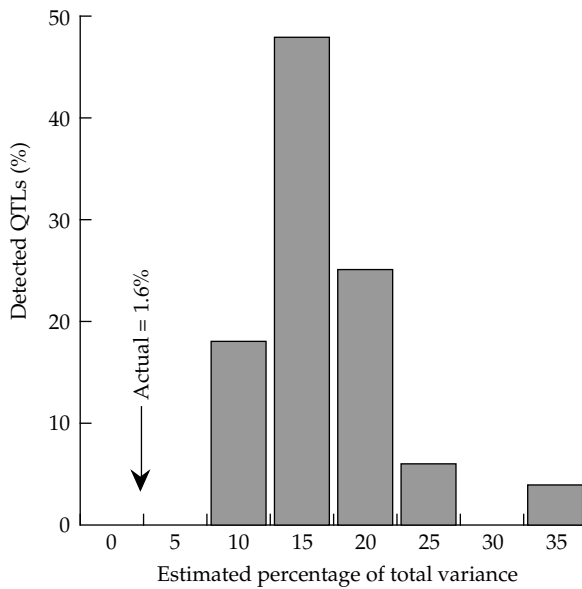
Figure 18.10 Distribution of the estimated effects of detected QTLs. Here 40 QTLs, each accounting for 1.6% of the variance, were assumed. Using 100 F_2 individuals, only 4% of such loci were detected. The average estimated fraction of total variation fraction accounted for by each detected QTLs was 16.3%, with the distribution of estimates skewed towards larger values. (From Beavis 1994.)

Here σ is the sampling variance for the test, $\varphi(z)$ is the value of a unit normal U at value z (Equation 2.11), and $x_{[p]}$ satisfies $\Pr(U \leq x_{[p]}) = p$. The weight on σ has values of 0.80, 1.27, 1.75, 2.07, and 2.67 when the power is 50%, 25%, 10%, 5%, and 1%, respectively. While this might appear to be a small effect, note that σ can often be $\gg \mu$, especially when the latter is small. As an aside, the factor multiplying σ is just the selection intensity associated with saving the upper β of a population (WL Chapter 14).

This inverse relationship between bias and power was seen in a classic set of simulations by Beavis (1994), shown in Figures 18.9 and 18.10. For example, a QTL accounting for 0.75% of the total F_2 variation had only a 3% chance of being detected with 100 F_2 progeny

with markers spaced at 20 cM. However, for cases in which such a QTL is detected, the average estimated total variance it accounts for is 15.8% (a 19-fold overestimate of the correct value), highly suggestive of a major gene. With 1,000 F_2 progeny, the probability of detecting such a increases to 25%, and each detected QTL on average accounts for approximately 1.5% of the total variance (only a twofold overestimate), and a more correct interpretation of the true genetic architecture. Further, these are the *average* values for the estimates. As shown in Figure 18.10, the distribution of observed effects is skewed, with a few loci having large estimated effects, and the rest small to modest effects. Such an observed distribution of effects, commonplace in QTL mapping studies, have usually been taken as being representative of the true distribution of effects. The simulation studies by Beavis showed that they can be *spuriously generated by a set of loci with equal effects*.

It is important to stress that the Beavis effect can be much more than a statistical bookkeeping error. In the extreme, it can result in a QTL experiment giving a *very misleading picture of the genetic architecture.* Consider the case where all of the QTL effects are small, so that a designs has low power ($1 - \beta \simeq 0$). If the number n of QTLs is large, it is likely that at least a few will be detected, $n(1 - \beta) \geq 1$. Such a scenario, wherein we detect a small number of QTLs that appear to account for a significant fraction of the total character variation, can lead to the false conclusion that character variation is largely determined by a few QTLs of major effect (Beavis 1994; Utz and Melchinger 1994). As both Figure 18.8 and Equation 18.43 show, when the power is very small, the bias among *significant* effects is very large.

Finally, a variety of approaches have been proposed to adjust for Beavis effects. The most straightforward (**subsampling**, e.g., Example 21.1) is to partition the original sample into detection and estimation subsamples. This is not an efficient use of an often limited sample size. One clever approach to accomplish this partition while still preserving power is based on a modification of bootstrap resampling (Sun and Bull 2005; Sun et al. 2011; Wu et al. 2005, 2006; Poirier et al. 2013; Huang et. al 2018). Recall that a bootstrap sample of size n is generated by drawing (with replacement) n individuals from the original sample. Some individuals will be sampled more than once, while others will not appear in that particular sample. The **out-of-sample** (or **bootstrap resampling bias reduction**, **BR-squared**) method uses the bootstrap sample for detection and the set of individuals *not* drawn (for a given sample) for estimation. This is a nonparametric approach, while number of authors have suggested parametic methods build around truncated distributions (Figure 18.8), such as truncated normals (Xu 2003; Palmer and Pe'er 2017; Panigraphi et al 2021), t random variables (Xiao and Boehnke 2011), or noncentral chi-squares (Xie et al. 2021).

Sample Size for a Desired Level of Mapping Resolution

Finally, there is a major difference between *detection* of a marker linked to one (or more) QTLs (the above issues of power), and the *precision* of the QTL location, i.e., the width of the confidence interval (CI) for its position. Darvasi et al. (1993) defined the **resolving power** of a design as the expected width of 95% CI for QTL position assuming a fully saturated map. The resolving power also places a lower limit on the effective marker density, as markers closer than the resolving power limit provide no additional information.

Darvasi and Soller (1997) performed extensive computer simulations with a known QTL whose position was estimated using ML mapping under either a BC or F_2 design. They assumed a QTL with effects of $-a : d : a$, where the variance about each genotype, σ_e^2, was set to one. Hence, the genetic variance σ_Q^2 contributed by this QTL is $(a + d)^2/4$ under the BC design and $(2a^2 + d^2)/4$ under the F_2 design, with the fraction r^2 of the *trait* variance accounted for by that QTL being

$$r^2 = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_e^2} = \frac{\sigma_Q^2}{\sigma_Q^2 + 1} \tag{18.44a}$$

Their simulations suggested a rather general approximation for the resolving power limit

(in cM) of

$$\text{CI} = \frac{3000}{m\,n\,\delta^2} = \frac{530}{n\,r^2} \qquad (18.44b)$$

Here the allelic substitution effect $\delta = a + d$ (BC design) and a (F_2, assuming $d = 0$), n is the sample size, and m is the number of **informative meioses** (those involving double heterozygous parents; Chapter 19), which is 1 for the BC design (one F_1 crossed to a homozygous P) and 2 for the F_2 design (two F_1s crossed). While the results of Darvasi and Soller are empirical observations extracted from simulations, Visscher and Goddard (2004) were able to derived them analytically, finding that using values of 3073 and 499, respectively, in Equation 18.44b are slightly more accurate. They also presented a general expression for other designs based on the noncentrality parameter (Equation A5.14a) of the test statistic.

Example 18.18.   Consider a QTL that accounts for 10 percent of the trait variance in the F_2, where 300 progeny are scored. What is the expected width of the 95% CI for its position? Applying Equation 18.44b with the Visscher-Goddard correction,

$$\text{CI} \simeq \frac{499}{300 \cdot 0.10} = 16.6$$

Hence, even for a major QTL (accounting for 10% of all variation) has a confidence interval length of roughly 17 cM (17 megabases assuming 1 Mb per cM). What sample size is required for a 95% CI of width 1 cM (~ 1 Mb)? Rearranging Equation 18.44b yields

$$n \simeq \frac{499}{\text{CI} \cdot r^2} = \frac{499}{1 \cdot 0.1} = 4990$$

Conversely, 300 F_2 progeny is a very high-power design for *detecting* a QTL accounting for $r^2 = 0.1$. Applying Equation 18.38 shows that a sample of around 100 has 90% power to detect such a QTL using a test with $\alpha = 0.05$. Hence, if the goal of the QTL experiment is mapping to a reasonably small interval, the typical QTL sample size is far too small. For QTLs accounting for 5% and 2.5%, the expected CIs are, 33.3 and 66.5 cM, respectively, while the corresponding sample size required for a 1 cM interval become 9980 and 19,960.

The Otto-Jones Estimator for the Number of QTLs

As was the case when estimating the number of factors in a line-cross analysis (the Castle-Wright estimator; Chapter 11), a QTL experiment underestimates the actual number of underlying loci generating a between-line divergence. Two factors cause this bias. The first is power, in that sample sizes are usually not sufficiently large to detect QTLs of small effect. In theory, this issue can be addressed by increasing the sample size or by using RILs and sufficient replication. Typically the number of detected QTL increases with the number of progeny in a cross (e.g., Schön et al. 2004; and examples below).

   The second factor is the size of an average linkage block in the cross progeny. As was the case for a traditional line-cross experiment, large fractions of any parental chromosome remain intact in most of their progeny due to insufficient recombination. Hence, a single QTL is often a linkage block containing multiple causative loci. Upon attempts to fine-map this QTL it **fractionates**, turning into a series of smaller QTLs. This is a common observation (reviewed by Flint and Mackay 2009; Mackay et al. 2009).

   Otto and Jones (2000) attempted use the detected number (n_d) and the average effect size ($\overline{a}$) of a detected QTL to obtain an estimate of the actual number of QTLs. Their estimate uses some feature of the Castle-Wright estimator, and makes two key assumptions. First, it assumes no antagonistic alleles, so that a high line has all + alleles relative to a low line. Second, it assumes that the QTL effects (a) are drawn from a specified distribution. The

logic for their estimator follows along similar lines to our above analysis of the Beavis effect. Assuming an exponential distribution for a, one can take the estimated average value ($\overline{a}$) above the detection threshold (γ) and use these values to estimate the probability remaining to the left of the threshold. There is a conservable body of theory on the distribution of effect sizes fixed by selection during an adaptive walk (WL Chapter 27), which suggests than an exponential distribution of such effects is not unreasonable assumption. As with the Castle-Wright estimator, Otto and Jones start with the observed difference (Δ) in trait mean between the crossed lines. For an exponential distribution of effects they obtain the estimator

$$\widehat{n}_{OJ} = \frac{\Delta}{\overline{a} - \gamma} \tag{18.45a}$$

The threshold value γ is a bit more delicate. It is certainly less than a_{min}, the smallest effect size for the n_d detected QTLs. Otto and Jones suggest the approximate estimate

$$\widehat{\gamma} \simeq \frac{a_{min}n_d - \overline{a}}{n_d - 1} \tag{18.45b}$$

provided that $\overline{a}$ is > 25% larger than a_{min} and that $n_d > 3$. Using simulations, they found that Equation 18.45b performed well under these conditions. Finally, they estimated the average effect of an undetected QTL as

$$E[a_{ud}] = \overline{a}\left(1 - \frac{\zeta}{1 - \exp[-\zeta/(1 - \zeta)]}\right), \quad \text{with} \quad \zeta = \frac{\gamma}{\overline{a}} \tag{18.45c}$$

Simulations suggested that Equation 18.45a is somewhat robust to the presence of antagonistic alleles, provided that their fraction is modest (less than 15%). A deeper concern would seem to be the Beavis effect, but Otto and Jones note that such effects would inflate *both* the estimates of $\overline{a}$ and γ, while it is the *difference* between these that is used in their estimator. Simulations suggest that unless the number of detected QTLs is small, this is not an unreasonable assumption. Confidence intervals, as well as an extension assuming the more general gamma distribution for effect sizes (of which the exponential is a special case; Appendix 7) are developed in Otto and Jones.

Finally, one of the more interesting findings of Otto and Jones was that when the detection threshold is small ($\gamma \leq 0.1 \cdot \Delta$), then as the true number of QTLs increases, the number of detected QTLs also initially increases, but this relationship is *not* monotonic. Rather, the expected number of detected QTLs eventually reaches some maximum, and then *declines* afterwards as the number of true QTLs increases.

ADVANCED DESIGNS

Biparental and Multiparental Mapping Populations

We have already introduced the use of a collection (or **panel**) of RILs and DH lines for QTL mapping. One of their advantages is that the line genotypic value for any trait can be measured with arbitrary precision through sufficient replication (Appendix 9). The issue of the optimal number of individuals to score within a line was examined by Belknap (1998) and Crusio (2004). Challenging characters, such as life span, which can have huge variation when based on single individual measurements, become well behaved when one considers the average value of a given line (e.g., Nuzhdin et al. 1997; Mackay et al. 2005). Similarly, replicated performance of a RIL or DHL over defined environments allows for very accurate estimates of QTL x environment interactions (e.g., Leips and Mackay 2000; Chapter 27). Further, a set of RILs has increased mapping accuracy relative to an F_2 panel due to the additional rounds of recombination during their formation (note that this is *not* true for most sets of DHLs). From Equation 18.5b, a panel of n RILs formed by selfing has roughly the same mapping resolution as a set of $2n$ F_2s, while RILs formed by brother-sister mating are equivalent of roughly $4n$ F_2s.
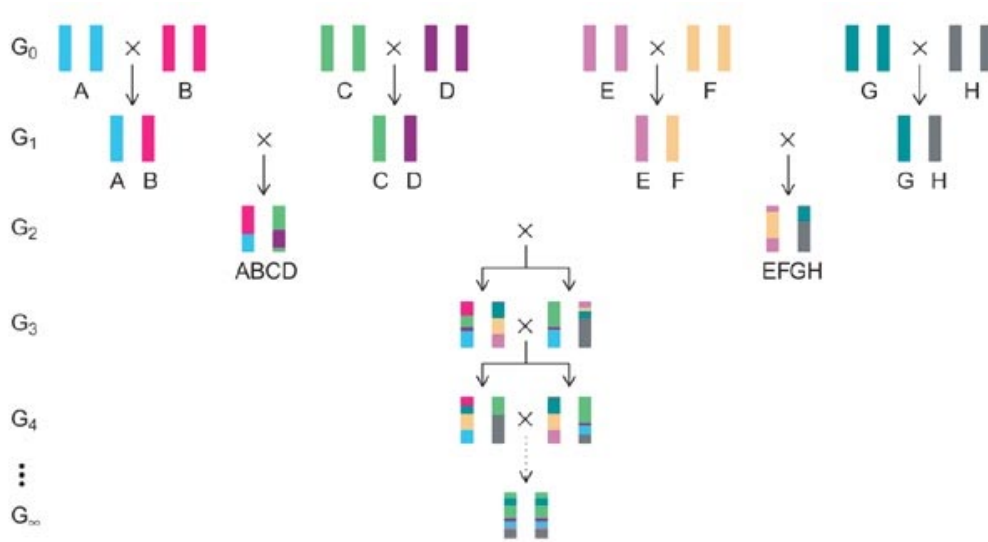
Figure 18.11 An eight-parent (A through H) cross leading to a set of multiparent advanced generation intercross (MAGIC) lines. Following generation 3 (the final wide-population cross), a series of brother-sister crosses are made within sublines until a collection of fully inbred lines are produced. Note the extra rounds for recombination relative to a standard biparental cross. Although there are any number of variations of this basic crossing scheme (such as additional rounds of recombination), the resulting lines are still called MAGICs.

A related strategy can be used in yeast, where meiosis generates haploid lines that can be asexually propagated, each instantly forming a clonal lineage (Nogami et al. 2007; Liti and Louis 2012; Cubilloe et al. 2013). One can also cross RILs, resulting in panels of **recombinant inbred intercross** (**RIX**) **lines** (Threadgill et al. 2002; Zou et al. 2005). In plants, an RIX panel is called an **immortalized F_2** (**IMF_2**) (Hua et al. 2002). Futher, one can use an advanced generation AIL as the starting population for RILs, resulting in **recombinant inbred advanced intercross lines** (**RIAILs**), with the notation (Valdar et al. 2006) of RIAIL10 implying RILs starting from an F_{10} AIL base population. RIAILs experience two stages of map expansion, one from the AIL, and the second from the RIL formation. To a good approximation, RIL formation via selfing adds roughly two additional generations of recombination, while brother-sister RILs adds around four. Hence, the final recombination value $\widetilde{c}$ for a RIAIL is given by Equation 18.4, using $t + 2$ and $t + 4$, for selfed and brother-sister RILs. See Winkler et al. (2003) and Teuscher et al. (2005) for a more exact treatment.

Finally, an important mapping resource for examining wild germplasm in an elite germplasm background are **advanced backcross lines** (Tanksley and Nelson 1996). While wild cultivars (such as landraces or even wild species) almost certainty contain a rich pool of exploitable genetic diversity, they also underperform current elite cultivars. The advanced backcross design of Tanksley and Nelson uses as the mapping panel a series of inbred lines (either RILs or DHs) constructed from the second or third backcross generation of a donor wild cultivar repeatedly backcrossed to a common parent elite line. For example, (wild x elite) progeny are crossed to elite (BC1), BC1 progeny are backcrossed to elite (BC2), and BC2 progeny are again backcrossed to elite (BC3). The expected fraction of wild alleles in these progeny are, respectively, 1/4, 1/8, and 1/16. Hence, the resulting lines (**AB-RILs** or **AB-DHs**, depending on whether RILs or DHs were used to extract inbred lines) contain small segments of wild material in an otherwise largely elite background. One of the first successful examples of QTL mapping using such panels is the work of Yun et al. (2006) for mapping fungal disease resistance traits in barley. Wang and Chee (2010) reviewed additional of applications of this approach in tomatoes, rice, barley, wheat, maize, and cotton.

Table 18.3   Relative strengths and weaknesses of biparental (e.g., F_2s, RILs, etc.), multiparental (e.g., MAGIC, NAM) and GWAS mapping approaches. (After Pascual et al. 2015.)

	Biparental	Multiparental	GWAS
Pop. developmental time	Intermediate	Long	Short
Mapping precision			
Common alleles	+	++	+++
Rare alleles	++	++	+
Total Recombination events	+	++	+++
Sample size	+	++	+++
Detection effect size	Large	Modest	Small
Population structure	No	No	Yes
Strength	Rare alleles	Multiple alleles	High precision
Libation	Poor precision	Large effort	High LD Pops

The advantage of any inbred line panel is that, once genotyped, any trait can be QTL mapped by simply measuring it across the lines and using the existing marker information. At this stage, QTL mapping is solely an issue of accurate phenotyping. Thus, while such panels are expensive and time consuming to produce, once generated they become an important community resource. The major limitation of most panels is that they are usually generated from a **biparent** (**two-parent**) **cross**, and thus have limited genetic diversity. Further, such panels are often produced with a specific trait in mind, influencing the set of parents crossed during line formation (e.g., crossing lines with extreme values of a focal trait). While other traits can be analyzed using the resulting panel, phenotypic diversity may be limited. As a result of these diversity limitations, a number of communities have made the expenditure (often as collaborative projects) to develop **multiple-parent crosses**, with the resulting progeny forming a **multiparental population** (**MPP**). A panel of RILs extracted from an MPP show greater diversity, both genetically and phenotypically, than a standard biparental panel. MPP panels are occasionally referred to as **second generation mapping resources** (Rakshit et al. 2012), and there are two different strategies for creating such panels: MAGIC and NAM lines.

As noted by Pascual et al. (2015), these various mapping strategies—biparental (BP; i.e., classical crosses) and multiparental (MP) linkage mapping, and association mapping (GWAS; Chapter 20)—can be regarded as complementary approaches. Table 18.3 summarizes key strengths and weaknesses of these different approaches. Linkage mapping has limited resolution due to a modest number of recombinations within the sample (with MP greater than BP) and generally much smaller samples than GWAS (again, with MP potentially having larger sample sizes than BP), and hence requiring larger effect sizes for detection. Casual allele diversity is lowest in BP and highest in GWAS. Conversely, GWAS (population-sample) approaches have lower power for rare-allele detection. When such alleles are fortunately included under an BP or MP design, their frequency is either 1/2 (BP) or ($1/k$) for MP designs, resulting in larger power. By sampling multiple families, MP designs have a larger chance (than BP) of capturing rare alleles.

MAGIC Lines

Multiparent advanced generation intercross (**MAGIC**) **lines** (Mackay and Powell 2006; Cavanagh et al. 2008; Cockram and Mackay 2018)—also called **recombinant inbred heterogeneous stocks, RIHS** (Valdar et al. 2003, 2006)—typically result from four- or eight-way crosses (Figure 18.11). In addition to greater diversity by using multiple founders, the additional rounds of recombination during parental crossing further expands the genetic map relative to a biparental population. Hence, one can think of these as multiple parent AILs, from which RILs are extracted. A few examples of this approach include the **mouse Collaborative Cross** (Threadgill et al. 2002; Churchill et al 2004; Aylor et al. 2011; Threadgill and Churchill 2012) involving eight diverse parental lines; the *Drosophila* **Synthetic**
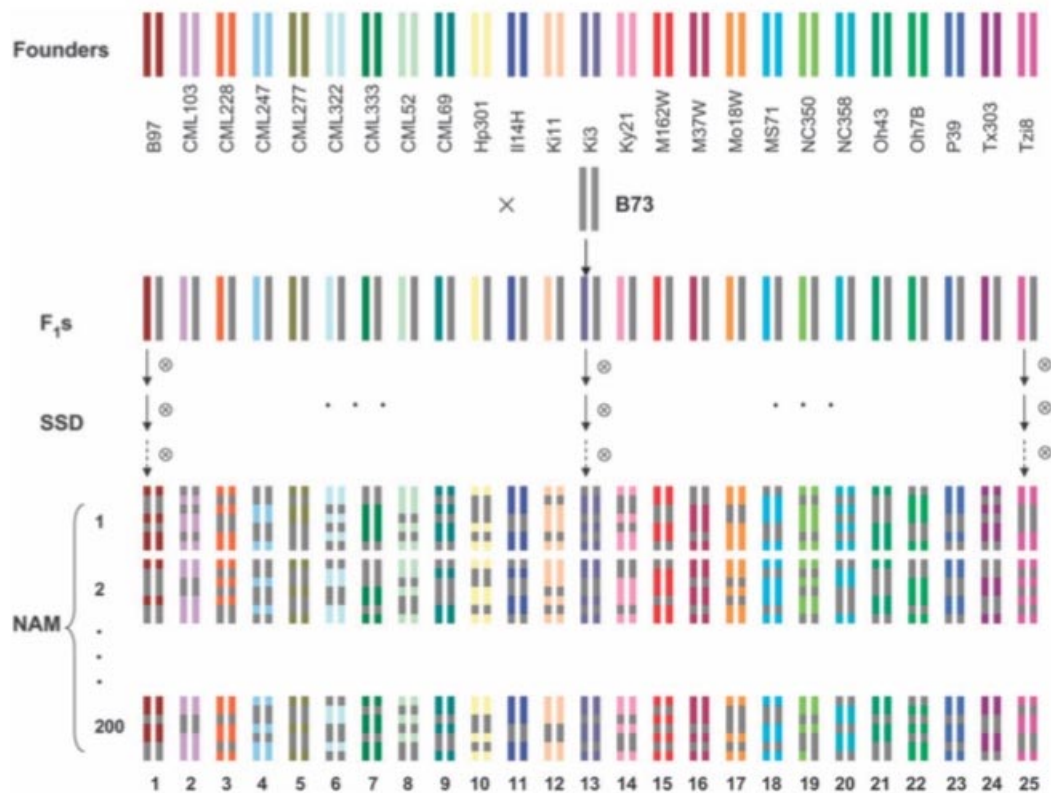
Figure 18.12 The maize NAM lines. Here 25 diverse donor parents were crossed to a common parental line (B73), resulting in 25 F_1 populations. For each cross, single seed descent (SSD) was used to form 200 RILs. The entire collection of 5000 lines forms the maize NAM panel. See Example 18.19 for more details. (From Yu et al. 2008.)

Population Resource (**DSPR**), a panel of over 1600 RILs formed from 16 diverse parents (King et al. 2012); the *Arabidopsis* MAGIC lines of Kover et al. (2009) of around 530 RILs from 19 founding parents; the *Arabidopsis* **multiparent RIL population** (**AMPRIL**), a set of over 500 RILs from four founder parents (Huang et al. 2011); a set of 1091 RILs with 8 founding parents in Winter-sown Wheat (Mackay et al. 2014), and a panel of haploid lines from a four-parent cross in yeast (Cubillos et al. 2013). Numerous other examples are given in Scott et al. (2020). One indication of the power of the MAGIC design is the result of Valder et al. (2006). They noted that 1000 MAGIC lines (starting the 8 founders) are sufficient to map a QTL accounting for 5% of the phenotypic variance to a region slightly less than one cM, a nearly ten-fold reduction relative to an F_2 mapping population (Example 18.18).

The construction of MAGIC lines involves three phases (Valdar et al. 2006), resulting in a large number of variations from the basic crossing structure shown in Figure 18.11.

1. Mixing. The first step is the choice of parental lines (by some criteria) followed by a subsequent crossing pattern to generate a **synthetic base population**. The crossing used is often called a **funnel**, distilling the variation from a number of input lines down into a final synthetic product. In addition to the crossing scheme shown in Figure 18.11, other schemes, such as a diallel (Chapter 25), namely all pairwise crosses are made and then subsequently intercrossed, have been used (e.g., the AMPRIL lines of Huang et al. 2011). Typically, the crossing pattern during the mixing phase is such that the all parental genomes are roughly equally represented in the synthetic base popula-

tion. However, if differential contributions are desirable, this can be accomplished by modifying the number of times a parent is crossed in this initial phase. Recombination during the mixing phase results in map expansion relative to a biparental cross. Huang et al. (2015) and Scott et al. (2020) discuss founder selection in more detail.

2. Maintenance.   Once a synthetic base populations has been constructed, one has the option of randomly mating this population for several generations to further expand the genetic map. Simulation studies by Yamamoto et al. (2014) found little additional gain by using less than two cycles of random mating following the formation of the base population, but using more than six generations significantly improved power under most settings. While random mating should not substantially alter the frequencies of genomic contribution from the founding parents, natural selection can (i.e., selection against deleterious genotypes formed during the mixing phase).

3. Inbreeding.   Once a sufficient amount of mixing and subsequent recombination has occurred, then either selfing or recurrent brother-sister mating is used to create the panel of RILs. Note that one could also use doubled haploids constructed from the final stage of the maintenance phase. Selection can be extremely important here, as lines are often lost during inbreeding, especially in species that are largely outbred in nature (Chapter 12). For example, 95% of all attempted RILs in the mouse Collaborative Cross became extinct during inbreeding (Shorter et al. 2017), and only 22% of F_2 plants in the maize NAM panel (Example 18.19) successfully produced lines that survived through five generations of selfing (McMullen et al. 2009).

While QTL detection and mapping in MAGIC lines follows the same basic logic as interval mapping, one complication is that there are more than two potential genotypes at a given site in a RIL. In particular, if k (typically 4 or 8) inbred lines are crossed to make a MAGIC line, then up to k distinct homozygotes can occur in any locus. At the core of most methods is some marker-based assignment of a given region in a MAGIC line to a founder line. Such assignments are typically probabilistic, e.g., for the marker haploid in a given region there is (say) a 57% chance that it came from parent 1, 20% from parent 2, etc. Following these assignments, then (as above) one uses regression or ANOVA of line value on the identity of the founder segment (e.g., Xu 1996; Mott et al. 2000; Valdar et al 2006; Verbyla et al. 2014; Wei and Xu 2016). See Huang et al. (2015) and Scott et al. (2020) for a more general review of such methods. Permutation testing and model selection issues are examined by Peirce et al. (2008) and Valdar et al. (2009). General reviews of MAGIC lines are given by Rakshit et al. (2012), Gatti et al. (2014); Huang et al. (2015), and Scott et al. (2020). Additional, and more technical, discussions are offered by Broman (2005; 2012a, 2012b).

NAM Lines

A second strategy for constructing an advanced mapping panel is a more structured approach, wherein a series of inbred **donor founder parents** (lines) are all crossed to a single **common founder parent**, resulting in a panel of **nested association mapping** (**NAM**) **lines** (Figure 18.12). This is loosely analogous to an NC II design wherein full sibs are nested within half sibs (Chapters 23 and 25). A NAM panel consists of series of RILs from each cross (the full sib analog), with a common parent over all of the crosses (the half sib analog). Originally developed in maize (Yu et al. 2008; McMullen et al. 2009; Giraud et al. 2014) and barley (Maurer et al. 2015), a number of crop species now have NAM panels (Gage et al. 2020; Gireesch et al. 2021). Design issues such as optimal family structure, replication, and crossing schemes have been examined by Yu et al. (2008), Stich (2009), Stich et al. (2010), and Liu et al. (2013).

As stressed by Yu et al. (2008), the power of the NAM design arises because it exploits *two time scales for recombination*. Within any given cross, we are using standard QTL linkage mapping on the associated collection of RILs, and hence relatively crude resolution due to just a few generations of recombination. As a result, the genome of any RIL from that

cross is a mosaic of chromosomal segments from the donor and common parent. However, associations between very tightly linked markers within a shared segment *across* donors reflect the impact of recombination over a much deeper time scale. Any such signal is due to population level LD among the donors, which is expected to persist only between markers that are sufficiently close that any LD between then has yet to fully decay. We examine how this signal can be exploited by association mapping in Chapter 20. The second powerful feature of NAMs is that any given QTL allele might be rare (i.e., confined to a single donor line). As such, there is not much of a signal *over* the donor lines. Conversely there is a very strong signal *within* the collection of RILs from its cross to the common parent, as its frequency is expected to be 50 percent. The NAM approach represents an attempt to maximize the advantages of linkage and LD mapping while minimizing their limitations. Yu et al. (2008) proposed using a simple regression analysis on marker allele copy number, with a cross-specific mean, for QTL detection, while Buckler et al. (2009) and Li et al. (2011) proposed an extension of CIM to the NAM design.

As with MAGIC populations, there are variations on this basic crossing structure, albeit still within the framework of a single common parent and a diverse collection of donor parents. The notation of **RIL-NAMs** and **DH-NAMs** is occasionally used (e.g., Li et al. 2016) to denote that the final lines were constructed from RILs (typically, using single seed descent in selfers, where a single seed from each plant is used to form the next selfed generation) or from DHs. Recall that RILs typically have effectively two rounds of recombination relative to DH lines, creating a tradeoff between a slightly more expanded map (RILs) versus potentially much faster (but often more laborious) formation of the final inbred lines (DHs). There are also modifications of NAMs using the advanced backcross approach mentioned above. Here instead of RILs being formed using the F_1s in a series of diverse lines crossed to a common parent, they are constructed using a series of BC2 or BC3 backcrosses to this common (elite) parent, yielding **advanced backcross NAMs**, **AB-NAM** (**backcross NAMs**, **BC-NAM** is also used). For example, in Figure 18.12, instead of the 25 starting crosses for RIL formation being F_1s, they would be (say) BC2s to the common parent. As example of this approach is Nice et al. (2016), where 25 wild barley accessions were twice backcrossed to an elite parent (Rasmusson) before RIL formation proceeds.

Example 18.19.   As shown in Figure 18.12, the maize NAM panel consists of 25 very diverse lines crossed to a single parent, B73 (Yu et al. 2008; McMullen et al. 2009). The common parent was chosen based on is widespread use, public physical map, and its available genome sequence. The donor lines consisted on nine temperate lines, more than half tropical lines, two sweet corns, and one popcorn line, capturing a good sample of global maize diversity. While the goal was for 5000 RILs, the eventual number was 4700, due (in part) to only modest success in using SSD to create inbred lines (a roughly 80% failure rate). A total of almost 140,000 recombination events occurred over the entire collection, corresponding to around three crossovers per gene (McMullen et al. 2009). Within each nested set of RILs, common-parent-specific (CPS) markers (alleles fixed in B73 that were otherwise rare in the donor parents) could quickly assign regions into having a B73 or a donor parent origin.

Other Multiple-cross Designs

As highlighted with both MAGIC and NAM lines, one can take any standard line-cross design (Chapters 11 and 25) and augment it with markers for QTL mapping. Various line-cross designs are usually employed for more than simply a matter of convenience, such as obtaining sufficient variation (e.g., dallelic crosses; Chapter 25), obtaining sufficient control over particular factors (such as maternal effects, e.g., reciprocal diallels; Chapter 25), and (most importantly), to generate desired informative contrasts, such as direct estimates of dominance or epistatic variances (e.g., NC III, triple test cross; Chapters 13 and 25). Many

of these same concerns arise in QTL mapping (e.g., Cockerham and Zeng 1996; Verhoeven et al. 2006). For example, some MAGIC lines are constructed using diallels to generate maximal diversity in the base population. Similarly, **test cross** designs are often used when heterotic groups are present (Chapter 13), wherein a mapping population (such as a set of F_2, RILs, etc.) are crossed to a common parent from a different heterotic group (e.g., Schön et al. 2004).

Methods for combinings QTL mapping information across multiple crosses have been developed by Rebai and Goffinet (1993; 2000), Xu (1998c), Liu and Zeng (2000), Zou et al. (2001), Yi and Xu (2001, 2002b), Li et al. (2005), and Xiao et al. (2016). These build on the fundamental idea of this chapter, conditional QTL probabilities given marker information (Equation 18.1), but involve more than a fair bit of bookkeeping and then more subtle questions such as whether to use a fixed or random analysis, and the appropriate framework (regression, likelihood, or Bayesian). See the aforementioned references for more details.

Segregation Distortion

Finally, many parental populations are chosen because of their wide difference in traits of interest, and marker loci in these inbred-line crosses often show strong depautres from Mendelian segregation, **segregation distortion** (e.g., Vallejos and Tanksley 1983; Edwards et al. 1987; Paterson et al. 1988; Bonierbale et al. 1988; Doebley and Stec 1991; Schön et al. 1993). As discussed in Chapter 17, this is a common feature when there is significant genetic divergence between parental lines. The impact of such distortion on QTL mapping, and designs to exploit some of its features, have been examined by Xu (2008), Xu and Hu (2009), Zhang et al. (2010b), and Cui et al. (2015).

SELECTED APPLICATIONS

The Nature of Transgressive Segregation

QTL mapping experiments provide insight into the nature of **transgressive segregation** (or **transgression**), wherein some F_2 individuals show more extreme trait values than are seen in either parental line. Note that this phenomena is distinct from heterosis (Chapter 13)—the overperformance of the *mean* in the F_1 (which typically declines in the F_2)—rather, it refers to the presence of outlier *individuals* in the F_2. One explanation for such outliers is nonadditive gene action, i.e., epistasis and/or overdominance. Alternatively, transgressive segregation could be caused by the parental lines being fixed for sets of *additive* alleles having opposite effects (**antagonistic alleles**), e.g., one line is fixed for $+-/+-$, the other $-+/-+$, which would generate more extreme genotypes in the F_2 than observed in either parent (e.g., $++/++$ and $--/--$). Such lines are said to contain **complementary QTL alleles.** This latter explanation is the one supported by most QTL studies (Rieseberg et al. 1999, 2003).

For example, Li et al. (1995) observed transgressive segregation for heading date in the cross of *Lemont* and *Teqing* strains of rice (*Oryza sativa*). Using 113 markers (with an average spacing 19 cM) and 2,418 F_4 lines, three regions that together accounted for 77% of the phenotypic variance in heading date were mapped. While the difference in average heading date between parental strains was just 6 days, one region from *Lemont* decreased heading date by 8 days (relative to its counterpart in *Teqing*), while another from *Teqing* decreased it by 7 days. Hence, these lines were fixed for alternative alleles at major loci, resulting in effects that largely canceled.

Transgressive segregation was also observed in 8 of 11 traits measured in a large F_2 population from a cross of *Lycopersicon esculentum* (cultivated tomato) and *L. pennellii*, its wild Peruvian relative (deVicente and Tanksley 1993). Of the 74 QTLs detected for these 11 traits, 36% showed alleles having effects on the character that were antagonistic to parental-line differences (alleles reducing a trait being found in parents from the large line, and vice versa). Pairwise epistasis was ruled out as a major cause for the observed transgressive seg-

regation, as the number of significant epistatic associations did not exceed that expected by chance. However, overdominance (or associative overdominance; Chapter 13) contributed in a few cases, with marker heterozygote means being more extreme than those for marker homozygotes. Likewise, Weller (1987) and Weller et al. (1988) observed that around 25% of the significant marker-QTL relationships in their tomato crosses were opposite in sign from the parental differences. A similar study based on a cross of two phenotypically similar cultivars of soybeans also noted that transgressive segregation due to complementary QTL alleles was quite common (Mansur et al. 1993).

Transgressive segregation has also been found when lines resistant for certain insect pests or plant pathogens have been crossed to sensitive lines. For example, of seven detected maize QTLs conferring increased resistance to the European corn borer in a resistant $\times$ sensitive cross, five came from the resistant parent, while two came from the sensitive parent (Schön et al. 1993). Dirlewanger et al. (1994) similarly found that a sensitive pea line carried a resistance allele for *Ascochyta* fungal blight that was not present in a resistant line.

A more general picture of the ubiquity of transgression was offered in two surveys by Rieseberg and colleagues. Rieseberg et al. (1999) examined 171 studies from a wide range of plant and animal crosses (both natural and domesticated), finding that 91% (155) of these had at least one transgressive trait in the study, and that 44% of the 1229 total traits showed transgression. Hence, transgression is *more the rule than the exception*. They noted that transgression tended to be more common in plants than in animals, and was more common in intraspecific, rather than in interspecific, crosses. They also made a series of predictions for when transgression is expected. One seems counterintuitive: *transgression should be more common in crosses between phenotypically similar, rather than phenotypically dissimilar, lines* (e.g., Cubillos et al. 2011). The logic is that more similar lines are likely be a mixture of different plus and minus alleles, while directional selection might drive down the fraction of antagonistic alleles in more diverged populations. WL Chapter 12 reviewed a number of QTL-based tests for detecting past selection on a trait that exploit this logic, such as an excess of plus alleles (Orr 1998). While Rieseberg et al. (1999) largely focused on the behavior of traits, Rieseberg et al. (2003) examined 3253 QTLs from 749 traits over 96 studies. Almost two thirds of the traits had at least one antagonistic allele, meaning that the potential for transgression is rather common, as suggested from the trait-only data. Consistent with the trait data, they also found that plants had a significantly higher fraction of antagonistic QTLs than did animals, and that such QTLs were more common in intraspecific, as opposed to interspecific, crosses.

Beyond obvious ramifications for breeding programs, transgressive segregation also has important evolutionary implications. Lewontin and Birch (1966) suggested that interspecies and wide-population hybrids can result in rapid adaptation to new environments (but also recall outbreeding depression; Chapter 13). If transgressive segregation in population crosses is the rule rather than the exception (as seems to be the case), then the (F_2) hybrids from such crosses possess the genetic variability to extend, perhaps considerably, the phenotypic range of a trait relative to either parental population. The results of Rieseberg et al. (1999, 2003) suggested mild caution with this idea in that intraspecific crosses inflate the importance of transgression over what might be expected in interspecific crosses. At a minimum, it is clear that mean phenotypic differences between lines are often very poor predictors of the architecture of their underlying genetic differences.

QTLs Involved in Protein Regulation

Any trait that can be *quantified* (assigned a numerical trait value, even simply as 0 or 1) can be subjected to a quantitative-genetic (QG) analysis. Historically, the traits considered were usually morphological or production phenotypes, such as body weight, height, measures of shapes, or yield. This view then evolved into considerations of aspects of physiology and behavior. As modern functional genomics blossomed, it was realized that the machinery of QG applies equally to quantifiable molecular traits, such as levels of mRNA transcripts, proteins, and metabolites. QTL that control the level of expression of a molecular target
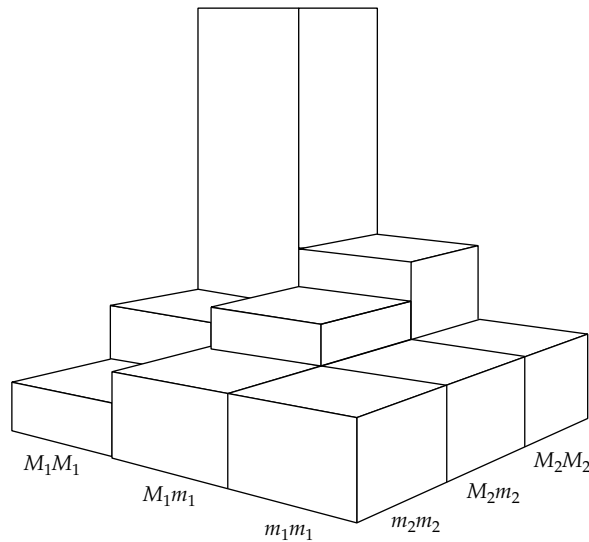
Figure 18.13   An example of epistasis for QTLs influencing protein volume in maize. Height indicates the amount of protein volume for each of the genotypes. Here the $M_1M_1M_2M_2$ marker genotype had the greatest effect on protein volume, while $M_1M_1m_2m_2$ had the smallest. (After Damerval et al. 1994.)

are denoted as **eQTLs**, for **expression QTLs**, and are a vibrant field of study (examined in more detail in Chapter 21). These have also been called **mQTLs**, for **metabolite QTLs**, where referring to control of concentration of metabolites (Breunig et al. 2014).

Here we focus on the seemly modest maize study of Damerval et al. (1994) that started the hunt for eQTLs. They analyzed the spot volumes of 72 anonymous (i.e., unknown) proteins (from a specific seed tissue) separated by high-resolution 2-D polyacrylamide gel electrophoresis. Genes controlling protein volume are, by definition, **regulatory genes** influencing the amount of that protein. Sixty F_2 individuals were scored with 76 RFLP markers, and both ML-interval mapping and single-marker ANOVA detected a total of 70 QTLs affecting 46 of the 72 proteins. Of these 46 proteins, 25 were influenced by two or more QTLs (up to a maximum of five). Of the 70 detected QTLs, 33 showed strict additivity, while the remaining 37 showed at least some dominance. The amount of variation in protein volume accounted for by a single QTL ranged from 16% (the lower detection limit for this sample size) to 67%, and the cumulative variation accounted for by all detected QTLs for each protein ranged from 37% to 90%. Perhaps the most striking observation was the presence of significant epistasis. Four proteins had QTLs that were only detected through epistasis (their single-locus effects were not significant). In all, 14% of the 72 proteins showed detectable epistasis (Figure 18.13).

QTLs Involved in Reproductive Isolation in *Mimulus*

Bradshaw et al. (1995, 1998) examined the genetic basis of floral differences between sibling species of monkey flower, *Mimulus lewisii* and *M. cardinalis*. Although the ranges of these species overlap and laboratory F_1 hybrids are completely interfertile, hybrid plants are not found in nature. Presumably, this is due to nonoverlap of pollinators. *Mimulus lewisii* shows characters typical of bumblebee pollinatiuon: pink flowers with yellow nectar guides, a wide corolla, small volume of highly concentrated nectar, and short anthers and stigma. *Mimulus cardinalis*, on the other hand, shows a typical suite of hummingbird-pollinated characters: red petals lacking nectar guides, a narrow tubular corolla, high nectar volumes, and long anthers and stigma.

Their intial study (denoted as experiment one; Bradshaw et al. 1995) used 93 F_2 plants and 159 markers in a cross between these species, and detected (using ML interval mapping)
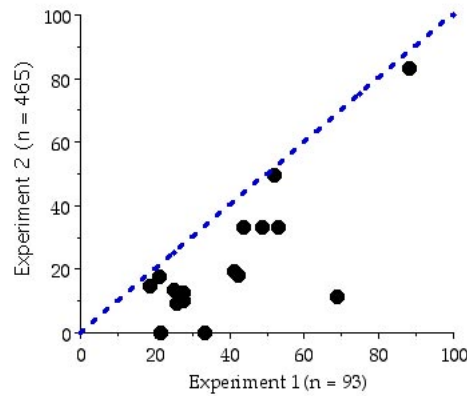
Figure 18.14 Relationship between the effects of detected QTLs for *Mimulus* pollination traits, expressed as percent of F_2 phenotypic variation (r^2), over two crosses with different sample sizes (Bradshaw et al. 1995, 1998). Experiment one measured 96 F_2 plants while experiment two measured 465. Note that all detected QTLs had larger estimated values in the smaller experiment, a clear example of the Beavis effect. The two values of zero from experiment two correspond to the petal anthocyanin QTLs that were not replicated.

Table 18.4 Number of detected QTLs influencing pollination characters involved in reproductive isolation between *Mimulus cardinalis* and *Mimulus lewisii* and their estimated individual effects (measured by % of variance explained). Due to sampling error, the sum of individual r^2 values exceeds 100% in a few cases. (After Bradshaw et al. 1995.)

	Number of QTLs	% Phenotypic Variance ($r^2 \times 100$)
Pollinator attraction characters		
Petal anthocyanins	2	33.5, 21.5
Petal carotenoids	1	88.3
Corolla width	3	68.7, 33.0, 25.7
Petal width	3	42.4, 41.2, 25.2
Pollinator reward		
Nectar volume	2	53.1, 48.9
Nectar concentration	2	28.5, 23.9
Pollination efficiency		
Stamen length	4	27.7, 27.5, 21.3, 18.7
Pistil length	2	51.9, 43.9

a number of QTLs for these pollination traits. As shown in Table 18.4, four of the characters appear to each have a QTL accounting for over 50% of the total F_2 variance, while all other characters had a QTL accounting for at least 25% of the total variance. Hence, it appeared that the bulk of the differences in pollination characters (and hence reproductive isolation) could be accounted for by one or two loci for each character. However, with these small sample sizes, some caution is in order, given our previous comments about overestimation of QTL effects when power is low (Beavis effects).

To address this concern, a follow up study (experiment two; Bradshaw et al. 1998) increased the sample size five-fold to 465 F_2 plants. The two major QTLs for petal anthocyanins from experiment one were not reproduced. However, all of the other QTLs were, albeit (in most cases) with a substantial reduction in the fraction of variation explained (Figure 18.14). Clearly, the Beavis effect inflated the initial importance of regions in experiment one due to its small sample size. Ignoring anthocyanin, of the initial major QTLs ($r^2 \geq 0.25$),

Table 18.5   Character differences between maize and teosinte (primitive maize race Reventado × *Zea mays parviglumis*). Listed are mean character values (Means), the number of detected QTLs (N), the r^2 value for the largest (Max) and smallest (Min) detected QTLs, and the total r^2 for a model containing all detected QTLs. Locations for the QTLs detected in this cross are plotted as the upper bars in Figure 18.15. (After Doebley and Stec 1993.)

|	Means		QTLs			
	Maize	Teos.	N	Max	Min	r^2
Plant Architectural Characters						
Lateral branch internode length (LBIL)	0.7	21.9	5	0.45	0.05	0.63
Number of branches (LIBN)	0.0	5.8	4	0.24	0.04	0.42
% male primary lateral inflorescences (STAM)	0.0	97	5	0.23	0.05	0.52
No. secondary ears/lateral branch (PROL)	1.0	8.4	7	0.25	0.04	0.63
Ear Characters						
Cupules along a single rank (CUPR)	37.4	5.3	6	0.25	0.04	0.61
Disarticulation: 1 = none, 10 = full (DISA)	1.0	10.0	6	0.42	0.04	0.60
Glume score: 1 = soft, 10 = hard (GLUM)	1.0	10.0	2	0.41	0.08	0.75
% cupules with only one spikelet (PEDS)	0.0	100	5	0.25	0.08	0.69
Number of rows of cupules (RANK)	5.6	2.0	6	0.36	0.05	0.87

less than half retained an $r^2 \geq 0.25$ in experiment two. However, most of the original traits still had large-effect QTLs ($r^2 \geq 0.15$). As might be expected, the larger sample size of experiment two detected 16 additional QTLs not seen in the original cross. In experiment one, the smallest r^2 for a detected QTL was 0.187, while it was 0.033 in experiment two. Further, 7 of the 16 additional QTLs had effects less than the smallest detected value from experiment one.

Following detection of these QTLs, the authors took F$_2$ hybrids into the field and estimated which traits were most predictive of bee versus hummingbird attraction (Schemske and Bradshaw 1999; Bradshaw and Schemske 2003). The most predictive traits still had significant single QTLs ($r^2 \geq 0.15$) in the more powerful (and less biased) experiment two. In an especially informative experiment, Bradshaw and Schemske (2003) introgressed alternative alleles at the QTL for yellow carotenoid pigment (which had an $r^2 = 0.83$ in experiment two) into otherwise pure background of the other species. When the yellow allele from *M. lewisii* (bee pollinated) was introduced into an otherwise *cardinalis* background (hummingbird pollinated), the resulting NIL received a 68-fold increase in bee visits relative to the wildtype, while a NIL with a *cardinalis* allele in an otherwise *lewisii* background showed a 74-fold increase in hummingbirds visits.

QTLs Involved in the Differences Between Maize and Teosinte

Maize and teosinte are dramatically different (see Figure 5.2), to the point that they were originally placed in separate genera. Hybrids, however, are fully interfertile and maize is believed to have resulted from domestication of teosinte (Beadle 1980; Doebley 1992, 2004). In an elegant series of papers, Doebley and colleagues (Doebley et al. 1990, 1994, 1995a, 1995b, 1997; Doebley and Stec 1991, 1993; Dorweiler et al. 1993; Hubbard et al 2002; Wang et al. 2005b; Studer and Doebley 2011; Whipple et al. 2011; Wills et al. 2013; Wang et al. 2015) characterized many of the genes involved in these dramatic differences.

Maize and teosinte have major differences in plant architecture (Figure 5.2; Table 18.5). Teosinte has multiple long lateral branches, topped with male inflorescences (tassels). In maize, these branches are very short and topped with ears (female inflorescences). These differences in plant architecture can be quantified by considering four characters: internode length on lateral branches (small in maize, long in teosinte), the number of branches (none to few in maize, many in teosinte), percentage of lateral branches topped with tassels as opposed to ears (mostly tassels in teosinte, ears in maize), and the number of secondary ears
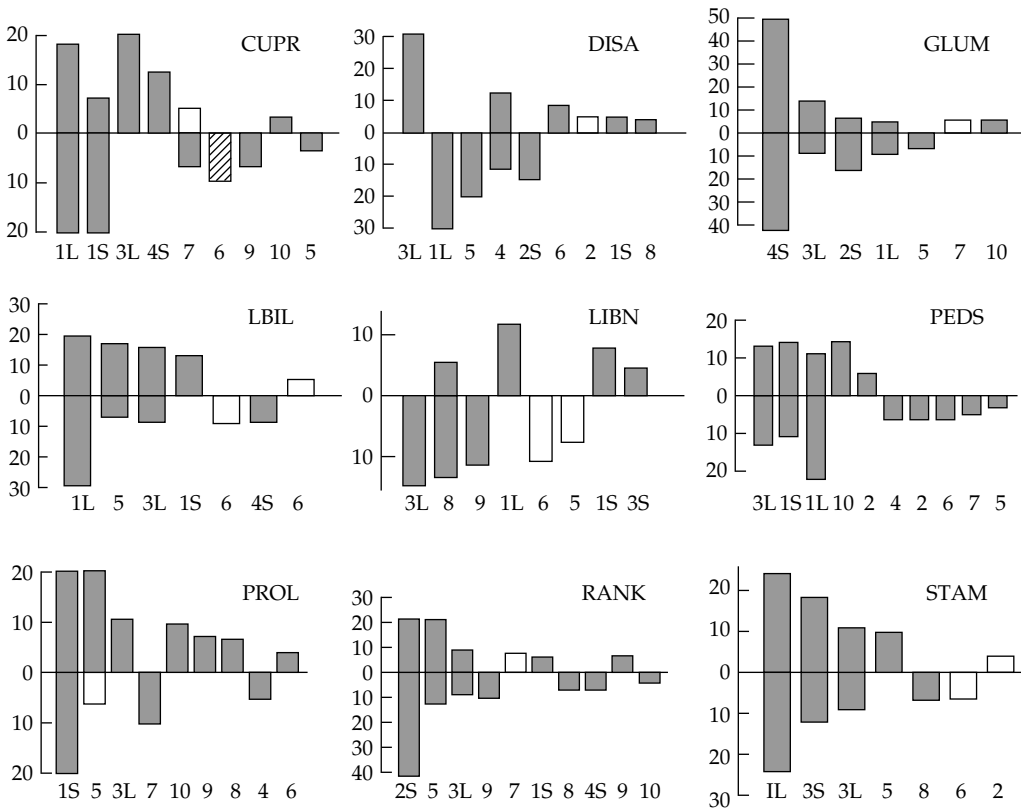
Figure 18.15 QTLs accounting for differences between maize and teosinte in the nine characters listed in Table 18.5. Each bar represents a detected QTL, with bar height indicating its r^2 value. Acronyms for characters are listed in Table 18.5. Chromosome position (chromosome number and short vs. long arm) is indicated under each bar. The upper bars refer to QTLs detected in a cross between a primitive maize strain and one subspecies of teosinte (*Z. m. parviglumis*), while the lower bars are for a cross involving a different primitive maize strain and another teosinte subspecies (*Z. m. mexicana*). Aligned bars denote QTLs mapping to very similar positions in both crosses. White bars indicate QTL effects in the opposite direction from parental phenotype (antagonistic effects), while the cross-hatched bar indicates apparent overdominance. (From Doebley and Stec 1993.)

on each lateral branch (few in maize, many in teosinte). Table 18.5 gives the mean values for these characters in maize and teosinte.

Differences in the structure of the female inflorescence (the ear) are even more dramatic (Figure 5.2; Table 18.5). The teosinte ear has 5–10 cupulate fruitcases arranged in pairs. Each of these has a single spikelet that gives rise to a kernel, resulting in 10–20 kernels per teosinte ear. Each mature fruitcase is covered by a hardened outer glume that seals in the kernel, making harvesting very difficult. In contrast, the maize ear is composed of 100 or more cupules (arranged in multiple rows rather than pairs of rows), each cupule containing two spikelets, leading to two kernels per spikelet. These changes result in the maize ear having an order of magnitude more kernels than the teosinte ear. The maize outer glume is soft, so the kernels remain exposed for easy harvesting. Finally, while the teosinte ear easily disarticulates (**shattering** to scatter seeds), kernels on the maize ear stay intact, further facilitating the harvesting of kernels.

In their initial studies, QTLs were mapped in two different crosses, each involving a different primitive maize race and a different subspecies of teosinte (Doebley et al. 1990;

Table 18.6   QTLs influencing age-specific weight and age-specific growth rates in mice, measured at weekly intervals. Early, Middle, and Late correspond to growth from 1 to 3 weeks, growth from 3 to 6 weeks, and growth from 6 to 10 weeks, respectively, while 6-week refers to growth from 1 to 6 weeks. (After Cheverud et al. 1996.)

	Age-specific weight (weeks)									
	1	2	3	4	5	6	7	8	9	10
No. of QTLs	7	10	16	13	15	15	14	14	16	17
Total r^2	0.29	0.30	0.56	0.52	0.59	0.63	0.64	0.56	0.67	0.76

	Age-specific growth			
	Early	Middle	6-week	Late
No. of QTLs	11	12	14	12
Total r^2	0.39	0.51	0.54	0.38

Doebley and Stec 1991, 1993). As shown in Table 18.5, a few QTLs of major effect account for most of the differences between characters. These QTLs are mostly in very similar positions in the two crosses (Figure 18.15), with both sets of crosses showing five regions of the maize genome that account for most of the differences. Such results are consistent with Beadle's hypothesis of five major genes accounting for the bulk of the obvious phenotypic differences between maize and teosinte (Beadle 1939). Beadle arrived at this figure by examining 50,000 F_2 maize $\times$ teosinte offspring, finding the frequency of all-maize or all-teosinte phenotypes to be $\simeq 1/500$. If n genes are involved, the expected F_2 frequency of either parental genotype is $(1/4)^n + (1/4)^n$. Setting this equal to $1/500$ and solving gives $n = 5$.

Focusing on the five major regions, marker-selected NILs (Chapter 17) were constructed to further characterize the QTLs. Doebley's first target was a QTL on chromosome 4, which accounted for 50% of the variance in glume score. A small maize segment containing this region was introgressed into a teosinte background by three generations of backcrossing and selection for flanking markers (Dorweiler et al. 1993). When NILs with the introgressed teosinte region were backcrossed to the maize recurrent parent, the resulting F_2 progeny showed two discrete classes for glume score, as would be expected with a single major gene. The putative gene was named *tga1*, for *teosinte glume architecture 1* (and later shown to result from a single amino acid substitution; Wang et al 2015). A similar analysis of two other regions, QTL-3L and QTL-1L, showed strong epistasis for a number of key traits separating maize and teosinte (Doebley et al. 1995a; see Example 5.1). By using marker-selected introgressed lines, QTL-1L was shown by complementation tests to be the locus *teosinte branched 1* (*tb1*). In maize, mutants at this locus result in teosinte-like features for inflorescence sex (tassels, not ears), and number (many instead of one) and length of lateral branches (long, not short). Doebley found that the joint effects of *tb1* and QTL-3L, by themselves, were sufficient to account for essentially all of the differences in plant architecture between teosinte and maize. Further, these two QTL result in substantial differences in ear architecture. Hence, there is direct evidence that just a few genes can account for a very significant amount of the dramatic differences between teosinte and maize.

Subsequent studies on the *tb1* locus noted an extended region of reduced variation just upstream of this gene in maize, but not teosinte. This is the classical hallmark of a selective sweep (WL Chapter 8), indicating selection on this region in maize. Remarkably, the target site was not the *tb1* coding region, but rather an upstream regulatory region. Using results from sweep theory (WL Chapters 8 and 9), it was estimated that this allele had around a 5% advantage during the domestication phase (Wang et al. 1999; Clark et al. 2004, 2006). This leveraging of QTL information to narrow a search for selection to a small genomic region is routinely done in the search for domestication genes (WL Chapter 9).
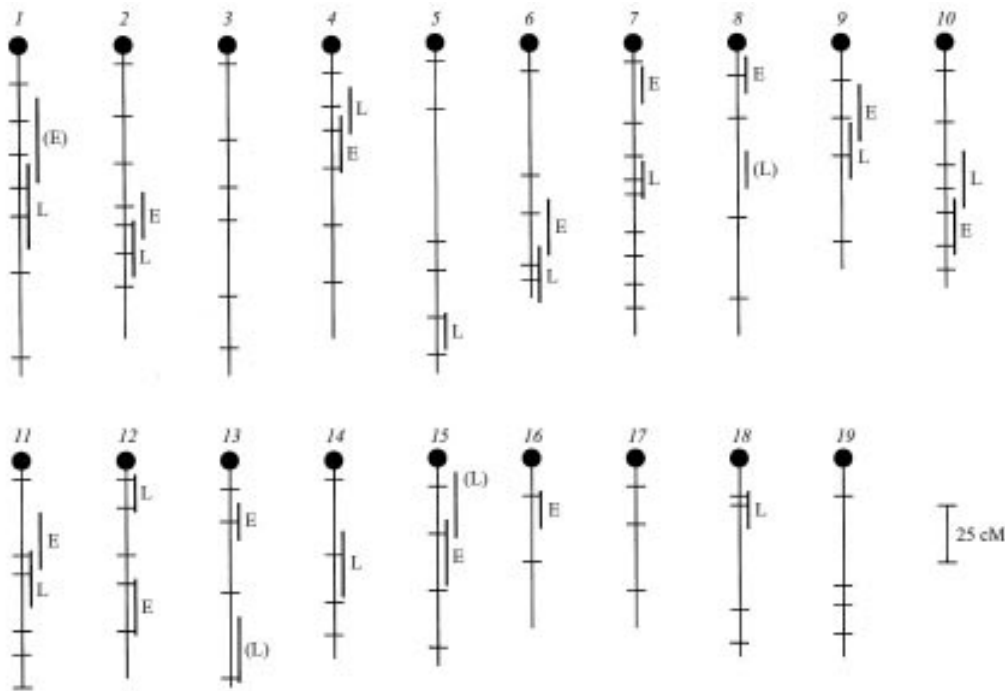
Figure 18.16   Locations for early, E, and late, L, growth QTLs, and early, (E), and late (L) weight QTLs on the 19 mouse autosomes. The marker locations are indicated by hatch marks and QTL confidence intervals by bars (From Vaughn et al. 1999.)

A final interesting study using *tb1* addressed the issue of whether this QTL fractionated upon fine mapping (showing that the initial QTL was, in reality, a group of linked loci). Through the use of NILs, Studer and Doebley (2011) found that *tb1* did not fractionate for plant architectural traits (the detected QTL appeared to be a single gene for these traits). However, the effects of *tb1* on ear morphology did appear to fractionate, with *tb1* still having an impact, but its isolated effect was diminished relative to its initial value.

QTLs for Age-specific Growth in Mice

Cheverud et al. (1996) and Vaughn et al. (1999) performed two replicated crosses of the SM/J and LG/J inbred lines (the small and large lines from the Jackson Laboratories). Interval mapping was used to examine age-specific weight (at ages 1 through 10 weeks) and age-specific growth rate. Intercross I (Cheverud et al. 1996) consisted of 535 F_2 mice scored for 75 microsatellite markers, generating 55 intervals that averaged around 28 cM in length. Intercross II (Vaughn et al. 1999) consisted of 510 F_2 mice scored for 96 markers, resulting in 72 intervals averaging around 23cM (Figure 18.16). As Table 18.6 (Intercross I results) shows, considerable numbers of QTLs were found for all characters. All detected QTLs had small effects, with the largest accounting for around 10% of the F_2 phenotypic variance, while the average (detected) effect was around 4% of the F_2 variance. Note also from Table 18.6 that as age increases, so does the number of QTLs for weight and the total r^2 value of these detected loci. Early vs. late weight and growth showed different genetic architectures. First, largely distinct sets of QTLs are involved (Figure 18.16). Second, dominance was found to be much more important in early weight and growth than in late weight and growth. The results from Intercross II yielded very similar conclusions, but only 70% of detected QTLs were seen in both studies. As expected, QTLs with larger LOD values in Intercross I were more likely to be replicated in Intercross II. This pair of papers highlights a key feature of many large-sample replicated QTL experiments: coarse characterization of the genetic architecture typically remains unchanged, but replication of specific QTLs is often modest.

QTLs in the Illinois Long-term Selection Lines of Maize

In 1896, C. Hopkins initiated a set of maize lines selected for high and low oil and high and low protein content (Hopkins 1899; Smith 1908). Selection on these lines continues today, and results from this remarkable study after 76, 90, 100, and 106 generations of selection were summarized by Dudley (1977), Dudley and Lambert (1992, 2004), Moose et al. (2004), and Dudley (2007). The smooth and continuous long-term response in these lines suggested that a large number of genes of relatively small effect underlie the impressive gains that were observed. Crosses of the divergently selected lines have been used in several QTL mapping studies, starting with basic F_2 analysis (Goldman et al. 1993, 1994; Berke and Rocheford 1995) and eventually using RIAILs for finer mapping (Dudley et al. 2004; Laurie et al. 2004; Clark et al. 2006; Willmot et al. 2006; Dudley et al. 2007), as well as backcrosses to one of the founding lines (Wassom et al. 2008a, 2008b).

The first studies were by Goldman et al. (1993, 1994), who crossed the selected (76-generation) high- and low-protein lines and then examined 100 F_3 families (formed by selfing F_2s) using 100 markers spanning the maize genome at an average spacing of about 20 cM. Using single-marker ANOVA, 22 markers on 10 chromosome arms were significantly associated with protein concentration, 19 markers on nine arms were associated with starch concentration, 26 on 13 arms with oil concentration, and 18 on 10 arms with kernel weight. Many of the marker-trait associations extended across clusters of linked markers. In these cases, single-marker ANOVA cannot distinguish between several linked markers all detecting the same linked QTL or multiple QTL.

A multiple regression involving only six (unlinked) markers accounted for 65% of the variation in protein concentration, and this increased to 84% when five significant pairwise epistatic interactions between these markers were incorporated into the regression (using Equation 18.14d). Seven markers accounted for 66% of the variation in starch, increasing to 78% when one significant pairwise epistatic interaction was included. Four marker loci accounted for 43% of the variation in oil concentration, while six markers accounted for 47% of the variation in kernel weight. It seemed very surprising that so few loci could account for such a significant fraction of the differences, especially given the long-term continuous and gradual change in the lines. As we detail shortly, this was almost certainly a Beavis effect, as more recent, and much higher-powered, studies found a large number of small effect QTLs. Berke and Rocheford (1995) examined a cross of two other variant lines from this experiment (High Oil with Low Oil), and found similar results, with six loci accounting for 58% of the genetic variation in oil concentration and seven markers accounting for 56% of the variation in starch concentration.

A biased view of the genetic architecture of a trait from a QTL analysis can arise from two sources: insufficient recombinant (lumping the composite effects of several causal loci into single factor) and low power (resulting in both an undercounting of the actual number of QTL and overestimation of the effects of detected QTL). Dudley et al. (2004) and Willmot et al. (2006) attempted to address linkage concerns by examining the impact of additional generations of random mating before selfing. Both the first (Generation 70 high x low protein) and second (Generation 70 high x low oil) study compared the results of inbred lines formed from F_2s with those from (random-mating) F_5s. An important limitation is that these studies were limited to 64 and 125 markers, respectively. Not surprisingly, both studies detected significantly *fewer* QTLs using F_5-derived inbreds relative to using F_2 inbreds. This is expected from a reduction in power due to the expanded map distance generated by additional generations of random mating. Recalling that power declines as $(1 - 2\tilde{c})^2$, the low marker density was simply not sufficient to adequately cover the expanded map.

Much more powerful designs were used by Laurie et al (2004), Clark et al. (2006), and Dudley et al. (2007). These authors took generation 70 high versus low crosses, followed by 7–10 generations of random mating before 500 lines were generated by selfing. The resulting lines were then replicated (multiple locations at two years each), and scored for around 500 markers. This improved design attempted to address both the linkage and power limitations of earlier studies. Laurie et al. and Clark et al. used oil crosses (10 generations), while Dudley

et al. used protein crosses (7 generations of random mating). Laurie et al. detected around 50 QTL for oil, which accounted for around 50% of the generic variance. Effects were small and largely additive. Surprisingly, around 20% had antagonistic effects (up alleles in the down line, and vise versa). Similar results for other traits were seen in the other experiments.

Limitations of Line-cross QTL Mapping

Despite the impressive statistical machinery developed above for linkage-based QTL mapping using inbred-line crosses, we close on a slightly somber note. Researchers typically perform a QTL mapping experiment to address two, not necessarily independent, questions: (i) what is genetic architecture underpinning the differences between two lines, and (ii) what are the locations of the actual genes underlying these differences. As highlighted in the previous examples, linkage-based QTL mapping is a substantial improvement over classic line-cross analysis (Chapter 11) for probing the genetic architecture of a trait, and can often provide important insight. However, it tends to underestimate the number of QTLs and overestimate the effects of detected QTLs, especially when sample size is modest (< 500, which is the norm). Hence, it is biased towards presenting a more major-gene view of the architecture (fewer genes of large effect).

Even more problematic is the very poor level of resolution of linkage-based mapping, which tends to be on the tens of megabase scale under the typical sample sizes used. This is an improvement over mapping to specific chromosome arms (Chapter 17), but only moderately helpful for fine mapping. Finally, there is the issue of repeatability, as replicated experiments (using identical line crosses) often fail to find some of QTLs detected in previous studies. Again, this is most pronounced when sample sizes are small to modest. A compounding issue is that an apparently large QTL detected in an initial study (whose actual value was significantly overestimated by the Beavis effect) might lead the investigator to use a comparably sized population in an attempt to replicate this finding, when in reality a much large sample size is needed.

A final issue is that an inbred-line cross is probing the nature of *fixed* differences between two lines, typically chosen because they are rather divergent in a focal trait. Hence, many of these genes may have been fixed by directional selection, which can result in large-effect alleles being disproportionately fixed (WL Chapters 25–27). The segregating genetic loci in the progeny of an inbred line cross are thus a rather *unnatural population*, with segregating alleles (or fixed alleles over a final RIL or DHL panel) having frequency one half, and likely enriched for large-effect alleles. In natural populations, segregating large-effect alleles tend to at very low frequencies (WL Chapters 24 and 28).

Despite these concerns, there still remain settings, especially in inbred crops, where line-cross mapping may have greater mapping resolution than a GWAS (especially when using second-generation mapping designs, such as MAGIC or NAM panels). The reason is that many elite inbred crops show LD over (at least) hundreds of kilobases. This places a limit on the resolution possible by LD (GWAS) mapping (Myles et al. 2009). Again, however, for linkage-based mapping to have greater resolution, the sample sizes need to be large, allowing more recombination events to be captured. In the words of Cockram and Mackay (2018), "we believe that days of mapping QTL in small populations must come to an end $\cdots$ we started with a target of developing 1,000 lines per population: that number now looks to be on the low side.". Many of the current users of QTL mapping suffer from what Tversky and Kahneman (1971) called a "belief in the law of small numbers," gambling their research hypotheses on small samples without realizing that the odds against them are unreasonably high.

Our next three chapters attempt to use segregating natural variation to map QTLs. As such, they are **very likely probing different genetic architectures than are inbred-line crosses.** Chapter 19 examines linkage mapping using family-based crosses, or more generally, pedigrees. These designs have much less power than inbred-line crosses, and have (at best) similar levels of mapping resolution. Chapter 20 examines association-based mapping, which can use *very* large population samples coupled with very dense marker maps

to exploit naturally occurring linkage disequilibrium. Such LD often extends over no more than a few tens of kilobase scale (and often much less), as opposed to linkage information (which tends to be on a tens of megabase scale unless the sample size is largess), offering substantially improved mapping resolution. Finally, Chapter 21 summarizes the impact of GWAS studies on our current understanding of the genetic architecture of segregating genetic variation.