Extending the Modified Bayesian Information Criterion (mBIC) to Dense Markers and Multiple Interval Mapping

Małgorzata Bogdan,[1,2,*] Florian Frommlet,[2,3] Przemysław Biecek,[4] Riyan Cheng,[2] Jayanta K. Ghosh,[2,5] and R.W. Doerge[2]

[1]Institute of Mathematics and Computer Science, Wrocław University of Technology, Wrocław 50-370, Poland

[2]Department of Statistics, Purdue University, West Lafayette, Indiana 47907-2066, U.S.A.

[3]Department of Statistics, University of Vienna, Vienna A-1010, Austria

[4]Institute of Mathematics, Polish Academy of Sciences, Wrocław 51-617, Poland

[5]Indian Statistical Institute, Kolkata 700108, India

[*] *email:* Malgorzata.Bogdan@pwr.wroc.pl

SUMMARY. The modified version of Bayesian Information Criterion (mBIC) is a relatively simple model selection procedure that can be used when locating multiple interacting quantitative trait loci (QTL). Our earlier work demonstrated the statistical properties of mBIC for situations where the average genetic map interval is at least 5 cM. In this work mBIC is adapted to genome searches based on a dense map and, more importantly, to the situation where consecutive QTL and interactions are located by multiple interval mapping. Easy to use formulas for the extended mBIC are given. A simulation study, as well as the analysis of real data, confirm the good properties of the extended mBIC.

KEY WORDS: Model selection criterion; Multiple QTL; QTL interaction.

1. Introduction

Many quantitative traits in plants, animals, and humans are, to a certain extent, determined genetically. Regions of the genome that influence such traits are called quantitative trait loci (QTL), and typically molecular markers are employed to detect and locate QTL using statistical models. These molecular markers are polymorphic (exhibiting variation) at identifiable locations on chromosomes, and their genotypes can be identified experimentally. From a statistical point of view, marker genotypes are qualitative explanatory variables and the task of locating QTL relies on the associations between marker genotypes and the trait values. Most of traditional techniques for QTL mapping (for a review see e.g., Doerge, 2002) identify parts of the genome with additive effects on the trait and are unable to detect epistatic effects (i.e., interactions). Epistasis is however a common phenomenon (see e.g., Doerge, 2002; Carlborg and Haley, 2004, and references given there) and neglecting epistatic effects may lead to oversimplified models for inheritance of complex traits and, as noted by Carlborg and Haley (2004), often results in a relatively low economic gain if such models are used for marker-assisted selection. Although simple, an approach that acknowledges epistasis is multiple regression or ANOVA models with interactions (see e.g., Kao, Zeng, and Teasdale, 1999). The most difficult part in fitting such models lies in the identification of the nonzero coefficients that could, in principle, be addressed by employing the popular Bayesian information criterion (BIC; Schwarz, 1978). In comparison with other model selection criteria BIC has a relatively large penalty for model dimension and is often considered to be conservative. However, in an exploration of BIC as applied to QTL mapping, Broman and Speed (2002) reported that it overestimates the number of QTL. This phenomenon was explained in Bogdan, Ghosh, and Doerge (2004), where a suitable modification of BIC was proposed. The modified version of BIC (mBIC) allows the incorporation of prior knowledge about QTL number. When prior knowledge is lacking Bogdan et al. (2004) proposed a standard version of mBIC which adjusts for multiple testing and controls the type I error. In Baierl et al. (2006), mBIC is further extended by a two-step procedure that adjusts the prior according to the results of the initial step. Simulations reported in Bogdan et al. (2004) and Baierl et al. (2006) demonstrate that both the standard mBIC and the two-step version retain good power when distances between genetic markers are larger than 5 cM. However, when distances between markers are smaller than 5 cM the penalty for mBIC becomes too large, and results in an unnecessary decrease of statistical power. Motivated by the fact that current QTL mapping populations are relatively large, and dense genetic maps are common, the mBIC is extended to situations where map distances are smaller than 5 cM. This extension also sets the stage for applying mBIC to interval mapping and association mapping (discussed elsewhere).

2. Methods

The current work is mainly devoted to detecting QTL in backcross populations where there are only two possible genotypes at any particular locus. Either an individual is homozygous (has both alleles from the same parental line) or heterozygous (has alleles from both parental lines). The QTL genotypes at each location are denoted by Q_{ij}: $Q_{ij} = -\frac{1}{2}$ if the ith individual is homozygous at the jth QTL, and $Q_{ij} = \frac{1}{2}$ if it is heterozygous. The relationship between quantitative trait values and multiple QTL genotypes is assumed to be a normal regression model,

$$Y_i = \mu + \sum_{j=1}^{m} \beta_j Q_{ij} + \sum_{1 \leqslant j < l \leqslant m} \gamma_{jl} Q_{ij} Q_{il} + \varepsilon_i, \qquad (1)$$

where m is the number of QTL and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ is environmental noise. The second summation corresponds to pairwise epistatic interactions. The coefficients β_j and γ_{ij} can both equal zero. The number of QTL with nonzero main effects is denoted as p, and the number of nonzero epistatic terms as q. Because locating so many interacting QTLs by multiple interval mapping poses a complex multidimensional computational problem, as a first step QTL location is restricted to marker positions. Specifically, detection and location of QTL is based on choosing the best model of the form $Y_i = \mu + \sum_{j \in I} \beta_j X_{ij} + \sum_{u,v \in U} \gamma_{uv} X_{iu} X_{iv} + \varepsilon_i$, where X_{ij} is the genotype of the ith individual at the jth marker, and I and U are sets of markers with significant main and epistatic effects, respectively. Note that this model allows the interaction terms to appear even when the corresponding markers do not exhibit additive effects. This choice of modeling strategy is motivated by the well-documented findings of genes that do not have main effects and influence the trait only by interactions with other genes (see e.g., Fijneman et al., 1996).

One of the most popular tools for selecting influential regressor variables is the BIC, which recommends choosing the simplest model for which $n \log RSS + k \log n$ obtains a minimal value. Here RSS is the residual sum of squares from regression, k is the number of regressor variables, and n is the sample size. As mentioned earlier, BIC has a tendency to overestimate the QTL number. To address this problem Bogdan et al. (2004) proposed the mBIC by following the Bayesian ideas of George and McCulloch (1993) and supplementing the BIC with additional terms that are based on a realistic binomial prior distribution for the number of QTL effects. Let N denote the number of available markers and let $N_e = N(N-1)/2$ be the number of possible pairwise interactions. Moreover, use c_1 and c_2 to denote the expected values of the number of main and epistatic effects, respectively. The mBIC is based on minimizing the quantity $mBIC = n \log RSS + (p+q)\log n + 2p \log(l-1) + 2q \log(u-1)$, where p is the number of main effects in the model, q is the number of epistatic terms, $l = N/c_1$ and $u = N_e/c_2$. For situations where the prior knowledge on the number of QTL is not available Bogdan et al. (2004) proposed the use of $c_1 = 2.2$ and $c_2 = 2.2$ which leads to a standard form of mBIC, $mBIC = n \log RSS + (p+q)\log n + 2p \log(N/2.2 - 1) + 2q \log(N_e/2.2 - 1)$. As demonstrated by theoretical calculations presented in Bogdan et al. (2004), the standard version of mBIC controls the overall type I error at the level below 8% if the sample size $n \geqslant 200$ and the number of markers $N > 20$. The criterion retains the consistency of BIC, so both type I and type II errors converge to 0 when the sample size n converges to infinity.

2.1 *Calibrating mBIC for Dense Markers*

Committing a family-wise type I error occurs when at least one QTL is detected when there is none. In this context, mBIC might either incorrectly detect a QTL with one of the simple regression models, including only one marker or one interaction term, or detect false QTL based on a multiple regression model when all simple regression models do not detect QTL. The extensive simulation studies reported in Bogdan et al. (2004) and Baierl et al. (2006) demonstrate that the probability of the second event is extremely small. Thus, the type I error of mBIC is mainly determined by the results of the initial search over single markers and interaction terms.

Consider the simple regression model M_j: $Y_i = \mu + \beta_i X_{ij} + \varepsilon_i$, where X_{ij} is the genotype of ith individual at jth marker. It is straightforward to check (e.g., see the calculations in the Appendix of Bogdan et al., 2004) that mBIC chooses M_j over the null model M_0 with no QTL if the likelihood ratio statistic $LRT_j = 2\log\{L(Y \mid M_j)/L(Y \mid M_0)\}$ is larger than $\log n + 2 \log(N/2.2 - 1)$. Similarly, mBIC selects the model M_{ij} with just one interaction between ith and jth marker if the corresponding likelihood ratio statistic $LRT_{ij} = 2\log\{L(Y \mid M_{ij})/L(Y \mid M_0)\}$ is larger than $\log n + 2\log(N_e/2.2 - 1)$. Thus, the choice of penalty coefficients for mBIC is closely related to the choice of thresholds for the likelihood ratio tests at individual markers. The additional penalty terms $2\log(N/2.2 - 1)$ and $2\log(N_e/2.2 - 1)$, which depend on the number of markers and potential interaction terms, play the role of correction for multiple testing. The increased penalty for interaction terms is related to a larger number of tests that need to be performed for every marker pair. Theoretical calculations presented in Bogdan et al. (2004), and in the Web Appendix C for this work, show that the proposed standard penalty coefficients divide the error about equally between main effects and interactions, and are similar to a Bonferroni correction.

Note that while the Bonferroni correction provides good control of the family-wise error rate (FWER) when individual tests are independent, it can be strongly conservative (i.e., give smaller FWER, which results in decreasing of power) when test statistics are correlated. Therefore, in the situations where markers are very densely spaced (less than 5 cM) the standard mBIC penalty coefficients can be relaxed. To provide a theoretically coherent approach to this problem we propose to adjust the penalty coefficients in such a way that the overall type I error of the resulting procedure is controlled at a level independent of the marker density. For this aim, given a marker map, the threshold values for the maximum of likelihood ratio statistics for individual main effects and interactions are estimated. Then, the effective numbers of independent chi-square tests, N_m^{eff} and N_e^{eff}, are calculated, yielding the same threshold values. As demonstrated by theoretical calculations in Appendix C, and reported via a simulation study, using N_m^{eff} and N_e^{eff} instead of N and N_e in mBIC criterion results in overall type I error that only slightly depends on the marker density, and for $n \geqslant 200$ and $N_m^{eff} \geqslant 20$ does not exceed 8%. Instead the power of the search over very

dense markers is substantially larger than for the standard mBIC.

The issue of computing the genome-wise threshold value for the single marker and interval QTL mapping has been intensively studied. In particular, Lander and Botstein (1989), Dupuis and Siegmund (1999), and Rebaï, Goffinet, and Mangin (1994) addressed the problem by approximating the distribution of the likelihood ratios at neighboring marker locations by the square of the Gaussian process. We use these theoretical results to estimate the chromosome-wise threshold values for main effects and compare them with values resulting from computer simulations. Computer simulations were also used to estimate the thresholds for the search over two-way interactions, for which the theoretical results are not available.

2.2 *Searching Over Markers*

Dupuis and Siegmund (1999) showed when a genome scan based on one chromosome with N markers spaced every δ cM is performed, the overall type I error can be approximated by

$$\alpha = P_{H_0}\left(\max_{i\in\{1,\ldots,N\}} LRT_i > c\right) \tag{2}$$

$$\approx 1-\exp\left[-2\left\{1-\Phi(\sqrt{c})\right\}-0.04L\sqrt{c}\phi(\sqrt{c})\nu\left(\sqrt{0.04c\delta}\right)\right],$$

where $L=(N-1)\delta$ is the length of the chromosome in cM, ϕ is the density and Φ is the distribution function of the standard normal distribution, and $\nu(t)$ is

$$\nu(t)=2t^{-2}\exp\left\{-2\sum_{n=1}^{\infty} n^{-1}\Phi(-|t|n^{1/2}/2)\right\}. \tag{3}$$

Alternatively, the overall type I error resulting from performing N_m^{eff} tests at unlinked markers is

$$\begin{aligned}\alpha &= P_{H_0}\left(\max_{i\in\{1,\ldots,N_m^{eff}\}} LRT_i > c\right)\\ &= 1-\prod_{i=1}^{N_m^{eff}}\{1-\Pr(LRT_i>c)\}\\ &\approx 1-[1-2\{1-\Phi(\sqrt{c})\}]^{N_m^{eff}}.\end{aligned} \tag{4}$$

Comparing equations (2) and (4) the effective number of independent tests corresponding to N markers spaced every δcM is computed as

$$N_m^{eff}=\log(1-\alpha)/\log\{2\Phi(\sqrt{c})-1\}, \tag{5}$$

where c depends on α, δ and N according to equation (2). It turns out that the dependence of N_m^{eff} on α is comparably small. Therefore values of N_m^{eff} for $\alpha=0.05$, reported in the present article, can serve as a very good approximation for all $\alpha\in(0,\,0.1]$.

To model the diminished effects of densely spaced markers we define a weight $w_{SM}^{add}(\delta,\,N)$, which will be assigned to additive effects if the average distance between markers is equal to δcM,

$$w_{SM}^{add}(\delta,N)=N_m^{eff}/N. \tag{6}$$

This weight (equation (6)) clearly depends on δ, but also on the number of markers on the chromosome. It is evident that for any fixed δ, $w_{SM}^{add}(\delta,\,1)=1$. Moreover, in the Web Appendix B it is shown that $\lim_{N\to\infty} w_{SM}^{add}(\delta,\,N)=1$. This result illustrates that the dependence between neighboring markers is of a "short range" and has a negligible influence on the behavior of the maximum of the likelihood ratio statistic when the length of the chromosome converges to infinity. However, Figure 1a demonstrates that the rate of increase of w_{SM}^{add} is very slow, and the weights remain relatively stable over a wide range of N.

Equation (2) from Dupuis and Siegmund (1999) is based on the assumption that the likelihood ratios at neighboring locations can be asymptotically described by the square of a Gaussian process. This fact was utilized when designing a simulation study to estimate the weights for the additive effects. In the first step the correlation matrix S for genotypes of markers uniformly spaced on a chromosome was calculated, and 10,000 instances were generated from a multivariate normal distribution with mean 0 and covariance matrix S. For each of these 10,000 random vectors the maximum over the squares of its coordinates was calculated and 0.95 quantile of the empirical distribution of these maxima was used to estimate the chromosome-wise threshold value c for the likelihood ratio statistic to detect main effects. Finally, weights w_{SM}^{add} were computed by applying equations (5) and (6). As seen in Figure 1b, the simulated values tend to be consistently slightly smaller than the theoretical counterparts, but the difference is so small that the effect on mBIC is negligible.

It is assumed that markers on different chromosomes are inherited independently. Therefore, when considering the whole genome N_m^{eff} is calculated separately for each chromosome. The effective numbers are then added and divided by the overall number of markers to achieve the appropriate weights. For chromosomes of equal length this procedure is the same as choosing the weight for a single chromosome. As a general rule, because the difference in weights is small in the range between 100 and 300 cM, we suggest using the weights computed for 150 cM for a standard adjustment of mBIC for correlated markers. When the distance between markers is larger than 30 cM then w_{SM}^{add} is larger than 0.90 and the modification of the penalty in mBIC is not necessary. Otherwise, w_{SM}^{add} can be very well approximated by a quantity not depending on N, namely,

$$\hat{w}_{SM}^{add}(\delta)=1-0.9\exp\{-10\delta/100+10(\delta/100)^2\}. \tag{7}$$

The accuracy of this approximation is illustrated in Figure 2 in the Web Appendix E.

Because our simulation approach led to reasonably good approximations of the weights for additive effects, we applied the same strategy to estimate the effective number of tests N_e^{eff} corresponding to interactions. Necessary formulas for the covariances between different interaction terms can be found in the Web Appendix A. The dependence of the simulated weights for epistatic effects $w_{SM}^{epi}=N_e^{eff}/N_e$, on both the number of markers and the distance δ is illustrated in Figure 1c. Note that $w_{SM}^{epi}(\delta,\,N)$ appears to have the same qualitative behavior as $w_{SM}^{add}(\delta,\,N)$ in the sense that for $N=2$, $w_{SM}^{epi}(\delta,\,2)=1$, and the simulation results suggest that $\lim_{N\to\infty} w_{SM}^{epi}(\delta,\,N)=1$. Figure 2 in Web Appendix E illustrates that the

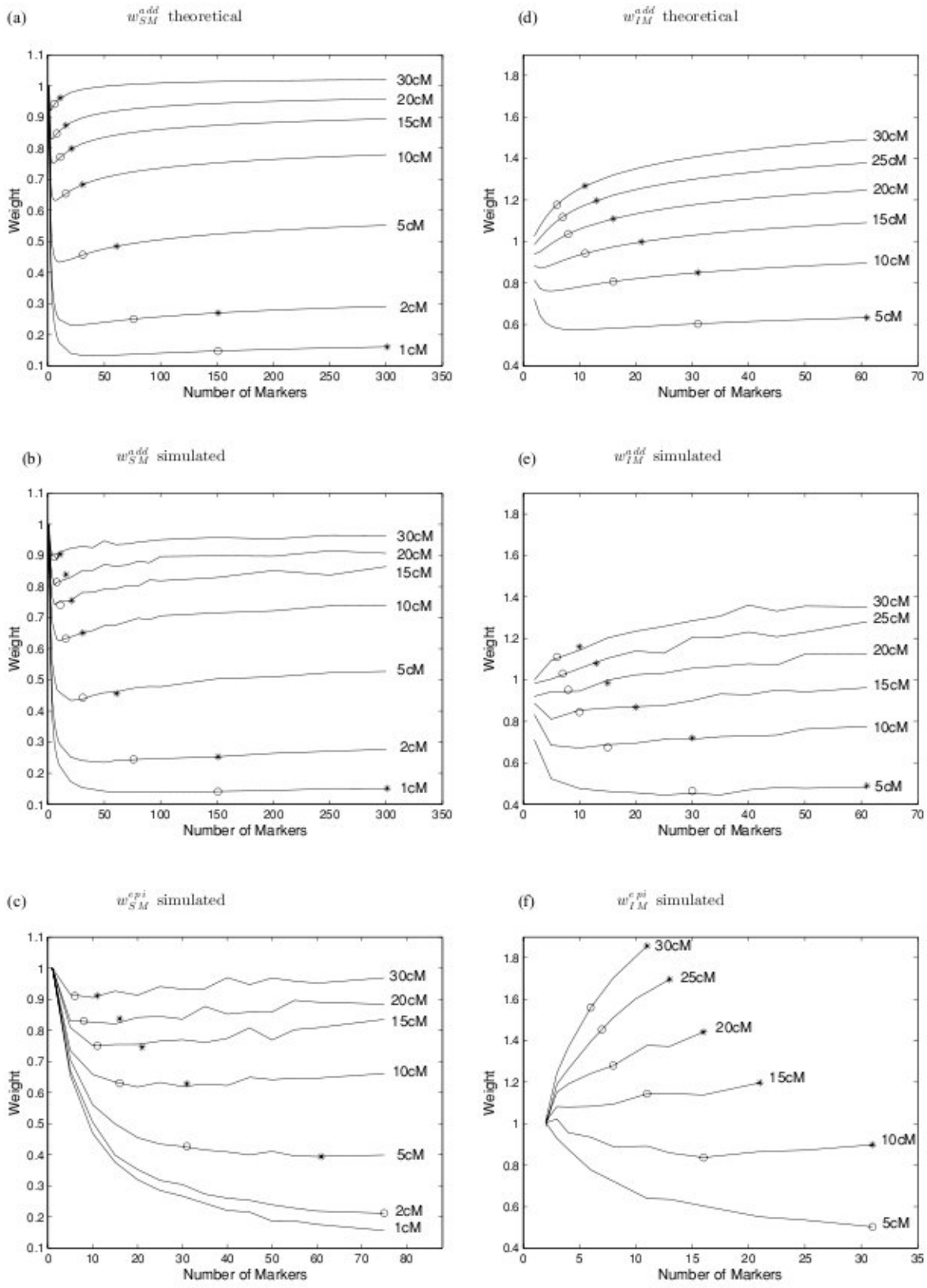

Figure 1. Dependence of the weights corresponding to $\alpha = 0.05$ on the number of markers (N) and the genetic distance between markers (δ). For additive effects both theoretical and simulation results are reported. Circles and stars denote chromosomes of length 150 and 300 cM, respectively.

weights of chromosomes of length 150 cM and 300 cM for different δ can be well approximated by

$$\hat{w}_{SM}^{epi}(\delta) = \exp\{(-10.7\delta/100 + 8.7(\delta/100)^2)\}. \quad (8)$$

For situations where the markers are not equally spaced the average intermarker distance $\bar{\delta}$ can be used with weights $w_{SM}^{add} = \hat{w}_{SM}^{add}(\bar{\delta})$ and $w_{SM}^{epi} = \hat{w}_{SM}^{epi}(\bar{\delta})$ according to equations (7) and (8). Based on these results the standard version of mBIC for dense and unequally spaced markers recommends choosing the model for which $mBIC_{SM} = n \log RSS + (p + q) \log n + 2p \log(w_{SM}^{add} N/2.2 - 1) + 2q \log(w_{SM}^{epi} N_e/2.2 - 1)$, obtains a minimal value.

2.3 *Multiple Regression Interval Mapping, MRIM*

Interval mapping (IM; Lander and Botstein, 1989), while originally based on a single QTL model, was extended to multiple interval mapping (Kao et al., 1999) for the purpose of locating multiple interacting QTL. In the present article, we apply a simplified version of multiple interval mapping based on the approach proposed by Haley and Knott (1992). The method relies on replacing missing genotypes of putative QTL by their expected values conditioned on the genotypes of neighboring markers. The procedure is very simple and quick and is particularly advantageous when there are multiple interacting QTL and many competing models need to be searched to estimate QTL number and their location. A comparison between the EM algorithm, applied in Kao et al. (1999), and Haley and Knott regression did not provide significant differences in the performance of mBIC in an interval mapping setting.

MRIM based on the Haley and Knott method relies on fitting the multiple regression model $Y_i = \mu + \sum_{j \in I} \beta_j G_{ij} + \sum_{u,v \in U} \gamma_{uv} G_{iu} G_{iv} + \varepsilon_i$, on a dense set of possible QTL locations. Here G_{ij} is the expected value of the genotype of ith individual at jth position on the genome (the formulas for G_{ij} are provided in Kao [2000]), and I and U are sets of locations corresponding to QTL with significant main effects and epistatic effects, respectively. To estimate the number of QTL and their locations the appropriate version of mBIC is used.

The main difference between MRIM and searching over individual markers is that a much larger set of possible QTL locations is investigated with MRIM. To accommodate the increased number of investigations the penalty of mBIC needs to be increased. Because the predictor variables corresponding to the given intermarker locations are completely specified by genotypes of neighboring markers the correlations between likelihood ratio statistics at neighboring locations are stronger than when searching over a dense map of markers. To adapt mBIC for MRIM the effective number of tests N_{IM}^{eff} corresponding to the genome search based on one-dimensional interval mapping is calculated. Here one option is to use the theoretical results from Rebaï et al. (1994), which state that the significance level α of the genome search based on interval mapping can be approximated by

$$\alpha = \Pr\left(\sup_{0 \leqslant x \leqslant \sum_{i=1}^{k} \delta_i} LTR_x > c \right) \quad (9)$$

$$\approx 2\Phi(-\sqrt{c}) + \frac{2}{\pi} \exp(-c/2) \sum_{i=1}^{k} \arctan\left(\sqrt{r_i/(1 - r_i)}\right),$$

where c is the threshold, k is the number of intervals, δ_i is the length of the ith interval, and r_i is the probability of recombination for the ith interval. In this case markers can be either equally or unequally spaced.

To find the related number of independent tests N_{IM}^{eff} equation (5) is used, with dependence between α and c provided by equation (9). The corresponding additive effect weight is $w_{IM}^{add} = N_{IM}^{eff}/N$. As might be expected, the behavior of $w_{IM}^{add}(\delta, N)$ differs from that of $w_{SM}^{add}(\delta, N)$. Specifically, in the Web Appendix B we prove that $\lim_{N \to \infty} w_{IM}^{add}(\delta, N) = \infty$, or more precisely for every fixed δ $w_{IM}^{add}(\delta, N) = O(\sqrt{\log(N)})$ when $N \to \infty$. Therefore, w_{IM}^{add} increases slowly with N, and in particular it does not need to be bounded by 1 (Figure 1d). This result demonstrates that while the additional test statistics between marker positions are strongly correlated with statistics at flanking markers they still significantly increase the maximum of the likelihood ratio statistics over the genome. Note, however, that this asymptotic result no longer holds in the situation when the interval mapping is performed by maximizing the likelihood function at a finite grid of locations spaced every 1 or 2 cM. In this case the effective number of tests is always bounded by the number of positions at which tests are performed. To deal with this more realistic situation, simulations are used to estimate the corresponding weights. Imputations that are equally spaced within each interval at distances equal to 2 cM are considered. When calculating covariance matrices for genotypes of imputed positions the Haley and Knott imputations are approximated by linear combinations of the flanking marker genotypes, $\tilde{G}_i = (\delta_1 X_{2i} + \delta_2 X_{1i})/(\delta_1 + \delta_2), i = 1, \ldots, n$, where δ_1 and δ_2 are the genetic distances between the putative QTL and the flanking markers, and X_{1i} and X_{2i} are the corresponding genotypes of flanking markers. This approximation, neglecting the possibility of a double crossover within a given interval, is quite accurate when the distance between neighboring markers does not exceed 30 cM. A subsequent simulation study for MRIM based on exact Haley and Knott regression demonstrates that the weights computed according to the presented simulation strategy perform very well. In Figures 1d and 1e simulation results are compared to theoretical values of Rebaï et al. (1994). As expected, the simulated weights are systematically smaller, which is due to the fact that they are computed assuming the more realistic scenario of a discrete set of imputations.

Simulations were also used to estimate the weights corresponding to interaction terms (Figure 1f). For the purpose of reducing the size of the covariance matrix required to simulate weights for interactions corresponding to a chromosome of the length of 300 cM, a "loose" grid of imputations, separated by approximately 4 cM was used. For a 150 cM chromosome both "medium" (approximately 2 cM) and "loose grid of imputations" were implemented. The difference in weights resulting from the "medium" and "loose" grid turned out to be negligible, while the difference between 150 and 300 cM chromosomes is substantial only when the distance between markers exceeds 20 cM (see the Web Appendix E). It is also worth mentioning that the observed differences between weights for 150 and 300 cM when $\delta > 20$ *cM* have a very small influence on mBIC, which depends on the weight through its logarithm. For the purpose of aiding the reader in applying the mBIC_{IM}

to real data the following empirical approximations for the additive weights can be used

$$\hat{w}_{IM}^{add}(\delta) = -0.15 + 3.1\sqrt{\delta/100} - 1.3\delta/100. \quad (10)$$

The corresponding interval mapping weights for interactions (for chromosome length of approximately 150 cM) can be well approximated by

$$\hat{w}_{IM}^{epi}(\delta) = -0.53 + 5.4\sqrt{\delta/100} - 2.7\delta/100. \quad (11)$$

The accuracy of these approximations is demonstrated in the Web Appendix E. In situations where markers are not equally spaced the average distance between markers $\bar{\delta}$ can be computed to yield weights $w_{IM}^{add} = \hat{w}_{IM}^{add}(\bar{\delta})$ and $w_{IM}^{epi} = \hat{w}_{IM}^{epi}(\bar{\delta})$ (equations (10) and (11), respectively). The adjusted version of mBIC for MRIM recommends choosing the model for which $mBIC_{IM} = n\log RSS + (p+q)\log n + 2p\log(w_{IM}^{add}N/2.2 - 1) + 2q\log(w_{IM}^{epi}N_e/2.2 - 1)$, obtains a minimal value.

3. Simulations

In the Web Appendix D we apply our method to the analysis of *Drosophila* data of Zeng et al. (2000). Results of this analysis, based on the interval version of mBIC, agree well with findings of Zeng et al. (2000) and demonstrate good properties of our method in the situation when the trait is influenced by many QTL. In this section we present the results of the extensive simulation study investigating the properties of the extended version of mBIC under different search scenarios.

We simulate marker and QTL genotypes for a backcross with three unique 100 cM chromosomes, under two sample sizes: $n = 200$ and $n = 500$. All simulations are based on 1000 replicates. Due to the computational complexity of large-scale simulations a simple forward selection with the standard version of mBIC is used. At each step of the forward selection, all the additive and epistatic terms not yet in the model are searched and the one leading to the largest improvement of mBIC is included. When the "best" of remaining terms does not improve mBIC the procedure is terminated.

3.1 *Type I Error*

For the purpose of examining whether the proposed modifications to mBIC allow the control of type I error at the desired level quantitative trait data were simulated under a null hypothesis represented by a normal distribution with $\mu = 0$ and $\sigma = 1$. Both a dense marker and multiple interval mapping setting were investigated. Specifically, densely spaced marker genotypes were simulated every 2, 5, 10, and 20 cM. Furthermore, MRIM based on a 2 cM grid of locations and markers spaced every 5, 10, 20, and 25 cM was considered. To adjust mBIC, weights based on equations (7) and (8) for densely spaced markers and equations (10) and (11) for MRIM were implemented. Finally, 30 unequally spaced markers (Figure 2) were simulated such that the average width of the inter-marker distance was 11.1 cM and the corresponding weights were equal to $w_{SM}^{add} = w_{SM}^{epi} = 0.66$ for single marker analysis and $w_{IM}^{add} = 0.74$, $w_{IM}^{epi} = 0.97$ for MRIM. Table 1 gives the estimated probabilities of incorrectly detecting both the main and the epistatic effects, as well as the total type I error. These results demonstrate that the type I errors for both single marker analysis and MRIM are comparable, and that both only slightly exceed the approximate theoretical levels of 7.3% for $n = 200$ and 4.4% for $n = 500$ (see the Web Appendix C). Because mBIC is a consistent model selection procedure type I error will further diminish as the sample size increases.

Table 1

Probabilities of false detections (in %). e_m and e_e denote the percentage of simulations for which main and epistatic effects were falsely detected. $e_t = e_m + e_e$ is the percentage of simulations for which at least one false signal was detected (i.e., weak sense FWER). SM denotes the search over markers and MRIM denotes the multiple regression interval mapping. The approximate theoretical significance levels are 7.3% for $n = 200$ and 4.4% for $n = 500$.

n	method	δ	e_m	e_e	e_t
200	SM	2 cM	4.4	2.4	6.9
	SM	5 cM	4.8	2.9	7.8
	SM	10 cM	3.7	3.6	7.3
	SM	20 cM	4.7	3.4	8.1
	SM	Figure 3	4.8	3.0	7.8
200	MRIM	5 cM	4.9	2.9	7.8
	MRIM	10 cM	4.1	4.0	8.1
	MRIM	20 cM	4.8	4.2	9.0
	MRIM	25 cM	3.9	4.3	8.2
	MRIM	Figure 3	4.8	3.1	7.9
500	SM	2 cM	1.7	1.5	3.2
	SM	5 cM	2.9	2.4	5.3
	SM	10 cM	2.8	2	4.8
	SM	20 cM	3.4	2.4	5.8
	SM	Figure 3	3.3	2.4	5.7
500	MRIM	5 cM	3	2.6	5.6
	MRIM	10 cM	2.4	2.6	5.0
	MRIM	20 cM	2.6	2.4	5.0
	MRIM	25 cM	3.1	2.2	5.3
	MRIM	Figure 3	3.2	2.1	5.3

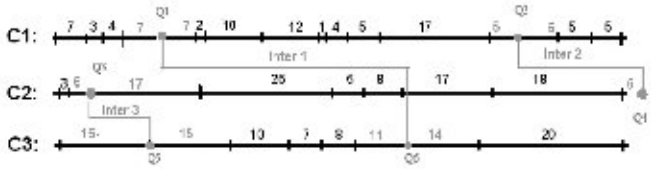

Figure 2. Marker and QTL locations on three chromosomes. Thirty unequally spaced markers are denoted by black vertical lines. Six QTL locations are represented as circles. The distance between markers is specified, and the average interval distance is 11.1 cM.

3.2 *Power and Accuracy of QTL Localization*

The statistical power and accuracy of QTL localization based on the adjusted mBIC are considered under three different search strategies: single marker analysis over a relatively sparsely spaced set of markers (scenario 1), MRIM with a 2 cM spacing (scenario 2), and single marker analysis over markers spaced every 2 cM (scenario 3).

QTL locations, as well as locations of markers used in scenario 1 are presented in Figure 2. Six QTL (Q1–Q6) are simulated, two on each of three chromosomes. Three of these QTL: Q1, Q4, and Q6 have main effects. The corresponding

Table 2

Estimates of power and the precision for QTL location (in cM) for the search over markers in a sparse map (scenario 1), multiple interval mapping (scenario 2), and the search over markers spaced every 2 cM (scenario 3). sc denotes the scenario, pow is the power in %, sl is the standard error of QTL location in cM and afp denotes the average number of false positives.

		Q1	Q4	Q6	Int 1	Int 2	Int 3	
		pow	*pow*	*pow*	*pow*	*pow*	*pow*	
n	sc	*sl*	*sl*	*sl*	*sl Q1*	*sl Q2*	*sl Q3*	*afp*
					sl Q6	*sl Q4*	*sl Q5*	
		67	76	35	24	47	7	0.14
200	1	10.2	6.9	14.8	10.9	9.2	10.1	
					14.5	7.8	16.6	
		69	76	39	28	47	10	0.18
200	2	9.8	8.1	13.2	9.7	8.0	9.0	
					12.1	8.6	11.4	
		71	75	53	40	45	19	0.17
200	3	7.8	9.6	9.8	8.3	7.3	7.1	
					8.2	9.7	8.2	
		96	98	86	79	94	40	0.32
500	1	8.0	5.0	12.9	8.2	5.8	9.2	
					13.0	5.0	15.0	
		97	99.7	83	80	97	48	0.33
500	2	5.6	7.0	6.3	6.2	4.7	5.7	
					6.1	6.9	6.8	
		99	98.5	93	96	97	84	0.22
500	3	3.6	6.3	4.3	3.8	3.2	3.8	
					3.4	6.2	4.4	

effect sizes, according to the model (1), are: $\beta_{Q1} = 0.6$, $\beta_{Q4} = 0.7$, and $\beta_{Q6} = 0.5$. Additionally, three interaction effects are simulated: interaction 1 involving Q1 and Q6, interaction 2 between Q2 and Q4, and interaction 3 between Q3 and Q5. The corresponding effect sizes are $\gamma_{Q1Q6} = 1.2$, $\gamma_{Q2Q4} = 1.4$, and $\gamma_{Q3Q5} = 1$. The standard deviation of the error term ε is 1. Define a single effect heritability as $h^2 = \sigma^2_{eff}/\sigma^2_Y$, where σ^2_{eff} is the variance of the trait due to a particular effect, and σ^2_Y is the total trait variance. The heritabilities corresponding to the simulated effects are equal to $h^2_{Q1} = h^2_{Q1Q6} = 0.058$, $h^2_{Q4} = h^2_{Q2Q4} = 0.08$, $h^2_{Q6} = h^2_{Q3Q5} = 0.04$. The overall broad sense trait heritability $H^2 = 0.355$.

Table 2 illustrates both the power of detection for each of the simulated effects, and the standard error of the estimate of QTL location. When $n = 200$, the standard errors of the estimates of location reach the level of 15 cM; therefore, an effect identified by mBIC is qualified as a true positive if it is within 30 cM of the true QTL. Epistatic effects are classified as true positives if both detected positions are within 30 cM of the true epistatic QTL. For the sample size $n = 500$ this detection window was decreased to within ±15 cM of the true QTL. If more than one effect was found in a detection window only one was classified as a true positive. All other effects were classified as false positives. Table 2 reports the average number of false positives in one simulation run. Table 2 demonstrates that, as expected, the power of mBIC increases with an increase of the sample size. For $n = 500$ the power of detecting the weakest main effect with $h^2 = 0.04$ reaches 93% and the corresponding, weakest epistatic effect is detected with a power of 84%. Note that lower power for epistatic effects results from using a larger penalty than for main effects.

In all simulations the average numbers of false positives significantly exceed 0.08, which is the assumed value for FWER. This effect is related to the lack of precision of estimating QTL location (see the estimated standard deviations in Table 2 and the discussion in Bogdan and Doerge, 2005). In the result of this phenomenon some of accurately detected QTL fall out of the assumed detection window and they are classified as false positives. Note, however, that in all simulation runs the average false discovery rate (the ratio of the number of false positives to the total number of detected effects) is below 7%.

Results demonstrated in Table 2 confirm that dense genetic maps greatly increase the power of detecting QTL. For $n = 200$ the power of detecting the weakest main effect, Q6, which was 11 cM from the closest marker in the sparse map, increased from 35% for the sparse map to 53% for a dense map. The advantage of using a dense map is, however, not so obvious when the QTL effects are large and located close to markers from a sparse map. In fact, in this situation the search over a dense map may even yield slightly lower power due to the necessity of adjusting the detection thresholds to the increased multiple testing problem. However, even in the situation where the dense map does not yield the highest power it usually allows for a more precise QTL localization. The only exception for these simulations is Q4, which is located outside the last marker on chromosome 2 (Figure 2). The dense map was constructed only within the markers on this chromosome and did not allow for a more precise localization of Q4.

The use of MRIM had a small influence on the power of QTL detection when compared to the sparse map case. However, as anticipated it substantially increased the precision of QTL localization. This difference is clearly visible for $n = 500$ where the standard deviation of QTL localization estimates obtained by interval mapping was smaller than the distance from the closest flanking marker. Again, Q4 is an exception to this, where interval mapping actually increases the standard deviation.

Our simulations demonstrate that the precision of QTL location increases significantly with the sample size. However, in case when the search is performed over markers, then the standard error of QTL location is always larger than the distance between the QTL and the closest marker. Therefore, in the sparse map case, the improvement of the precision with increasing sample size is limited.

4. Discussion

The mBIC of Bogdan et al. (2004) is adapted to two unique and applicable situations, namely the search over markers from a dense marker maps and multiple interval mapping. Simulation results demonstrate that the proposed method of relaxing the penalty in the standard version of mBIC allows for the control of the type I error and the proportion of false positives at the assumed level. The need to relax the penalty in the standard version of mBIC is apparent when markers are spaced closer than 5 cM. In fact, using mBIC with such dense marker maps may help to increase the power and precision of QTL location. For multiple interval mapping the implemented penalty keeps the type I error at the assumed level and only slightly depends on the number of tests performed

within intermarker intervals. Based on our results multiple interval mapping only slightly increases the power of QTL detection when compared to the search over relatively distant markers, but can substantially increase the precision of QTL localization.

The particular weights to calibrate the penalty as presented in Figure 1 are specific for the backcross design. A general method for computing the "efficient" number of markers (equation (5)) can be used for other experimental designs. In situations where the structure of correlations between the regressor variables is known, the critical value c can be simulated using the approach presented in this article. When the structure of correlations is not known the critical value c can be approximated using the permutation approach described in Churchill and Doerge (1994).

Following majority of papers on QTL mapping the simulations presented in this article were performed under the ideal model with a normal error distribution. However, an extensive simulation study reported in Żak et al. (2007) demonstrates that due to the central limit theorem mBIC often preserves its properties even when the distribution of the error term is different from normal. In situations where the distribution of the error term substantially differs from normality, we cannot rely on the central limit theorem. Therefore, the version of mBIC based on ranks, instead of original trait values, can be used (see Żak et al., 2007). The results reported in Kruglyak and Lander (1995) and Żak et al. (2007) demonstrate that for reasonable sample sizes $n \geqslant 200$, and independent of the distribution of the error term, the threshold values for the rank version of the interval mapping and the distribution of the rank version of mBIC do not differ substantially from the corresponding properties of classical methods under normality. Thus, the weights proposed in this article can be directly used to relax the penalty coefficients for the rank version of mBIC.

The `R` code with implementation of the methods discussed in this article is available at `http://www.stat.purdue.edu/~doerge/software/mBIC.html`.

5. Supplementary Materials

Web Appendices, referenced in Sections 2 and 3, are available under the Paper Information link at the *Biometrics* website `http://www.biometrics.tibs.org`.

Acknowledgements

We thank two anonymous reviewers and the associate editor for helpful suggestions. This work is partially funded by NSF Plant Genome Grant 0501712-DBI to RWD and by grant 1 P03A 01430 of the Polish Ministry of Science and Higher Education to MB.

References

Baierl, A., Bogdan, M., Frommlet, F., and Futschik, A. (2006). On locating multiple interacting quantitative trait loci in intercross designs. *Genetics* **173,** 1693–1703.

Bogdan, M. and Doerge, R. W. (2005). Biased estimators of quantitative trait locus heritability and location in interval mapping. *Heredity* **95,** 476–484.

Bogdan, M., Ghosh, J., and Doerge, R. W. (2004). Modifying the Schwarz Bayesian information criterion to locate multiple interacting quantitive trait loci. *Genetics* **167,** 989–999.

Broman, K. W. and Speed, T. P. (2002). A model selection approach for the identification of quantitative trait loci in experimental crosses. *Journal of the Royal Statistical Society, Series B* **64,** 641–656.

Carlborg, O. and Haley, C. (2004). Epistasis: Too often neglected in complex trait studies? *Nature Reviews Genetics* **5,** 618–625.

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

Doerge, R. W. (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nature Reviews Genetics* **3,** 43–52.

Dupuis, J. and Siegmund, D. (1999). Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics* **151,** 373–386.

Fijneman, R. J. A., De Vries, S. S., Jansen, R. C., and Demant, P. (1996). Complex interactions of new quantitative trait loci, sluc1, sluc2, sluc3, and sluc4, that influence the susceptibility to lung cancer in the mouse. *Nature Genetics* **14,** 465–467.

George, E. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **88,** 881–889.

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

Kao, C. (2000). On the differences between maximum likelihood and regression interval mapping in the analysis of quantitative trait loci. *Genetics* **156,** 855–865.

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

Kruglyak, L. and Lander, E. S. (1995). A nonparametric approach for mapping quantitative trait loci. *Genetics* **139,** 1421–1428.

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

Rebaï, A., Goffinet, B., and Mangin, B. (1994). Approximate thresholds of interval mapping test for QTL detection. *Genetics* **138,** 235–240.

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

Żak, M., Baierl, A., Bogdan, M., and Futschik, A. (2007). Locating multiple interacting quantitative trait loci using rank-based model selection. *Genetics* **176,** 1845–1854.

Zeng, Z. B., Liu, J., Stam, L. F., Kao, C.-H., Mercer, J. M., and Laurie, C. C. (2000). Genetic architecture of a morphological shape difference between two Drosophila species. *Genetics* **154,** 299–310.