Review of Statistical Methods for QTL Mapping in Experimental Crosses

Karl W. Broman, PhD

Identification of quantitative trait loci (QTLs) in experimental animals is critical for understanding the biochemical bases of complex traits, and thus for the identification of drug targets. The author reviews the basic statistical methods for mapping QTLs in experimental crosses and comments on a number of the statistical issues to consider in the application of these methods.

Broman is an Assistant Professor in the Department of Biostatistics, Johns Hopkins University, 615 N. Wolfe St., Baltimore, MD 21205. Please send reprint requests to the author at the above address or email: kbroman@jhsph.edu.

Traditional genetic studies have concentrated on dichotomous traits such as the presence or absence of a disease. Such traits are often the result of a mutation at a single gene. However, many interesting traits, like blood pressure or survival time after an infection, are quantitative in nature, and are affected by many genes and environmental factors. There are several reviews of the statistical methods for mapping quantitative trait loci (QTLs, the genes responsible for variation in quantitative traits) in experimental crosses[1-3]. Here, the attempt is to describe these methods to the reader who may have little detailed knowledge of statistics. This paper will sidestep most of the mathematical details, but the hope is to impart the important statistical concepts and issues.

Experimental Crosses

Repeated sibling matings of various experimental organisms—including mice, on which this review focuses—have led to the establishment of panels of well-defined strains. The process of inbreeding has fixed a large number of biomedically relevant traits in these strains. If two strains show consistent phenotypic differences, despite being raised in a common environment, the investigator may be confident that the strain difference has a genetic basis. The identity of the genes underlying such phenotypic differences may be revealed by performing a series of crosses, the simplest of which is the backcross (**Fig. 1**). The statistical methods used, and the statistical issues that arise, are largely the same for the different types of crosses; however, the backcross has the advantage of simplicity: at each locus in the genome, the backcross progeny have one of only two possible genotypes (genetic composition).

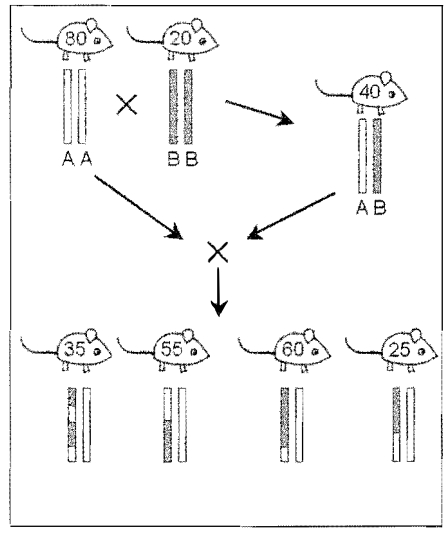

FIGURE 1. A backcross experiment begins with two inbred strains that differ in the trait of interest (*e.g.*, the response to an invasive procedure; the numbers on the mice indicate phenotype values). The two strains are crossed to produce the F_1 generation, which is then crossed back to one of the parental strains to obtain the backcross generation. The backcross generation exhibits genetic variation. The objective of the experiment is to identify genomic regions for which genotype predicts phenotype.

In a backcross, the investigator chooses two inbred strains, referred to here as the A and B parental strains, that differ in the trait of interest. The parental strains are crossed to obtain the first filial (F_1) generation. F_1 individuals receive a copy of each chromosome from each of the two parental strains; wherever the parental strains differ, the F_1 generation is heterozygous. The F_1 individuals are crossed to one of the two parental strains. For

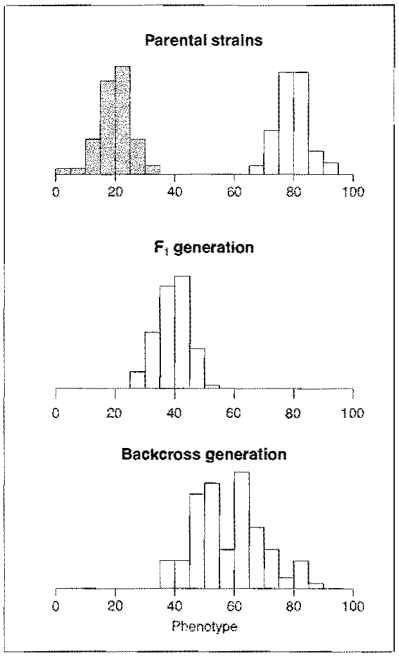

FIGURE 2. Histograms of the phenotype distributions in the parental strains, the F_1 generation, and the backcross generation.

example, if an F_1 individual is crossed with its A strain parent, the backcross progeny receive one chromosome from the A strain, and one from the F_1. Thus, at each locus, they have genotype AA or AB. The chromosome received from the F_1 parent is a mosaic of the two parental chromosomes, as a result of recombination during meiosis.

The investigator produces a number of backcross progeny (generally around 100 individuals) and determines the phenotype (the trait value) for each individual. Since this review considers quantitative phenotypes, rather than dichotomous ones, the phenotype will be a number, such as blood pressure, tumor mass, or survival time. Each individual is genotyped at a number of genetic markers, generally 10–20 centiMorgans (cM) apart, chosen to cover the genome uniformly. (The cM is the unit of genetic distance, and is equivalent to 1% recombination.) For each marker and each individual, it is observed whether the F_1 parent transmitted the A or the B allele. A genetic map specifying the order of the markers and the intermarker distances will be known or estimated based on the data. The objective is to identify genomic regions for which there is an association between the phenotype of a backcross individual and whether it received the A or B allele from the F_1 parent.

Fig. 2 contains histograms of the phenotype distributions for the parental strains, the F_1 generation, and the backcross generation for an imaginary backcross experiment. The parental strains were chosen to have markedly different phenotype distributions; the A and B strains have average phenotypes of 80 and 20, respectively. (To make this more concrete, think of the phenotype as time-to-death, in hours.) While individuals within each strain are genetically identical, there is some variation in the phenotypes due to environmental (nonheritable) differences and measurement error. Here, the phenotype distribution for the F_1 generation is intermediate between the two parental strains, but shows approximately the same degree of variation, with a standard deviation (SD) of about 5. (The SD may be interpreted as the typical difference from the average. Individuals in the F_1 generation have an average phenotype of about 40, but they typically deviate from that by about 5, having a phenotype between 35 and 45.) The F_1 generation need not be intermediate; for many traits, one observes heterosis (also known as hybrid vigor), wherein the F_1 hybrid exhibits greater fitness than either of the parental strains.

It is often assumed, though not always observed, that the degree of environmental variation will be independent of genotype, as is seen in Fig. 2—the SDs in the parental strains and in the F_1 generation are all about 5. The backcross generation, however, shows greater variation in phenotype because of genetic variation.

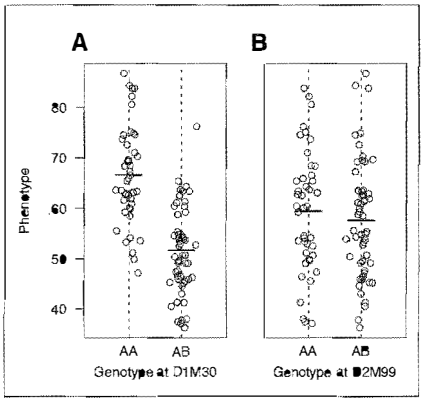

FIGURE 3. Dotplots of phenotypes of backcross progeny, split according to the genotypes (A) at a marker that appears to be linked to a QTL (D1M30) and (B) at a marker that appears not to be linked to any QTLs (D2M99). Horizontal line segments indicate the group averages.

The aim of QTL mapping is to identify regions of the genome that are contributing to variation in the trait of interest. In agricultural experiments, this knowledge may be used to design crosses leading to improved products. In biomedical experiments, the goal is to enhance understanding of the biochemical basis of the trait and to identify new drug targets.

Methods for Mapping QTLs

Analysis of Variance

The simplest method for QTL mapping is analysis of variance (ANOVA, sometimes called "marker regression") at the marker loci[4]. At each typed marker, one splits the backcross progeny into two groups, according to their genotypes at the marker, and compares the phenotype distributions of the two groups. For example, in Fig. 3A, we see that the individuals with genotype AA at marker D1M30 have somewhat higher phenotype values than those with genotype AB at that marker, indicating that the marker is linked to a QTL. In contrast, when the individuals are split according to their genotype at marker D2M99 (Fig. 3B), the phenotype distributions are approximately the same; this marker does not appear to be linked to a QTL.

The assessment of the strength of evidence for the presence of a QTL will be described in detail below. Briefly, in a backcross, one may calculate a t-statistic[5] to compare the averages of the two marker genotype groups. For other types of crosses (such as the intercross), where there are more than two possible genotypes, one uses a more general form of ANOVA, which provides a so-called F-statistic. These are both equivalent to the LOD score statistic (described below).

The chief advantage of ANOVA at the marker loci is its simplicity. In addition, a genetic map for the markers is not required, and

the method may be easily extended to account for multiple loci. A further advantage is the easy inclusion of covariates, such as sex, treatment, or an environment effect. Many phenotypes show marked sex differences, and these must be accounted for in QTL mapping. In addition, one may apply a treatment to some individuals but not others, or raise some individuals in one environment and others in a different environment.

The ANOVA approach for QTL mapping has three important weaknesses. First, we do not receive separate estimates of QTL location and QTL effect. QTL location is indicated only by looking at which markers give the greatest differences between genotype group averages, and the apparent QTL effect at a marker will be smaller than the true QTL effect as a result of recombination between the marker and the QTL. Second, we must discard individuals whose genotypes are missing at the marker. Third, when the markers are widely spaced, the QTL may be quite far from all markers, and so the power for QTL detection will decrease.

TABLE 1. Conditional probabilities for the QTL genotypes, given the genotypes at two flanking markers.

Marker Genotype		QTL Genotype	
left	right	AA	AB
AA	AA	$(1-r_L)(1-r_R)/(1-r)$	$r_L r_R/(1-r)$
AA	AB	$(1-r_L)r_R/(1-r)$	$r_L(1-r_R)/(1-r)$
AB	AA	$r_L(1-r_R)/(1-r)$	$(1-r_L)r_R/(1-r)$
AB	AB	$r_L r_R/(1-r)$	$(1-r_L)(1-r_R)/(1-r)$

r_L = recombination fraction between the left marker and the QTL
r_R = recombination fraction between the QTL and the right marker
r = recombination fraction between the two markers

Interval Mapping

Lander and Botstein[6] developed interval mapping, which overcomes the three disadvantages of analysis of variance at marker loci. Interval mapping is currently the most popular approach for QTL mapping in experimental crosses. The method makes use of a genetic map of the typed markers, and, like analysis of variance, assumes the presence of a single QTL. Each location in the genome is posited, one at a time, as the location of the putative QTL.

Given the marker genotype data (and assuming that the recombination process in meiosis exhibits no interference), one may calculate the probability that an individual has genotype AA (or AB) at a putative QTL (recall that we are considering a backcross to parental strain A). These QTL probabilities depend only on the genotypes at the nearest flanking typed markers, and are displayed in Table 1. As an example, consider two markers 20 cM apart, and suppose there is a QTL 7 cM from the left marker. A mouse that has genotype AA at both markers will also have genotype AA at the QTL, with a probability of ~99%. A mouse that has genotypes AA and AB at the left and right markers, respectively, will have genotype AA at the QTL with a probability of ~65% (and so will have genotype AB at the QTL with a probability of ~35%).

In interval mapping, we assume that, given the QTL genotype, the phenotype follows a normal distribution (a "bell curve"), with mean phenotype μ_A or μ_B, according to whether the QTL genotype is AA or AB, respectively, and common SD, σ. Given the genotypes at the markers flanking the QTL, the conditional phenotype distribution is then a mixture of the two normal distributions. For the example described above, the conditional phenotype distributions are displayed in Fig. 4. Consider mice with genotype AB at both markers: they will very likely have genotype AB at the QTL as well, and so their phenotypes will approximately follow a normal distribution with mean μ_B. Of the mice with genotype AB at the left marker and AA at the right marker, approximately 65% will have genotype AB at the QTL, while the other 35% will have QTL genotype AA. The distribution of phenotypes will then be as shown in Fig. 4, with 65% coming from a normal distribution with mean μ_B and 35% coming from a normal distribution with mean μ_A. The dashed curves in Fig. 4 correspond to the components of this mixture.

Fig. 5A displays the LOD ("logarithm of the odds favoring linkage," a score that measures the strength of evidence for the presence of a QTL) curve for a chromosome of length 100 cM, for an imaginary backcross of 100 individuals typed at markers every 10 cM. The LOD curve achieves its maximum at position 35 cM, indicating the presence of a QTL at this position. Note that the dots correspond to the typed marker loci. In the case of complete genotype data, the analysis of variance approach, described previously, provides exactly these dots. Interval mapping links these points togeth-

The Case of a Single QTL

Suppose that the mice with QTL genotype AA have average phenotype μ_A, while the mice with QTL genotype AB have average phenotype μ_B. The QTL thus has effect $\Delta = \mu_B - \mu_A$. Consider a marker locus that is a recombination fraction r away from the QTL. Of the individuals with marker genotype AA, a fraction $(1 - r)$ of them will have QTL genotype AA, while the remainder will have QTL genotype AB, and so these individuals have average phenotype $(1 - r)\,\mu_A + r\,\mu_B = \mu_A + r\,\Delta$. Similarly, the individuals with marker genotype AB have average phenotype $(1 - r)\,\mu_B + r\,\mu_A = \mu_B - r\,\Delta$. Thus, the difference between the phenotype averages for the two marker genotype groups is $(\mu_B - r\,\Delta) - (\mu_A + r\,\Delta) = \Delta\,(1 - 2r)$.

Note that when $r = 1/2$, the marker and QTL are unlinked, $\Delta\,(1 - 2r) = 0$, and the two marker genotype groups will have the same phenotype average. When $r < 1/2$, the marker and QTL are linked, $\Delta\,(1 - 2r) \neq 0$ (provided that that the QTL really does have an effect, $\Delta \neq 0$). Thus, a nonzero difference between the marker genotype groups indicates linkage between the marker and a QTL. Note that the difference between the phenotype averages for the two marker genotype groups will always be smaller (in absolute value) than the true QTL effect, Δ, unless there is complete linkage between the marker and the QTL ($r = 0$).

Geneticists often measure the effect of a QTL as the proportion of the phenotypic variance that is attributable to the QTL. The variance induced by a QTL is the variance in the trait that would be observed if there were no environmental variation nor measurement error, and no other QTLs. For a backcross and a QTL with an effect Δ (the difference in the phenotype averages, described above), this variance is $\Delta^2/4$, and so the proportion of the phenotypic variance attributable to the QTL is $\Delta^2/(4\tau^2)$, where τ^2 is the total phenotypic variance in the backcross generation.

er, and indicates that the best estimate for the QTL position is at a location between the markers at 30 and 40 cM. The 1.5-LOD support interval (the region where the LOD score is within 1.5 of its maximum) indicates the most plausible location for the QTL. A plot of the LOD score, re-centered so that its maximum is at 0 (see Fig. 5B), assists in identifying the evidence for QTL location.

Interval mapping has several advantages over analysis of variance at the marker loci. First, it provides a curve, such as that in Fig. 5A, which indicates the evidence for QTL location. Second, it allows for the inference of QTLs to positions between markers. Third, it provides improved estimates of QTL effects. (The apparent effect at a marker locus is attenuated as a result of recombination between the marker and the QTL). Fourth, and perhaps most important, appropriately performed interval mapping makes proper allowance for incomplete marker genotype data. In the calculation of an individual's QTL genotype probabilities, conditional on its marker genotype data, one considers the closest flanking typed markers for that individual. If an individual is missing the marker genotype for a flanking marker, one moves to the next flanking marker for which genotype data are available. Allowance may even be made for the presence of genotyping errors[7].

The key disadvantage to interval mapping, in comparison to analysis of variance, is that it requires some increase in computation time, and the use of specially designed software. Haley and Knott[8] described a method that approximates interval mapping quite well, but requires much less computation and can be carried out with standard statistical software. When possible, one should use the exact interval mapping method, but the Haley-Knott approach gives a remarkably good approximation that provides a reasonable first look at the linkage results. (The LOD curves displayed in this paper were actually calculated using the Haley-Knott approximation to interval mapping.)

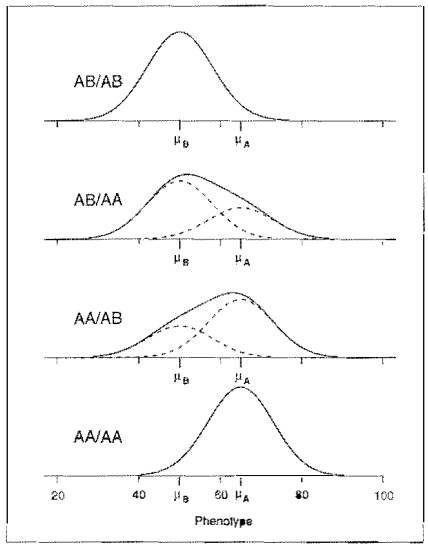

FIGURE 4. Conditional phenotype distributions given marker genotypes at positions flanking a QTL. Markers are assumed to be 20 cM apart, with the QTL located 7 cM from the left marker. For the recombinant genotypes, dashed curves indicate the components of the phenotype distributions.

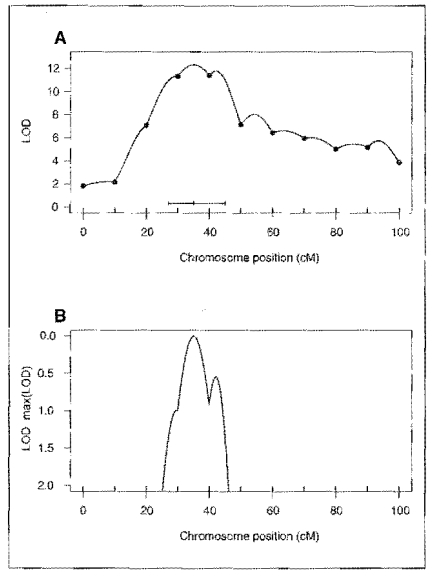

FIGURE 5. (A) LOD curve for an imaginary backcross composed of 100 mice genotyped at markers every10 cM; the 1.5-LOD support interval is indicated. (B) The LOD curve re-centered so that the maximum is at 0 to better indicate the evidence for QTL location.

LOD Scores

Given a putative QTL at location z, if we knew the three QTL parameters, μ_A, μ_B, and σ, we could write down (although we won't do so here) the probability of the observed data, Pr(data | QTL at z, μ_A, μ_B, σ). (The vertical bar is to be read as "given.") Considered as a function of the three unknown parameters, this is called the *likelihood*. In interval mapping, we obtain *maximum likelihood estimates* of the three parameters, defined to be the values for which this probability achieves its maximum; we denote these estimates $\hat{\mu}_A$, $\hat{\mu}_B$, $\hat{\sigma}$. (Note that these estimates will be different for different putative QTL locations, z.) We then form the LOD score:

$$\text{LOD}(z) = \log_{10}\left\{\frac{\Pr(\text{data}|\text{QTL at } z, \hat{\mu}_A, \hat{\mu}_B, \hat{\sigma})}{\Pr(\text{data}|\text{no QTL})}\right\}$$

The LOD score measures the strength of the evidence for the presence of a QTL at the location z, compared to there being no segregating QTL in the backcross. Larger LOD scores correspond to greater evidence for the presence of a QTL. The LOD score is calculated at each position of the genome (or, in practice, just every 0.5 cM or so).

Statistical Significance

When confronted with a LOD curve (or, for the mouse, with 19 or 20 such curves, one for each chromosome), a natural question arises: is an observed peak actually a QTL? As mentioned above, the LOD score indicates the strength of evidence for the presence of a QTL, with larger LODs corresponding to greater evidence. The question is, how large is large? The standard approach to answering this question has been to formulate the problem as one of hypothesis testing.

Consider the *null* hypothesis, that there are no QTLs segregating in the backcross. We determine the distribution of the LOD score in this situation. The probability of obtaining a LOD score as large or larger than that which was observed, if there were no QTLs, is called the P value. Large LOD scores give small P values; very small P values indicate that either the null hypothesis is false (there really is a QTL) or a very rare event occurred.

The solid curve in Fig. 6 corresponds to the approximate distribution of the LOD score for a particular genomic position, given that there are no QTLs segregating in the backcross. LOD scores above 1 appear to be rare (giving a P value of about 3%), and so provide reasonably strong evidence for the presence of a QTL. Unfortunately, our story is not yet complete. When one performs a genome scan to identify QTLs, one examines the LOD score at 100

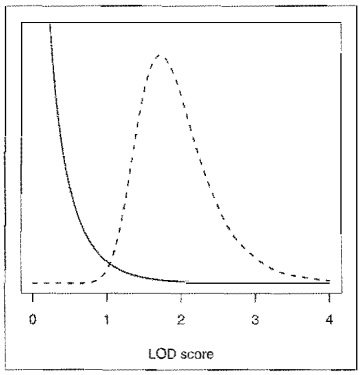

FIGURE 6. Null distribution of the LOD score at a particular genomic position (solid curve) and of the maximum LOD score from a genome scan (dashed curve).

or more marker loci (and, using interval mapping, at all locations between markers). Thus, the null distribution of the LOD score at a single location is not appropriate for forming an overall threshold. Some adjustment must be made for our examination of multiple putative QTL locations over the whole genome. What we are after is the distribution, assuming that there are no QTLs segregating in the cross, of the maximum LOD score across the entire genome. Stated another way, we seek the chance of obtaining this large a LOD score somewhere in the genome, if there were no QTLs anywhere.

The approximate distribution of the maximum LOD score, given no QTLs, is shown as the dashed curve in Fig. 6. Even if there are no QTLs, one will typically see a LOD score of 2 or more somewhere in the genome. LOD scores must be closer to 3 before they will generally be deemed interesting. (A LOD score of 3 indicates that the chance of obtaining the observed data, given that there is a QTL at the specified position, is 1,000 times more likely than if there are no QTLs.) The efforts of many statisticians have been expended in estimating this dashed curve (or the LOD threshold, which corresponds to a P value of 5%). This curve, the null distribution of the maximum LOD score, depends on the type of cross (backcross or intercross), the size of the genome, the number and spacing of genetic markers, the amount and pattern of missing genotype information, and the true phenotype distribution.

Lander and Botstein[6] performed extensive computer simulations to estimate the appropriate LOD threshold for various genome sizes and marker densities, and gave analytical calculations for the case of a very dense marker map. These guidelines should suffice for most uses. Another approach (which requires somewhat hefty computation, but provides an estimate of the null distribution of the maximum LOD score, given the observed phenotype distribution, marker spacing, and pattern of missing genotype data) is to perform a permutation test[9]. One permutes (randomizes) the phenotype data, keeping the genotype data intact; performs interval mapping; and identifies the maximum LOD score, across the genome. This process is repeated 1,000 times. The observed LOD score (with the phenotypes in the correct order) is compared to the 1,000 LOD scores obtained from permuted versions of the data. The proportion of these 1,000 LOD scores that exceed the actual, observed LOD score, is reported as an approximate P value. This provides a customized threshold, tailor-made for the individual experiment.

How Many Markers? How Many Mice?

In considering how many mice to obtain and how many markers to type, one thinks about both the chance of detecting QTLs and the resolution of localization of QTLs. (For further discussion of these issues, see van Ooijen[10].) The chance of detecting a QTL is called the "power." Suppose that under the null hypothesis of no segregating QTLs, one obtains a maximum LOD score, genomewide, of at least 3 only 5% of the time, so the threshold of 3.0 may be used to define significant evidence for the presence of a QTL. In this case, the power to detect a QTL is the chance that one will obtain a LOD score above 3 in the region of the QTL.

Power depends on the type of cross, the size of the effect of the QTL, the number of mice obtained, the density of typed markers in the region of the QTL, and the stringency of the chosen LOD threshold (*i.e.*, the significance level). Fig. 7 provides an illustration of the power to detect a QTL in a backcross with phenotypic SD 11.5, as a function of the size of the QTL's effect and the number of mice. The dashed curve corresponds to the null distribution of the maximum LOD score, genomewide (see Fig. 6); a LOD score of 3 corresponds to a P value of approximately 5% because the area under the dashed curve, to the right of 3, is approximately 5% of the total area. The solid curves correspond to the distribution of the LOD score at the QTL, for QTLs with different sizes of effect, for a backcross with 100 mice. (Note that the proportion of the phenotypic variance attributable to the QTL is 5, 12, and 22%, respectively, for the three examples in Fig. 7.) The shaded area under this curve, to the right of 3, is the power to detect the QTL. As the size of the QTL effect increases, the distribution of the LOD score shifts to the right, and the power to detect the QTL increases. The dotted curves correspond to the distribution of the LOD score with 200 mice, rather than 100; the curves shift appreciably to the right as the power to detect the QTL increases.

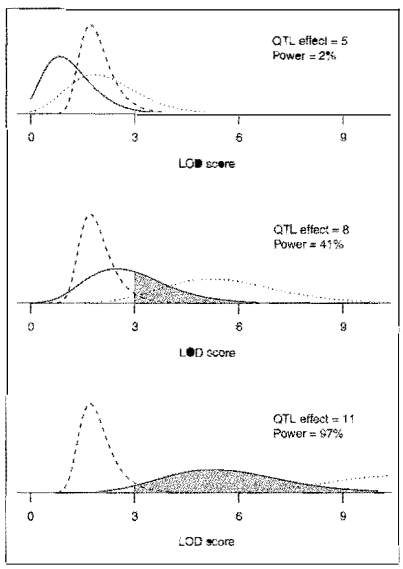

FIGURE 7. Distribution of LOD score in the presence of a QTL. The dashed curve corresponds to the null distribution of the maximum LOD score, genomewide. The solid and dotted curves correspond to the distributions of the LOD score for a QTL with various sizes of effects for crosses with 100 and 200 mice, respectively. The shaded regions correspond to the cases where significant genomewide evidence for the presence of a QTL would be obtained.

Fig. 7 illustrates the power to detect a specific QTL. When a QTL has an effect of only moder-

ate size, this power can be extremely low. It is possibly more interesting to consider the power to detect at least one QTL. If there are 10 unlinked QTLs segregating in a cross, and for each of them, the power is only 20%, one will still have approximately 90% power to detect at least one of them. This has implications for the replication of experiments; if there are many moderate-sized QTLs segregating in a particular cross, then from one group of mice to the next, the set of QTLs for which one will obtain strong evidence may be quite different. Of course, QTLs with quite strong effect will be detected with high power, and so will be seen with each group of mice.

Fig. 8A presents the LOD curves (re-centered to have a maximum at 0) obtained for 100 mice typed at markers every 10 cM (solid curve) and every 1 cM (dashed curve). In this example, the additional typed markers provided no assistance in localizing the QTL. In Fig. 8B, we have increased the number of mice to 200. When typed at a 10 cM spacing, the additional mice do little to improve the localization of the QTL. (The evidence for a QTL in the region, as reflected in the maximum LOD score, increased tremendously, although that is not shown in the figure.) However, with 200 mice typed at a 1 cM spacing (Fig. 8B, dashed curve), the precision of localization of the QTL is greatly improved. But these results are not necessarily typical. The author recommends that initial genotyping in an experimental cross be performed with markers at a 10-15 cM spacing. Typing additional markers in the region of an inferred QTL may improve the resolution of its localization, but such improvement will likely only occur if one has typed many mice or the QTL has a relatively large effect.

Selection Bias

An important, yet often ignored, issue in QTL mapping concerns selection bias in the apparent (estimated) effects of QTLs. Such estimated effects are often too large. Consider a single QTL with an effect of moderate size, and imagine there is a marker very near the QTL. In a particular experiment, the estimated effect of the QTL will be somewhat different from its true effect—the observed difference between the phenotype averages for the two QTL genotype groups will not be the same as the true difference. Nevertheless, to produce a LOD score sufficiently large for us to declare the presence of a QTL, the estimated effect must be large. This introduces bias in the estimated effect. (Bias is also introduced in the maximization over possible QTL locations; the inferred location for a QTL is the one that gives the largest estimated QTL effect.) Because this bias is the result of the selection of only those loci for which there is sufficient evidence for the presence of a QTL, we call it "selection bias."

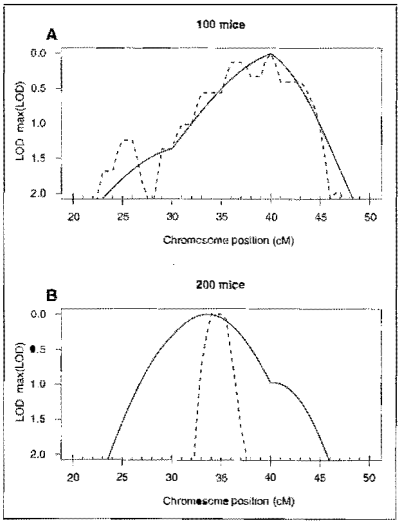

FIGURE 8. Effect of increasing marker density and increasing the number of mice—from (A) 100 to (B) 200—on the precision of localization of a QTL. Solid and dashed curves correspond to 10 and 1 cM marker spacings, respectively.

Fig. 9 displays the distribution of the estimated effect of a QTL, given its true effect, for a backcross with 100 mice, where the phenotype SD is 11.5. To produce a LOD score above 3.0, the estimated effect of the QTL must be at least 8.3. In the case that the true effect of a QTL is 5, the power to detect the QTL will be only ~2%, and in the cases where it is detected, its estimated effect will be nearly twice too large. (For this case, the average estimated effect, given that significant evidence for a QTL is obtained, is 8.93, and so the bias is 8.93 – 5 = 3.93. Viewed as a proportion of the true effect, the bias is 79%.)

The power to detect QTLs with a larger effect is higher, and the bias in their estimated effects will be lower, but may still be substantial. QTLs with very large effect are always detected, and so the bias in their estimated effects will be minimal.

For a particular inferred QTL with an estimated effect of moderate size, we will not know whether it is a weak QTL that we were extremely lucky to detect, and whose true effect is really rather small, or a QTL of truly large effect that happened to appear to be not so strong in this particular set of mice; the estimated effects of QTLs can often be overly optimistic.

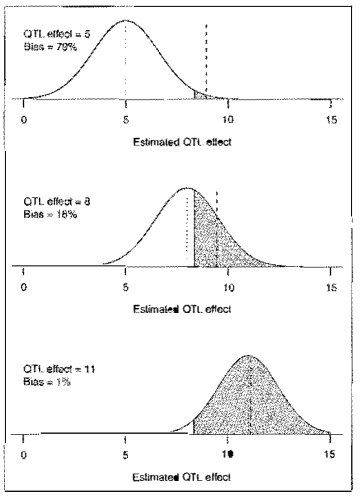

FIGURE 9. Illustration of selection bias in the estimated QTL effect. The curves correspond to the distribution of the estimated QTL effect for different values of the true effect (indicated by the dotted lines). The shaded regions correspond to the cases where significant genomewide evidence for the presence of a QTL would be obtained. The dashed vertical lines indicate the average estimated QTL effect, conditional on the detection of the QTL.

Multiple QTLs

Interval mapping assumes the presence of a single QTL. One may use interval mapping to identify multiple QTLs, especially when they are on separate chromosomes, but there are several advantages to using methods that model multiple QTLs simultaneously. First, by controlling for the presence of a QTL, one may reduce the residual variation and obtain greater power to detect additional QTLs. Second, one may better separate linked QTLs. Third, the identification of interactions between QTLs (called epistasis[11-13]) requires the joint modeling of multiple QTLs.

Before describing the major approaches for mapping multiple QTLs, let us examine the phenomenon of epistasis. (Strickberger[14] provides an interesting discussion of epistasis.) Fig. 10A displays the phenotype averages when splitting mice into four groups according to their joint genotypes at two QTLs, that are acting additively. For example, the average phenotype for mice with genotype AA at both QTLs is 40, while that for mice that are AB at QTL 1 and AA at QTL 2 is 60. The effect of a QTL is the change in the phenotype average when the QTL genotype is changed from AA to AB. If we consider only those mice whose genotype at QTL 2 is AA (the circles in Fig. 10A), the effect of QTL 1 is 20. Similarly, for those mice whose genotype at QTL 2 is AB (the triangles in Fig. 10A), the effect of QTL 1 is again 20. The QTLs are acting additively if the effect of QTL 1 is the same, irrespective of the genotype at QTL 2. (The two lines in Fig. 10A are parallel.) Similarly, the effect of QTL 2 is the same, irrespective of the genotype at QTL 1.

Fig. 10B displays the phenotype averages for two QTLs that are not additive. QTL 1 has effect 10 when the genotype at QTL 2 is AA, but has effect 30 when the genotype at QTL 2 is AB. Similarly, QTL 2 has no effect when the genotype at QTL 1 is AA, but has an effect of 20 when the genotype at QTL 1 is AB. In this case, we say that the QTLs exhibit epistasis. Such an interaction may arise when two genes are part of a common biochemical pathway, with gene 1 upstream of gene 2, so that in individuals homozygous for a null mutation at gene 1, mutations in gene 2 have no effect. This is the origin of the term epistasis[15], which means literally "to stop." Statistical geneticists now apply the term more widely, to indicate any deviation from additivity between QTLs[12].

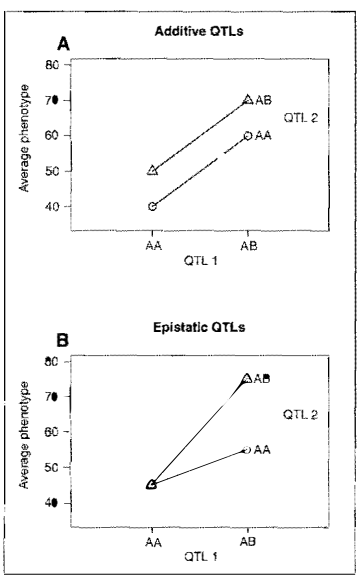

FIGURE 10. Phenotype averages, conditional on the joint genotypes at two additive QTLs (A) and two epistatic QTLs (B).

Model Selection

Consider the case of dense markers and relatively complete genotype data. In this situation, it is appropriate to use the immediate extension of analysis of variance, multiple regression[16]. For the moment, imagine that the QTLs are acting additively. Let y denote the phenotype for a mouse, and let $x_i = 1$ or 0, according to whether the mouse has genotype AB or AA at marker i. We assume

$$y = \mu + \sum_i \Delta_i x_i + \varepsilon$$

where μ is the average phenotype for mice with genotype AA at all loci, Δ_i is the effect of marker i, and ε follows a normal distribution with mean 0 and SD σ. We are left with the question: which of the markers (x's) should be included in the model (as proxies for QTLs)? Whereas for interval mapping and analysis of variance we discussed hypothesis testing, here we are confronted with the problem of model selection[17]. The model is the set of inferred QTLs and, possibly, their interactions.

Model selection entails four distinct steps: 1) select a class of models (*e.g.*, additive models or models including pairwise interactions between QTLs); 2) search through the space of models (there may be more possible models than may be inspected individually); 3) compare models; and 4) assess the performance of a model selection procedure.

Consider the class of models composed of a finite number of QTLs, acting additively, where QTLs may occur only at the marker loci. If there are 100 markers, then there are $2^{100} \approx 10^{30}$ such models. There are approximately 10^{13} models that include 10 or fewer markers. If one allows only three or fewer QTLs, one may perform a simultaneous search to consider each such model. But if one wishes to consider the possibility of many more QTLs, it is impossible to inspect each possible model individually, so one must form some procedure for searching through this space of models to pick out the best ones without looking at all of them.

The simplest approach for searching through models is forward selection. First, look at all models with one marker and pick the best one. Next, search through all two-marker models, which include the first selected one. Continue adding additional markers, one at a time, to obtain a nested sequence of increasingly larger models. Alternatively, one may use backward deletion: consider the model including all markers, drop the worst of them, and repeat to create a decreasing, nested sequence of models. Stepwise selection (alternating between addition and deletion of markers) and randomized searches may improve on these simple searches by looking at a larger portion of the space of models, but we have shown that, in the case of QTLs acting additively, the simplest approach, forward selection, behaves as well as the more complex search methods[18]. However, for more complex classes of models, such as the inclusion of pairwise interactions, more-extensive searches may be necessary.

The more difficult problem is that of comparing models. For models of the same size (*i.e.*, including the same number of markers), one may look at the residual sum of squares (RSS, *see* Box). For models of different sizes, however, one must place a penalty on the size of the model: the inclusion of an additional marker is allowed only if it gives an appreciable increase in the explanatory ability of the model.

Finally, it is important to consider how one may assess the performance of a model selection procedure. In making decisions about the appropriate criteria for comparing models (and of searching through the space of models), it is important to consider the performance characteristics of different possible procedures: What proportion of QTLs will be detected and how often will extraneous loci be included? Decisions should be guided by the aims of the study. In a study seeking to use marker-assisted selec-

RSS: Residual Sum of Squares

Let y_j denote the phenotype of mouse j, and let $\hat{y}_j$ denote the fitted phenotype value under some model. Then RSS = $\Sigma_j (y_j - \hat{y}_j)^2$, which measures the amount of variation in the phenotype that has not been "explained" by the model. A model that explains a greater proportion of the variation in the phenotype (*i.e.*, has a smaller RSS) is preferred. For models of different sizes, however, one finds that the inclusion of additional markers always reduces the RSS. Thus one must place a penalty on the size of the model: the inclusion of an additional marker is allowed only if it results in a decrease in the RSS above some specified amount. This approach is equivalent to considering a conditional LOD score; we compare the probability of the data under a model including a set of markers, M, plus an additional marker, x, to the probability of the data given just the set, M:

$$\text{LOD}(x|M) = \log_{10}\left\{\frac{\Pr(\text{data}|M + x)}{\Pr(\text{data}|M)}\right\}$$

In deciding whether to allow the inclusion of x as an inferred QTL, one may require that the conditional LOD score be above some threshold. How to choose such thresholds is a matter of research; statisticians have not yet adequately solved this problem.

tion to improve an agricultural product, one may be willing to allow a few extraneous loci in an effort to identify a reasonably large number of QTLs. A scientist wishing to positionally clone a QTL may be satisfied only with a small number of strongly supported QTLs—this avoids wasting expensive and time-consuming efforts on extraneous loci. These sorts of aims should guide the researcher in framing the desired performance characteristics for a procedure, which may then be used in choosing an appropriate mapping method. One will need to rely on experience, educated guesses, and large computer simulation studies, because, unfortunately, the appropriate mapping method will vary with the context.

Multiple QTL Methods

We now turn to a description of the major statistical approaches for QTL mapping that makes use of multiple QTL models. We have just described the simplest such method, multiple regression. The aim was principally to frame the problem as one of model selection and to describe the key issues in model selection (the most important of which was the choice of criteria for comparing models). While this simple approach should be more widely used, it shares many of the disadvantages of analysis of variance at marker loci; most important, it requires complete marker genotype data.

The simplest multiple QTL method that makes allowance for missing genotype data is the use of forward selection in interval mapping. One identifies a putative QTL in a genome scan using interval mapping, and then obtains the residuals, $y - \hat{y}$, where y is the observed phenotype and $\hat{y}$ is the predicted phenotype, given the individual's marker genotypes. These residuals are then used as new phenotypes, and interval mapping is performed again. This procedure is appropriate when QTLs may be assumed to act additively. Its advantages are in increasing the power to detect additional QTLs and in separately linked QTLs. Recall the LOD curve in Fig. 5A, where the maximum LOD occurred at position 35 cM. If we perform interval mapping on this chromosome a second time using the residuals obtained assuming a QTL at 35 cM, we obtain the conditional LOD curve shown in Fig. 11. The maximum conditional LOD slightly exceeds 3.0, with a 1.5-LOD support interval covering the latter half of the chromosome. There is evidence for the presence of a second QTL on this chromosome, although the location of this second QTL is not well resolved.

An approach that has received much attention and has been widely applied in practice is composite interval mapping (CIM)[19-22]. In this method, one performs interval mapping using a subset of marker loci as covariates. These markers serve as proxies for other QTLs to increase the resolution of interval mapping, by accounting for linked QTLs and reducing the residual variation. The key problem with CIM concerns the choice of suitable marker loci to serve as covariates; once these have been chosen, CIM turns the model selection problem into a single-dimensional scan. The choice of marker covariates has not been solved, however. Not surprisingly, the appropriate markers are those closest to the true QTLs, and so if one could find these, the QTL mapping problem would be complete anyway. The author recommends against the use of CIM.

An interesting development is multiple interval mapping (MIM)[23,24]. MIM is the extension of interval mapping to multiple QTLs, just as multiple regression extends analysis of variance. MIM allows one to infer the location of QTLs to positions between markers, makes proper allowance for missing genotype data, and can allow interactions between QTLs. This is not the final solution to the QTL mapping problem; one is still confronted with comparing models and searching through models. Statistical researchers have much work to do in this area.

We have described the major approaches to QTL mapping in experimental crosses. Several other approaches are available, including Bayesian methods[25-28] and the use of a genetic algorithm[29]. These new methods may become important in the future, but are beyond the scope of this elementary description of statistical methods for QTL mapping.

Summary

The simplest statistical method for QTL mapping is analysis of variance at marker loci. This approach suffers when there is appreciable missing marker genotype data and when the markers are widely spaced. Interval mapping, though more complicated and more computationally intensive, allows for missing genotype data. LOD scores are used to measure the strength of evidence for the presence of a QTL; the LOD curve for a chromosome indicates whether a QTL may be present and where it is likely to be located. The region where the LOD score is

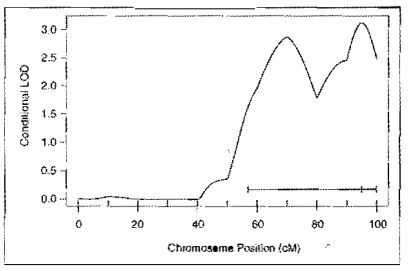

FIGURE 11. Conditional LOD curve, obtained by performing interval mapping using the residuals after fitting a model with a QTL at 35 cM. The 1.5-LOD support interval is indicated.

within 1.5 of its maximum may be taken as the plausible region for the location of the QTL. (A plot of the LOD curve, re-centered so that its maximum is at 0, is a valuable tool for depicting the evidence for QTL location.)

In determining whether a LOD score is sufficiently large for one to be confident of the presence of a QTL, consider the distribution of the LOD score under the null hypothesis of no segregating QTL. Adjustment must be made for the genomewide search for QTLs, so consider the distribution of the maximum LOD score genomewide. Permutation tests are valuable for determining significance landmarks for the LOD score; although computationally intensive, permutation tests allow for the observed phenotype distribution, marker density, and pattern of missing genotype data.

In determining how many animals to obtain and what density of markers to genotype, consider the power to detect a particular QTL and the precision with which QTLs may be localized. Typing additional markers in the region of an inferred QTL to improve its localization may not be successful unless there are many individuals or the QTL has a relatively strong effect.

Selection bias is an important issue that has been largely ignored in considering the estimated effects of inferred QTLs. The estimated effects of QTLs are generally optimistically large.

Interval mapping and analysis of variance make use of a single-QTL model. Methods that consider multiple QTLs simultaneously have three advantages: greater power to detect QTLs, greater ability to separate linked QTLs, and the ability to estimate interactions between QTLs. These more complex methods may facilitate the identification of additional QTLs and assist in elucidating the complex genetic architecture underlying many quantitative traits.

Model selection is the principal problem in multiple QTL methods; the chief concern is the formation of appropriate criteria for comparing models. The simplest multiple QTL method, multiple regression, should be used more widely, although, like analysis of variance, it suffers in the presence of appreciable missing marker genotype data. A forward selection procedure using interval mapping (*i.e.*, the calculation of conditional LOD curves) is appropriate in cases of QTLs that act additively, and makes proper allowance for missing genotype data. MIM is an improved method, that, although computationally intensive, can, in principle, map multiple QTLs and identify interactions between QTLs. The important aspects of the model selection problem require much further study, and will not have general solutions.

Acknowledgments

Victor Boyartchuk, Gary Churchill, Mark Neff, Sara Okamura, and Saunak Sen generously provided comments for improvement of the manuscript.

Received 3/14/01; accepted 4/2/01.

References:

1. Doerge, R.W., *et al.* Statistical issues in the search for genes affecting quantitative traits in experimental populations. *Stat. Sci.*; **12:**195-219, 1997.
2. Lynch, M. and Walsh, B. *Genetics and Analysis of Quantitative Traits.* Sinauer Associates, Sunderland, MA, pp. 431-89, 1998.
3. Broman, K.W. and Speed, T.P. A review of methods for identifying QTLs in experimental crosses. *In:* Seillier-Moiseiwitsch, F., *ed., Statistics in Molecular Biology and Genetics.* Vol. 33 of IMS Lecture Notes – Monograph Series, pp. 114-42, 1999.
4. Soller, M., *et al.* On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theor. Appl. Genet.*; **47:**35-9, 1976.
5. Bailey, N.T.J. *Statistical Methods in Biology,* 3rd Edition. Cambridge University Press, 1995.
6. Lander, E.S. and Botstein, D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*; **121:**185-99, 1989.
7. Lincoln, S.E. and Lander, E.S. Systematic detection of errors in genetic linkage data. *Genomics*; **14:**604-10, 1992.
8. Haley, C.S. and Knott, S.A. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*; **69:**315-24, 1992.
9. Churchill, G.A. and Doerge, R.W. Empirical threshold values for quantitative trait mapping. *Genetics*; **138:**963-71, 1994.
10. van Ooijen, J.W. Accuracy of mapping quantitative trait loci in autogamous species. *Theor. Appl. Genet.*; **84:**803-11, 1992.
11. Shrimpton, A.E. and Robertson, A. The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster.* I. Allocation of third chromosome sternopleural bristle effects to chromosome sections. *Genetics*; **118:**437-43, 1988.
12. Frankel, W.N. and Schork, N.J. Who's afraid of epistasis? *Nat. Genet.*; **14:**371-3, 1996.
13. Mackay, T.F.C. The nature of quantitative genetic variation revisited: Lessons from *Drosophila* bristles. *BioEssays*; **18:**113-21, 1996.
14. Strickberger, M.W. *Genetics,* 3rd Edition. Macmillan, New York, pp.180-8, 1985.
15. Bateson, W. Facts limiting the theory of heredity. *Science*; **26:**649-60, 1907.
16. Cowen, N.M. Multiple linear regression analysis of RFLP data sets used in mapping QTLs. *In:* Helentjaris, T. and Burr, B., *eds. Development and Application of Molecular Markers to Problems in Plant Genetics.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp.113-6, 1989.
17. Miller, A.J. *Subset Selection in Regression.* Chapman and Hall, New York, 1990.
18. Broman, K.W. Identifying quantitative trait loci in experimental crosses. Ph.D. thesis, Department of Statistics, University of California, Berkeley; 1997.
19. Jansen, R.C. Interval mapping of multiple quantitative trait loci. *Genetics*; **135:**205-11, 1993.
20. Jansen, R.C. and Stam, P. High resolution of quantitative traits into multiple loci via interval mapping. *Genetics*; **136:**1447-55, 1994.
21. Zeng, Z.-B. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc. Natl. Acad. Sci. USA*; **90:**10972-6, 1993.
22. Zeng, Z.-B. Precision mapping of quantitative trait loci. *Genetics*; **136:**1457-68, 1994.
23. Kao, C.-H., *et al.* Multiple interval mapping for quantitative trait loci. *Genetics*; **152:**1203-16, 1999.
24. Zeng, Z.-B., *et al.* Estimating the genetic architecture of quantitative traits. *Genet. Res.*; **74:**279-89, 1999.
25. Hoeschele, I. and VanRaden, P.M. Bayesian analysis of linkage between genetic markers and quantitative trait loci. I. Prior knowledge. *Theor. Appl. Genet.*; **85:**953-60, 1993.
26. Satagopan, J.M., *et al.* A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. *Genetics*; **144:**805-16, 1996.
27. Uimari, P. and Hoeschele, I. Mapping-linked quantitative trait loci using Bayesian analysis and Markov chain Monte Carlo algorithms. *Genetics*; **146:**735-43, 1997.
28. Sillanpää, M.J. and Arjas, E. Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. *Genetics*; **151:**1605-19, 1999.
29. Carlborg, O., *et al.* The use of a genetic algorithm for simultaneous mapping of multiple interacting quantitative trait loci. *Genetics*; **155:**2003-10, 2000.