

A Model Selection Approach for the Identification of Quantitative Trait Loci in Experimental Crosses, Allowing Epistasis

Ani Manichaikul,* Jee Young Moon,[†] Saunak Sen,[‡] Brian S. Yandell^{†,§} and Karl W. Broman^{**1}

^{*}Department of Biomedical Engineering, University of Virginia, Charlottesville, Virginia 22908, [†]Department of Statistics, [§]Department of Horticulture and ^{**}Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, Wisconsin 53706 and [‡]Department of Epidemiology and Biostatistics, University of California, San Francisco, California 94143

Manuscript received July 29, 2008

Accepted for publication December 22, 2008

ABSTRACT

The identification of quantitative trait loci (QTL) and their interactions is a crucial step toward the discovery of genes responsible for variation in experimental crosses. The problem is best viewed as one of model selection, and the most important aspect of the problem is the comparison of models of different sizes. We present a penalized likelihood approach, with penalties on QTL and pairwise interactions chosen to control false positive rates. This extends the work of Broman and Speed to allow for pairwise interactions among QTL. A conservative version of our penalized LOD score provides strict control over the rate of extraneous QTL and interactions; a more liberal criterion is more lenient on interactions but seeks to maintain control over the rate of inclusion of false loci. The key advance is that one needs only to specify a target false positive rate rather than a prior on the number of QTL and interactions. We illustrate the use of our model selection criteria as exploratory tools; simulation studies demonstrate reasonable power to detect QTL. Our liberal criterion is comparable in power to two Bayesian approaches.

QUANTITATIVE traits, such as blood pressure and fasting glucose, are often affected by multiple genetic factors, called quantitative trait loci (QTL). The identification of such QTL can lead to improved understanding of molecular mechanisms behind such traits and is the central goal of many experimental crosses involving inbred lines.

We focus on the case of a backcross or an intercross derived from two inbred parental lines and on a continuously varying quantitative trait with normally distributed residual variation. The goals of a QTL mapping experiment include the identification of QTL and epistatic interactions, the derivation of interval estimates for the locations of the QTL, and the estimation of QTL effects. We focus strictly on the identification of QTL and their interactions.

The simplest and most commonly used approach for QTL mapping is interval mapping (LANDER and BOTSTEIN 1989). One posits the presence of a single QTL and considers each genomic location, one at a time, as the putative location for the QTL. At each location, a LOD score is calculated, comparing the hypothesis of a single QTL at the given position to the null hypothesis of no QTL. Much of the focus has been on statistical significance correcting for the genome scan (that is, for the multiplicity of statistical tests that are performed). A

significance threshold or set of corrected *P*-values is derived on the basis of the distribution of the genome-wide maximum LOD score, under the null hypothesis of no QTL. This null distribution is commonly estimated via a permutation test (CHURCHILL and DOERGE 1994).

Despite being based on a single-QTL model, interval mapping has been remarkably useful. However, the consideration of multiple-QTL models has several advantages: (1) controlling for a QTL with large effect reduces the residual variation and increases the power to detect additional QTL of modest effect; (2) one can better separate linked QTL; and (3) epistatic interactions among QTL can be inferred.

The exploration of multiple-QTL models goes beyond hypothesis testing and is better viewed as *model selection* (also known as variable selection). We seek to identify the set of QTL (and epistatic interactions) that are best supported by the data. Special features of the model selection problem in QTL mapping make it unique. First, our primary goal is identifying a good model, not minimizing prediction error; this involves balancing the omission of important loci (false negatives) against the inclusion of extraneous loci (false positives). Second, we have a continuum of ordinal-valued covariates (the genetic loci). Finally, the collinearity of genetic loci is well understood: genotypes for different chromosomes are independent, and for linked loci the correlation decays rapidly with genetic distance.

We split the model selection problem into four distinct parts: choice of a class of models, model fit, model

¹Corresponding author: Department of Biostatistics and Medical Informatics, University of Wisconsin, 1300 University Ave., Room 6770 MSC, Madison, WI 53706. E-mail: kbroman@biostat.wisc.edu

search, and model comparison. The class of models may contain only additive QTL models or may allow pairwise interactions and possibly higher-order interactions. Model fit with complete genotype data at the putative QTL uses linear regression. With missing data at QTL, several methods are available, including multiple-interval mapping (MIM) (KAO *et al.* 1999), Haley–Knott regression (HALEY and KNOTT 1992), and multiple imputation (SEN and CHURCHILL 2001). Model search with a large class of possible models requires an efficient search procedure. Possible procedures include forward selection, backward elimination, stepwise selection, and randomized algorithms including simulated annealing and Markov chain Monte Carlo (MCMC).

Model comparison is arguably the most important aspect of the model selection problem. Models of the same size can be compared directly by maximum likelihood. However, models of varying size must be compared by seeking a balance between model fit and model complexity. A common approach is to use a penalized likelihood criterion, with a penalty on the number of terms in the model. Many classical criteria have this form, including the Akaike information criterion (AIC) (AKAIKE 1969) and the Bayesian information criterion (BIC) (SCHWARTZ 1978).

In their original forms, neither the AIC nor the BIC is ideal for model selection in QTL mapping, due to the large numbers of potential covariates. However, numerous modifications to the BIC have been proposed. BALL (2001), BOGDAN *et al.* (2004), and BAIERL *et al.* (2006) suggested incorporating a prior on the model size, while SIEGMUND (2004) arrived at a similar criterion by formulating QTL mapping as a change-point problem. BROMAN and SPEED (2002) modified the penalty term to control the false positive rate.

BOER *et al.* (2002), ZHANG and XU (2005), and TANCK *et al.* (2006) described penalized likelihood methods for linkage analysis, with penalties on the sizes of effects, in a form similar to ridge regression, to shrink both the model and the estimated effects. The approach we follow is quite different.

Another important stream of methods for multiple-QTL mapping makes use of a Bayesian analysis via MCMC (*e.g.*, SATAGOPAN *et al.* 1996; YI *et al.* 2003, 2005). An advantage of the Bayesian methods is that an expression of uncertainty in the final inference is integral to the framework. However, the Bayesian framework requires a specification of prior distributions on all aspects of the model (including the number of QTL and the number of interactions), implementing the MCMC algorithm requires great care, and the interpretation of the MCMC results can be difficult.

In this article, we extend the work of BROMAN and SPEED (2002) to allow for pairwise interactions among QTL. BROMAN and SPEED (2002) considered strictly additive QTL models and used the null distribution of the genomewide maximum LOD score to derive a pen-

alty on the number of QTL that provided appropriate control on the rate of inclusion of extraneous (*i.e.*, false positive) loci. We apply the same logic, considering the results of a two-dimensional, two-QTL scan, to derive a penalty for the interaction terms. We focus on a class of models that includes pairwise interactions, but with an imposed hierarchical structure in which the inclusion of an interaction term requires the inclusion of the corresponding main effects.

Our goal is to develop a model selection procedure that identifies as many true QTL as possible, while controlling the rate of inclusion of extraneous loci. We generally view the inclusion of an extraneous interaction (or the failure to identify an interaction) as less severe than the inclusion of an extraneous locus (or the failure to identify a locus). This target is most appropriate for biomedical research and may not be appropriate for agricultural or evolutionary studies.

We illustrate our methods through application to data from a mouse backcross and validate the performance of these methods through computer simulation.

METHODS

We consider the case of a backcross or an intercross derived from two inbred lines and of a continuously varying quantitative trait with normally distributed residual variation. We consider QTL models with possible pairwise interactions among QTL and impose a hierarchy on the models, with the inclusion of a pairwise interaction requiring the inclusion of both corresponding main effects. In the case of an intercross, we always include both the additive and the dominance terms for a QTL, and a pairwise interaction requires the inclusion of all four parameters.

We extend the BIC_8 criterion of BROMAN and SPEED (2002) to allow pairwise interactions among QTL, defining two new criteria, $pLOD_H$ and $pLOD_L$, with heavy and light penalties on the pairwise interactions.

BROMAN and SPEED (2002) considered strictly additive QTL models. Their BIC_8 criterion is equivalent to choosing the model that maximizes the penalized LOD score

$$pLOD_a(\gamma) = LOD(\gamma) - T_m|\gamma|, \quad (1)$$

where γ denotes a model, $LOD(\gamma)$ is the \log_{10} -likelihood ratio for the model γ relative to the null model (of no QTL), $|\gamma|$ is the number of QTL in the model γ , and T_m is a penalty on QTL. (While one would usually count model parameters in such a criterion, note that here we are counting QTL.)

The penalty on QTL, T_m , was chosen as the $1 - \alpha$ quantile of the genomewide maximum LOD score under the null hypothesis of no QTL (derived, for example, from a permutation test). With this choice, the false positive rate is maintained at the rate α in the case

of no QTL and with the search restricted to models with no more than one QTL. BROMAN and SPEED (2002) showed, through computer simulation, that the false positive rate is also maintained at or near α in the presence of QTL and for wider searches.

We rename the BIC_{δ} criterion as $pLOD_a$ for consistency with the two new criteria, defined below, and because the term “BIC” is a misnomer: in the traditional BIC, the penalty increases with the sample size, on the order $\log n$, but in the BIC_{δ} criterion, the penalty is approximately constant with sample size.

We now turn to the case of pairwise interactions among QTL. We consider the penalized LOD score

$$pLOD(\gamma) = LOD(\gamma) - T_m |\gamma|_m - T_i |\gamma|_i, \quad (2)$$

where $|\gamma|_m$ and $|\gamma|_i$ are the number of QTL and the number of pairwise interactions, respectively, in the model γ . We thus allow separate penalties on the main effects and interactions and note again that our penalties are on the number of QTL and interactions, and not on the individual degrees of freedom in such terms.

For consistency, we use the same penalty on the main effects as in Equation 1: the significance threshold from a single-QTL genome scan. Thus, if one restricts the search to additive QTL models, this more general criterion reduces to the $pLOD_a$ criterion.

To derive the penalty for the interaction terms, we follow the same logic as in BROMAN and SPEED (2002), but here we consider the results of a two-dimensional, two-QTL genome scan (HALEY and KNOTT 1992; SEN and CHURCHILL 2001). This is the obvious extension of the single-QTL genome scan of interval mapping: we consider each pair of genomic positions (λ_1, λ_2) as putative QTL and fit both a full model, with the two QTL allowed to interact, and an additive model. We also consider the results of a single-QTL genome scan. We consider three sets of LOD scores. $LOD_f(\lambda_1, \lambda_2)$ is the \log_{10} likelihood ratio comparing the hypothesis of two interacting QTL, with one at position λ_1 and the other at position λ_2 , to the null hypothesis (no QTL). $LOD_a(\lambda_1, \lambda_2)$ is the analogous LOD score comparing the hypothesis of two additive QTL to the null hypothesis. $LOD_1(\lambda)$ is the LOD score from the single-QTL scan, at position λ .

In considering the appropriate penalty for interaction terms, imagine that there are two additive QTL and that one performs the two-dimensional, two-QTL scan outlined above. We might then choose the penalty as follows:

$$T_i^H = (1 - \alpha) \text{quantile of} \left\{ \max_{\lambda_1, \lambda_2} LOD_f(\lambda_1, \lambda_2) - \max_{\lambda_1, \lambda_2} LOD_a(\lambda_1, \lambda_2) \right\}. \quad (3)$$

With this choice, the penalized LOD score in Equation 2 has the property that, if the truth is a pair of additive QTL, and one restricts the search to models with no more than two QTL, we will falsely choose an interacting

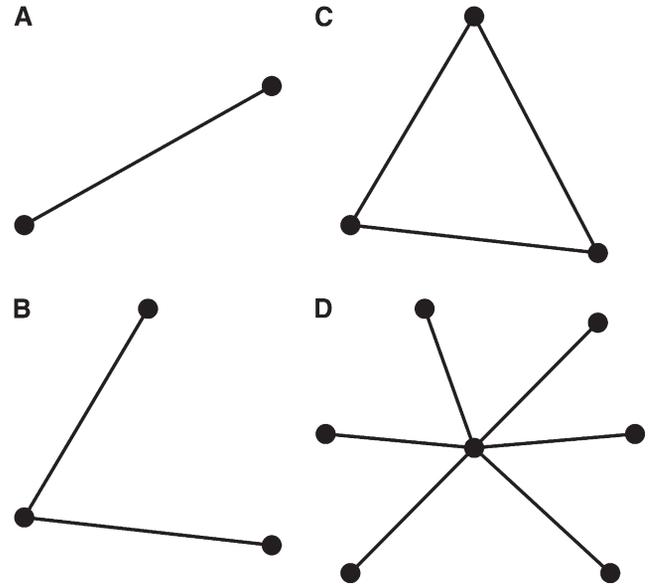


FIGURE 1.—Example graphical representations of QTL models. Nodes represent QTL and edges represent pairwise interactions. (A) A model containing two QTL and their interaction. (B) A model containing three QTL and two pairwise interactions. (C) A model containing three QTL and all possible pairwise interactions. (D) A model containing seven QTL, with one central locus interacting with the other six.

pair over the additive model at the target rate, α . We call this the heavy penalty (to distinguish it from a lighter penalty described below).

The quantile in Equation 3 is ideally for the distribution in the presence of two additive QTL. Such a quantile would be difficult to derive and would likely depend on the locations and effects of the two QTL. However, it is likely to be well approximated by the quantile under the null hypothesis of no QTL, as the interaction terms are approximately orthogonal to the main-effect terms. Thus, we again recommend the use of a permutation test (CHURCHILL and DOERGE 1994) to derive this penalty.

For the class of models under consideration, allowing pairwise interactions but with our enforced hierarchy, one may represent a QTL model as an undirected graph, with nodes representing QTL and edges representing pairwise interactions. Four example models are displayed in Figure 1. In Figure 1A, there is a pair of interacting QTL. In Figure 1B, there are three QTL, with two of the three possible pairwise interactions. In Figure 1C, there are three QTL with all possible pairwise interactions. In Figure 1D, there are seven QTL, one of which shows an interaction with each of the other QTL.

The heavy penalty in Equation 3 seeks to control the rate of inclusion of an extraneous interaction (an extraneous edge in the graph). With this approach, we have low power to detect interacting loci with limited marginal effects. Moreover, the inclusion of a false interaction (or the exclusion of a true interaction) is

generally not so bad as the inclusion of a false locus (or the exclusion of a true locus). Our primary goal should be to identify the major players. The correct identification of the detailed interactions can be useful (particularly for guiding subsequent fine-mapping experiments), but it is not so critical as the correct identification of the loci themselves.

With this in mind, we consider a variation on our logic for deriving the penalty on interactions. Consider the case that the truth is a single QTL and that we perform a two-dimensional, two-QTL genome scan and seek to control the rate of inclusion of an additional interacting locus. The idea is that the truth is a single node, and we wish to control the rate of inclusion of an extraneous “pin” (that is, an additional node with a pairwise interaction edge). An additional interacting QTL would give an additional main-effect penalty (T_m) plus an interaction penalty (T_i). Thus, one could choose the interaction penalty according to the following equation:

$$T_m + T_i^L = (1 - \alpha) \text{ quantile of } \left\{ \max_{\lambda_1, \lambda_2} \text{LOD}_f(\lambda_1, \lambda_2) - \max_{\lambda} \text{LOD}_1(\lambda) \right\}. \quad (4)$$

This gives our light interaction penalty: the difference between the quantile specified on the right-hand side of Equation 4 and the main-effect penalty, T_m . The penalized LOD in Equation 2, with the light interaction penalty, T_i^L , has the property that, if the truth is a single QTL and one performs a search over models with no more than two QTL, the rate of inclusion of a second (false) QTL is maintained at a rate no more than α .

While we would hope that the false positive rates would be maintained at their target rates irrespective of the size of the model, computer simulation experiments indicated that, in the presence of multiple QTL, use of the light interaction penalty will often lead to a model with an extraneous QTL exhibiting interactions with multiple true QTL (similar to the model depicted in Figure 1D, with the central node being an extraneous QTL). The light interaction penalty controls the rate of inclusion of an extraneous pin (a QTL with a single interaction) and improperly penalizes the case of a multipronged pin (as in the central locus in Figure 1D).

Therefore, we compromise between the exclusive use of the heavy interaction penalty and the exclusive use of the light interaction penalty. We consider the graphical representation of a model (for example, imagine that the four panels in Figure 1 represent a single model, with 15 QTL and 12 pairwise interactions). For each connected component in the model, we apply one light interaction penalty, T_i^L , and give all other pairwise interactions the heavy interaction penalty, T_i^H . (Thus, for the model consisting of the entirety of Figure 1, the penalty would be $15T_m + 4T_i^L + 8T_i^H$.)

We thus define two penalized LOD scores. The first, more conservative criterion, denoted pLOD_H , is of the

form in Equation 2, with the exclusive use of the heavy interaction penalty, T_i^H , defined to control the rate of inclusion of a false interaction in the presence of two additive QTL. A second, more liberal criterion, denoted pLOD_L , does not fit the form of Equation 2, but requires consideration of the connected components of QTL; each component is assigned a single light interaction penalty, T_i^L (except, of course, if the component consists of a single QTL), with all additional interactions being assigned the heavy penalty, T_i^H .

As mentioned earlier, we prefer to separate the definition of a model comparison criterion from the procedures to search through the space of models. In constructing a model comparison criterion, we imagine the case that one could consider all possible models exhaustively. In forming a model search procedure, the task is then to optimize the chosen model comparison criterion.

For the purpose of identifying interactions, we propose searching the model space using a two-at-a-time version of forward selection. At each step of forward selection, we consider adding up to two QTL plus their pairwise interaction, and we choose the model with the maximum penalized LOD (even if it is lower than the penalized LOD for the current model). After a predetermined number of steps forward, we proceed with backward elimination as usual. The final model chosen is the one with the highest penalized LOD among all models visited in the search space.

After each step of forward selection (and, potentially, after each backward elimination step), we use an additional model refinement step, based closely on the search algorithm described by ZENG *et al.* (1999): we iteratively refine the location of each QTL along a chromosome, keeping positions of all other QTL fixed. In the case of multiple QTL on a chromosome, the position for a given QTL is scanned between the nearest flanking QTL (or ends of the chromosome), so that the order of the QTL is not modified.

APPLICATION

We illustrate our multiple-QTL mapping methods through application to the data of SUGIYAMA *et al.* (2001), on salt-induced hypertension in 250 backcross mice. A selective genotyping strategy was used (for most markers, only 92 individuals with extreme phenotypes were genotyped). To deal with missing genotype data, we used multiple imputation (SEN and CHURCHILL 2001).

We begin by calculating appropriate significance thresholds for this data set. To account for selective genotyping, we perform stratified permutations, preserving the relationship between the missing genotype pattern and the phenotypes (MANICHAIKUL *et al.* 2007). The LOD penalties of $T_m = 2.56$, $T_i^H = 2.39$, and $T_i^L = 1.01$ were obtained as 95% quantiles of distributions from 10,000 permutations with a two-dimensional, two-

QTL scan on a 2.5-cM grid with 64 imputed sets of genotypes.

Previous analysis by SUGIYAMA *et al.* (2001) suggested a six-QTL model with a pair of linked loci on chromosome 1; unlinked QTL on chromosomes 4, 5, 6, and 15; and a pairwise interaction between the QTL on chromosomes 6 and 15. We use this model as a starting point and proceed to explore neighboring models, to see how well the data support models with additional or fewer QTL. In accordance with our two-at-a-time forward selection strategy, we considered neighboring models by adding (1) an interaction between an existing pair of QTL, (2) an additional additive QTL, (3) a new QTL interacting with a QTL already in the model, (4) a pair of additional additive QTL, and (5) a pair of additional QTL with a pairwise interaction.

Data analysis was performed on a 2.5-cM grid with 256 imputed sets of genotypes. The best model of each type is summarized in Figure 2. An interaction between the loci on distal chromosome 1 and chromosome 5 from the original model increased the model LOD score by 0.76, missing the light interaction penalty of 1.01. An additional locus on proximal chromosome 5 improved the LOD by 1.56, missing the main-effect threshold of 2.56 and decreasing the penalized LOD score by 1.0. Fitting an interaction between this new locus on chromosome 5 and the existing locus on proximal chromosome 1 further increased the model LOD score by 1.53, exceeding the corresponding light interaction threshold of 1.01 and increasing the value of $pLOD_L$ by 0.52 compared to the model without this interaction. However, the extended model with a second locus on chromosome 5 interacting with proximal chromosome 1 still had a lower penalized LOD than the original model of SUGIYAMA *et al.* (2001) because the improvement due to interaction did not outweigh the decrease from addition of the main effect on chromosome 5. In considering whether to add a pair of loci to the model, the best additive pair we identified consisted of a new locus on chromosome 2, together with the aforementioned additional locus on chromosome 5. However, adding this pair of loci did not yield the required LOD increase of twice the main-effect penalty, or 2×2.56 . The best pair of interacting loci contained a new QTL on chromosome 4 and the same locus on proximal chromosome 5. The increase in LOD from the addition of the new term on chromosome 4 and its interaction with proximal chromosome 5 was 2.61, missing the corresponding penalty of $2.56 + 1.01 = 3.57$. In summary, our search through neighboring models did not identify any additional model terms that improved the penalized LOD beyond that of SUGIYAMA *et al.* (2001).

We also performed backward elimination to explore the possibility of a smaller model. We found that removing the loci on proximal chromosome 1 and chromosome 5 improved the penalized LOD score from 7.40 to 9.38. So, our $pLOD_L$ criterion would support removing

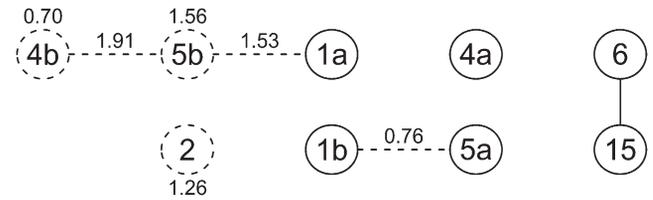


FIGURE 2.—The final model for hypertension data in SUGIYAMA *et al.* (2001) contained a pair of QTL on chromosome 1, unlinked QTL on chromosomes 4 and 5, and a pair of interacting QTL on chromosomes 6 and 15 (solid line). We explored the possibility of additional interactions between QTL in the original model of SUGIYAMA *et al.* (2001) and also considered additional QTL on chromosomes 2, 4, and 5, with possible interactions involving these loci (dashed lines). Each new model term is annotated with its corresponding contribution to the model LOD score. For example, an interaction between model terms 1b and 5a increased the model LOD score by 0.76 compared to the original model.

these two QTL from the model, resulting in a model with QTL on chromosomes 1, 4, 6, and 15 and including the 6×15 interaction. This more parsimonious model was also chosen by application of our fully automated search algorithm. It reflects the strict control over false positives in our approach to QTL model selection. But note that we are not limited to choosing a single model; our proposed penalized LOD scores can be used as an exploratory tool to compare relative support of the data for different models.

SIMULATIONS

While the application of a QTL mapping method to specific data sets can be informative, a more complete exploration of the performance of a method requires computer simulation experiments. To characterize the performance of our proposed penalized LOD criteria and to compare them to other approaches, we simulated intercrosses of 500 individuals and compared the results of our proposed criteria ($pLOD_L$ and $pLOD_H$), the criterion of BROMAN and SPEED (2002) ($pLOD_a$), a Bayesian model selection method using MCMC (YI *et al.* 2005), and the modified BIC criteria of BAIERL *et al.* (2006). To create a realistic generating QTL model, we considered that model estimated via the $pLOD_L$ criterion with data from a mouse intercross on self-selected diet intake (SMITH RICHARDS *et al.* 2002). We considered the total food intake weight, in grams, adjusted for postdiet body weight (TgB) phenotype. The estimated model included QTL on chromosomes 5, 7, 8, 10, 16, 17, 18, and 19, and interactions between QTL on the chromosome pairs 7×10 , 8×18 , and 5×16 . QTL effect parameters (estimated by multiple imputation with 64 sets of imputed genotypes) were reduced by a factor of 0.6 to more clearly differentiate the relative performance of different methods. The resulting heritability values for each of the model terms (the proportions of the phenotypic variance explained by each term)

are displayed in Figure 4B. As seen in the figure, the heritabilities corresponding to the main-effect terms of QTL on chromosomes 5 and 16 are relatively weak, so we would expect our power to detect these QTL without allowing for epistasis to be markedly reduced. We used a marker map modeled after the mouse genome with markers at a 10-cM spacing; we considered the autosomes only.

Model selection strategies: For each simulated data set, we applied a number of different model selection approaches, described below.

For each method and each simulation replicate, QTL main effects were considered to be correctly identified if the final selected model contained a QTL within 50 cM of the true position. The large window for picking up true QTL gave us a generous view of power, but we complemented this measure with a summary of the precision in QTL positions.

pLOD_L: In analyses with *pLOD_L* as the criterion to choose a final model, a search of the model space was performed using two-at-a-time forward selection, followed by backward elimination. The search was performed on a 2.5-cM grid, and models were fitted using Haley–Knott regression (HALEY and KNOTT 1992). The LOD penalties of $T_m = 3.58$, $T_i^H = 2.81$, and $T_i^L = 4.41$ were obtained as 95% thresholds of the appropriate \log_{10} -likelihood-ratio statistics, estimated via 10,000 simulation replicates under the null hypothesis of no QTL. The maximum number of model parameters (excluding the grand mean and error variance) was limited to 50; forward selection was stopped upon reaching the first model with at least 50 parameters. Further, we required at least two markers within the interval between any two linked QTL fitted at non-marker positions and at least one marker between linked QTL where one QTL was at a marker and the other was not. This restriction is justified because essentially all information in a QTL position in between markers can be summarized by genotype data at the flanking markers (WHITTAKER *et al.* 1996), and so we should not fit linked QTL at nonmarker positions with flanking markers in common.

pLOD_H: Analysis using the *pLOD_H* criterion was the same as for *pLOD_L* above, but using heavy interaction penalties only.

pLOD_a (BROMAN and SPEED 2002): Analysis using the *pLOD_a* criterion was the same as described for the *pLOD_L*, except that interactions were excluded from the search space.

Bayesian model selection (YI *et al.* 2005): Bayesian model fitting was performed, with priors specified on number of QTL, QTL locations, genetic effects, overall phenotypic mean, and residual variance as described in YI *et al.* (2005). The prior on the number of QTL in the model was chosen as a Poisson distribution with mean 6, and the upper bound on the number of QTL allowed in the model was 13.

Inclusion of QTL main effects was assumed independent of epistatic effects, and QTL locations were assumed independent and uniformly distributed. Genetic effects of included main and epistatic terms were assumed to be normally distributed. The prior on the overall phenotypic mean was specified as normally distributed with mean and variance taken as their observed sample values. Finally, residual variance was assigned the improper noninformative prior, $p(\sigma^2) \propto 1/\sigma^2$.

Pseudomarkers were spaced approximately every 2.5 cM. The MCMC sampler was run for 404,000 steps, with 4000 burn-in steps removed. The posterior distribution was estimated on the basis of the results from every 20th iteration and so contained 20,000 MCMC samples.

Model selection was performed by choosing the pattern of chromosomes and interactions with the highest posterior probability. Estimated positions of the QTL were obtained as the median positions from among all visited models whose pattern either matched the selected pattern exactly or was a superset of the selected pattern.

Modified BIC with empirical Bayes (BAIERL *et al.* 2006): To control the problem of model overfitting encountered in allowing for epistasis, mBIC (BOGDAN *et al.* 2004) incorporates a prior distribution to account for the large number of model parameters considered. Extending this work to deal with intercrosses, BAIERL *et al.* (2006) proposed QTL model selection by minimizing the following general criterion:

$$\text{mBIC} = n \log \text{RSS} + (p + q) \log n + 2p \log(l - 1) + 2q \log(u - 1). \quad (5)$$

Here p is the number of main-effect parameters and q is the number of epistatic parameters in the model. The mBIC criterion is very similar to the original BIC, with the addition of two separate penalty terms on main effects and interactions to account for the large number of possible predictors of these two types. Specifically, $l := m_v / \text{EN}_v$ and $u := m_e / \text{EN}_e$, where m_v and m_e are the number of main and epistatic parameters in the model, while EN_v and EN_e are the expected number of main and epistatic parameters, according to the chosen prior. Performing model selection on the m marker positions only, we take $m_v = 2m$ and $m_e = 2m(m - 1)$ and $\text{EN}_v = \text{EN}_e = 2.2$, which gives mBIC obtained as a direct extension of that presented in BOGDAN *et al.* (2004).

As an alternative to fixed priors, BAIERL *et al.* (2006) presented an empirical Bayes type strategy for setting the expected number of true QTL model terms. This practical approach begins with an initial scan using mBIC as described above. The observed numbers of main and epistatic terms, \hat{N}_v and \hat{N}_e , are then used to set the expected number of QTL parameters as $\text{EN}_v = \max(2.2, \hat{N}_v)$ and $\text{EN}_e = \max(2.2, \hat{N}_e)$. Plugging these new expected values into the general version of mBIC in Equation 5 yields an empirical Bayes version of the

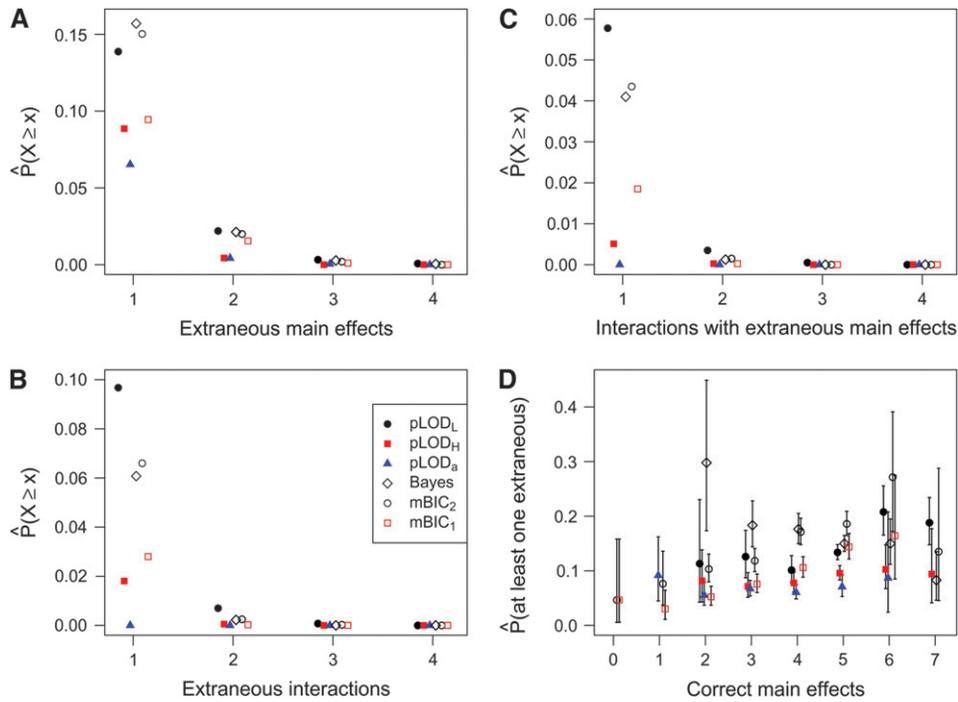


FIGURE 3.—Estimated error rates for intercross simulations under a model based on data from SMITH RICHARDS *et al.* (2002). Plots display observed distribution summaries of (A) the number of extraneous main effects, (B) the number of extraneous interactions, (C) the number of interactions associated with extraneous main effects, and (D) the probability of one or more extraneous unlinked terms as a function of the number of correct main effects, shown only when at least 30 simulation replicates had the specified number of correct main effects using a given method. Ninety-five percent confidence intervals in D are obtained by inverting an exact binomial test. Results are shown for pLOD_L (solid black circle), pLOD_H (solid red square), pLOD_a (solid blue triangle), Bayes (open black diamond), mBIC₂ (open black circle), and mBIC₁ (open red square).

criterion, called mBIC₁. This modified version always sets the expected number of model parameters to be as large as or larger than that specified by mBIC and so should be more powerful in detecting QTL effects.

Modified BIC with additional scan for interactions: A second extension of mBIC was proposed in BAIERL *et al.* (2006) to improve power to detect interactions. This modification, called mBIC₂, is designed to perform a focused scan for interactions involving QTL terms already identified by mBIC₁. The idea is that mBIC₁ could miss real interactions because it allows interactions to be added to a model without corresponding main effects, making the model space more noisy and requiring the search criterion to be more stringent. By performing a scan targeted at a reduced set of interactions, the extended criterion, mBIC₂, is designed to improve power to pick up epistatic terms.

For simulations, mBIC₁ and mBIC₂ are implemented to search models with QTL at marker positions. Also, note that BOGDAN *et al.* (2004) and BAIERL *et al.* (2006) do not impose the same hierarchy as we do on the model space: they allow interactions to be included in a model without corresponding main effects. They also allow additive effects in an intercross model without corresponding dominance terms and vice versa.

Software: Simulations were performed using the statistical software R (IHAKA and GENTLEMAN 1996) and the add-on package R/qtl (BROMAN *et al.* 2003). Some computationally intensive functions were coded in C or Matlab and called from R to improve speed. The Bayesian analysis was performed using R/qtlbim (YANDELL *et al.* 2007). Analysis with the mBIC criteria was performed with Matlab code obtained from Andreas Baierl.

Results: The results are summarized in Figures 3 and 4. Figure 3 displays the error rates for each method, with the frequency of extraneous loci in Figure 3A, the frequency of extraneous interactions in Figure 3B, and the frequency of extraneous interactions that involve an extraneous locus in Figure 3C. Figure 3D shows the frequency of extraneous loci as a function of the number of correctly identified QTL. Figure 4A shows the power to detect each individual QTL and interaction; Figure 4B shows the proportion of the phenotypic variance due to the individual terms, for comparison purposes. Figure 4D shows the distribution of the number of correctly identified QTL. Figure 4C shows the mapping resolution of the individual QTL.

pLOD_L detected more true QTL than pLOD_H and pLOD_a (Figure 4D), with notably higher power to detect the loci on chromosomes 5 and 16, which had strong interaction but weak main effects. At the same time, pLOD_L also had notably higher error rates than pLOD_H and pLOD_a (Figure 3A).

The Bayesian approach detected correct main effects approximately in the same range as pLOD_L (Figure 4D). Similarly, power for individual model terms was comparable for these two methods, although pLOD_L performed noticeably better in picking up the interaction terms (Figure 4A). Extraneous main effects were included in models selected by the Bayesian approach at a rate comparable to that seen for pLOD_L (Figure 3A).

The detection of correct main effects for the mBIC methods was roughly in between that of pLOD_H and pLOD_a (Figure 4D), with mBIC₂ having the higher power of the two mBIC approaches. The power to detect interactions was quite varied across the pLOD and mBIC methods. The pLOD_L and pLOD_H criteria had the

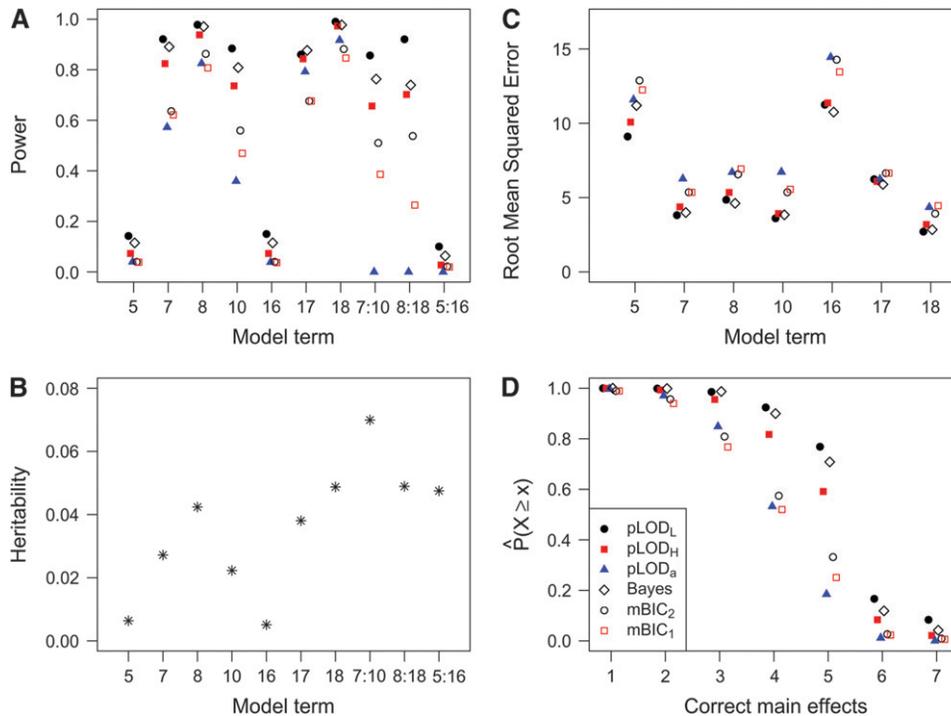


FIGURE 4.—Estimated power to detect true QTL in intercross simulations under a model based on data from SMITH RICHARDS *et al.* (2002). Plots display (A) the power to detect each individual model term, (B) the heritability of the QTL main effects and interactions, (C) the root mean squared error of the estimated positions among correctly identified QTL, and (D) the observed distribution of the number of correct main effects identified in each simulation replicate. A, C, and D show results for pLOD_L (solid black circle), pLOD_H (solid red square), pLOD_a (solid blue triangle), Bayes (open black diamond), mBIC₂ (open black circle), and mBIC₁ (open red square).

highest power to detect interactions, followed by mBIC₂, with mBIC₁ having the lowest power among methods designed to detect epistasis (Figure 4A). In terms of overall rates of including extraneous main effects, mBIC₂ was slightly higher than pLOD_L, while mBIC₁ was a bit higher than pLOD_H (Figure 3A). Since the mBIC approaches do not constrain additive and dominance terms to be at the same marker positions, we adjusted our reported error rates by dropping one of each pair of neighboring markers from our count of extraneous main effects, but the rates displayed for mBIC may still be somewhat inflated.

Looking at the error rates stratified by the number of correctly identified main effects, we saw that pLOD_L tended to add more extraneous terms when more QTL were identified correctly (Figure 3D). The same pattern was also noticeable to a lesser extent for mBIC₂. This phenomenon makes sense for mBIC₂ and pLOD_L because both of these criteria were designed to allow more interaction terms into the model in the presence of already identified main effects, and these interaction terms often carry in extraneous main effects (Figure 3C). The Bayesian approach seemed to show a slight decrease in error rate as the number of correctly identified main effects increased, but this may simply reflect the default prior of six main-effect QTL. Further, the small number of realizations with two or fewer QTL identified correctly by the Bayesian method makes it difficult to characterize this relationship completely.

Regarding computation time, the analysis with pLOD_L took an average of 8.0 min per simulation replicate, the analysis with pLOD_H took 9.2 min, and the analysis with pLOD_a took 5.7 min. (Analysis with pLOD_a was fastest,

since it did not involve a search for interactions; analysis with pLOD_L was faster than pLOD_H, as it generally reached a model with the maximum allowed size in fewer steps.) The two mBIC analyses, combined, took an average of 3.1 min per simulation replicate, but recall that model search included only the marker locations. The average analysis time for the Bayesian MCMC approach was 42.4 min per simulation replicate. These estimated computation times are based on runs on a common machine.

DISCUSSION

QTL mapping methods that consider models with multiple QTL have a number of advantages over methods based on single-QTL models, such as interval mapping. However, it can be difficult to establish the support for QTL in the context of multiple-QTL models.

Bayesian methods provide the most natural expression of model uncertainty: the posterior distribution on QTL models and, as a corollary, the posterior probability that an individual locus or interaction term is involved in the phenotype. Bayesian methods have the further advantages of more completely capturing uncertainty in the inference and of tying together all aspects of the problem. However, the Bayesian approach requires a specification of prior distributions; particularly difficult is the specification of the prior on QTL models, including the number of QTL and the number of interactions. The posterior can be quite sensitive to the choice of prior. Additional difficulties include the construction of efficient MCMC samplers and the diagnosis of deficiencies in a sampler.

We have described a penalized likelihood approach to the problem of comparing multiple QTL models. Our approach extends the work of BROMAN and SPEED (2002) to allow for epistatic interactions. We also allow QTL to reside at nonmarker locations. The key advantage of the new approach is that one needs only to specify the target false positive rate.

The choice of a prior on the number of QTL and interactions in a Bayesian approach and in the modified BIC criterion is essentially equivalent to a choice of penalties on model complexity; modification of the prior will affect the performance characteristics of the methods. In QTL mapping for biomedical research, control of the false positive rate is particularly important. We view our system for selecting penalties, which considers such false positive rates directly, as more natural than the specification of a prior. It may be possible to derive priors for which our penalized LOD criteria correspond approximately to the log posterior for a model. This deserves further investigation, as it would enable a simple system for specifying the prior in Bayesian QTL analysis.

As expected, we observed notably different frequencies of extraneous QTL using $pLOD_L$ vs. $pLOD_H$. In fact, the two extensions to $pLOD_a$ are useful toward different goals. With the interaction penalty in the $pLOD_H$ criterion, we seek to control the rate of any extraneous interaction, while with $pLOD_L$ we seek to control the inclusion of an extraneous interacting locus and are relatively permissive of extraneous interactions, providing that they do not bring in a false locus. The penalized LOD criteria had false positive rates exceeding their targets, and given that there are two separate penalties, it should probably not be surprising that the error rates are on the order of 2α , when the target is α . Moreover, the error rate for the $pLOD_L$ criterion was seen to grow with the number of correctly identified QTL. This again is not unexpected, as with a larger number of true QTL, there are more ways in which an extraneous interacting locus may enter the model. This behavior is likely to be acceptable; one should be less concerned about a single extraneous QTL among six correctly identified QTL than about an extraneous QTL when only one QTL was correctly identified.

Unlike BOGDAN *et al.* (2004), Yi *et al.* (2005), and BAIERL *et al.* (2006), we have required the hierarchy that interactions are considered together with their corresponding main effects, and we have also restricted additive and dominance terms of an intercross to join any model as a pair. This makes a comparison between our methods and these others somewhat difficult. As with any choice, our restriction of model space has both advantages and disadvantages. On the one hand, we have reduced flexibility in terms of the types of models we consider and so should expect lower power in certain situations. On the other hand, the restriction on model space can lead to lower penalties on model terms, and the

results may be less noisy. Furthermore, we find results with additive and dominance terms at the same position to be easier to interpret and more biologically plausible than models with additive and dominance terms at neighboring positions, which often arise in the method of BAIERL *et al.* (2006), and so required special consideration in our summaries of the simulation results.

Possible extensions of our proposed criteria include special consideration of linked QTL, treatment of the X chromosome, and consideration of QTL \times covariate interactions. Proper treatment of the X chromosome might be obtained by considering X chromosome-specific significance thresholds, along the lines of BROMAN *et al.* (2006). Penalties on QTL \times covariate interactions could be derived through thresholds obtained by single-QTL scans with and without the interaction term.

We have employed a relatively simple model search algorithm. Improvements in this algorithm, to more exhaustively search model space and to reduce computation time, deserve additional investigation. Our two-at-a-time forward selection algorithm can be computationally intensive and may have little advantage over one-step forward selection to a large model followed by backward elimination. A more adaptive search algorithm, in place of forward selection to a model of predetermined size, may further reduce computation time. In addition, randomized search methods, such as MCMC and simulated annealing, could allow a more exhaustive consideration of the space of interactions.

While our penalized LOD criteria are based on fixed penalties that are derived from fixed targets for the false positive rates, we generally disapprove of strict significance thresholds. Precisely defined criteria can be useful to the novice QTL mapper, and particularly for the assessment of the performance of a method in computer simulations, but one should keep in mind that there will generally be an array of similarly plausible models. Thus, in practice, it is important to explore the set of models whose penalized LOD is similar to the chosen model. The simplest approach may be to consider a single working model (such as that for which the penalized LOD score is optimized) and focus on the change in likelihood that accompanies the omission of a single term or the addition of one or two QTL or interactions. Such evidence can be usefully represented in a graph, such as that in Figure 2.

In summary, we have presented a straightforward and versatile approach to QTL model selection in the presence of pairwise interactions among QTL. The model selection criteria directly control false positive rates, and so the support for QTL identified through these procedures can be interpreted in a sense similar to traditional hypothesis testing. Since the model selection technique is closely tied to permutation testing, the criteria can be applied quite generally.

The authors thank Gary Churchill for inspiring conversations leading to this work, Andreas Baierl for sharing his Matlab code for mBIC, Bev Paigen for making the hypertension data publicly available, and Brenda Smith Richards for the use of the TgB data. This work was supported in part by National Institutes of Health grants GM074244 (to K.W.B.), GM078338 (to S.S.), and GM069430 (to B.S.Y. and J.Y.M.) and by a National Science Foundation graduate research fellowship (to A.M.).

LITERATURE CITED

- AKAIKE, H., 1969 Fitting autoregressive models for prediction. *Ann. Inst. Stat. Math.* **21**: 243–247.
- BAIERL, A., M. BOGDAN, F. FROMMLET and A. FUTSCHIK, 2006 On locating multiple interacting quantitative trait loci in intercross designs. *Genetics* **173**: 1693–1703.
- BALL, R. D., 2001 Bayesian methods for quantitative trait loci mapping based on model selection: approximate analysis using the Bayesian information criterion. *Genetics* **159**: 1351–1364.
- BOER, M. P., C. J. TER BRAAK and R. C. JANSEN, 2002 A penalized likelihood method for mapping epistatic quantitative trait loci with one-dimensional genome searches. *Genetics* **162**: 951–960.
- BOGDAN, M., J. K. GHOSH and R. W. DOERGE, 2004 Modifying the Schwarz Bayesian information criterion to locate multiple interacting quantitative trait loci. *Genetics* **167**: 989–999.
- BROMAN, K. W., and T. P. SPEED, 2002 A model selection approach for the identification of quantitative trait loci in experimental crosses. *J. R. Stat. Soc. B* **64**: 641–656.
- BROMAN, K. W., H. WU, S. SEN and G. A. CHURCHILL, 2003 R/qt: QTL mapping in experimental crosses. *Bioinformatics* **19**: 889–890.
- BROMAN, K. W., S. SEN, S. E. OWENS, A. MANICHAIKUL, E. M. SOUTHARD-SMITH *et al.*, 2006 The X chromosome in quantitative trait locus mapping. *Genetics* **174**: 2151–2158.
- CHURCHILL, G. A., and R. W. DOERGE, 1994 Empirical threshold values for quantitative trait mapping. *Genetics* **138**: 963–971.
- HALEY, C. S., and S. A. KNOTT, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**: 315–324.
- IHAKA, R., and R. GENTLEMAN, 1996 R: a language for data analysis and graphics. *J. Comput. Graph. Stat.* **5**: 299–314.
- KAO, C. H., Z. B. ZENG and R. D. TEASDALE, 1999 Multiple interval mapping for quantitative trait loci. *Genetics* **152**: 1203–1216.
- LANDER, E. S., and D. BOTSTEIN, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185–199.
- MANICHAIKUL, A., A. A. PALMER, S. SEN and K. W. BROMAN, 2007 Significance thresholds for quantitative trait locus mapping under selective genotyping. *Genetics* **177**: 1963–1966.
- SATAGOPAN, J. M., B. S. YANDELL, M. A. NEWTON and T. C. OSBORN, 1996 A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. *Genetics* **144**: 805–816.
- SCHWARTZ, G., 1978 Estimating the dimension of a model. *Ann. Stat.* **6**: 461–464.
- SEN, S., and G. A. CHURCHILL, 2001 A statistical framework for quantitative trait mapping. *Genetics* **159**: 371–387.
- SIEGMUND, D., 2004 Model selection in irregular problems: applications to mapping quantitative trait loci. *Biometrika* **91**: 785–800.
- SMITH RICHARDS, B. K., B. N. BELTON, A. C. POOLE, J. J. MANCUSO, G. A. CHURCHILL *et al.*, 2002 QTL analysis of self-selected macronutrient diet intake: fat, carbohydrate, and total kilocalories. *Physiol. Genomics* **11**: 205–217.
- SUGIYAMA, F., G. A. CHURCHILL, D. C. HIGGINS, C. JOHNS, K. P. MAKARITSIS *et al.*, 2001 Concordance of murine quantitative trait loci for salt-induced hypertension with rat and human loci. *Genomics* **71**: 70–77.
- TANCK, M. W., J. W. JUKEMA and A. H. ZWINDERMAN, 2006 Simultaneous estimation of gene-gene and gene-environment interactions for numerous loci using double penalized log-likelihood. *Genet. Epidemiol.* **30**: 645–651.
- WHITTAKER, J., R. THOMPSON and P. VISSCHER, 1996 On the mapping of QTL by regression of phenotype on marker type. *Heredity* **77**: 23–32.
- YANDELL, B. S., T. MEHTA, S. BANERJEE, D. SHRINER, R. VENKATARAMAN *et al.*, 2007 R/qtlbim: QTL with Bayesian interval mapping in experimental crosses. *Bioinformatics* **23**: 641–643.
- YI, N., S. XU and D. B. ALLISON, 2003 Bayesian model choice and search strategies for mapping interacting quantitative trait loci. *Genetics* **165**: 867–883.
- YI, N., B. S. YANDELL, G. A. CHURCHILL, D. B. ALLISON, E. J. EISEN *et al.*, 2005 Bayesian model selection for genome-wide epistatic quantitative trait loci analysis. *Genetics* **170**: 1333–1344.
- ZENG, Z.-B., C.-H. KAO and C. J. BASTEN, 1999 Estimating the genetic architecture of quantitative traits. *Genet. Res.* **74**: 279–289.
- ZHANG, Y. M., and S. XU, 2005 A penalized maximum likelihood method for estimating epistatic effects of QTL. *Heredity* **95**: 96–104.

Communicating editor: L. M. McINTYRE