21 Quantitative Trait Loci

The genetic basis of quantitative variation described in the preceding chapters has, of necessity, considered only the aggregate effects of all the genes causing the variation. A complete description needs to take account of the properties of the genes individually – their gene frequencies and the magnitude of their effects on the trait of interest. The genes cannot be studied individually using the methods of classical Mendelian genetics because their effects are lost in the statistical 'fog' of all other background variation. In the absence of knowledge about the genes' individual properties theoretical work has had to make some unrealistic assumptions, such as the gene frequencies at all loci are more or less the same, that the genes' effects and dominance relations are all about the same, and in some contexts there are an indefinitely large number of genes affecting the trait. Recently, however, methods have become available for studying the individual genes; these genes are known as *Quantitative Trait Loci*, or *QTLs*. This chapter will explain the methods for identifying QTLs and of estimating their effects on quantitative traits. We shall see, however, that what is identified as a QTL is a segment of chromosome affecting the trait, not necessarily a single locus.

Identification of the individual genes could lead to several useful applications. First, it could improve the efficacy of selective breeding, especially for traits with low heritability or that can only be measured in one sex (see Soller and Beckmann, 1988; Lande and Thompson, 1990). Second, transgenic technology might be applied to quantitative traits. Third, in medicine, the identification of alleles causing predisposition to common multifactorial diseases, such as heart disease or diabetes, could lead to improved methods of prevention. And fourth, quantitative genetic theory will be made more realistic when the numbers and properties of the genes are known, and the more realistic theories will improve our understanding of evolution.

Major genes

Many loci affecting quantitative traits have been identified fortuitously, by chance discovery of alleles with effects on the trait that were large enough to be recognized by their individual segregation. These include the major morphological mutations of classical genetics that arose by spontaneous or induced mutation, in, for example, *Drosophila*, mice, and maize. Some genes of large effect have been found to be segregating in selected lines. Examples of these are alleles of the *scabrous*, *bobbed*, and *scute* bristle loci discovered in lines of *Drosophila* selected for bristle number (reviewed by Mackay, 1989), alleles of the *pygmy* locus (Example 7.1), the

obese and dwarf alleles (reviewed by Roberts and Smith, 1982) found in lines of mice selected for body size, and a gene (*hg*) causing rapid post-weaning weight gain in a line of mice selected for weight gain (Bradford and Famula, 1984). Other examples of alleles with major effects on commercial traits in domestic animals are given in Table 21.1. In many cases the alleles of large effect probably arose *de novo* by mutation in the selected lines (Chapter 12). Such major-effect alleles are at very low frequencies in unselected populations and contribute little to segregating quantitative variation. This is almost certainly due to their adverse pleiotropic side effects on fitness. *Drosophila* females homozygous for the *bobbed* mutation are less viable and have longer developmental times than wild type (Lindsley and Zimm, 1992); the dwarf, obese, and high growth genes in mice are either sterile or have reduced fertility when homozygous (Roberts and Smith, 1982; Bradford and Famula, 1984); and the halothane gene in pigs and the double muscling gene in cattle reduce viability and fertility (Hanset, 1982; Webb *et al.*, 1982).

Table 21.1 Major genes affecting quantitative traits.

Species	*Trait*	*Gene*	*Reference*
Poultry	Body size	Dwarf	Merat and Ricard (1974)
Pig	Leanness, porcine stress syndrome	Halothane sensitivity	Smith and Bampton (1977) Webb *et al.* (1982)
Cattle	Leanness, muscular hypertrophy	Double muscling	Rollins *et al.* (1972) Hanset and Michaux (1985*a*, *b*)
Sheep	Prolificacy	Booroola F	Piper and Bindon (1982) Piper *et al.* (1985)

Given that alleles with very large effects on quantitative traits exist, there has been considerable interest in ascertaining whether other alleles with large effects but without deleterious fitness effects segregate in natural populations or have contributed to selection response. Note that 'large' in this sense does not necessarily mean an effect over three standard deviations from the population mean, as would be required to qualify as a classical Mendelian mutation; effects of 0.5–1.0 σ_P are 'large' in this context. For example, the additive genetic variance attributable to a strictly additive allele with a standardized effect of one phenotypic standard deviation ($2a = 1.0\ \sigma_P$) at a frequency of $p = q = 0.5$ is 12.5 per cent of the total phenotypic variance (equation [8.5]). An allele of this effect would not, however, be apparent in a segregating population because of the confounding effects of variation at other loci.

Methods of detection

There are several ways of finding major genes affecting quantitative traits; they are, in outline only, as follows.

(1) *Multimodal distribution* If a gene has an effect large enough relative to the background genetic and environmental variation (in excess of about $3\sigma_P$) it will produce a multinomial distribution in a segregating population. The best way to look for multimodality is in the generations derived from a cross between two

divergent strains. The distribution will be trimodal in the F_2 if the alleles are partially dominant, and bimodal in the cross to the more recessive strain; the F_1 will have shown which strain this is.

(2) *Backcrossing with selection* This method is a way of making the bimodal distribution in the backcross clearer, by reducing the background genetic variation. It is one of the earliest methods used and was proposed by Wright (1952). Two divergent strains produced by selection (or inbreeding) are crossed and repeated backcrosses are made to the more recessive strain. In every generation selection is made for individuals with the more dominant phenotype; i.e., if the backcrosses are made to the low strain, selection is made for high phenotypes. The selection keeps segregating any allele with a large effect in the selected direction. Its effect is detected by the bimodal distribution of heterozygotes *versus* homozygotes of the allele from the low strain. At the same time, the frequencies of other genes of small effect from the high strain are halved in each generation of backcrossing (Table 5.1), and the background genetic variation is thereby reduced, making the bimodal distribution clearer. This procedure was the one used, although unintentionally, in the development of the Booroola Merino sheep, in which the F allele with a large fecundity effect (Table 21.1) segregates (Piper and Bindon, 1988).

(3) *Non-normal distribution* A gene whose effect is not large enough to cause a multimodal distribution may nevertheless cause a detectable departure from normality. If the gene's frequency is intermediate the distribution will be platykurtic (more flat than normal); if it is extreme (near 0 or 1) the distribution will be skewed and leptokurtic (more peaked than normal).

(4) *Heterogeneity of variance* If a major gene is segregating there will be heterogeneity of the variance within families, because the major gene will be segregating in some families but not in others. This test and the preceding one, however, have little power and require very large sample sizes to detect departure from normality or heterogeneity of sibship variance, and also suffer from the problem that factors other than segregating major genes can cause departure from normality or heterogeneous within-family variance (reviewed by Hill and Knott, 1990; Mayo, 1989). Hammond and James (1970) used some of these methods to detect major genes affecting *Drosophila* abdominal bristle number in a segregating population and concluded they are ineffective, particularly if the heritability of the trait is low.

(5) *Offspring–parent resemblance* A different group of tests have been proposed by Karlin, Carmelli, and Williams (1979) and Karlin, Williams, and Carmelli (1981), collectively termed 'structured exploratory data analysis'. These are based on the intuitively obvious principle that, with polygenic inheritance and no major gene the mean of offspring will resemble more closely the mid-parent value than single parents, while if a major gene is segregating the reverse will be true. One of these tests, the 'major gene index', was applied to families of mice in a population known to be segregating for the *hg* gene and the test succeeded in detecting the major gene (Famula, 1986). These tests, however, are sensitive to non-normality of

the distributions and can lead to erroneous conclusions (Mayo, Eckert, and Nugrobo, 1983; Kammerer, MacCluer, and Bridges, 1984).

(6) *Complex segregation analysis* The most powerful approach for detecting major genes affecting quantitative variation is complex segregation analysis (Morton and MacLean, 1974; reviewed by Hill and Knott, 1990), developed specifically for human pedigrees of parents and full sibs. Complex segregation analysis tests whether the inheritance of a trait is best explained by the segregation of a single major gene, by strictly polygenic inheritance, or by a major gene plus multiple loci with smaller effects. A full mixed model is specified that includes the allele frequency and additive and dominance effects at a single major locus, additive genetic effects from multiple polygenic factors, common environmental and random environmental effects. Maximum likelihood estimates of parameters are made for a series of increasingly complicated hypotheses: a pure environmental model, single gene model, polygene model, and the full model. The significance of each hypothesis is tested by comparing the likelihood of the data, given maximum likelihood estimates of model parameters, with that calculated assuming the appropriate null hypothesis for which the tested parameters are set to zero, using a likelihood ratio test. The likelihood functions are very complicated and their evaluation is computationally demanding, but the method is generally extensible to more complex genetic hypotheses including pleiotropy and linkage to a known marker. Complex segregation analysis has been applied to the inheritance of many known human diseases and risk factors for disease; in all cases there was a significant polygenic component and in some cases the additional presence of a major gene effect was indicated (reviewed by Sing *et al.*, 1988).

Inferring the presence of major genes affecting a quantitative trait by any of the above methods still does not tell us what these genes are, and the loci contributing to the polygenic fraction of the variation remain invisible. For this reason current approaches to resolving QTLs are directed towards identifying all relevant loci that may have a range of effects (from major gene down to the limit of experimental resolution), placing these loci on linkage maps, and, ultimately, molecular cloning of the relevant DNA sequences.

Methods for mapping QTLs

Experimental designs for estimating effects and map positions of QTLs are extensions of standard methods for mapping single genes, and are based on linkage disequilibrium between alleles at a marker locus and alleles at the linked QTL. The requirements for mapping QTLs are thus (i) a linkage map of polymorphic marker loci that adequately covers the whole genome, and (ii) variation for the quantitative trait within or between populations or strains.

Marker loci

Ideally, marker loci should be (i) highly polymorphic, so that pairs of individuals or lines are likely to carry different alleles at each locus; (ii) abundant, so comprehensive marker coverage of the genome is achieved; (iii) neutral, both with respect to

the quantitative trait of interest and to reproductive fitness; and (iv) co-dominant, so all possible genotypes at a marker locus can be identified. This final requirement is less stringent, as dominant/recessive markers can be used successfully in some designs, as described below.

Until recently, mapping the loci underlying quantitative variation was seriously hampered by the paucity of suitable markers. Cryptic protein variation, such as blood group antigens and electrophoretically distinguishable enzyme alleles, often satisfies the criteria of neutrality and co-dominance, but is neither sufficiently polymorphic nor sufficiently abundant to mark entire genomes. The situation has changed with the molecular biology revolution and the discovery of DNA-based markers that satisfy all essential criteria outlined above. The restriction fragment length polymorphisms (RFLPs), variable number of tandem repeat (VNTR), or minisatellite, loci, and microsatellite (or simple sequence repeat, SSR) loci described in Chapter 1 are all suitable genetic markers. RAPD (randomly amplified polymorphic DNA) markers primarily detect DNA sequence variation in an arbitrary 10-base sequence used as a primer in a polymerase chain reaction (PCR). The marker DNA is amplified whenever the 10-base sequence flanks a genomic sequence of appropriate size. RAPD markers are dominant, as the PCR product is typically present or absent. RAPD and SSR markers are highly abundant and polymorphic, and their detection using PCR means linkage maps can be constructed more rapidly and efficiently than is possible using RFLP markers. In some organisms, such as *Drosophila* and mice, the genomic locations (sites of insertion) of transposable element or ecotropic retrovirus sequences are highly variable and their presence at an insertion site can be used as a dominant polymorphic marker. Dense molecular marker linkage maps are currently available for the genomes of human, mouse, rat, *Drosophila*, and many livestock and plant species.

QTL genotypes

There are two general sorts of methods for identifying and mapping QTLs: those based on crosses between lines that differ for the trait of interest, and methods based on segregating populations. The most efficient experimental designs for locating QTLs use crosses between lines that are fixed for alternate alleles at both the QTL and the marker loci, because of the maximum linkage disequilibrium between the loci in the F_1. Preferably, all alleles should be in *association*; that is to say, alleles that increase the value of the trait should be homozygous in one parental line, and the alleles that decrease the value of the trait fixed in the other parental line. Homozygosity and the arrangement of QTLs in association are most likely to be met if parental populations have resulted from divergent artificial selection for the trait of interest, and have been inbred subsequently. If each line has some increasing and some decreasing alleles fixed, the alleles are in *dispersion*; in these circumstances it is still possible to locate QTLs, but fewer will be found than when the alleles are in association, for the reason to be explained later.

Most commonly, the parental inbred populations are crossed to produce the F_1 generation, which is then either backcrossed to one or both parental lines (the BC design), or crossed *inter se* to produce the F_2 generation (the F_2 design). In species

that tolerate inbreeding, the F_2 can be inbred to produce recombinant inbred lines (Oliverio, 1979), and these are also useful for mapping. For species that either do not tolerate inbreeding and/or that have very long generation intervals, the principles of line cross analysis can still be applied, provided lines are available that segregate for marker loci, but are nevertheless divergent for the quantitative trait of interest, and thus are likely to be fixed for alternate QTL alleles. In this case parents must be screened for marker configurations that produce informative segregations (i.e., backcross or F_2 ratios) in the progeny. This is the procedure used in human genetic mapping. Detecting QTLs by linkage to marker loci is least efficient in randomly breeding populations, because all but very tightly linked loci are likely to be in linkage equilibrium.

The principle underlying identification of QTLs by linkage to marker loci is conceptually simple: individuals are scored for their genotype at the marker locus, and their phenotype for the quantitative trait. If there is a difference in mean phenotype among marker genotype classes, then we can infer the presence of a QTL linked to the marker. Marker loci can be considered singly or simultaneously. The number of QTLs detected by linkage with markers is always an underestimate of the number of loci because two QTLs closely linked to each other may appear as only one if in association, or may not be detected at all if in dispersion. In most experiments, a map distance of, very roughly, 20 centimorgans (cM) is the limit of resolution, so what is detected as a QTL is a segment of chromosome of this length, which may contain several loci affecting the trait, not necessarily in the same direction. Thus more QTLs are likely to be detected when the alleles are in association than when they are in dispersion. With more refined methods, however, the limit of resolution can be reduced to about 3 cM.

Single marker analysis

To illustrate the method for detecting QTLs by association with single markers, consider a marker locus (M) and a QTL (A) with c the recombination frequency between them. Let the genotypes of one parental line be M_1A_1/M_1A_1 and of the other parental line be M_2A_2/M_2A_2. Following Fig. 7.1, the genotypic values of the A_1 and A_2 homozygous parents are a and $-a$, respectively. The genotypic value of the F_1 individuals (M_1A_1/M_2A_2) is d. Parental F_1 gametes (M_1A_1 and M_2A_2) are each produced with frequency $(1 - c)/2$, and recombinant F_1 gametes (M_1A_2 and M_2A_1) are each produced with frequency $c/2$. Random mating of the F_1 gives 10 possible F_2 genotypic classes. The contribution of each marker genotype class to the F_2 mean is obtained by multiplying the frequency of each genotype by its genotypic value, then summing within marker genotype classes. The procedure is illustrated in Table 21.2. We are not interested here in the contributions of the marker classes to the mean of the F_2, but in the differences between the marker classes. We therefore need the actual means, which are obtained by dividing the contribution to the F_2 mean by the frequency of that marker class, which is the Mendelian segregation ratio of $\frac{1}{4}$ for the homozygotes and $\frac{1}{2}$ for the heterozygotes. The actual means are given in the last column of the table. The means of marker classes in backcrosses are calculated in the same way, but are simpler because there are only four genotypes and two marker classes.

Table 21.2 Genotypes in an F_2 with one marker locus, M, and a linked QTL, A.

F_2 genotype	*Frequency*	*Value*	*Marker class*	*Frequency*	*Contribution to F_2 mean*	*Actual mean*
M_1A_1/M_1A_1	$(1-c)^2/4$	a	M_1/M_1	$\frac{1}{4}$	$a(1-2c)/4 + dc(1-c)/2$	$a(1-2c) + 2dc(1-c)$
M_1A_1/M_1A_2	$c(1-c)/2$	d				
M_1A_2/M_1A_2	$c^2/4$	$-a$				
M_1A_1/M_2A_1	$c(1-c)/2$	a	M_1/M_2	$\frac{1}{2}$	$d[(1-c)^2 + c^2]/2$	$d[(1-c)^2 + c^2]$
M_1A_1/M_2A_2	$(1-c)^2/2$	d				
M_1A_2/M_2A_1	$c^2/2$	d				
M_1A_2/M_2A_2	$c(1-c)/2$	$-a$				
M_2A_1/M_2A_1	$c^2/4$	a	M_2/M_2	$\frac{1}{4}$	$-a(1-2c)/4 + dc(1-c)/2$	$-a(1-2c) + 2dc(1-c)$
M_2A_1/M_2A_2	$c(1-c)/2$	d				
M_2A_2/M_2A_2	$(1-c)^2/4$	$-a$				

From Table 21.2 we can see that if the marker locus is unlinked to the QTL (i.e., $c = 0.5$), then all the marker classes have the same expected mean, $0.5d$. However, if the QTL is linked to the marker locus, the following contrasts of marker class means are functions of a, the genotypic value or additive effect, and d, the dominance deviation:

$$(M_1/M_1 - M_2/M_2)/2 = a(1 - 2c) \quad \ldots [21.1a]$$

and, after some simplification,

$$M_1/M_2 - [(M_1/M_1 + M_2/M_2)/2] = d(1 - 2c)^2 \quad \ldots [21.1b]$$

Thus a significant difference in the mean value of a quantitative trait between homozygous marker genotype classes can be taken as evidence of linkage of a QTL and the marker locus. However, estimates of a and d/a from single marker analysis are confounded with recombination frequency, and will generally underestimate the true values by $(1 - 2c)$. For example, a mean difference in phenotype of 0.2 standard deviation between the homozygous marker classes could be due to a QTL of this effect completely linked to the marker, or to a QTL of effect $a = 0.5$ standard deviation 46 cM away (using Haldane's mapping function $c = [1 - exp(-2x)]/2$ to relate distance (x) in morgans to recombination fraction).

Example 21.1

The first example of an association between a marker locus and a quantitative trait was reported by Sax (1923), between a pigment locus and seed size in the bean, *Phaseolus vulgaris*. One of the parental lines, Improved Yellow Eye 1317, was homozygous for the dominant pigmentation factor, *P*, and had seeds that weighed on average 48 centigrams (cg). The other parental line, White 1228, was homozygous for the recessive pigmentation factor, *p*, and had an average seed weight of 21.0 cg. The genotypes at the pigmentation locus and average seed weights in the F_2 of the cross were:

Genotype	*PP*	*Pp*	*pp*
Seed weight (cg)	30.7	28.3	26.4

(The genotypes of the pigmented F_2 were deduced from the presence or absence of unpigmented *pp* progeny in the F_3.) Equation [21.1*a*] gives the estimate of $a(1 - 2c)$ as $(30.7 - 26.4)/2 = 2.15$ cg. The effect is perfectly additive because the marker heterozygote is almost exactly midway between the two homozygotes. Equation [21.1*b*] gives $d(1 - 2c)^2 = -0.25$ cg. The difference in seed weight between the *PP* and *pp* genotypes in the F_2 (4.3 cg) accounts for 16 per cent of the total difference in seed weight between the two parental strains (27 cg). This is a large effect associated with the marker, but the conclusions we can draw about a QTL are limited. First, part, or even all, of the effect may have been due to the *P* locus itself having a pleiotropic effect on seed weight. Second, if there is a QTL linked to the *P* locus, its effect ($2a$) is correctly estimated as 4.3 cg only if there is no recombination with the marker ($c = 0$). The effect could equally have been due to a less closely linked QTL with a larger effect. And, third, the QTL identified may have been not one but two or more linked loci.

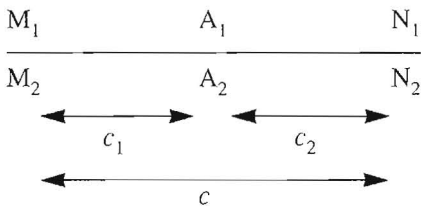

Fig. 21.1. Recombination frequencies between two marker loci, M and N, and a QTL, A.

Interval mapping analysis

The problem of confounding estimates of additive and dominance effects of a QTL linked to a single marker locus with recombination frequency can be solved by considering pairs of linked marker loci (M and N), separated by recombination fraction c, both of which are fixed for alternate alleles in the parental strains. In this case the map positions of the marker loci, and hence c, are known. Assume there is a QTL, A, between the two marker loci, with the recombination frequency c_1 between M and A, and c_2 between N and A; assuming no interference, $c_1 + c_2 = c$. The relationship between the marker loci in the F_1 is depicted in Fig. 21.1.

The genotypes of the two parental lines are thus $M_1A_1N_1/M_1A_1N_1$ (with genotypic value a) and $M_2A_2N_2/M_2A_2N_2$ (with genotypic value $-a$), and of the F_1 is $M_1A_1N_1/M_2A_2N_2$ (with genotypic value d). The analysis can be made in a backcross or an F_2 generation. The expected marker class means in a backcross to the $M_1A_1N_1$ homozygous parent are set out in Table 21.3, calculated in the same manner as in Table 21.2. Here the expected marker genotype frequencies depend on the recombination frequency between the two markers, so the means of the marker classes with parental F_1 gametes are divided by $(1 - c)/2$ and the means of the marker classes with recombinant F_1 gametes are divided by $c/2$. It is assumed that the two markers are closely enough linked that double recombination can be ignored. The contrasts between backcross marker class means, that give estimates of the effects of the QTL and its map position relative to the flanking markers, are given in equations [21.2]:

$$M_1N_1/M_1N_1 - M_1N_1/M_2N_2 = a - d \quad \ldots [21.2a]$$
$$M_1N_1/M_1N_2 - M_1N_1/M_2N_1 = (a - d)(c_2 - c_1)/c \quad \ldots [21.2b]$$

A disadvantage of the backcross design is that the estimate of the additive effect of the QTL is unbiased only if $d = 0$, and recessive or partly recessive QTLs may not be detected. This problem can be overcome by backcrossing to both parental lines, or by using an F_2 design. The expected genotypic values of the nine marker classes produced in an F_2 are given by Haley and Knott (1992).

Example 21.2

Before molecular markers became available most QTL mapping was done with *Drosophila* bristle numbers because (1) selected lines with divergent values were available, (2) there are major mutant genes spread over all three major chromosomes

Continued

Table 21.3 Genotypes in a backcross with two linked markers, M and N, and a linked QTL, A. The F_1 is backcrossed to $M_1A_1N_1$.

F_1 Gamete type	*Frequency*	*Value*	*Marker class*	*Frequency*	*Contribution to BC mean*	*Actual mean*
$M_1A_1N_1$	$(1 - c)/2$	a	M_1N_1/M_1N_1	$(1 - c)/2$	$a(1 - c)/2$	a
$M_1A_1N_2$	$c_2/2$	a	M_1N_1/M_1N_2	$c/2$	$(ac_2 + dc_1)/2$	$(ac_2 + dc_1)/c$
$M_1A_2N_2$	$c_1/2$	d				
$M_2A_1N_1$	$c_1/2$	a	M_1N_1/M_2N_1	$c/2$	$(ac_1 + dc_2)/2$	$(ac_1 + dc_2)/c$
$M_2A_2N_1$	$c_2/2$	d				
$M_2A_2N_2$	$(1 - c)/2$	d	M_1N_1/M_2N_2	$(1 - c)/2$	$d(1 - c)/2$	d

Example 21.2 *continued*

to use as markers, and (3) chromosomes can be studied one at a time. The early work is reviewed by Thoday (1979). The data below are taken from Wolstenholme and Thoday (1963).

The parental chromosomes used were a third chromosome from a line selected for high numbers of sternopleural bristles, and a tester third chromosome, unselected for bristle number, with several morphological markers. Two of the markers, the recessive gene *clipped* (*cp*) and the dominant gene *Stubble* (*Sb*), are separated by a distance of 12.9 cM. To determine whether there was a QTL between *cp* and *Sb* affecting sternopleural bristle number, flies with the selected chromosome (HH) were crossed to flies from the marker tester chromosome strain. The doubly heterozygous F_1 flies (*cpSb/HH*) were then backcrossed to a *cp* + tester stock. The mean bristle numbers of the four marker genotype classes in the backcross were as follows:

	Genotype	*Bristle number*
(1)	HH/*cp* +	20.62
(2)	H*Sb*/*cp* +	19.19
(3)	*cp*H/*cp* +	18.95
(4)	*cpSb*/*cp* +	18.00

Inspection shows clearly that there is a QTL affecting bristle number linked to the markers; the H alleles from the selected line are associated with higher numbers. From the differences in bristle numbers between the marker classes we can get the following estimates:

by equation [21.2*a*], $(a - d) = (1) - (4) = 2.62$ bristles, and
by equation [21.2*b*], $(a - d)(c_2 - c_1)/c = (2) - (3) = 0.24$ bristle,

where c_1 is the recombination frequency between the QTL and *cp*, and c_2 the recombination frequency between the QTL and *Sb*. The known distance between the markers gives $c = 0.129$. Substitution of $(a - d) = 2.62$ and $c = 0.129$ into the second equation above gives $c_2 - c_1 = 0.0118$. So, from $c_1 + c_2 = c$ we can estimate $c_1 = 0.059$ and $c_2 = 0.070$. Thus, the QTL is located between the markers, nearer to *cp* than *Sb*. When there is only one backcross we cannot separate a and d to estimate the effect and gene action of the QTL.

Genetical and statistical considerations

Although the principles of mapping QTLs are straightforward, many problems arise in practice regarding optimum methods of statistical analysis and the genetic interpretation. We now need to consider issues involved in deciding sample size, number and density of markers, optimal experimental design, and appropriate statistical analysis as well as limitations regarding genetic interpretation of experimental data.

Experimental design

Our description of mapping methods has been overly simplified in the sense that we have not specified how marker class means (or a contrast of two marker class

means such as equation [21.1*a*] or [21.2*a*]) are judged to be significantly different. Assume for the moment that the *t*-test is used for this purpose, and that we wish to design a single marker experiment to detect a difference of some arbitrary value δ between the two homozygous marker genotypes caused by a QTL completely linked to the marker ($c = 0$) in an F_2 cross. Given that the number of individuals (n) scored for the quantitative trait in each marker genotype class will be sufficiently large that we can assume the phenotypic values to be normally distributed, the n required is given by standard statistical theory (Sokal and Rohlf, 1981, p. 263) as:

$$n \geq 2\,(z_\alpha + z_{2\beta})^2/(\delta/\sigma_W)^2 \qquad \ldots [21.3]$$

In this expression δ is the smallest difference between marker classes that the experimenter wants to detect as significant, and σ_W is the phenotypic standard deviation within marker classes; α and β are the significance levels set, α being the acceptable error rate of false positives (Type I errors) and β the acceptable error rate of false negatives (Type II errors); z is the ordinate of the normal distribution corresponding to its subscript. Suppose, for example, that the error rates are set at $\alpha = 0.05$ and $\beta = 0.1$, giving $z_\alpha = 1.96$ and $z_{2\beta} = 1.28$. Then if standardized effects (δ/σ_W) of 1 or over are to be detected, equation [21.3] shows that the number of individuals needed in each marker class is $n = 21$; if, however, δ/σ_W is set at 0.25 then $n = 336$. Thus very large sample sizes are required to detect QTLs with moderate or small effects; with small sample sizes only large-effect QTLs will be found. Similar power calculations give rise to the following general guidelines for the relative efficiencies of different experimental designs for QTL mapping.

1. Interval mapping is preferable to single marker analysis because, with a single marker, the QTL effects are confounded with the map distance of the QTL from the marker, as we have already seen. Interval mapping is also more efficient than single marker analysis. The sample size needed to detect a given standardized effect is increased by a fraction of $1/(1 - 2c)^2$ for single marker analysis (Soller, Brody, and Genizi, 1976).

2. The F_2 design is more powerful than the backcross design. Backcrosses to a single parent only detect heterozygous effects, which for the case of additive QTLs are half the homozygous effects detected in the F_2, so four times as many individuals need to be scored to achieve the same power. With dominance, one backcross will be more efficient than the F_2 and the other less efficient (Soller, Brody, and Genizi, 1976). Backcrosses to both parents are necessary to estimate homozygous effects, which is also less efficient than the F_2 design.

3. If two parental populations are fixed for alternate alleles at the QTL but are segregating for marker alleles, the sample size to achieve the same power as the case for fixed marker alleles is increased. For example, the increase is by a factor of $1/(q_1 - q_2)^2$ for single marker analysis, where q_1 and q_2 are the gene frequencies of the marker in the two strains (Soller, Brody, and Genizi, 1976). Even for a favourable case with $q_1 = 0.8$ and $q_2 = 0.3$ (i.e., marker allele frequencies are relatively extreme), a four-fold increase in sample size would be necessary, assuming all F_2 individuals are scored. A considerable saving of effort, however, can be achieved if only those F_1 matings that give informative progeny are used. Informative matings

are those between two individuals that are heterozygous for the marker, and between a heterozygote and a homozygote. For details of how many individuals are needed see Beckmann and Soller (1988).

4. The power to detect a difference in mean between two marker genotypes does not depend on the absolute value of the difference (δ), but rather on the difference scaled by the within-marker-class standard deviation (δ/σ_W). Therefore, strategies to reduce σ_W can yield increased power. Reducing σ_W requires more accurate estimates of phenotypic values. This can be achieved by progeny testing individuals of the segregating generation, F_2 or backcross, or by inbreeding to produce recombinant inbred lines (Thoday, 1961; Lander and Botstein, 1989; Soller and Beckmann, 1990). Making recombinant inbred lines has the disadvantage that the linkage between the QTL and the markers is reduced by further recombination in the subsequent generations. In *Drosophila*, however, recombinant isogenic lines can be synthesized in a few generations with no further recombination, so for this species this method is best (Long *et al.*, 1995).

Multiple tests

The above discussion has referred to a QTL linked to a single marker or pair of markers. In reality, many markers are distributed throughout the genome and each one (or pair) is in turn tested for linkage to a QTL. Furthermore, the parental lines used typically are divergent for multiple traits, all of which are scored in the segregating generation, so the tests for linkage to the markers are repeated for each trait. This means that some 'significantly' positive associations will occur by chance, and a more stringent level of significance must be set. The number of false positives increases rapidly with the number of tests. Suppose that the desired significance level for a set of n independent tests is $\alpha = 0.05$, then the level for each separate test must be set at α/n. For example, if single marker associations were tested for 50 independent markers, the significance level for each test should be $\alpha = 0.001$. Re-evaluation of equation [21.3] with $z_{0.001} = 3.291$ shows that the sample size must be doubled to achieve the same power. The tests, however, are not independent because there will be linkage between some of the markers and some of the tests will therefore be partially redundant. So, treating the tests as independent sets too stringent a significance level. When the data are analysed by maximum likelihood, as explained below, proper allowance is made for the non-independence of markers.

Maximum likelihood estimation

The relative efficiencies of different designs were discussed above in the context of t-tests for purposes of illustration. However, the use of this test is technically not appropriate since it is based on an assumed normal distribution of phenotypes within marker class genotypes. Inspection of Table 21.1 shows this is not true: F_2 homozygous marker genotypes each contain the three QTL genotypes, so the distributions are mixtures of three normal distributions (assuming further that the environmental variance is normally distributed). The use of the t-test (or analysis of variance) to detect QTLs linked to markers is robust to all but extreme departures from normality, such as would be caused by very few QTLs of very large effect

distinguishing the two parental strains. More sophisticated methods based on maximum likelihood are, however, more appropriate for parameter estimation and significance tests since they take into account the correct distributional properties of the marker genotypes with respect to the segregating QTL. For details see Lander and Botstein (1989) and Knott and Haley (1992).

The procedure, in outline, is as follows. A likelihood function is specified in terms of the observed data (the numbers of individuals and their phenotypes in each marker class) and the parameters to be estimated (c, and the means and variances of the QTL genotypes). Trial values are assigned to the unknown parameters and an iterative computer program finds the likelihood function, L, for each trial value. The trial values that maximize L are the maximum likelihood estimates of the unknown parameters. Maximum likelihood estimation is thus advantageous for the single marker case since the recombination fraction between the marker and QTL can be put in the model and estimated. The test of significance is the logarithm (base 10) of the ratio L/L_0 (which is distributed as χ^2), where L is the observed maximum likelihood, and L_0 is the likelihood computed for the null hypothesis that there is no QTL segregating.

In the context of interval mapping the log likelihood ratio is called the LOD score (for 'log odds'), following the convention of human linkage mapping. See Lander and Botstein (1989) for details. The LOD score is computed for varying positions of the QTL within the interval; the maximum likelihood estimates of c and a are the values for which the LOD score is maximized. The estimates are usually presented graphically, as a plot of LOD score against chromosome position, in cM. Figure 21.2 shows an example. The value of the LOD score above which the presence of a QTL in the interval is judged significant for an overall error rate of 5 per cent must be computed from the number of independent intervals tested, which in turn depends on marker density, number of chromosomes and genome size. Typical LOD threshold values over a wide range of assumed values for these parameters are between 2 and 3, roughly corresponding to $\alpha = 0.001$ for each interval tested (Lander and Botstein, 1989). In the study of grain yield in maize shown in Fig. 21.2 there were eight linked markers, whose positions are shown on the x-axis, with the intervals between them in cM. The position marked by the triangle is the position of the QTL that maximizes the LOD score, and this is the most probable position for a QTL. The horizontal line at a LOD score of 2 marks the threshold of significance ($\alpha = 0.05$) for the experiment as a whole. The existence of a QTL affecting yield is clearly established.

Multiple QTLs

The parental lines chosen for mapping are clearly likely to differ at many loci affecting the traits of interest. Up to this point we have ignored the presence of QTLs that are unlinked to the markers under consideration, but that nevertheless segregate in the F_2 or backcross generation. Thus the variance within marker genotype classes, σ_W, will contain a genetic component due to segregation of unlinked QTLs. Furthermore, QTLs identified by linkage to marker loci are not loci in the usual genetic sense (i.e., segments of DNA involved in producing a polypeptide chain) but are rather effective factors (Chapter 12) and may contain several loci

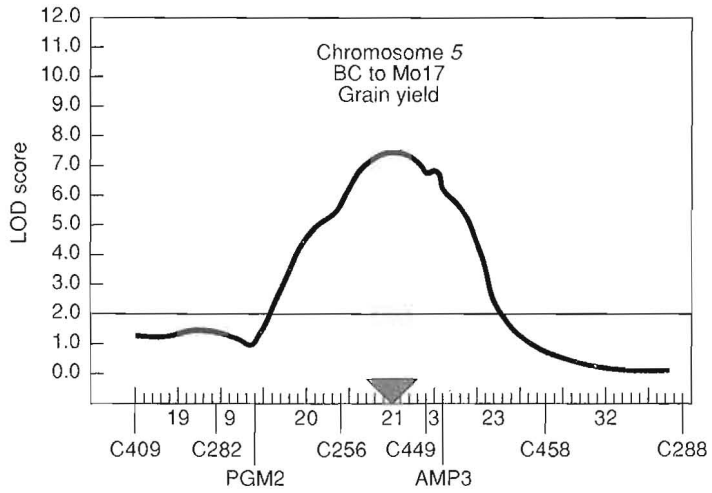

Fig. 21.2. Plot of LOD score against chromosomal position for QTLs affecting grain yield on maize Chromosome *5*, for a backcross of the F_1 between two élite maize inbred lines, B73 and Mo17, to the Mo17 parent. The positions of isozyme (PGM2 and AMP3) and RFLP (labelled C———) markers are given on the *x*-axis, with the distances between markers given in cM. The most likely position of a QTL for yield is indicated by the shaded triangle, corresponding to the peak in LOD score. (*The unpublished figure and data have kindly been provided by C.W. Stuber.*)

affecting the trait. Therefore when a test reveals a QTL linked to a marker, the effect observed may be due to two or more loci, not just one locus. The effect observed is then the aggregate of the effects of the two or more loci. The alleles at the linked loci may be in association (all increase the trait) or in dispersion. Consequently, the effects of the separate loci may be smaller, if in association, or larger, if in dispersion, than the observed aggregate effect.

There are methods for dealing with multiple QTLs simultaneously which improve the estimates of map positions and of effects. For descriptions of these methods, see Jansen and Stam (1994) and Zeng (1994). Methods that account for multiple QTLs are optimal for analysis of experimental data.

Experimental results

The general result, firmly established by experimental work on many traits in many organisms, is that there are multiple QTLs scattered throughout the genome. Thus the polygenic model of quantitative variation is clearly confirmed. Furthermore, the 'infinitesimal' model, in which there is a nearly infinite number of loci each with a vanishingly small effect, is clearly disproved. The experimental results are summarized below.

Number of loci

All estimates of the number of QTLs are minimum estimates of the true number of loci affecting the trait, for three reasons: (i) experiments are limited in their power

to separate closely linked loci, (ii) there must always be other loci with effects too small to be detected by an experiment of a particular size, and (iii) the loci found are those differentiating the two strains compared; other loci would probably be found in other strains.

Some examples of the number of QTLs affecting bristle number of *Drosophila melanogaster* are given in Table 21.4. There are three major chromosomes and one very small one; most of the experiments looked for QTLs on only one or two of the chromosomes. Pairs of parental chromosomes tested were either from a line selected for high bristle number and an unselected tester stock (Wolstenholme and Thoday, 1963; Spickett and Thoday, 1966), or from lines selected for high and low bristle number (Shrimpton and Robertson, 1988*a*, *b*; Long *et al.*, 1995). All or most of the difference in bristle number between the tested parental chromosomes is accounted for by the QTLs found. This means that all the important QTLs on the chromosomes tested were detected.

Table 21.4 Numbers of QTLs affecting sternopleural (ST) and abdominal (AB) bristle numbers in *Drosophila melanogaster*. Transposable element insertion sites were used as markers by Long *et al.* (1995); in the other studies morphological mutant markers were used.

Trait	*Chromosomes tested*	*Number of markers*	*Number of QTLs*	*% parental difference*	*Reference*
ST	*3*	4	2	114	Wolstenholme and Thoday (1963)
ST	*1, 2, 3*	10	5	87.5	Spickett and Thoday (1966)
ST	*3*	7	18	103	Shrimpton and Robertson (1988*b*)
AB	*1, 3*	45	7	89.9	Long *et al.* (1995)

More recently, molecular marker maps of tomato, maize, mice, and other species have been used to map QTLs affecting a wide variety of characters. Some examples are given in Table 21.5; for a more comprehensive review, see Tanksley, 1993. The QTLs detected account for large fractions of the phenotypic variance in the F_2 or BC generations. Without knowing the heritabilities of the traits in the populations used for mapping we cannot say how much of the genetic variance was accounted for, but most of the important QTLs must have been detected.

Gene effects

Additive effects In most studies, QTLs with large additive effects have been found. However, not all QTLs have equal effects, and the general pattern emerging for traits as different as *Drosophila* bristles and maize vegetative and reproductive characters is one in which a few QTLs with large effects account for most of the divergence between the parental strains, with a larger number of QTLs with smaller effects accounting for the remainder of the difference. Figure 21.3 shows the distribution of the effects of QTLs affecting sternopleural bristle number in *Drosophila* (Shrimpton and Robertson, 1988*b*). Eighteen QTLs on the third chromosome were found, and their effects (*a*) ranged from 0.4 to 1.9 σ_P. Only three had effects greater than 1.0 σ_P. The small number with effects below 0.6 σ_P does not mean that the real distribution falls off at the lower levels of effect. Those with smaller effects have a lower probability of being detected and therefore only some of them appear in the data. The real distribution is probably more highly skewed, with more QTLs of small effect.

Table 21.5 Examples of mapping QTLs by linkage to molecular markers. The phenotypic variance is for the F_2 or BC population.

Species	*Number of markers*	*Trait*	*Number of QTLs*	*% phenotypic variance*	*Reference*
Tomato (*Lycopersicon* sp.)	70	Fruit mass	6	58	Paterson *et al.* (1988)
		Soluble solids	4	44	
		Fruit pH	5	48	
	71	Fruit mass	7, 9	72	Paterson *et al.* (1991) (1)
		Soluble solids	4, 3, 5	44	
		Fruit pH	5, 8, 2	34	
Maize (*Zea mays*)	76	Grain yield	6, 8	61, 59	Stuber *et al.* (1992) (2)
		Plant height	3, 5		
		Ear number	2, 2		
Human	289	Type I diabetes	18		Davies *et al.* (1994)

(1) Estimates from plants grown in two, or three, environments. % Phenotypic variance is for the first value listed.
(2) Estimates from backcrosses to each of the parental lines, each averaged over six environments.

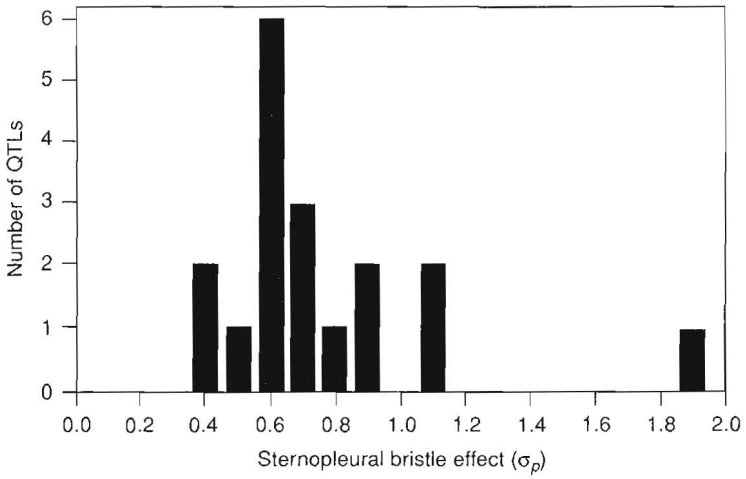

Fig. 21.3. Distribution of effects (a) of QTLs affecting sternopleural bristle number on Chromosome *3* of *Drosophila melanogaster*, in phenotypic standard deviation units. (*Adapted from Shrimpton and Robertson, 1988b.*)

Degree of dominance The degree of dominance of QTLs spans the entire range from additivity to complete dominance, and even to overdominance. This is illustrated in Fig. 21.4, which shows the distribution of the degree of dominance (d/a) among 74 QTLs identified in the F_2 from an interspecific cross of tomatoes. Dominance of the increasing and of the decreasing allele were equally frequent. The ratio d/a is difficult to estimate precisely, because it is subject to the sampling errors of both a and d. Estimates taking extreme values are especially open to suspicion because they result from very low values of a, and the ratio is then very sensitive to the sampling error of a. Cases of overdominance in particular need to be further confirmed.

The range of degree of dominance observed for QTLs contrasts with the complete dominance or recessivity normally shown by mutations with major phenotypic effects. Presumably, then, the alleles of QTLs responsible for quantitative variation actually produce quantitative differences in, rather than total absence of, the protein produced by the locus. This could be, say, altered activity if the gene product is an enzyme, or altered efficiency of binding if the gene product is a transcription factor. Few enzymes, however, act in isolation. Rather, they are steps in a metabolic pathway leading to an end product, and it is the amount, or the rate of production, of this end product that affects the phenotypes that we see. Considerations of enzyme kinetics in metabolic pathways lead to the conclusion that when there is a large difference in enzyme activity between two alleles, the allele with the higher activity will tend to be dominant, but when there is a small difference the alleles will tend to act additively (Kacser and Burns, 1981; see also Hartl, Dykhuizen, and Dean, 1985; Dean, Dykhuizen, and Hartl, 1988).

Epistasis Interaction between QTLs is difficult to detect because experiments with large numbers are needed. Strong epistatic interactions have been observed

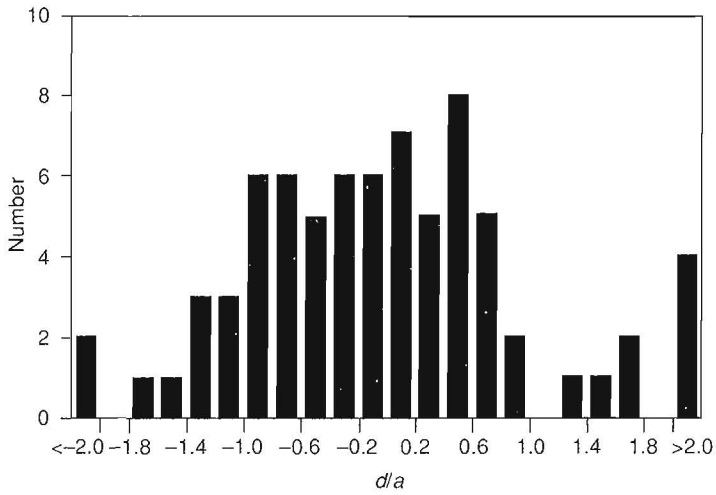

Fig. 21.4. Distribution of degrees of dominance (*d*/*a*) among 74 QTLs affecting vegetative traits of tomatoes, detected in the F_2 of a cross between the cultivated tomato and a wild species. (*Adapted from deVicente and Tanksley, 1993.*)

between QTLs affecting *Drosophila* bristle number (Spickett and Thoday, 1966; Shrimpton and Robertson, 1988*a*; Long *et al.*, 1995). The epistatic interaction effects detected by Long *et al.* (1995) were about the same magnitude as the main effects. Though epistasis may be common, this does not necessarily mean that it contributes a large proportion of the genetic variance in a random breeding population.

Correlated effects In view of the widespread occurrence of genetic correlation between traits, it is not surprising that many QTLs have been found to affect more than one trait. For example, several experiments with *Drosophila* (e.g., Shrimpton and Robertson, 1988*b*; Long *et al.*, 1995) have found that QTLs with the largest effects on the bristle trait for which the parental strains in the test had been selected also affected other bristle traits. And, in maize, a single chromosome region was found to affect as many as 78 of the 82 traits examined (Edwards, Stuber, and Wendel, 1987).

Some experiments have tested the same cross in different environments (Paterson *et al.*, 1991; Stuber *et al.*, 1992). Some of the QTLs were detected in more than one environment, as would be expected if performance in the different environments were genetically correlated.

Consistency

Some 'significant' QTLs may be false positives, and QTLs responsible for significant variation within and between populations can be missed if the tested strains are fixed by chance for alleles with similar effects. Therefore, QTLs should be confirmed by repeated experiments using the same and different strains. QTLs affecting *Drosophila* bristle number have been mapped using strains derived from

different base populations (Breese and Mather, 1957; Thompson and Thoday, 1974; Shrimpton and Robertson, 1988*b*; Long *et al.*, 1995); as have QTLs affecting tomato fruit traits (Weller, Soller, and Brody, 1988; Paterson *et al.*, 1988, 1991) and maize inflorescence traits (Doebley and Stec, 1991, 1993). While accord between different strains is far from perfect, there is a clear trend for estimated QTL map positions and effects to cluster in the same genomic regions. Of more interest, perhaps, are comparisons between species. Independent crosses mapping QTLs affecting seed weight in the cowpea (*Vigna unguiculata*) and mungbean (*Vigna radiata*) using a comparative RFLP linkage map revealed the QTL with the largest effect on seed weight in both species mapped to the same location (Fatokun *et al.*, 1992). It is possible that allelic variation at a restricted number of common loci is responsible for variation within and between populations, and even between species. Thus the use of experimentally tractable model systems, such as mouse models of human disease or of production traits in domestic animals, is well justified by the likely homology between QTLs in the model organism and the species of more practical interest.

From QTL to gene

Experiments to map QTLs typically succeed in localizing them to approximately 20 cM regions that potentially contain many loci affecting the trait. Linked QTLs within an interval can be separated by further recombination and their map positions pinpointed within roughly 3 cM by progeny testing, if no markers are available within the interval (Thoday, 1961; Lander and Botstein, 1989; but see McMillan and Robertson, 1974), or by fine-scale mapping, if the interval contains markers (Paterson *et al.*, 1990). This level of resolution may be sufficient to use the QTLs in selective breeding programmes, but is still several orders of magnitude away from identifying allelic differences at a single locus responsible for the difference in quantitative trait phenotypes. Identifying the actual loci affecting quantitative traits is necessary for risk assessment of polygenic human diseases, for application of transgenic technology to traits of agricultural importance, and for describing the genetic basis of quantitative variation in terms of allele frequencies and effects. There are two approaches for identifying a gene detected as a QTL in a particular genomic region; they are positional cloning, and association of variation in the quantitative trait phenotype with polymorphic markers at 'candidate' loci in the same region.

Positional cloning requires that the map position of the locus of interest is known to within a 0.3 cM interval, which is approximately the size of genomic inserts that can be contained in currently available cloning vectors. This can be achieved either by very high resolution meiotic mapping in experimental organisms, or by screening randomly mating populations for polymorphic markers in the region to which the QTL has been mapped that are in strong linkage disequilibrium with the quantitative trait phenotype. From this point it is conceptually straightforward but technically extremely arduous to identify the gene of interest and determine what are the polymorphisms associated with alleles of different effect. This method is feasible for loci defined by mutant alleles of large effect (reviewed by Takahashi.

Pinto, and Vituterna, 1994), and has been used to identify single loci affecting human diseases (for example, cystic fibrosis and Huntington's disease; reviewed by McKusick, 1994). The difficulty will be compounded significantly for loci with smaller effects, and the method has not yet been used to resolve QTLs to the level of single loci.

The most common strategy for going from mapped region to gene is the candidate gene approach. Often many loci of known function have been identified and cloned from the region to which the unknown locus maps. Known loci that could potentially give rise to the phenotype associated with the unknown locus are *candidate* loci, and the procedure is to search for associations of the phenotype with molecular polymorphisms at each candidate locus in the region. An approximate location of a QTL is not a prerequisite for proposing candidate loci known to be functionally related to a trait. Examples of the successful application of this approach to human diseases are the association of Apolipoprotein E (ApoE) alleles with total serum cholesterol and heart disease (reviewed by Sing *et al.*, 1988); the association of ApoE–4 alleles with late-onset Alzheimer's disease (Corder *et al.*, 1993), and the association of Type 1 diabetes with variation in the major histocompatibility HLA region and the insulin gene region (reviewed by Davies *et al.*, 1994).

If we are to apply the candidate locus approach to quantitative traits in general, we need to be able to propose relevant candidate loci functionally related to the traits. For the human diseases described above, detailed knowledge of the biological differences between affected and unaffected individuals guided the choice of candidate loci. For other quantitative traits, potential candidate loci may be those involved in the biochemical and developmental pathways leading to the phenotype of interest. For example, candidate loci that might account for variation in milk components in dairy cattle are casein and lactoglobulin loci (Pirchner, 1988); and candidate loci for quantitative variation in *Drosophila* bristle number may be loci that are necessary for bristle development (Mackay and Langley, 1990). Many candidate loci for quantitative traits thus have been identified by virtue of alleles with major mutant effects. Applying the candidate locus approach to QTLs is based on the assumption that segregating 'isoalleles' with small (i.e., not sufficiently large to qualify as Mendelian mutant) effects at these loci give rise to quantitative variation (Thompson, 1975; Mackay, 1985*b*; Robertson, 1985).

Examples that illustrate the power of this approach are the demonstration that the bovine *β-Lg* locus is associated with approximately half of the genetic variance of whey protein concentration in milk (Pirchner, 1988) and that the *Drosophila achaete-scute* and *scabrous* gene regions each account for over 20 per cent of the *X* and second chromosome genetic variation, respectively, of both abdominal and sternopleural bristle number (Mackay and Langley, 1990; Lai *et al.*, 1994). Furthermore, map positions of genomic regions identified by QTL mapping often roughly coincide with the location of a known major gene affecting the measured trait. QTLs affecting *Drosophila* bristle number map approximately to the locations of the major bristle loci *achaete-scute*, *Notch*, *bobbed*, *daughterless*, *scabrous*, *extramacrochaetae*, *hairy*, *malformed abdomen*, *Delta*, and *Enhancer of split* (Shrimpton and Robertson, 1988*b*; Long *et al.*, 1995). A dwarf locus, *compact*, is

located in the region associated with large effects on multiple maize yield traits (Edwards, Stuber, and Wendel, 1987), and the maize locus *teosinte-branched* maps to the region with large effects on inflorescence traits in two maize–teosinte crosses (Doebley and Stec, 1991, 1993).

The future for understanding quantitative traits in terms of complex genetics rather than statistical descriptions is bright. The various genome projects are yielding very dense linkage maps for humans, model organisms and species of agricultural importance that often show remarkable conservation of linkage groups across taxa. With the development of improved statistical methods for analysis of experimental crosses and pedigrees to detect segregating QTLs associated with molecular markers, and with the potential to resolve QTLs to the level of single genes, the description of the Mendelian genetic basis of quantitative variation is within reach.

Problem

21.1 The data below come from an experiment (Long *et al.*, 1995) in which QTLs affecting bristle numbers in *Drosophila melanogaster* were mapped on two of the three major chromosomes. The data here refer only to Chromosome *3*. Two-way selection for abdominal bristles was applied to a population derived from a large sample of flies from a natural population. After 25 generations of selection the high and low lines were each made homozygous for all its chromosomes, except the very small 4th, which was ignored. (This was done by special techniques for manipulating *Drosophila* chromosomes.) Then, in order to reduce unwanted variation and to have a common background on which to compare the effects of Chromosome *3*, Chromosomes *1* and *2* of the high line were replaced by their low-line homologues. These homozygous lines then differed only in their 3rd chromosome; they were the 'parental lines' for analysis. The parental lines were crossed. By means of the special techniques, 3rd chromosomes from F_1 females in which recombination had occurred were made homozygous, while the low-line homologues of Chromosomes *1* and *2* were retained as homozygotes. The recombinant 3rd chromosomes were then propagated in isogenic lines, in each of which one homozygous recombinant chromosome was replicated in many individuals. Forty individuals in each isogenic line were scored for abdominal bristle number and also for sternopleural bristle number.

	Abdominal bristles		*Sternopleural bristles*	
Parental lines	High	Low	High	Low
Mean	21.4	7.7	20.9	16.3
Interval 1	HH	LL	HH	LL
n	13	31	13	31
Mean	14.3	11.9	19.5	16.7
SS (Corrected)	277.2	374.6	17.4	11.2
Variance	23.1	12.5	1.4	0.4

The markers used to locate QTLs were the insertion sites of a transposable element known as *roo*. There were 29 sites on Chromosome *3* at which the parental lines differed, with an average distance of 3.8 cM between adjacent markers. The data here refer to one of the intervals, interval 1, at one end of the chromosome. Only the lines that had not recombined in this

interval are considered. There are thus two marker classes, HH and LL homozygotes, of which there were 13 and 31 lines respectively. In addition to these there were 5 lines that had recombined in the interval.

What conclusions can be drawn about a QTL in this interval? What is the source of the variance within marker classes? What might be the reason for sternopleural bristles having a much lower variance than abdominal bristles? [Solution 141]

Apper

Appendix Tab
of population v
the mean, in
values exceedi
values of p gre
negative sign;
linear interpola
+0.001 when p

$p\%$	x
0.01	3.719
0.02	3.540
0.03	3.432
0.04	3.353
0.05	3.291
0.06	3.239
0.07	3.195
0.08	3.156
0.09	3.121
0.10	3.090
0.10	3.090
0.12	3.036
0.14	2.989
0.16	2.948
0.18	2.911
0.20	2.878
0.22	2.848
0.24	2.820
0.26	2.794
0.28	2.770
0.30	2.748
0.32	2.727
0.34	2.706
0.36	2.687
0.38	2.669
0.40	2.652
0.42	2.636
0.44	2.620
0.46	2.605
0.48	2.590
0.50	2.576
0.50	2.576
0.55	2.543
0.60	2.512
0.65	2.484
0.70	2.457
0.75	2.432