Genet. Res., Camb. (1999), **74**, *pp.* 279–289. Printed in the United Kingdom © 1999 Cambridge University Press.

Estimating the genetic architecture of quantitative traits

ZHAO-BANG ZENG[1]*, CHEN-HUNG KAO[2] AND CHRISTOPHER J. BASTEN[1]
[1] *Program in Statistical Genetics, Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA*
[2] *Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan, ROC*

(*Received 13 May 1999 and in revised form 5 August 1999*)

Summary

Understanding and estimating the structure and parameters associated with the genetic architecture of quantitative traits is a major research focus in quantitative genetics. With the availability of a well-saturated genetic map of molecular markers, it is possible to identify a major part of the structure of the genetic architecture of quantitative traits and to estimate the associated parameters. Multiple interval mapping, which was recently proposed for simultaneously mapping multiple quantitative trait loci (QTL), is well suited to the identification and estimation of the genetic architecture parameters, including the number, genomic positions, effects and interactions of significant QTL and their contribution to the genetic variance. With multiple traits and multiple environments involved in a QTL mapping experiment, pleiotropic effects and QTL by environment interactions can also be estimated. We review the method and discuss issues associated with multiple interval mapping, such as likelihood analysis, model selection, stopping rules and parameter estimation. The potential power and advantages of the method for mapping multiple QTL and estimating the genetic architecture are discussed. We also point out potential problems and difficulties in resolving the details of the genetic architecture as well as other areas that require further investigation. One application of the analysis is to improve genome-wide marker-assisted selection, particularly when the information about epistasis is used for selection with mating.

1. Introduction

The essence of quantitative genetics is to understand the genetic basis and architecture of quantitative trait variation within and between populations. The genetic architecture of quantitative traits links genotypes to phenotypes: it consists of the number, genomic locations, frequencies and effects of quantitative trait loci (QTL), as well as the interactions of QTL alleles within (dominance) and between (epistasis) loci, pleiotropic effects of QTL, QTL by environment interactions, and so forth. The accurate mapping of a few significant QTL depends critically on appropriately identifying and estimating the architectural parameters, and similarly the study of the genetic architecture depends critically on identifying the genomic positions of individual QTL.

Because of the potential complexity of the genetic architecture of traits in natural populations, most QTL mapping studies are performed via designed experiments. One popular experimental design is to cross two widely separated inbred lines, populations or species, to create a heterozygous F1 population, and then backcross the F1 to parental lines to create backcross populations, or alternatively to intercross F1 to create an F2 population. Recombinant inbred lines are also popular for QTL mapping. For these standard experimental designs, the number of segregating QTL alleles is restricted to two, and the allelic frequencies of the QTL (as well as markers) and their linkage phases are known, thus greatly simplifying the genetic architecture of the traits. The inference of the genetic architecture of a quantitative trait is then restricted to the number, genomic locations, main and interaction effects of the QTL. If an experiment contains multiple traits and/or multiple environments, the issues of pleiotropy and QTL by environment interaction need be addressed in the analysis.

* Corresponding author.

Currently, many statistical methods for mapping QTL focus predominantly on detecting individual QTL effects and assigning these effects to genomic regions. Both interval mapping (IM) (Lander & Botstein, 1989) and composite interval mapping (CIM) (Zeng, 1994; Jansen & Stan, 1994) detect QTL effects at different genomic regions separately. Jiang & Zeng (1995) have extended CIM to multiple traits and multiple environments in order to study pleiotropy and QTL by environment interactions. These methods have been shown to be successful in detecting a few significant QTL for a number of traits in a number of organisms (Tanksley, 1993; Lynch & Walsh, 1998). However, the methods are still not designed to study the whole genetic architecture of quantitative traits; indeed, many architectural parameters, such as those describing epistasis of QTL, have not been taken into account.

Epistasis is a very important component of the genetic architecture of a trait. The pattern of epistasis for a trait can be very complex. It is necessary to take the whole genome into account for mapping multiple QTL and in inferring the epistatic pattern of QTL. This dictates that the search for, and mapping of, multiple epistatic QTL need be performed at multiple intervals simultaneously: it motivated the development of the multiple interval mapping (MIM) method introduced by Kao & Zeng (1997) and Kao *et al.* (1999). The MIM approach can improve the fit of the model and aid the discovery of more QTL. More importantly, MIM permits the study of the genetic architecture by fitting multiple QTL parameters (including epistasis) in a comprehensive framework for model identification and parameter estimation.

Previously, Lander & Botstein (1989, appendix A6) and Knott & Haley (1992) discussed briefly similar approaches for mapping a few linked QTL. Satagopan *et al.* (1996) and Sillanpaa & Arjas (1998) used a Bayesian approach relying on a Markov chain Monte Carlo simulation to map multiple QTL, although they did not discuss epistasis. Broman (1997) treated the issue of model selection in the context of fitting multiple marker effects on a quantitative trait. In this paper, we review the theory and methodology of MIM, with particular emphasis on its potential power and advantages in studying the genetic architecture of quantitative traits. We also point out areas that need more detailed study.

2. Composite interval mapping and its limitations

As motivation for MIM, we begin with an explanation of the idea behind CIM and its practical limitations. Zeng (1993, 1994) introduced CIM to disassociate linkage effects of multiple linked QTL during the identification of individual QTL. This is accomplished by testing for a QTL at a particular genomic region conditioned on other selected markers, known as co-factors. The purpose of using these co-factors is to minimize the effects of QTL in the remainder of the genome when attempting to identify a QTL in a particular region.

For a backcross population of n individuals, the composite interval mapping model is

$$y_i = \mu + b^* x_i^* + \sum_k b_k x_{ik} + e_i = b^* x_i^* + \mathbf{X}_i \mathbf{B} + e_i, \quad i = 1, \ldots, n, \quad (1)$$

where y_i is the phenotypic value of individual i and x_i^* is a coded variable for the genotype of a putative QTL at the test site: x_i^* takes on a value of $\frac{1}{2}$ or $-\frac{1}{2}$ corresponding to the two QTL genotypes. Parameters in the model include the mean (μ), effect of a putative QTL (b^*) and the variance (σ^2) of a residual effect (e_i) assumed to have a normal distribution with mean zero. Co-factors are included in the regression model via the sum in (1). The kth co-factor has associated with it an effect b_k that is the partial regression coefficient of the phenotypes at the co-factor. The datum for co-factor k in individual i is encoded by x_{ik}, which also takes on values of $\frac{1}{2}$ and $-\frac{1}{2}$. A test is performed at every possible genomic position for significance of the QTL effect, b^*, conditioned on the co-factor effects (b_k's): If b^* is found to be statistically significant, a QTL is declared to be in the region. If multiple genomic regions show significant QTL effects, multiple QTL are indicated based on independence of the tests in different genomic regions (Zeng, 1994). See Basten *et al.* (1995) for practical issues involved in data analysis with CIM.

This method creates a relatively simple and systematic procedure to map multiple QTL. It detects and estimates each individual QTL by conditioning the test on other selected linked and unlinked markers ($\mathbf{X}_i$). This means that the test statistic is affected primarily by QTL near the test site (Zeng, 1994). However, there are some limitations to CIM. One is that the analysis can be affected by an uneven distribution of markers in the genome, meaning that the test statistic in a marker-rich region may not be comparable to that in a marker-poor region. Another limitation concerns the difficulty of estimating the joint contribution to the genetic variance of multiple linked QTL. Thirdly, CIM is not directly extendable to analysing epistasis. Finally, the use of tightly linked markers as co-factors can reduce the statistical power to detect a QTL.

3. Multiple interval mapping

To address the limitations of CIM, Kao & Zeng (1997) and Kao *et al.* (1999) proposed and imple-

mented a MIM procedure for mapping multiple QTL simultaneously. The idea of MIM is to fit multiple putative QTL effects and associated epistatic effects directly in a model to facilitate the search, test and estimation of positions, effects and interactions of multiple QTL. MIM thus combines QTL mapping analysis with the analysis of the genetic architecture of quantitative traits.

MIM consists of four components:

1. *An evaluation procedure* designed to analyse the likelihood of the data given a genetic model (number, positions and epistatic terms of QTL).
2. *A search strategy* optimized to select the best genetic model (among those sampled) in the parameter space.
3. *An estimation procedure* for all parameters of the genetic architecture of the quantitative traits (number, positions, effects and epistasis of QTL; genetic variances and covariances explained by QTL effects) given the selected genetic model.
4. *A prediction procedure* to estimate or predict the genotypic values of individuals and their offspring based on the selected genetic model and estimated genetic parameter values for marker assisted selection.

For m putative QTL in a backcross population, the MIM model is defined by

$$y_i = \mu + \sum_{r=1}^{m} \alpha_r x_{ir}^* + \sum_{r \neq s \in (1,\ldots,m)}^{t} \beta_{rs}(x_{ir}^* x_{is}^*) + e_i. \quad (2)$$

As before, y_i is the phenotypic value of individual i while x_{ir}^* is a coded variable denoting the genotype of putative QTL r (defined by $\frac{1}{2}$ or $-\frac{1}{2}$ for the two genotypes: see Kao & Zeng (2000) for the reason). The variable x_{ir}^* is unobserved, but its conditional probability given observed marker phenotypes can be analysed (Jiang & Zeng, 1997). Parameters include the mean (μ), the marginal effects of the putative QTL (α_r's), the variance (σ^2) of the residual effect (e_i, assumed to be normally distributed with mean zero) and epistatic effects. The epistatic effect between putative QTL r and s is denoted β_{rs}. We use a subset of all QTL pairs, indicated by $r \neq s \in (1,\ldots,m)$, to avoid the over-parameterization that could result when using all pairs. The m putative QTL are chosen based on either their significant marginal effects or their epistatic effects, while t is the number of significant pairwise epistatic effects. The MIM model can be extended to an F2 population. See Kao & Zeng (2000) for an appropriate genetic model with epistasis in an F2 population.

Since the genotypes of an individual at many genomic locations are not observed (but marker phenotypes are), the model contains missing data and thus the likelihood function of the data given the model is a mixture of normal distributions:

$$L(\mathbf{E}, \mu, \sigma^2 \mid \mathbf{Y}, \mathbf{X}) = \prod_{i=1}^{n} \left[\sum_{j=1}^{2^m} p_{ij}\, \phi(y_i \mid \mu + \mathbf{D}_j \mathbf{E}, \sigma^2) \right]. \quad (3)$$

The term in square brackets is the weighted sum of a series of normal density functions, one for each of the 2^m possible multiple-QTL genotypes. p_{ij} is the probability of each multilocus genotype conditioned on marker data. The QTL parameters (α's and β's) are contained in the column vector $\mathbf{E}$ while the row vector $\mathbf{D}_j$ specifies the configuration of x^*'s associated with each α and β for the jth QTL genotype (see Kao & Zeng, 1997). Finally, $\phi(y \mid \mu, \sigma^2)$ denotes a normal density function for y with mean μ and variance σ^2.

Clearly, the probability density of each individual is a mixture of 2^m possible normal densities with different means, $\mu + \mathbf{D}_j \mathbf{E}$, and mixing proportions, p_{ij}, which are calculated using marker information (Jian & Zeng, 1997).

Kao & Zeng (1997) described a procedure to obtain maximum likelihood parameter estimates using an expectation/maximization (EM) algorithm. The EM algorithm is an iterative procedure involving an E-step (Expectation) and an M-step (Maximization) in each iteration. In the $[t+1]$th iteration, the E-step is

$$\pi_{ij}^{[t+1]} = \frac{p_{ij}\, \phi(y_i \mid \mu^{[t]} + \mathbf{D}_j \mathbf{E}^{[t]}, \sigma^{2[t]})}{\sum_{j=1}^{2^m} p_{ij}\, \phi(y_i \mid \mu^{[t]} + \mathbf{D}_j \mathbf{E}^{[t]}, \sigma^{2[t]})} \quad (4)$$

and the M-step is

$$E_r^{[t+1]} = \frac{\sum_i \sum_j \pi_{ij}^{[t+1]} D_{jr} \left[(y_i - \mu^{[t]}) - \sum_{s=1}^{r-1} D_{js} E_s^{[t+1]} - \sum_{s=r+1}^{m+t} D_{js} E_s^{[t]} \right]}{\sum_i \sum_j \pi_{ij}^{[t+1]} D_{jr}^2}$$

$$\text{for } r = 1, \ldots, m+t, \quad (5)$$

$$\mu^{[t+1]} = \frac{1}{n} \sum_i \left(y_i - \sum_j \sum_r \pi_{ij}^{[t+1]} D_{jr} E_r^{[t+1]} \right), \quad (6)$$

$$\sigma^{2[t+1]} = \frac{1}{n} \left[\sum_i (y_i - \mu^{[t+1]})^2 - 2 \sum_i (y_i - \mu^{[t+1]}) \sum_j \sum_r \pi_{ij}^{[t+1]} D_{jr} E_r^{[t+1]} + \sum_r \sum_s \sum_i \sum_j \pi_{ij}^{[t+1]} D_{jr} D_{js} E_r^{[t+1]} E_s^{[t+1]} \right], \quad (7)$$

where E_r is the rth element of $\mathbf{E}$ and D_{jr} is the rth element of $\mathbf{D}_j$.

Previously, Kao & Zeng (1997) expressed the M-step in matrix notation as

$$\mathbf{E}^{[t+1]} = diag(\mathbf{V}^{[t+1]})^{-1} [\mathbf{D}' \mathbf{\Pi}^{[t+1]\prime} (\mathbf{Y} - \mu^{[t]}) - nondiag(\mathbf{V}^{[t+1]})\, \mathbf{E}^{[t]}], \quad (8)$$

$$\mu^{[t+1]} = \frac{1}{n} \mathbf{1}' [\mathbf{Y} - \mathbf{\Pi}^{[t+1]} \mathbf{D} \mathbf{E}^{[t+1]}], \quad (9)$$

$$\sigma^{2[t+1]} = \frac{1}{n}[(\mathbf{Y}-\mu^{[t+1]})'(\mathbf{Y}-\mu^{[t+1]}) - 2(\mathbf{Y}-\mu^{[t+1]})'\mathbf{\Pi}^{[t+1]}\mathbf{D}\mathbf{E}^{[t+1]} + \mathbf{E}^{[t+1]\prime}\mathbf{V}^{[t+1]}\mathbf{E}^{[t+1]}] \quad (10)$$

with

$$\mathbf{V} = \{\mathbf{1}'\mathbf{\Pi}(D_r \# D_s)\}_{r,s=1,\ldots,m+t} \quad \text{and} \quad \mathbf{\Pi} = \{\pi_{ij}\}, \quad (11)$$

where # denotes the Hadamard product, which is the element-by-element product of corresponding elements of two matrices of the same order, and $'$ denotes transposition of a matrix or vector.

Aside from notation, there is a subtle difference between (5) and (8). In (5), we use $E_s^{[t+1]}$ for $s = 1, \ldots, (r-1)$, not $E_s^{[t]}$ as implied in (8). It turns out that this is very important to ensure the convergence of the algorithm, particularly when $m+t$ is large. Thus for numerical stability and convergence of the algorithm, (4)–(7) should be used for coding the algorithm.

To clarify the meaning of, and contrast the difference between, p_{ij} and π_{ij}, note that p_{ij} is the probability of each multilocus QTL genotype conditioned on marker genotypes while π_{ij} is the probability of each multilocus QTL genotype conditioned on marker genotypes and also phenotypic value of a trait. If we let g denote a QTL genotype, M a marker genotype and y a phenotype, $\pi_{ij} = Prob(g \mid M, y) = Prob(g \mid M)\, Prob(y \mid g)/\sum_g Prob(g \mid M)\, Prob(y \mid M)$ as shown in (4). Compare this with $p_{ij} = Prob(g \mid M)$ and $Prob(y \mid g) = \phi(y_i \mid \mu + \mathbf{D}_j\mathbf{E}, \sigma^2)$, which is the probability of observing the phenotype given a genotype, within the context of the model defined in (2).

We also emphasize that when m is large, the number of possible mixture components (QTL genotypes) can become prohibitively large for efficient numerical analysis. However, since the probabilities of different genotypes for each observation sum to unity, and as the number of genotypes increases, an increasingly large proportion of genotypes have zero or very small probabilities, and many need not be evaluated. In a practical implementation of the algorithm, as an alternative option, we have also adopted a selection procedure to choose a subset of 'significant' mixture components for evaluation. The procedure uses mixture components that have $p_{ij} > \delta$ (defaults $\delta = 0{\cdot}001$) and requires that the sum of these 'significant' p_{ij} be larger than 0·95 (otherwise the criterion (δ) for the p_{ij} will be lower). After selection, the 'significant' p_{ij}'s will be normalized so that they sum to unity. Using this selection procedure, we discovered that the number of 'significant' mixture components is usually on the order of tens, and occasionally hundreds, depending on the marker density, and the number and positions of the putative QTL selected. This selection procedure greatly reduces the burden of numerical analysis with little loss in the accuracy of the likelihood evaluation.

The test for each QTL effect, say E_r, is performed by a likelihood ratio test conditioned on the other selected QTL effects:

$$LOD = \log_{10} \frac{L(E_1 \neq 0, \ldots, E_{m+t} \neq 0)}{L(E_1 \neq 0, \ldots, E_{r-1} \neq 0, E_r = 0, E_{r+1} \neq 0, \ldots, E_{m+t} \neq 0)}. \quad (12)$$

For given positions of m putative QTL and $m+t$ QTL effects, the likelihood analysis can proceed as outlined above. The task is then to search for and select the genetic model (number, positions and interaction of QTL) that best fits the data.

4. Model selection

Model selection is the key component of the analysis. It is the basis for genetic parameter estimation and data interpretation. Because our analysis is based on genomic positions, not necessarily on the markers, the search for QTL positions becomes more complicated. Kao *et al.* (1999) outlined a stepwise approach to search for QTL positions. Although not ideal, it is a very practical solution. Here we describe a modified procedure that we are currently using for data analysis.

In order to save computation time, we simplify our methods to select a good initial model for MIM analysis. For example, we can first use a backward stepwise regression or a combined forward/backward stepwise regression (Stuart & Ord, 1991) to select a subset of significant markers. For this purpose, we have found that using a stopping rule based on an F-to-enter statistic with $\alpha = 0{\cdot}01$ is generally satisfactory. Then, we can use the results from marker selection to perform CIM to scan the genome for candidate positions.

To identify candidate epistatic terms for the initial model, we use a procedure that pools markers and marker pairs together in a combined forward stepwise regression analysis. This combined analysis treats marker marginal effects and pairwise interaction effects equally during the selection phase. Thereafter, it is appropriate to compare the results of this analysis with CIM results to reach a consensus initial model that includes m marginal effects in m positions and t epistatic effects.

Each parameter in this initial model is then tested for significance using MIM. Those estimates that turn out to be non-significant are dropped stepwise from the model. After the first evaluation of the initial model, we perform the following stepwise selection analysis to finalize the search for a genetic model under MIM:

1. Begin with a model that contains m QTL and t epistatic effects.
2. Scan the genome to determine the best position of the $(m+1)$st QTL. When found, perform a likelihood ratio test for the marginal effect of this putative QTL. If the test statistic exceeds the critical value, this effect is retained in the model.
3. Search for the $(t+1)$st epistatic effect among those pairwise interaction terms not yet included in the model, and perform the likelihood ratio test on the effect. If the LOD exceeds the critical value, the effect is retained in the model. Repeat the process until no more significant epistatic effects are found.
4. Re-evaluate the significance of each QTL effect currently fitted in the model. If the LOD for a QTL (marginal or epistatic) effect falls below the significance threshold conditioned on the other fitted effects, the effect is removed from the model. However, if the marginal effect of a QTL that has a significant epistatic effect with other QTL falls below the threshold, this marginal effect is still retained. This process is repeated until the test statistic for each effect is above the significance threshold.
5. Optimize the estimates of the QTL positions based on the currently selected model. Instead of performing a multi-dimensional search around the regions of the current estimates of the QTL positions (which is a option), the estimates of the QTL positions are updated in turn for each region. For the ith QTL in the model, the region between its two neighbour QTL is scanned to find the position that maximizes the likelihood (conditioned on the current estimates of positions of other QTL and QTL epistasis). This refinement process is repeated sequentially for each QTL position until there is no change in the estimates of the QTL positions.
6. Return to step 2 and repeat the process until no more significant QTL effects can be added into the model and the estimates of the QTL positions are optimized.

It may be worthwhile to attempt to search for significant epistatic effects between selected and unselected QTL positions. This is accomplished in a stepwise manner by searching for the largest epistatic effect between a current QTL position and an unselected genomic position at 1 or 2 cM intervals, and testing for significance. Of course, numerical calculation is very intensive for this analysis.

Sometimes, the stepwise search may fail to uncover QTL in close repulsion linkage or to identify complex epistatic patterns that involve multiple components. When this happens, we may need to add more than one component at a time in order to improve the fit of the model significantly. Chunkwise selection (Kaol *et al.*, 1999) accomplishes this, although it is difficult to perform automatically, and thus needs to be done on an interactive basis.

The analysis of model selection in a high and unknown dimension is very complicated, particularly when it is performed on the whole genome (not just on the observed markers). The multi-dimensional likelihood landscape could have numerous peaks separated by valleys or connected by ridges. A model selected from this landscape may well be just a local peak, and there is no guarantee that a global peak can be found. There are also issues concerning the appropriate criteria used in model selection for an analysis (see below) and appropriate strategies to search for epistatic QTL and to estimate QTL epistasis. Clearly, more detailed and in-depth analyses of these issues are needed. These analyses need to be globally (i.e. genome wide) and architecturally (i.e. multiple components and multiple levels of genetic effects) oriented to be informative for the study of the genetic architecture of quantitative traits.

5. Stopping rules

Two important issues associated with model selection are the stopping rule for the model search algorithm as well as the criterion for comparing different models. In regression analysis, there exist many variable/model selection procedures including the Akaike information criterion (AIC) (Akaike, 1969), the C_p method (Mallows, 1973), the Bayes information criterion (BIC) (Schwartz, 1978; Hannan & Quinn, 1979), the final prediction error method (Shibata, 1984), the generalized information criterion (Rao & Wu, 1989) and its analogues (Potscherm, 1989), the delete-one cross-validation (Allen, 1974; Stone, 1974), the generalized cross-validation (Craven & Wahba 1979), the delete-d cross-validation (Shao, 1993; Geisser, 1975; Burman, 1989; Zhang, 1993), the bootstrap model selection (Shao, 1996) and the minimizing posterior predictive loss (Gelfand & Ghosh, 1998). As pointed out by Broman (1997), QTL mapping analysis is also a model selection analysis. However, unlike many model selection problems in regression analysis, the independent variables (QTL genotypes) are not observed, but the markers are. Model selection practised on markers only (Broman, 1997) is very informative, but insufficient in achieving the main objective of QTL mapping: locating QTL positions. Currently, many statistical analyses for mapping QTL use likelihood ratio or F statistic to test each genetic effect fitted in the model as a basis of model selection, with adjustment on significance value for each test to account for multiple tests practised in searching for QTL (Lander & Botstein, 1989; Zeng, 1994).

Many of these methods and criteria, though based on different principles and considerations, are in fact intimately related. For example, the final production error (FPE) criterion is based on minimizing $S_k = (n+k)RSS_k/(n-k)$, where $RSS(k)$ is the residual sum of squares and k is the number of parameters fitted in the model. The information criterion of the general form is based on minimizing $\mathrm{IC} = -2(\log L_k - kc(n)/2)$, where L_k is the likelihood of data given a genetic model with k parameters and $c(n)$ is a penalty function. This is approximately equivalent to $\mathrm{IC} = \log[RSS_k/n] + kc(n)/n$ in regression analysis.

Akaike (1969) suggests $c(n) = 2$, whereas Schwarz (1978) recommends $c(n) = \log(n)$ and Hannan & Quinn (1979) considered $c(n) = 2\log(\log n)$. It has been shown that S_k and Akaike's IC produce equivalent results asymptotically (Shibata, 1984). Breiman & Freedman (1983) showed that S_k is asymptotically optimal in the sense of minimizing the prediction error. The Schwarz and Hannan–Quinn criteria produce consistent estimators in the sense that the probability of selecting the true model (components) approaches one as $n \to \infty$: FPE and AIC do not achieve this, and typically include too many terms. These asymptotic results, however, give no indication of the behaviour of statistics for finite sample sizes.

The IC criteria can also be related to *F*-to-enter statistic (for regression analysis) or *LR*-to-enter statistic (for likelihood analysis) in the stepwise selection procedure. (It needs to be pointed out that the IC criteria can be used to compare any models, not necessarily the nested models under stepwise regression analysis.) For example, it can be shown that the likelihood ratio test statistic can be expressed as $LR_k = -2\log(L_k/L_{k+1}) \leqslant n\log(c(n)/n+1) \approx c(n)$ (Miller, 1990, p. 208). The criterion is basically defined by $c(n)$ if interpreted under BIC. Using AIC with $c(n) = 2$ would yield a final LOD threshold of 0·43!

In reference to QTL analysis on markers, Broman (1997) suggested using $c(n) = \delta \log n$ and recommended that δ be between 2 and 3. For $n = 100 \sim 500$, the LOD threshold would be $2 \sim 2{\cdot}7$ for $\delta = 2$ and $3 \sim 4$ for $\delta = 3$. This, on the surface, appears to be in line with some current practice in interval mapping (Lander & Botstein, 1989; Zeng, 1994) based on different arguments.

However, this argument is still rather arbitrary and is not related to the experimental design parameters, such as the genetic length of linkage map and the number and distribution of the markers. In fact, the threshold $c(n) = \delta \log n$ is somewhat strange for finite sample size, being higher for larger n. This points out the insufficiency of using asymptotic results in finite samples. Also, the problem of a stopping rule inherently depends on the underlying genetic model in question, the information obviously not available for the analysis. From simulation studies (S. Wang & Z.-B. Zeng, unpublished), we found that some genetic parameters, such as the heritability, play a very important role in the stopping rule and model selection. Clearly more studies on stopping rules are needed in the context of estimating the genetic architecture of quantitative traits.

Also it needs to be emphasized that model selection and stopping rules depend on the purpose of the analysis. If the main purpose of an experiment is to infer genetic parameters, such as number of QTL, model selection based on BIC-type criteria tend to be more appropriate. However, if the main purpose of an experiment is for marker-assisted selection (see below), model selection based on FPE or cross-validation is more appropriate.

6. Permutation test

For mapping QTL, Churchill & Doerge (1994) proposed using a permutation test to estimate empirically the threshold for a test statistic for detection of a QTL. First, a permuted sample is generated from the data by randomly pairing phenotypes and genotypes in the sample, stimulating the null hypothesis of no intrinsic association between genotypes and phenotypes (no QTL). The statistical test is then performed over the whole genome on the permuted sample for QTL, and the maximum test statistic is recorded. This permutation analysis is repeated for a number of replicates to obtain a distribution of the maximum test statistic, and from the distribution to obtain the threshold value. One then compares this threshold with the test statistic from the original sample, and declares the existence of a QTL if the peak test statistic in a region exceeds the threshold.

Subsequently, Doerge & Churchill (1996) extended the permutation method for detecting multiple QTL in two ways. One, called CET (conditional empirical threshold), is based on stratification. After a QTL is identified, a marker close to the QTL is selected and the data are permuted within each marker class. The analysis is then performed on other chromosomes in permuted samples to estimate the threshold for testing for more QTL. The other method is called RET (residual empirical threshold). This method subtracts estimated effects of a QTL from the phenotypic value once a QTL is identified, and then regards the estimated residuals as new trait data for a subsequent permutation test. Both procedures can be performed sequentially in multiple steps for detecting multiple QTL. Compared with the standard permutation test, these methods tend to have greater statistical power to find more (unlinked) QTL, because in later steps the test statistic (for minor QTL) tends to be relatively higher (as the residual variance becomes smaller) and

the threshold tends to be lower (as the search is performed over a smaller genome size). However, the methods are not designed for detecting multiple linked QTL, as linkage effects of possible multiple QTL are not properly safeguarded in the analysis, which may result in under-counting linked QTL and biasing the estimation of QTL positions.

7. Bootstrap test

Alternatively, we could use a bootstrap resampling method for hypothesis testing (Efron, 1979; Mammen, 1993). (Shao (1996) also described a bootstrap model selection procedure based on minimizing prediction error.) Hypothesis testing can be performed sequentially either forwards or backwards. However, in general, the backward search and test is preferred. In our situation, we could first perform a forward search using a lower threshold to overselect candidate QTL positions, and then perform backward elimination to eliminate some less significant candidate QTL. At some point, we may want to compare a model of k QTL with that of $(k+1)$ QTL. We assume that the k QTL positions are contained in the model of $(k+1)$ QTL, and we want to test whether the smallest QTL, say the $(k+1)$th, is significant. Thus, our hypotheses are:

$$\left.\begin{aligned} H_0\colon\ \alpha_i \neq 0\,(i=1,\ldots,k) \quad &\text{and} \quad \alpha_{k+1}=0 \\ &(k \text{ QTL model}), \\ H_1\colon\ \alpha_i \neq 0\,(i=1,\ldots,k) \quad &\text{and} \quad \alpha_{k+1}\neq 0 \\ &(k+1 \text{ QTL model}). \end{aligned}\right\} \qquad (13)$$

Let $\hat{y}_{i|H_0}$ and $\hat{y}_{i|H_1}$ be the estimated phenotypic value of individual i by (14) under H_0 and H_1, respectively. There are generally two ways to generate bootstrap samples under H_0 (Efron, 1979; Mammen, 1993).

1. *Paired bootstrap*: Sample pairs of genotypes and phenotypes $\{(X_i^*, Y_i^*)\}$ with replacement from $\{(X_j, y_j-\hat{y}_{j|H_1}+\hat{y}_{j|H_0}),\ \ j=1,\ldots,n\}$ with equal probability.
2. *Residual bootstrap*: Let $\hat{e}_i = y_i-\hat{y}_{i|H_0}$ and $\hat{\epsilon}_i = \hat{e}_i-\bar{e}$ with $\bar{e}=\sum_i \hat{e}_i/n$. To generate a bootstrap sample $\{(X_i^*, Y_i^*)\}$, we first generate a random sample of residuals $\{\epsilon_i^*\}$ from $\{\hat{\epsilon}_j, j=1,\ldots,n\}$ with replacement, and define $X_i^*=X_i$ and $Y_i^* = \hat{y}_{i|H_0}+\epsilon_i^*$.

The bootstrap test can be performed as follows:

1. Draw a bootstrap sample $\{(X_i^*, Y_i^*)\}$.
2. Search for the best position in the genome (other than the positions of the k QTL) for the hypothetical $k+1$ QTL and perform the likelihood ratio test (12) for the hypotheses in (13).
3. Repeat steps 1 and 2 for a predetermined number of times to obtain an empirical bootstrap distribution of the test statistic, T^*.
4. Reject H_0 if the test statistic for (13) in the original data exceeds $\hat{T}_\alpha$, where $\hat{T}_\alpha$ is the $(1-\alpha)$th quantile of the bootstrap distribution of T^*.

Mammen (1993) proved that the bootstrap test is asymptotically correct. Rayner (1990) showed that this bootstrap test is correct to second order in the sense that its type I error, the probability of rejecting H_0 when H_0 is true, differs from α by a term of order $O(n^{-1})$.

If the sampling for residuals (under residual bootstrap) is without replacement, the sample is a kind of residual permutation sample, and the corresponding test is a kind of residual permutation test. This test is closely related to, but somewhat different from RET of Doerge & Churchill (1996). We (S. Wang and Z.-B. Zeng, unpublished) have performed simulations to evaluate the performance of this residual bootstrap/permutation test in relation to QTL mapping. There are some interesting observations from the simulation study. As expected, there is little difference in the threshold between the residual bootstrap and the residual permutation tests. Obviously, the threshold value depends significantly on the size of the genome, among other things. However, it seems that the threshold value is relatively independent of what and how many (k) QTL are currently fitted in the model (13). Detailed results of the simulation study will be presented elsewhere. However, one difficult problem as yet unresolved is how to account for the tests in multiple steps in the search for QTL.

8. Estimating the genotypic value and marker-assisted selection

Given estimates of the QTL parameters, one can estimate genotypic values of an individual for marker-assisted selection (MAS). This estimation is complicated by the fact that QTL genotypes are not observed directly. Only marker genotypes are observed. Thus, the estimation for an individual is the weighted mean of the genotypic values for all possible genotypes, weighted by the probability ($\hat{\pi}_{ij}$) of each QTL genotype conditioned on the marker and phenotypic data. From (2) and (6), this estimator is

$$\hat{y}_i = \hat{\mu} + \sum_{j=1}^{2^m} \sum_{r=1}^{m+t} \hat{\pi}_{ij} D_{jr} \hat{E}_r, \qquad (14)$$

where the first summation is over all possible 2^m QTL genotypes (in numerical analysis, only those 'significant' QTL genotypes may be analysed and summed here: see above) and the second summation is over all effects of the model (m main effects and t epistatic effects). $\hat{\mu}$ is the maximum likelihood estimate (MLE) of μ obtained from (6) at the equilibrium of the final model, and $\hat{E}_r$ is MLE of QTL effect E_r obtained from (5). $\hat{\pi}_{ij}$ is obtained from (4).

Equation (14) is directly related to MAS. Lande & Thompson (1990) outlined a framework for MAS based on a selection index that combines phenotypic information and molecular genetic marker information. Their selection index uses a weighted mean of phenotypic value and a molecular score to estimate the breeding value of an individual for MAS. The molecular score is constructed based on a multiple regression of phenotypic value on a number of (selected) marker alleles (Gimelfarb & Lande, 1994*a*, *b*), and the weight is constructed based on the heritability and the proportion of the additive genetic variance explained by the molecular score. If an experiment contains molecular genetic marker information that is saturated throughout the whole genome, nearly all the genetic variance can be explained by a selected MIM model. Thus, in this case, the selection index is essentially determined by the molecular score. We feel that model (2) with (14) is a more appropriate way to estimate the molecular score or genotypic value than a multiple regression on markers.

To predict the genotypic value of an individual in a different sample or population based on marker information, such as in the cases of cross-validation and early selection, we need to use

$$\hat{y}_i = \hat{\mu} + \sum_{j=1}^{2^m} \sum_{r=1}^{m+t} \hat{p}_{ij} D_{jr} \hat{E}_r \tag{15}$$

as $\hat{\pi}_{ij}$ is a function of the phenotype y_i, that is unavailable in the early selection.

There is a question on whether to use interaction effects in model (2) for MAS. If mating is random after selection, interaction effects may not be used to estimate the breeding value of an individual for MAS, as the allelic combination of an individual will not be transmitted to the next generation. However, if interaction of QTL is significant and contributes a significant part to the genetic variance, we should take mating into account in MAS by selecting pairs of individuals (a male and a female) based on the predicted genotypic value of offspring of the two individuals. For that purpose, we need to use model (2) with interaction effects for mapping QTL and estimating QTL parameters, and then use equation (15) for predicting the genotypic value of offspring of two individuals. In this case, $\hat{p}_{ij}$ is the estimated probability of multilocus QTL genotype of an offspring, estimated based on its parental QTL genotype distributions, estimated positions of QTL and recombination probabilities between QTL pairs. This selection scheme can maximally utilize the genetic marker and phenotypic information in the current population for predicting selection response in the next generation. The performance of this selection scheme for quantitative traits with epistasis for MAS will be explored further and published elsewhere.

9. Estimating the variance explained by QTL

The genetic variances and covariances explained by each QTL effect can be estimated directly from the likelihood analysis. At the convergence of the EM algorithm, (8) leads to

$$\hat{\mathbf{E}} = \hat{\mathbf{V}}^{-1} \mathbf{D}' \hat{\mathbf{\Pi}}' (\mathbf{Y} - \hat{\mu}). \tag{16}$$

From (10), this means that

$$\hat{\sigma}^2 = \frac{1}{n} [(\mathbf{Y} - \hat{\mu})'(\mathbf{Y} - \hat{\mu}) - \hat{\mathbf{E}}' \hat{\mathbf{V}} \hat{\mathbf{E}}] \tag{17}$$

or

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n} \left[\sum_{i=1}^{n} (y_i - \hat{\mu})^2 - \sum_{r=1}^{m+t} \sum_{s=1}^{m+t} \sum_{i=1}^{n} \sum_{j=1}^{2^m} \hat{\pi}_{ij} D_{jr} D_{js} \hat{E}_r \hat{E}_s \right] \\ &= \frac{1}{n} \left[\sum_{i=1}^{n} (y_i - \bar{y})^2 - \sum_{r=1}^{m+t} \sum_{s=1}^{m+t} \sum_{i=1}^{n} \sum_{j=1}^{2^m} \hat{\pi}_{ij} (D_{jr} - \bar{D}_r) \right. \\ &\qquad \left. \times (D_{js} - \bar{D}_s) \hat{E}_r \hat{E}_s \right], \end{aligned} \tag{18}$$

where $\bar{y} = \sum_{i=1}^{n} y_i / n$ and $\bar{D}_r = \sum_{i=1}^{n} \sum_{j=1}^{2^m} \hat{\pi}_{ij} D_{jr} / n$. In this form $\hat{\sigma}^2$ is expressed as a difference between the MLE of the total phenotypic variance, $\hat{\sigma}_p^2$, in the first term of (18), and that of genetic variance $\hat{\sigma}_g^2$, in the second term of (18).

The estimated genetic variance, $\hat{\sigma}_g^2$, can be further partitioned into

$$\begin{aligned} \hat{\sigma}_g^2 &= \sum_{r=1}^{m+t} \left[\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{2^m} \hat{\pi}_{ij} (D_{jr} - \bar{D}_r)^2 \hat{E}_r^2 \right] \\ &\quad + \sum_{r=2}^{m+t} \sum_{s=1}^{r-1} \left[\frac{2}{n} \sum_{i=1}^{n} \sum_{j=1}^{2^m} \hat{\pi}_{ij} (D_{jr} - \bar{D}_r)(D_{js} - \bar{D}_s) \hat{E}_r \hat{E}_s \right] \\ &= \sum_{r=1}^{m+t} \hat{\sigma}_{E_r}^2 + \sum_{r=2}^{m+t} \sum_{s=1}^{r-1} \hat{\sigma}_{E_r, E_s} \end{aligned} \tag{19}$$

with $\hat{\sigma}_{E_r}^2$ estimating genetic variance due to QTL effect E_r and $\hat{\sigma}_{E_r, E_s}$ estimating genetic covariance between QTL effects E_r and E_s.

There has been some concern about the inflated estimate of the variance explained by QTL conditional on the detection of QTL in a small sample (Beavis, 1994). Estimation conditional on testing or model selection is a typical problem in statistical inference. From simulation studies we have observed, as did Beavis (1994), that when the heritability (the proportion of the variance due to QTL) is small, say 0·2 with $n = 300$, the estimate based on forward-then-backward model selection can be inflated – about 25% in our analysis. However, when the heritability is high, say 0·5 or higher, the estimate is usually close to the parameter value. Also, part of the bias can be corrected by using adjusted R^2 to estimate the total variance explained by QTL (Miller, 1990).

To obtain the sampling variances of architectural parameter estimates, Kao & Zeng (1997) described a

procedure based on the Fisher information matrix allowing for missing data. The confidence intervals of parameter estimates can also be estimated from the bootstrap analysis (Efron & Tibshirani, 1993; Shao & Tu, 1995; Visscher *et al*., 1996). One advantage of using the bootstrap for model testing and selection is that the same bootstrap samples can be used for estimating confidence intervals once a model is selected (Shao, 1996).

10. Application

We have applied MIM to data derived from several QTL mapping experiments. Kao *et al*. (1999) analysed QTL mapping data for three traits in a sample of 134 radiata pine. Based on MIM analysis, seven QTL were mapped for cone number, six for three diameter and five for branch quality. There are significant epistatic effects between four pairs of QTL in two traits. Together, these QTL explains 56%, 52% and 38% of the total phenotypic variances for the three traits, respectively. The MIM analysis indicated strong repulsion linkage effects of closely linked QTL. This was missed by the CIM analysis. Since three traits were analysed, we also estimated the phenotypic, genotypic and environmental (residual) correlations between the three traits in the context of the identified genetic models. It is interesting to note that, although the phenotypic correlation between cone number and branch quality is very small (0·013), the genetic correlation is estimated to be significantly negative ($-0{\cdot}196$) due to linkage effects of QTL, and the environmental correlation is estimated to be significantly positive (0·189).

We also applied the method to a much larger mapping experiment in *Drosophila* (Zeng *et al*., 1999). Two *Drosophila* species, *D. simulans* and *D. mauritiana*, were crossed to make F1 hybrids. Because F1 males are sterile, females of the F1 population were backcrossed to each of the parental lines to produce two backcross populations, each of about 500 individuals. The trait is the morphology of the posterior lobe of the male genital arch analysed as the first principal component in an elliptical Fourier analysis (Liu *et al*., 1996). Both parental difference (35 environmental standard deviations) and the heritability of the trait in backcross populations (93·2% in the *simulans* backcross and 91·6% in the *mauritiana* backcross) are quite large, providing a very favourable situation for QTL mapping. The use of MIM analysis gives evidence of 19 QTL (based on the joint analysis in two backcrosses) distributed on the three *Drosophila* major chromosomes: X, II and III. The additive effect estimates range from 1·0% to 11·4% of the parental difference. The greatest additive effect estimate equals 4·0 environmental standard deviations, but could represent multiple, closely linked QTL. Dominance parameter estimates vary among loci from essentially no dominance to complete dominance, and *mauritiana* alleles tend to be dominant over *simulans* alleles. Epistasis appears to be relatively unimportant as a source of variation. All but one of the additive effect estimates have the same sign, which means that one species has nearly all the plus alleles and the other nearly all the minus alleles. This result is unexpected under most evolutionary scenarios, and suggests a history of strong directional selection acting on the posterior lobe.

An analysis of the data using CIM yielded only 14 of the above 19 QTL. In this case, the use of MIM and the incorporation of the genetic complexity in the mapping analysis (two backcrosses and epistasis) helped to identify minor QTL, and this in turn aided the estimation of the parameters of the trait's genetic architecture.

Our third application is to a set of 519 recombinant isogenic lines in *Drosophila melanogaster* (Weber *et al*., 1999) derived from a cross between two divergently selected lines on wing shape. Only genes on the third chromosome are segregating in the RI lines: genotypes on the first and second chromosomes are identical. The trait is a shape index based on two wing dimensions. Using 65 *in-situ*-labelled transposable elements as markers, 11 QTL were estimated by MIM analysis with additive effect estimates ranging from 2·3% to 18·9% of the parental line difference. Again, all but one of the additive effect estimates have the same sign. Together, the 11 additive effects explain 0·947 of the total phenotypic variance with 0·274 due to the variance of additive effects and 0·673 due to the covariances between additive effects. There are nine QTL pairs that show significant additive by additive interaction effects. However, epistatic effect estimates are about equally positive and negative, and the nine epistatic effects explain only 0·012 of the total variance (0·072 due to the variance of epistatic effects and $-0{\cdot}060$ due to the covariance between epistatic effects). The covariances between additive and epistatic effects, expected to be zero asymptotically (Kao *et al*., 1999), are negative and very small ($-0{\cdot}004$) due to sampling. Thus the model explains 0·955 of the total phenotypic variance.

11. Discussion

As pointed out by Kao *et al*. (1999), there are several advantages to using MIM for QTL mapping studies. First, by directly using multiple QTL components and QTL epistasis in the analysis, MIM can aid the identification of QTL. It can improve the statistical power to identify more minor and complex QTL, and also improve the precision of estimating QTl positions.

A second advantage is that MIM can help to identify patterns and individual elements of QTL

epistasis, and to provide appropriate and integral estimation of individual QTL effects, variances and covariance contributions. These parameter estimates tend to be more stable statistically. This estimation can also help us to assess the relative contribution and importance of different genetic components, and to understand the genetic architecture of quantitative trait values and variation in a population.

Thirdly, with improved identification and estimation of QTL parameters, MIM can provide appropriate and powerful estimates of the genotypic values of individuals that can be used for marker-assisted selection. In particular, with identification of QTL epistasis, marker-assisted selection can be performed on pairs of individuals based on the predicted genotypic values of offspring of two individuals. This selection scheme can effectively utilize the information of QTL epistasis to maximize selection response in the next generation.

MIM helps to bring the three important studies (QTL mapping, genetic architecture and marker-assisted selection) together and provides a unified approach to study the genetic basis of quantitative traits.

The search algorithm and the stopping rule are at the heart of an MIM-based analysis. This is the basis for us to select a genetic model to interpret the genetic architecture of the quantitative traits in the data. The efficiency, reliability and robustness of the search algorithm are the key to the applicability, reliability and utility of this multiple gene based approach. We have studied many issues related to this general question. But many questions still remain to be answered, such as: What is an efficient algorithm to guide the search process to maximization? How robust is the search process? What is the appropriate procedure to determine the stopping rule for the search process? We are aware of many current developments using Markov chain Monte Carlo approaches to guide the search process to sample likelihoods in QTL mapping analysis, and will incorporate many such techniques in our search algorithms. The need to perform large-scale simulation studies to evaluate the reliability and robustness of the method in the genetic model identification is apparent and will be pursued.

This study was partially supported by grants GM 45344 from the National Institutes of Health and no. 9801300 from USDA Plant Genome Program, USA. C. H. K. is supported by grant NSC 88-2313-B-324-001 from National Science Council, Taiwan, ROC.

References

Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics* **21**, 243–247.

Allen, D. M. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometrics* **16**, 125–127.

Basten, C., Weir, B. S. & Zeng, Z.-B. (1995). *QTL Cartographer: A Reference Manual and Tutorial for QTL Mapping*. Department of Statistics, North Carolina State University, Raleigh, NC (http://statgen.ncsu.edu/qtlcart/cartographer.html).

Beavis, W. D. (1994). The power and deceit of QTL experiments: lessons from comparative QTL studies. In *49th Annual Corn and Corghum Industry Research Conference* (ed. ASTA), pp. 250–266. Washington, DC.

Breiman, L. & Freedman, D. (1983). How many variables should be entered in a regression? *Journal of the American Statistical Association* **78**, 131–136.

Broman, K. W. (1997). Identifying quantitative trait loci in experimental crosses. PhD thesis, Department of Statistics, University of California, Berkeley.

Burman, P. (1989). A comparative study of ordinary cross-validation, *v*-hold cross-validation and repeated learning-testing methods. *Biometrika* **76**, 503–514.

Churchill, G. A. & Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–971.

Craven, P. & Wahba, G. (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerical Mathematics* **31**, 377–403.

Doerge, R. W. & Churchill, G. A. (1996). Permutation tests for multiple loci affecting a quantitative character. *Genetics* **142**, 285–294.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7**, 1–26.

Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association* **70**, 320–328.

Gelfand, A. E. & Ghosh, S. K. (1998). Model choice: a minimum posterior predictive loss approach. *Biometrika* **85**, 1–11.

Gimelfarb, A. & Lande, R. (1994*a*). Simulation of marker assisted selection in hybrid populations. *Genetical Research* **63**, 39–47.

Gimelfarb, A. & Lande, R. (1994*b*). Simulation of marker assisted selection for non-additive traits. *Genetical Research* **64**, 127–136.

Haley, C. S. & Knott, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**, 315–324.

Hannan, E. J. & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society*, *Series B* **41**, 190–195.

Jansen, R. C. & Stam, P. (1994). High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136**, 1447–1455.

Jiang, C. & Zeng, Z.-B. (1995). Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* **140**, 1111–1127.

Jiang, C. & Zeng, Z.-B. (1997). Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. *Genetica* **101**, 47–58.

Kao, C.-H. & Zeng, Z.-B. (1997). General formulae for obtaining the MLEs and the asymptotic variance–covariance matrix in mapping quantitative trait loci when using the EM algorithm. *Biometrics* **53**, 653–665.

Kao, C.-H. & Zeng, Z.-B. (2000). Modeling epistasis of

quantitative trait loci using Cockerham's model. *Theoretical Population Biology* (in press).

Kao, C.-H., Zeng, Z.-B. & Teasdale, R. (1999). Multiple interval mapping for quantitative trait loci. *Genetics* **152**, 1203–1216.

Knott, S. A. & Haley, C. S. (1992). Aspects of maximum likelihood methods for the mapping of quantitative trait loci in line crosses. *Genetical Research* **60**, 139–151.

Lande, R. & Thompson, R. (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* **124**, 743–756.

Lander, E. S. & Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199.

Liu, J., Mercer, J. M., Stam, L. F., Gibson, G. C., Zeng, Z.-B. & Laurie, C. C. (1996). Genetic analysis of a morphological shape difference in the male genitalia of *Drosophila simulans* and *D. mauritiana*. *Genetics* **142**, 1129–1145.

Lynch, M. & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates.

Mallows, C. L. (1973). Some comments on C_p. *Technometrics* **15**, 661–675.

Mammen, C. L. (1973). Asymptotics with increasing dimension for robust regression with applications to the bootstrap. *Annals of Statistics* **17**, 382–400.

Miller, A. J. (1990). *Subset Selection in Regression*. London: Chapman and Hall.

Potscherm, B. M. (1989). Model selection under non-stationary autoregressive models and stochastic linear regression models. *Annals of Statistics* **17**, 1257–1274.

Rao, C. R. & Wu, Y. (1989). A strongly consistent procedure for model selection in a regression problem. *Biometrika* **76**, 469–374.

Rayner, R. E. (1990). Bootstrap tests for generalized linear squares regression models. *Economics Letters* **34**, 261–265.

Satagopan, J. M., Yandell, B. S., Newton, M. A. & Osborn, T. C. (1996). A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. *Genetics* **144**, 805–816.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461–464.

Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American Statistical Association* **88**, 486–494.

Shao, J. (1996). Bootstrap model selection. *Journal of the American Statistical Association* **91**, 655–665.

Shao, J. & Tu, D. (1995). *The Jackknife and Bootstrap*. New York: Springer.

Shibata, R. (1984). Approximate efficiency of a selection procedure for the number of regression variables. *Biometrika* **71**, 43–49.

Sillanpaa, M. A. & Arjars, E. (1998). Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. *Genetics* **148**, 1373–1388.

Stone, M. (1974). Cross-validation choices and assessment of statistical predictions. *Journal of the Royal Statistical Society*, *Series B* **36**, 111–147.

Stuart, A. & Ord, J. K. (1991). *Kendall's Advanced Theory of Statistics*, vol. 2, 5th edn. New York: Oxford University Press.

Tanksley, S. D. (1993). Mapping polygenes. *Annual Review of Genetics* **27**, 205–233.

Visscher, P. M., Thompson, R. & Haley, C. S. (1996). Confidence intervals in QTL mapping by bootstrapping. *Genetics* **143**, 1013–1020.

Weber, K., Eisman, R., Morey, L., Patty, A., Sparks, J., Tausek, M. & Zeng, Z.-B. (1999). An analysis of polygenes affecting wing shape on chromosome three in *Drosophila melanogaster*. *Genetics* **153**, 773–786.

Zeng, Z.-B. (1993). Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. *Proceedings of the National Academy of Sciences of the USA* **90**, 10972–10976.

Zeng, Z.-B. (1994). Precision mapping of quantitative trait loci. *Genetics* **136**, 1457–1468.

Zeng, Z.-B., Liu, J., Stam, L. F., Kao, C.-H., Mercer, J. M. & Laurie, C. C. (1999). Genetic architecture of a morphological shape difference between two *Drosophila* species. *Genetics* (in press).

Zhang, P. (1993). Model selection via multifold cross-validation. *Annals of Statistics* **21**, 299–313.