Theor Appl Genet (2009) 119:899–911
DOI 10.1007/s00122-009-1098-y

ORIGINAL PAPER

Look before you leap: a new approach to mapping QTL

B. Emma Huang · Andrew W. George

Received: 21 January 2009 / Accepted: 21 June 2009 / Published online: 8 July 2009

Abstract In this paper, we present an innovative and powerful approach for mapping quantitative trait loci (QTL) in experimental populations. This deviates from the traditional approach of (composite) interval mapping which uses a QTL profile to simultaneously determine the number and location of QTL. Instead, we look before we leap by employing separate detection and localization stages. In the detection stage, we use an iterative variable selection process coupled with permutation to identify the number and synteny of QTL. In the localization stage, we position the detected QTL through a series of one-dimensional interval mapping scans. Results from a detailed simulation study and real analysis of wheat data are presented. We achieve impressive increases in the power of QTL detection compared to composite interval mapping. We also accurately estimate the size and position of QTL. An R library, DLMap, implements the methods described here and is freely available from CRAN (http://cran.r-project.org/).

Introduction

Composite interval mapping (CIM) is a statistical technique for the simultaneous detection and localization of quantitative trait loci (QTL). In a procedure analogous to interval mapping, a test statistic measuring the evidence for the presence of a QTL is calculated at various hypothesized positions of a putative trait locus. A QTL curve or profile is then formed. Three types of profile features serve to identify QTL: the height of the curve is used for detection; distinguishable local maxima correspond to detectable QTL; and the positions of these peaks give estimates of QTL locations. A variety of competing QTL mapping strategies exist for the analysis of data from experimental populations, but for many practitioners CIM (Jansen 1993; Zeng 1993) is the method of choice.

There are two major challenges in using CIM to map QTL. First, CIM involves a preanalysis step where markers are selected to act as cofactors in the CIM model. These marker cofactors increase mapping precision. The difficulty lies in selecting the "best'' set of cofactors, which are those markers closest to the unknown QTL. Numerous variable selection strategies have been proposed (Broman and Speed 2002), but inclusion of too few or too many marker cofactors adversely affects performance (Broman and Speed 2002). Furthermore, QTL mapping programs such as R/qtl (Broman et al. 2003) and QTL Cartographer (Wang et al. 2005) require users to arbitrarily specify a fixed number of cofactors once the set of candidate markers has been identified. Second, using CIM to detect multiple linked QTL demands considerable skill and experience. The QTL curve may contain several peaks and each peak does not necessarily correspond to a putative QTL.

The weaknesses of CIM have led to the development of a number of alternate approaches to mapping multiple QTL. Direct modifications of CIM include multiple interval mapping (Kao et al. 1999) and inclusive CIM (Li et al. 2008), which allow for simultaneous estimation of additive and epistatic effects. The classical variable selection methods discussed in Broman and Speed (2002) have been extended by various modifications to the Bayesian information criterion to select epistatic effects (Bogdan et al.

Communicated by M. Sillanpaa.

B. E. Huang · A. W. George (✉)
CSIRO Mathematical and Information Sciences,
Queensland Bioscience Precinct, 306 Carmody Road,
Brisbane, QLD 4067, Australia
e-mail: andrew.george@csiro.au; geo047@csiro.au

2004; Baierl et al. 2006; Manichaikul et al. 2009). Mixed models have also been employed to model QTL as random effects rather than fixed effects (George et al. 2000; Lee and Van der Werf 2006; Verbyla et al. 2007). Finally, several Bayesian approaches use MCMC algorithms to model interacting QTL but can be computationally intensive (Yi et al. 2005; Yi and Shriner 2008).

In this paper, we advance QTL mapping in experimental populations by developing a statistical approach that avoids the difficulties associated with CIM. In our approach, we separate the detection of QTL from the localization of QTL. That is, we look for the presence of QTL before leaping along a chromosome as in CIM. In the detection stage, we have developed an iterative procedure for automating the discovery of QTL along a genome. This iterative procedure is implemented within a linear mixed model framework where the markers on a chromosome are included in the mixed model as separate random effects but with common chromosomal variance. In the localization stage, we position the detected QTL by performing a series of one-dimensional interval mapping scans along chromosomes with detectable QTL. That is, if a chromosome contains multiple QTL, a separate scan positions each trait locus while accounting for the extraneous effects of previously mapped and unmapped QTL.

The strengths of our approach are: (1) we avoid using a QTL profile to infer number of QTL, (2) we account for background genetic variation when detecting and localizing QTL, (3) we eliminate the arbitrary nature in which cofactors are included in an analysis, and (4) we do not rely on a single one-dimensional scan as in CIM to map multiple linked QTL, but instead perform a series of interval mapping scans to more accurately position linked QTL. Also by basing our mapping procedure on mixed models, our approach easily handles fixed and random effects. This affords us the opportunity to map QTL while simultaneously accounting for extraneous sources of variation such as design effects, spatial variation and polygenic effects as done by Piepho (2000) and Eckermann et al. (2001). We will show through a detailed simulation study and a real example in wheat that our QTL mapping strategy gives results superior to CIM.

Methods

Detection localization mapping (DLMapping) algorithm

We begin by providing readers with an overview of DLMapping, our QTL mapping strategy. For readers interested in the technical details, a more extensive exposition is given in "Detection" and "Localization". Our algorithm consists of two parts: a detection stage and a localization stage. Both stages are iterative, formulated within a linear mixed model framework, and invariant to the order of the chromosomes in the map.

Detection stage

- *Step D1: Specify linear mixed models.* A full model and a reduced (or nested) model for each chromosome under investigation are constructed. These models contain fixed and random marker effects to simultaneously account for the extraneous effects of detected and undetected QTL, respectively.
- *Step D2: Identify chromosomes containing undetected QTL.* A likelihood based test statistic is calculated for each chromosome under investigation. This test statistic measures the strength of evidence for the presence of undetected QTL on a chromosome. The genomewide significance of the test statistic is determined via permutation.
- *Step D3: Identify markers to treat as fixed effects.* For each chromosome found to contain significant evidence for undetected QTL in the previous step, the following procedure is performed. First, we construct a linear mixed model for each marker on the chromosome. The marker is treated as a fixed effect. Second, we calculate a Wald statistic for the fixed marker effect. Third, we identify the marker with the largest Wald statistic on a chromosome. This marker is the one most strongly associated with the QTL, and is incorporated into subsequent models as a fixed marker effect.

These three steps are repeated until chromosomes no longer contain detectable QTL. Upon completion of the detection stage, r_j QTL have been detected on chromosome j. We then perform r_j interval mapping scans on chromosome j to localize these QTL.

Localization stage

Perform interval mapping scan on a chromosome containing unmapped QTL First, we compute the expected genotype of a QTL conditional on its hypothesized position and the genotypes of the flanking markers. Second, we construct a linear mixed model for each hypothesized position. This model, analogous to the models used in the detection stage, contains fixed and random effects to account for the confounding effects on localization of mapped and unmapped QTL, respectively. We also include a fixed effect for the QTL in the model, formed from the expected QTL genotypes. From this fixed effect, we are able to compute the size of the QTL. Third, we calculate the Wald statistic of the QTL effect. The hypothesized

QTL position yielding the mixed model with the highest Wald statistic is the estimated location of the QTL. The QTL effect for this position is included as a fixed effect in subsequent scans. These steps are repeated for each detected QTL on a chromosome. Once the detected QTL have been iteratively positioned, we construct a final multiple regression model to more accurately estimate the sizes of the QTL.

Notation

We use the following notation in this paper. Suppose data are collected from an experimental family with n progeny on a quantitative trait ($\mathbf{y}$) and M genetic markers spanning C chromosomes. We assume the genetic markers $\mathbf{m}_1$, $\mathbf{m}_2$,…, $\mathbf{m}_c$ have known order and location where $\mathbf{m}_j$ is the set of M_j markers on the jth chromosome. The marker data $\mathbf{Z} = [\mathbf{Z}_1 \ldots \mathbf{Z}_C]$ are partitioned as data collected on C chromosomes where $\mathbf{Z}_j$ is a $(n \times M_j)$ matrix of observed genotypes on $\mathbf{m}_j$. Furthermore, let $\mathbf{Z}_{-j} = [\mathbf{Z}_1 \ldots \mathbf{Z}_{j-1}\ \mathbf{Z}_{j+1} \ldots \mathbf{Z}_C]$ denote marker data collected on C chromosomes except the jth chromosome.

Detection

Suppose $i-1$ iterations of our detection procedure have been performed where r QTL have been identified. We explain the additive effects of these QTL on the trait via r selected markers. We assume C^{i-1} chromosomes are still under investigation and potentially contain undetected QTL. We arbitrarily index these chromosomes as 1, 2,…, C^{i-1}.

The steps needed to complete an iteration of our algorithm are as follows. First, a full model and C^{i-1} reduced models are constructed. Here, the r markers are included in the linear mixed models as fixed effects. These fixed effects reduce the potential for the detected QTL to obfuscate the discovery of unknown QTL. Our models also include random components where the markers on the chromosomes being tested are treated as separate random effects but with common chromosomal variance. These random effects model a chromosome's genetic contribution.

The full (linear mixed) model is:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}^{i-1}\mathbf{u}^{i-1} + \mathbf{e} \tag{1}$$

where $\mathbf{X}$ is an $(n \times r)$ design matrix and the columns of this matrix contain the marker genotypes for the associated selected marker, β is an $(r \times 1)$ vector of additive QTL effects for QTL located at the markers, $\mathbf{Z}^{i-1} = [\mathbf{Z}_1^{i-1}\mathbf{Z}_2^{i-1}\ldots\mathbf{Z}_{C^{i-1}}^{i-1}]$ denotes the marker data on the C^{i-1} chromosomes under investigation excluding data on any of the r selected markers, $\mathbf{u}^{i-1}$ is a vector of random QTL effects associated with the markers on the C^{i-1} chromosomes under investigation excluding any of the r selected markers, and $\mathbf{e}$ is a residual vector. The random effects $\mathbf{u}^{i-1}$ and $\mathbf{e}$ are assumed to be uncorrelated and distributed as multivariate normal densities with $\mathbf{e} \sim \mathbf{N}_n\left(\mathbf{0}, \sigma_e^2\mathbf{I}_n\right)$ and $\mathbf{u}^{i-1} \sim \mathbf{N}_{C^{i-1}}(\mathbf{0}, \sigma^2\mathbf{S})$ where $\mathbf{S}$ is a block diagonal (co)variance matrix with $\mathbf{S}_{jj} = \gamma_j\mathbf{I}_{M_j^{i-1}}$ $(j = 1, \ldots, C^{i-1})$.

The kth reduced (linear mixed) model is:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}_{-k}^{i-1}\mathbf{u}_{-k}^{i-1} + \mathbf{e} \tag{2}$$

where $\mathbf{u}_{-k}^{i-1}$ is a vector of random QTL effects associated with the markers on the C^{i-1} chromosomes except the kth indexed chromosome. This model is formed by only removing the random marker effects associated with the kth indexed chromosome from the full model.

Second, we test for the presence of QTL on the C^{i-1} chromosomes. We begin by computing the residual log likelihood (logL) of the full model and the C^{i-1} reduced models using REML. We then measure the evidence for the existence of QTL on a chromosome by calculating the residual log likelihood ratio (lr). That is, for the kth indexed chromosome, we calculate $\mathrm{lr}_k = -2(\mathrm{logL}_{\mathrm{full}} - \mathrm{logL}_k)$ where $\mathrm{logL}_{\mathrm{full}}$ and logL_k are the residual log likelihoods under the full and kth reduced models, respectively. To assess the significance of lr_k, we compare it to the distribution of the test statistic under the null hypothesis of no QTL on the kth indexed chromosome. If only a single chromosome is under investigation, the distribution of the test statistic is well known, following a mixture of chi-squared distributions (Self and Liang 1987). However, if QTL are being detected on multiple chromosomes ($C^{i-1} > 1$), multiple hypothesis tests are performed and the distribution of lr_k under the null is not known. To circumvent this difficulty, we use permutation (Churchill and Doerge 1994) to empirically assess the genomewide significance of lr_k. We begin by permuting the marker random effects in the full and reduced models while preserving the structure of the trait and fixed effects. That is, we permute only the rows of $\mathbf{Z}^{i-1}$. A residual likelihood ratio statistic lr_k is then computed for each of the C^{i-1} chromosomes. The maximum $\mathrm{lr}_{\max}$ is recorded, and the process is repeated a large number of times (we perform 1,000 permutations). We can then easily calculate the genomewide significance of our observed test statistics from the ordered values of $\mathrm{lr}_{\max}$.

Third, for each chromosome found in the ith iteration to contain a QTL, we identify the marker on this chromosome that explains the most variation. We do this by constructing a separate linear mixed model for each marker on the chromosome. Suppose the kth indexed chromosome has been found to contain undetected QTL. A linear mixed model is constructed for each marker on this chromosome where the marker being assessed is treated as a fixed effect,

together with the r markers already identified as being associated with QTL. The markers on the other $C^{i-1} - 1$ chromosomes are treated as random effects where as before, markers on the same chromosome have common variance. We found our procedure worked best if the other markers on the kth indexed chromosome were not included as random effects in the model.

The linear mixed model for the xth marker on the kth indexed chromosome is

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{z}_{k;x}\beta_x + \mathbf{Z}_{-k}\mathbf{u}_{-k} + \mathbf{e} \quad (3)$$

where $\mathbf{z}_{k;x}$ is a $(n \times 1)$ vector of marker data collected on the xth marker on kth indexed chromosome, and β_x represents an additive fixed effect for a QTL at the marker. We then calculate the Wald statistic of β_x. We select the marker associated with the model giving the largest Wald statistic. This fixed effect is then incorporated into the full model in subsequent iterations. Note we are not using the Wald statistic to assess the significance of the fixed effect. We are only using this test statistic to identify the "best'' marker.

These three steps are repeated. After each iteration, the number of fixed effects increases and the number of chromosomes under consideration remains the same or is reduced to only those chromosomes containing QTL. In this way, we are able to account for the confounding effects of detected QTL (through the fixed marker effects) and reduce background genetic noise (by only considering chromosomes containing QTL). That is, each iteration builds on the knowledge gained from previous iterations to detect QTL.

Localization

In the localization stage of our QTL mapping strategy, we use an iterative procedure to position multiple trait loci on a chromosome. Each QTL is positioned in turn using a linear mixed model approach analogous to regression-based interval mapping. In our procedure, once a QTL has been positioned, it is included in subsequent scans of that chromosome as a fixed effect. We also include markers on other chromosomes as random marker effects (as we did in our detection procedure). After all QTL have been positioned along a genome, we construct a final multiple regression model to estimate the sizes of the QTL.

We make the following assumptions before detailing our localization procedure. We assume C^* chromosomes have been found to contain QTL in our detection stage where $C^* \leq C$. We arbitrarily index these chromosomes $1, 2, \ldots, C^*$. We assume $i - 1$ interval mapping scans (i.e., iterations) have been performed on the kth indexed chromosome where $i - 1$ QTL have been positioned.

The steps needed to complete the ith iteration of our localization procedure are as follows. First, for the hypothesized positions of the trait locus along the kth indexed chromosome, we compute the expected genotype of the QTL conditional on its position θ_Q and the genotypes of the flanking markers. That is, we compute $E(\mathbf{g}|\theta_Q, \mathbf{z}_L, \mathbf{z}_R)$ at each hypothesized trait position where $E(\cdot)$ is a $(n \times 1)$ vector of expected genotypes, $\mathbf{g}$ is a $(n \times 1)$ vector of QTL labels for the n progeny, $g_i \in \mathbf{g}$ is the ith individual's QTL label taking on values 1 and 0 (-1) for QTL genotypes QQ and Qq (qq), respectively, and $\mathbf{z}_L$ and $\mathbf{z}_R$ are $(n \times 1)$ vectors of genotypes of the left and right flanking markers, respectively. See Whittaker et al. (1995) for further details.

Second, we construct a linear mixed model for each hypothesized position θ_Q of the trait locus. The linear mixed model is similar in structure to Eq. 3 that we used in our detection procedure. The mixed linear model is

$$y = \mathbf{X}\beta + E(\mathbf{g}|\theta_Q, \mathbf{z}_L, \mathbf{z}_R)\beta_{\theta_Q} + \mathbf{Z}^*_{-k}\mathbf{u}^*_{-k} + \mathbf{e} \quad (4)$$

where $\mathbf{X}$ is a $(n \times i - 1)$ matrix of expected QTL genotypes for the $i - 1$ positioned QTL, β_{θ_Q} is the additive effect of a putative QTL at position θ_Q, $\mathbf{Z}^*_{-k}$ is the matrix of marker genotypes on the C^* chromosomes except for the markers on the kth indexed chromosome and any of the markers flanking the $i - 1$ previously positioned QTL, and $\mathbf{u}^*_{-k}$ is the vector of random marker effects associated with the QTL. We assume e and $\mathbf{u}^*_{-k}$ are uncorrelated and distributed as multivariate normal deviates, analogous to what we described above in the detection stage.

Third, for each linear mixed model, we calculate the Wald statistic of β_{θ_Q}. The hypothesized position yielding the mixed model with the highest Wald statistic is the estimated location of the QTL. We then repeat this process until all QTL have been positioned along the genome.

Once the C^* chromosomes have been iteratively scanned and the QTL positioned, we construct a final model to estimate the additive effects of the QTL. Estimates of the QTL effect sizes could have been obtained from Eq. 4; however, we found the following model gave superior parameter estimates. The multiple regression model we use to estimate the additive effects of the located QTL is

$$\mathbf{y} = \mu + \sum_{k=1}^{C^*}\sum_{m=1}^{t_k} E(\mathbf{g}|\theta_{k;m}, \mathbf{z}_{k;m-1}, \mathbf{z}_{k;m+1})\beta_{k;m} + \mathbf{e} \quad (5)$$

where μ is the intercept, $\theta_{k;m}$ is the position of the mth QTL on the kth indexed chromosome, t_k is the number of detected QTL on the kth indexed chromosome, and $\mathbf{z}_{k;m-1}$ and $\mathbf{z}_{k;m+1}$ are the $(n \times 1)$ vectors of genotypes collected on the markers flanking the mth QTL on the kth indexed chromosome.

Implementation of QTL mapping approach

We have implemented the DLMapping approach presented in this paper in an R (R Development Core Team 2008) library called DLMap. With DLMap, we can analyze data from double haploids, backcrosses, and by modifying the models presented in this paper to account for dominance, F2 designs. Our implementation also easily accommodates additional sources of fixed and random variation as are often encountered in animal and plant studies. In building this library, we have made use of the R/qtl library (Broman et al. 2003) for marker map estimation, imputation of missing marker genotypes and the calculation of expected QTL genotypes. Our R library contains two functions: dlm.asreml and dlm.lme. The function dlm.asreml requires access to the commercial package ASReml-R, an R front-end to the well known REML estimation package ASReml (Gilmour et al. 2006). The function dlm.lme is built on the freely available R library nlme (Pinheiro et al. 2008). Both implementations give identical results, but dlm.asreml is computationally more expedient and affords users greater flexibility when building models that also accommodate environmental effects. DLMap is freely available from CRAN (http://cran.r-project.org/), the official R package archive.

CIM analysis

CIM results are obtained via QTL Cartographer (Wang et al. 2005). Forward–backward selection (the default) is used for marker variable selection. K marker cofactors are included in the CIM analyses where K is specified by the user prior to analysis. We explore the impact of different values of K on the QTL findings. CIM analyses with $K = 1$, 5, and 9 are performed and referenced by CIM1, CIM5, and CIM9, respectively. Genomewide significance thresholds are computed empirically by performing 1,000 permutations.

As previously discussed, inferring multiple linked QTL from CIM profiles is challenging. Various decision rules for automating this process have been described. One strategy is to declare the local peaks of a profile that are above a (permutation) threshold to be associated with segregating QTL. However, we found that this results in many false positives. Alternately, we could infer linked QTL from a profile if the minimum score between two local peaks is below a threshold or if the minimum score is less than half the larger of the peak scores (Verbyla et al. 2007). However, the decision rule that we use in this paper is as follows. Suppose a CIM profile contains two peak scores above a significance threshold. The peak scores are interpreted as being caused by two linked QTL if there is at least a drop of 1.5 lod score units between the peak scores. This strategy was also employed by Broman and Speed (2002).

Simulation study

To compare the performance of our new QTL mapping approach with CIM, marker and trait data are simulated and analyzed on backcross populations under four different scenarios. For the first three scenarios, marker data are realized on nine chromosomes, each of chromosomal length 100 cM and containing 11 equally spaced markers. For the fourth scenario, we follow the setup of Zeng (1994) where genetic data are simulated on four chromosomes of length 150 cM with 16 equally spaced markers. In this simulation study, we assume the chromosomal order of the genetic markers is known, but the intermarker distances are unknown. Hence, prior to performing a QTL analysis of a data replicate, we use these data to estimate the marker positions. A background phenotypic variation $\left(\sigma_e^2\right)$ of 1.0 is used in generating all quantitative trait data.

When detection is based on a LOD profile as in CIM, a common approach is to label a QTL a true detection if it is within a certain distance of a true QTL. However, we fix the detected number of QTL before localization, so we assess the performance of each stage separately. We judge the performance of both CIM and DLMapping based on the separate areas of detection and localization. The detection stage determines the number of QTL on each chromosome. We calculate the power as the proportion of replicates for which the true number of QTL is correctly identified on a chromosome. The localization stage positions the QTL on each chromosome and constructs a final model with all QTL. We estimate the QTL position and additive size effect for each chromosome in models where the correct number of QTL was detected.

Under the first scenario, we assess the type I error (wrongly inferring the presence of QTL) of our approach and CIM. In this scenario, the quantitative trait is not influenced by QTL. That is, the quantitative trait is randomly drawn from a normal distribution with mean 0 and variance $\sigma^2 = 1.0$. Marker and trait data are generated on backcross populations with 100, 250, and 500 offspring. For each population size, 1,000 data set replicates are generated.

Under the second scenario, we examine the ability of our approach and CIM to detect linked QTL of varying size. We place two QTL on each of the first four chromosomes with the other five chromosomes not carrying QTL. The distance between the linked QTL is fixed at 40 cM. One of the QTL on each of the four chromosomes has an additive effect of 0.76 (equivalent to the QTL sizes used in Broman and Speed (2002)). The remaining QTL on chromosomes 1, 2, 3, and 4 have additive sizes as given in

Table 1 QTL locations, QTL sizes, and amount of phenotypic variance explained by the QTL (%QTL) in Scenarios 2, 3, and 4 for the chromosomes carrying QTL

	Chromosome									
	1			2			3			4	
Scenario 2: QTL of varying size											
Position	40	80		40	80		40	80		40	80
Size	0.76	0.25		0.76	0.354		0.76	0.5		0.76	0.76
%QTL	6.7	0.72		6.7	1.5		6.7	2.9		6.7	6.7
Scenario 3: QTL of varying position											
Position	11	31		11	41		11	51		11	61
Size	0.76	0.76		0.76	0.76		0.76	0.76		0.76	0.76
%QTL	5.3	5.3		5.3	5.3		5.3	5.3		5.3	5.3
Scenario 4: QTL of varying size and position											
Position	16	48	108	3	43	77	33	68	129	26	
Size	0.42	0.75	0.58	0.51	−0.615	−0.63	−0.46	0.805	0.88	0.74	
%QTL	2.0	6.3	3.8	2.9	4.3	4.5	2.4	7.3	8.7	6.2	

Table 1. The total amount of phenotypic variation explained by these QTL is 53.5%. Population size is varied from 100 to 500 in steps of 50 to evaluate the effect of family size on power. For each population size, 500 data set replicates are generated.

Under the third scenario, we examine the ability of the two QTL mapping approaches to detect QTL of varying chromosomal position. As in the previous scenario, we place two linked QTL on the first four chromosomes with the other five chromosomes not carrying QTL. All QTL have additive sizes of 0.76. The distances between QTL range from 20 to 50 cM, with the chromosomal positions of the QTL given in Table 1. The total amount of phenotypic variation explained by these QTL is 63.5%. As in Scenario 2 the population size is varied from 100 to 500 in steps of 50. For each population size 500 data set replicates are generated.

For the fourth scenario, we closely follow the setup described in Zeng (1994). This scenario is included in our simulation study to challenge DLMapping and CIM. Here, multiple linked QTL of varying size are considered where QTL are in coupling and repulsion. The positions of the QTL match those in Zeng (1994); however, we have halved the size of all QTL with size greater than 1.0 to increase the difficulty of QTL detection (Table 1). A backcross population of 300 individuals is assumed where 500 data set replicates are generated.

Wheat data analysis

To further demonstrate the strengths of our new QTL mapping approach over CIM, we analyzed data from a two-phase wheat field trial performed in Griffith, Australia. The trial involved 175 double haploid lines derived from crossing bread wheat varieties Chara and Glenlea. In the first phase, the double haploid lines were planted in a rectangular array of 28 rows by nine columns of plots. Plants were randomized to plots and duplicated according to a p/q partial replicated design (Smith et al. 2006). One-third of the double haploid lines (p) were duplicated in this phase and carried through to the next phase. In the second phase the grain harvested from the plots was milled to produce flour. One quarter of the genotypes (q) were duplicated in this phase. Genotypes were allocated to positions within the milling process using an incomplete block design with blocks being the milling day. In this study, quality measurements were collected on a number of economically important traits, but for ease of exposition we focus on flour yellowness, which was scored using a Minolta color meter (Min-b*).

We have adopted a two-stage strategy for analyzing these data. In the first stage, a mixed linear model is constructed to model the relationship between flour yellowness and the environmental effects. This multi-phase model accounts for field spatial effects, design factors and the milling process and has been described by Smith et al. (2006). We also include a fixed genotype (or variety) effect, from which predicted means are computed for the second stage. In the second stage, QTL analyses were performed using the predicted means as the (adjusted) trait. CIM was run with 3, 5, 7, and 9 marker cofactors and compared to DLMapping. Genomewide P values were calculated using 1,000 permutations for each method.

For these QTL analyses, marker data were collected on the double haploid lines. A total of 422 DArT, SSR, and RFLP markers were scored, spanning the 21 chromosomes, giving good coverage of the A, B, and D wheat genomes. On average, 17, 28, and 12 markers per chromosome

covered the A, B, and D chromosomes, respectively. Approximately 5% of the marker data were missing and were replaced with expected values. These expected values are calculated from the observed flanking marker data and the known inter-marker map distances (Martinez and Curnow 1992).

Results

Scenario 1: type I errors

The probability of wrongly inferring the presence of QTL (i.e., type I error) under DLMapping and CIM is given in Table 2. The type I errors are calculated from the percentage of replicates where at least one QTL is detected on any of the nine chromosomes. A 5% significance level is assumed. From Table 2, it is clear that permutation is effective in guarding against excess type I errors for DLMapping and CIM.

Table 2 Empirical type I error for CIM and DLMapping in Scenario 1

n	CIM1	CIM5	CIM9	DLMapping
100	0.029	0.044	0.044	0.038
250	0.047	0.043	0.043	0.057
500	0.037	0.037	0.037	0.045

Data are generated under Scenario 1 on a backcross family with 100, 250, and 500 progeny. Results are based on the analysis of 1,000 data replicates. A genomewide significance level of 0.05 is assumed

Scenario 2: linked QTL of varying size

The power curves for DLMapping and CIM are shown in Fig. 1. Here, power is plotted against family size for the four chromosomes containing QTL. A point on this curve is calculated as the proportion of replicates for which the two QTL are correctly identified on a chromosome for a given family size. A genomewide significance of 0.05 is assumed.

In Fig. 1, as expected, the power of the two QTL mapping approaches increases with family size and QTL size. Thus, the power for both approaches is lowest on the first chromosome, because this chromosome contains the QTL with the smallest effects. DLMapping has consistently higher power to detect QTL than CIM. In some cases, there is more than a 100% increase in power. It is also interesting to note that the performance of CIM varies unpredictably with the number of cofactors. For chromosomes 1 and 2, CIM with five marker cofactors (CIM5) has the highest power of the CIM analyses. However, for chromosomes 3 and 4, CIM with nine marker cofactors (CIM9) almost always has highest power.

In Table 3, the expected positions of the QTL and the expected additive QTL sizes are given for DLMapping and CIM5 for scenarios 2 and 3. The standard errors are in parentheses. Results are reported per chromosome and are

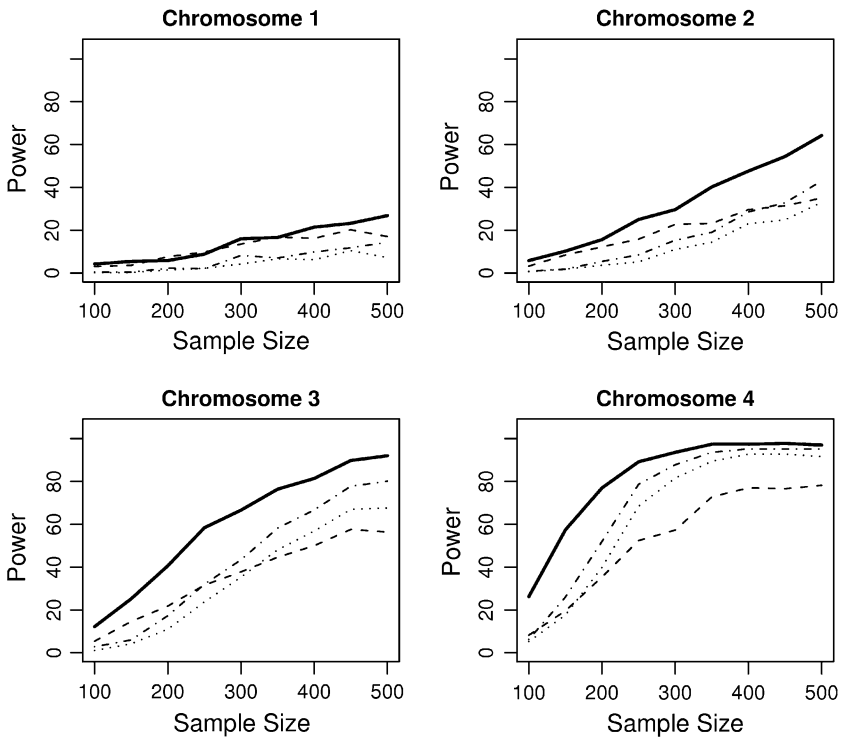

Fig. 1 QTL power curves for DLMapping and CIM in Scenario 2. A point on this curve is the probability of the correct number of QTL being detected on a chromosome. Data are generated under Scenario 2 where each chromosome contains two linked QTL of varying size. 500 replicates are analyzed for each family size. The *solid line* indicates the power for DLMapping; the *dashed*, *dotted*, and *dotted-dashed lines* indicate CIM analyses with 1, 5, and 9 cofactors, respectively. A genomewide significance level of 0.05 is assumed

Table 3 Position and QTL size estimates for DLMapping and CIM in Scenarios 2 and 3

	Chromosome							
	1		2		3		4	
Scenario 2: QTL of varying size								
True Size	0.76	0.25	0.76	0.35	0.76	0.50	0.76	0.76
DLMapping								
Detection (%)	9		25		58		89	
Position	33.0 (16.4)	71.9 (18.3)	39.3 (11.1)	79.2 (13.6)	40.7 (8.1)	80.7 (9.8)	40.8 (7.3)	79.3 (8.4)
Effect	0.51 (0.41)	0.61 (0.30)	0.68 (0.29)	0.55 (0.20)	0.75 (0.21)	0.59 (0.15)	0.77 (0.16)	0.78 (0.16)
CIM5								
Detection (%)	2		5		24		69	
Position	41.5 (12.4)	66.7 (15.8)	43.8 (12.3)	78.8 (13.9)	45.5 (7.9)	87.8 (10.4)	44.7 (7.1)	88.8 (8.2)
Effect	0.93 (0.15)	0.87 (0.16)	0.85 (0.37)	0.88 (0.20)	0.91 (0.21)	0.81 (0.19)	0.87 (0.18)	0.86 (0.17)
Scenario 3: QTL of varying position								
True position	11	31	11	41	11	51	11	61
DLMapping								
Detection (%)	44		70		85		89	
Position	13.0 (6.2)	39.8 (20.1)	12.2 (6.1)	41.5 (9.5)	12.8 (6.0)	50.0 (8.8)	12.3 (6.2)	59.9 (8.4)
Effect	0.90 (0.32)	0.69 (0.35)	0.80 (0.22)	0.76 (0.22)	0.76 (0.17)	0.77 (0.17)	0.75 (0.16)	0.78 (0.16)
CIM5								
Detection (%)	12		22		31		35	
Position	15.0 (6.3)	34.6 (11.3)	14.6 (6.3)	43.8 (8.6)	14.9 (7.7)	54.1 (11.6)	15.0 (9.1)	64.4 (8.2)
Effect	1.32 (0.26)	1.29 (0.27)	1.18 (0.25)	1.08 (0.27)	1.08 (0.24)	1.09 (0.25)	1.02 (0.23)	1.01 (0.24)

The mean and standard error (in parentheses) of position and QTL size estimates are given for replicates in which the correct number of QTL are detected on a chromosome. The percentage of these replicates (detection %) is for an assumed genomewide significance level of 0.05. Results are given for the analyses of data simulated under Scenario 2 (QTL of varying position) and Scenario 3 (QTL of varying linkage) on a backcross population of 250 progeny. For Scenario 2, QTL are simulated at positions 40 and 80 cM. For Scenario 3, QTL are of size 0.76. CIM results are obtained from analyses involving five marker cofactors (CIM5). QTL position is in centimorgans

based on analyses where both QTL are correctly detected. We have focused on QTL analyses of data generated on a backcross of moderate size (250 individuals).

For Scenario 2 in Table 3, the challenges of positioning and accurately estimating the sizes of QTL of small effect are obvious. Both approaches on average correctly localize QTL, but the standard errors associated with position estimates on chromosomes 1 and 2 are large. Also, difficulties in localization impact the estimation of QTL size effect. Overall, DLMapping does position QTL marginally better than CIM5. Moreover, it is superior to CIM5 for estimating the size of QTL effect. CIM5 overestimates QTL size which in turn inflates the importance of a linkage finding.

When the sample size is increased to a backcross population of 500 progeny, position and QTL size estimates are greatly improved (results not shown). For chromosome 1, the chromosome containing the most difficult QTL to map, DLMapping accurately positions the two QTL at 41.4 (6.9) cM and 80.3 (11.5) cM. Standard errors are given in parentheses. These are very close to the true positions of 40 and 80 cM. CIM5 also, on average, accurately positions the two QTL at 44.7 (5.8) and 81.8 (13.0) cM. However, as seen previously, DLMapping is superior to CIM5 when estimating the size of a QTL effect. For chromosome 1, DLMapping estimates the size of the first QTL effect as 0.71 (0.24) where the true size is 0.76. The size of the second QTL effect is estimated as 0.38 (0.17) where the true value is 0.25. CIM5 estimates the sizes of the QTL effects as 0.88 (0.14) and 0.57 (0.17).

Scenario 3: linked QTL of varying position

The power curves for DLMapping and CIM are shown in Fig. 2 for chromosomes 1, 2, 3, and 4. The four chromosomes each contain two linked QTL where QTL size is fixed across chromosomes but QTL position is varied. From Fig. 2, DLMapping is clearly more powerful than CIM. For a backcross with 250 progeny, CIM5 only detects the two QTL on the first chromosome in 12% of the data replicates. DLMapping detects the two QTL on chromosome 1 in 44% of the replicates. As the distance between the two linked QTL increases, this contrast in performance between the two QTL mapping approaches decreases. However, even when the two QTL are separated by 50 cM, DLMapping is still over 40% (150%) more

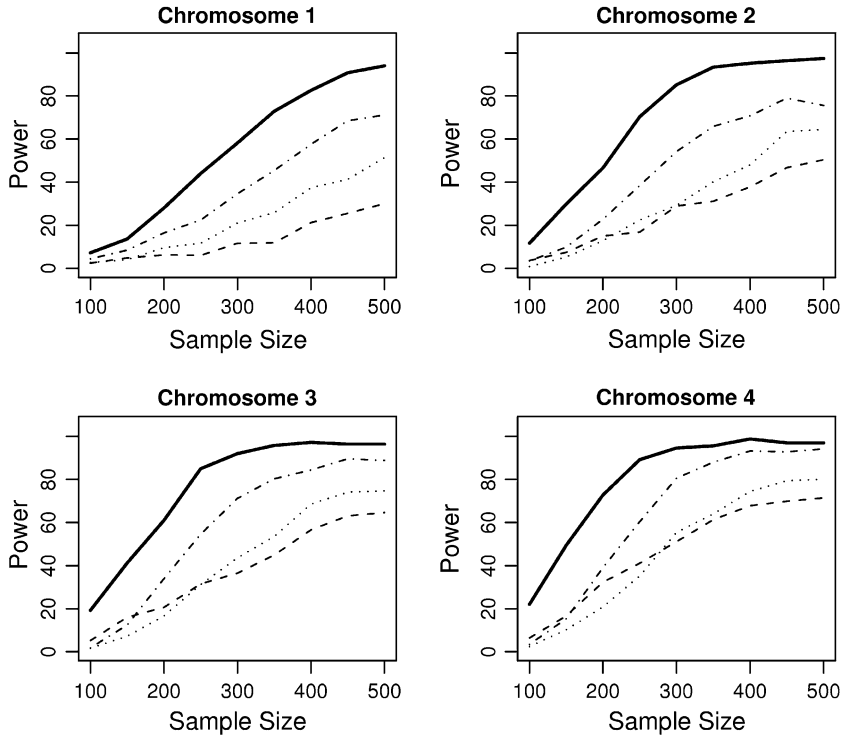

Fig. 2 QTL power curves for DLMapping and CIM in Scenario 3. A point on this curve is the probability of the correct number of QTL being detected on a chromosome. Data are generated under Scenario 3 where each chromosome shown contains two linked QTL of varying position. The distance between the two QTL is 20 cM on chromosome 1 and increases by 10 cM up to 50 cM on chromosome 4. 500 replicates are analyzed for each family size. The *solid line* indicates the power for DLMapping; the *dashed*, *dotted*, and *dotted-dashed lines* indicate CIM analyses with 1, 5, and 9 cofactors, respectively. A genomewide significance level of 0.05 is assumed

Table 4 Distribution and average number of QTL detected for CIM and DLMapping in Scenario 4

Chr	True No. of QTL	Method	No. of QTL detected					
			0	1	2	3	>3	Average
1	3	**DLMapping**		3.0	60.4	**30.4**	6.2	2.40
		CIM1	0.6	29.7	56.2	11.8	1.6	1.84
		CIM5	0.6	25.3	62.2	11.5	0.4	1.86
		CIM9	0.8	22.6	59.5	16.6	0.4	1.93
2	3	**DLMapping**		36.8	30.8	**26.4**	6.0	2.02
		CIM1	1.8	86.3	11.6	0.2		1.10
		CIM5	1.8	87.5	9.5	1.2		1.10
		CIM9	3.6	55.5	29.7	10.8	0.4	1.49
3	3	**DLMapping**		8.0	48.6	**36.2**	7.2	2.43
		CIM1	0.8	38.8	57.0	3.4		1.63
		CIM5	0.6	41.0	51.3	6.9	0.2	1.65
		CIM9	0.8	14.6	66.7	17.6	0.2	2.02
4	M	**DLMapping**	1.2	**97.6**	1.0	0.2		1.00
		CIM1	17.9	79.7	2.4			0.85
		CIM5	7.9	91.5	0.6			0.93
		CIM9	4.0	95.2	0.8			0.97

The approach with the highest power for detecting the correct number of QTL on each chromosome is in bold. 500 data replicates were generated and analyzed on a backcross population of 300 progeny. A genomewide significance level of 0.05 is assumed

likely to detect the correct number of QTL than CIM9 (CIM5) for a population of size 250.

For completeness, the position and size estimates of QTL generated under Scenario 3 are given in Table 3 for DLMapping and CIM. As seen previously, both approaches are comparable in terms of their accuracy of localizing QTL. However, when estimating the size of QTL effects, DLMapping is superior. CIM can overestimate the size of a QTL effect by as much as 70%. This bias toward overestimation by CIM has also been noted by Li et al. (2007).

Scenario 4: mapping multiple linked QTL of varying sizes and positions

In Table 4, the ability of DLMapping and CIM to detect linked QTL of varying size is reported for data generated

under Scenario 4. Here, ten QTL of varying size in coupling and repulsion are distributed along four chromosomes. Again we see the superiority of DLMapping for detecting multiple linked QTL compared to CIM. DLMapping has the highest power on all four chromosomes for detecting the true number of QTL. Furthermore, on chromosomes 1, 2, and 3, DLMapping has more than double the power of CIM in most cases. DLMapping occasionally detects too many QTL on chromosomes containing multiple QTL, but the probability of CIM failing to detect QTL is far greater than the probability of DLMapping inferring too many QTL.

It is also interesting to note that the commonly used average number of QTL is not always the best measure of detection performance. For chromosome 2, DLMapping detects 2.02 QTL on average. Similarly, CIM9 detects 2.02 QTL on average for chromosome 3. Based solely on these averages, we would conclude that DLMapping and CIM9 are performing equally well. However, CIM9 only detects the correct number of QTL on the third chromosome in 17.6% of the replicates, whereas DLMapping detects three QTL on the second chromosome in 26.4% of the replicates. By also reporting the distribution of the QTL, we uncover the fact that DLMapping is outperforming CIM.

In the interests of conserving space, we do not report the position estimates and QTL effect size estimates from our analyses. However, as we saw previously, DLMapping gives marginally better localization estimates than CIM and superior QTL effect size estimates.

Analysis of wheat data

A considerable amount has been published on the genetics of flour color, particularly pigment content. QTL have been confirmed on chromosomes 2A, 4B, 7A, and its homeologous chromosome 7B (Mares and Campbell 2001; Zhang and Dubcovsky 2008; Raman et al. 2007). Furthermore, the candidate gene *PSY-1* and candidate genes responsible for polyphenol oxidase activity (i.e., causes browning) have been mapped in these regions. Hence, this example affords us the opportunity to compare DLMapping and CIM for the analysis of experimental data where the location of QTL is well documented.

Shown in Fig. 3 are the QTL profiles for CIM analyses of the data with 3, 5, 7, and 9 marker cofactors included in the model. Chromosomes 2A, 4A, 4B, 7A, and 7B are labeled as they contained the five most significant LOD scores. We track their significance through the different analyses and calculate the permutation-adjusted genome-wide *P* values (Table 5). QTL on chromosomes 4B, 7A and 7B are highly significant for any number of cofactors but the confirmed QTL on chromosome 2A is never detected.

In contrast, DLMapping detected all four of the published QTL with permutation-adjusted *P* values under 0.01 (Table 5). DLMapping also detected a fifth QTL at a 10%

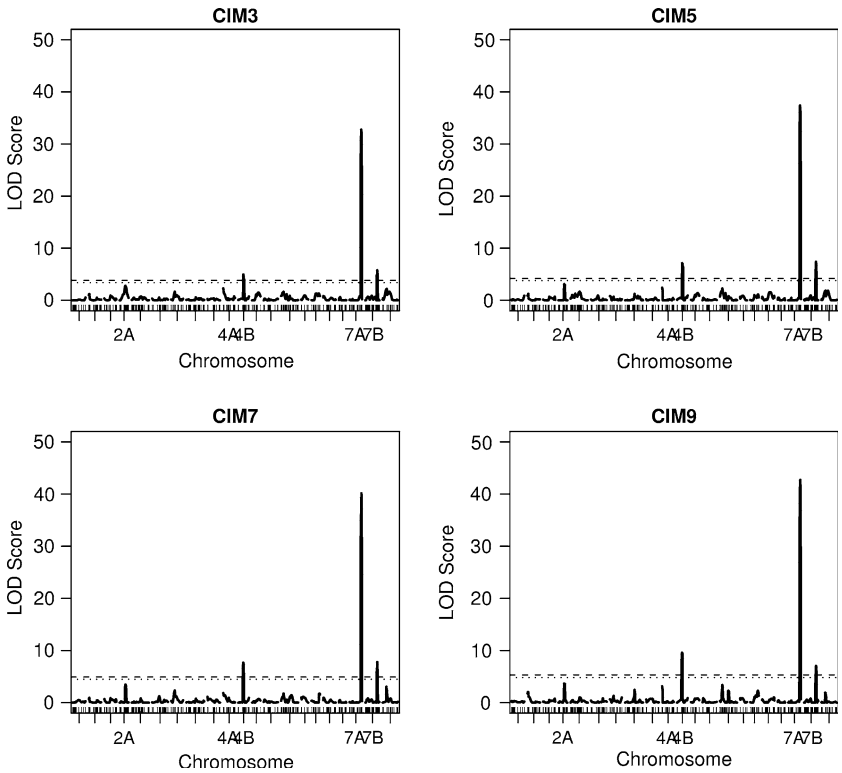

Fig. 3 QTL profiles for data from Chara × Glenlea wheat cross. Profiles are given for CIM analyses of the data involving 3, 5, 7, and 9 cofactors labeled CIM3, CIM5, CIM7, and CIM9, respectively. The *dashed* and *dotted lines* denote the 0.05 and 0.10 empirical thresholds obtained with permutation. The five chromosomes with the most significant results are labeled

Table 5 Selected genomewide *P* values for analysis of Chara × Glenlea wheat cross

Method	Chromosome				
	2A	4A	4B	7A	7B
CIM3	0.277	0.507	0.004	<0.001	0.001
CIM5	0.297	0.606	<0.001	<0.001	<0.001
CIM7	0.322	0.963	<0.001	<0.001	<0.001
CIM9	0.365	0.578	<0.001	<0.001	<0.001
DLMapping	0.008	0.063	<0.001	<0.001	<0.001

CIM and DLMapping *P* values are calculated from 1,000 permutations. Results are shown for the five most significant chromosomes where for CIM the significance refers to the largest peak on that chromosome

significance level on chromosome 4B. Mares and Campbell (2001) previously found evidence for a QTL on this chromosome associated with the same trait, flour yellowness (b*). DLMapping thus succeeded at detecting confirmed QTL where CIM failed. Furthermore, DLMapping identified an additional locus inviting further investigation.

Discussion

In this paper, we have presented a new approach for mapping QTL in experimental populations. In contrast to competing strategies, we treat the detection and localization of QTL as two separate tasks. This has allowed us to create customized iterative algorithms specific to QTL detection and QTL localization. Unlike standard approaches, we judge the performance of the method for these two stages separately. We demonstrate in the simulation study that our approach has increased power of detecting QTL, especially when there are multiple linked QTL on a chromosome. In fact, we achieved greater than a 100% increase in power over CIM in some instances. We also saw that our approach accurately positions QTL and gives more precise estimates of QTL size.

Many variations on the models and algorithms presented were considered in the process of developing and refining our QTL mapping approach. In the detection stage, we considered using intervals rather than markers as random effects and using one variance component for the entire genome rather than separate variance components for each chromosome. The formulation of the intervals incorporated the genetic distance between adjacent markers into the model. We explored the benefits of including marker information on all chromosomes in each iteration as opposed to including information only on those chromosomes still under investigation. We also evaluated the performance of forward, backward, and forward–backward variable selection for identifying chromosomes containing undetected QTL. In the localization stage, we experimented with including the additive effects of unlinked QTL in the linear mixed models, not including any random marker effects, including the markers on the chromosome being processed as random effects, including the flanking markers of localized QTL, and positioning linked QTL using a multiple dimensional interval mapping scan. The final modifications were selected on the basis of power, computational efficiency, and statistical simplicity. Interestingly, the abovementioned modifications all had improved performance over CIM. This confirmed our belief that we have been able to achieve impressive gains in power from the structural elements shared by all these modifications: a mixed model framework and the separation of QTL detection and localization.

One example of the refinement process is the differences between using a multi-dimensional scan to localize QTL rather than a series of single-dimension scans. Localizing in multiple dimensions should increase the precision of multiple QTL position estimates since the position of the first QTL is not fixed in subsequent scans.

To investigate the loss in precision in using single-dimension scans over multidimensional scans, we reanalyzed 100 replicates from Scenario 3 using the two strategies. We then calculated the mean squared error (MSE) of the estimate and true position of the QTL for both scan strategies. We found very little difference in accuracy between the two approaches, only a 0.7 cM difference in root MSE. However, there is a considerable difference in the amount of time taken to localize the QTL. For example, for one replicate, the time taken to position two QTL each on five chromosomes was 110 s for the single-dimension scans. In contrast, the multidimensional scans took 1,700 s, an order of magnitude longer. Furthermore, the time difference increases exponentially with the number of QTL detected on a chromosome. Thus, the small increase in precision did not justify the additional computational burden of the more extensive scan.

There are several aspects to our QTL detection algorithm worth noting. First, the detection stage of our QTL mapping approach is unaffected by errors in the marker map. Map information is only used to group linked markers. More traditional QTL mapping approaches can be seriously compromised by map misspecification (Daw et al. 2000). Second, our detection strategy could be used as a novel approach to analyze whole genome association data. Association analyses are most commonly performed one marker at a time to assess the strength of association between a trait and the marker. This ignores the possible confounding effects of extraneous QTL. With our detection approach, we correctly account for extraneous linked and unlinked QTL. Furthermore, in our approach, the strengths of association between an entire genome of markers and a

trait can be simultaneously assessed within a single analysis. We are currently exploring opportunities for the implementation of this approach in association studies conducted in wheat. However, it would be necessary to extend the method to properly control for false positive associations due to population structure in this type of analysis in a similar fashion to the mixed model approach of Yu et al. (2006). Finally, our detection approach can easily analyze data collected on outbred crosses and/or general pedigrees. It is only the localization stage that restricts our QTL mapping strategy to data collected on experimental populations.

A possible limitation of our approach is the computational burden imposed by permutation in the detection stage. Our approach may take up to three orders of magnitude longer than CIM for the same number of permutations. We could argue that the additional computational cost is justified given the increase in power. We could also argue that QTL data often takes years to collect; so it is reasonable to allow more than a few minutes for an analysis. A faster option, however, is to use an analytical threshold. We reanalyzed data generated in Scenario 1 (no QTL) and Scenario 2 (QTL of varying size) for a backcross with 250 progeny with such a significance threshold. The analytical threshold was obtained assuming that the residual log likelihood ratio test (used to assess if a chromosome contained undetected QTL) follows a mixture of chi-squared distributions (Self and Liang 1987) but with a significance level set according to the Bonferroni correction.

For analyses of Scenario 1 data, we found the type 1 error halved when Bonferroni was used instead of permutation, testifying to the well-known conservativeness of Bonferroni adjusted tests. This conservative behavior then reduced power. For analyses of Scenario 2 data, there were 45, 38, 16, and 6% drops in power on chromosomes 1, 2, 3, and 4, respectively, when Bonferroni instead of permutation was used. However, even with this loss in power, our approach remains more powerful than CIM. This is further supported by the analysis of the wheat data using the Bonferroni threshold. In this case, all four confirmed QTL were detected at the 0.01 level whereas CIM only detected three. By using an analytical threshold, our approach was substantially faster than CIM which is dependent upon permutation to determine the significance of its findings.

In conclusion, we have shown the "look before you leap'' philosophy to be a powerful and effective strategy for mapping QTL. Work is under way to extend its application to more advanced experimental crosses such as those aimed at developing 2^n-way recombinant inbred lines. It would also be valuable to extend the model formulation to allow for detection of epistatic effects. We aim to capitalize on its linear mixed model framework in joint spatial and marker analysis of sugarcane data. Results are preliminary but extremely encouraging. We are confident that DLMapping will become a valuable computational tool in the search for QTL.

References

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

Bogdan M, Ghosh JK, Doerge RW (2004) Modifying the Schwarz Bayesian information criterion to locate multiple interacting quantitative trait loci. Genetics 167:989–999

Broman KW, Speed TP (2002) A model selection approach for the identification of quantitative trait loci in experimental crosses. J R Stat Soc B 64:641–656

Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889–890

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

Daw EW, Thompson EA, Wijsman EM (2000) Bias in multipoint linkage analysis arising from map misspecification. Genet Epid 19:366–380

Eckermann PJ, Verbyla AP, Cullis BR, Thompson R (2001) The analysis of quantitative traits in wheat mapping populations. Aust J Agric Res 52:1195–1206

George AW, Visscher PM, Haley CS (2000) Mapping quantitative trait loci in complex pedigrees: a two step variance component approach. Genetics 156:2081–2092

Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2006) ASReml User Guide Release 2.0. VSN International Ltd., Hemel Hempstead, HP11ESJ, UK

Jansen RC (1993) Interval mapping of multiple quantitative trait loci. Genetics 135:205–211

Kao CH, Zeng ZB, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. Genetics 152:1203–1216

Lee SH, Van der Werf JHJ (2006) Simultaneous fine mapping of multiple closely linked quantitative trait loci using combined linkage disequilibrium and linkage with a general pedigree. Genetics 173:2329–2337

Li H, Ye G, Wang J (2007) A modified algorithm for the improvement of composite interval mapping. Genetics 175:361–374

Li H, Ribaut JM, Li Z, Wang J (2008) Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. Theor Appl Genet 116:243–260

Manichaikul A, Moon JY, Sen S, Yandell BS, Broman KW (2009) A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. Genetics 181:1077–1086

Mares DJ, Campbell AW (2001) Mapping components of flour and noodle colour in Australian wheat. Aust J Agric Res 52:1297–1309

Martinez O, Curnow RN (1992) Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor Appl Genet 85:480–488

Piepho HP (2000) A mixed-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. Genetics 156:2043–2050

Pinheiro J, Bates D, DebRoy S, Sarkar D, The R Core Team (2008) nlme: linear and nonlinear mixed effects models. R package version 3.1-88

Raman R, Raman H, Martin P (2007) Functional gene markers for polyphenol oxidase locus in bread wheat (triticum aestivum l). Mol Breed 19:315–328

Self SG, Liang K (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc 82:605–610

Smith AB, Lim P, Cullis BR (2006) The design and analysis of multi-phase plant breeding experiments. J Agric Sci 144:393–409

R Development Core Team (2008) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org, ISBN 3-900051-07-0

Verbyla AP, Cullis BR, Thompson R (2007) The analysis of QTL by simultaneous use of the full linkage map. Theor Appl Genet 116:95–111

Wang S, Basten CJ, Zeng ZB (2005) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC

Whittaker CJ, Curnow RN, Haley CS, Thompson R (1995) Using marker-maps in marker-assisted selection. Genet Res 66:255–265

Yi N, Shriner D (2008) Advances in Bayesian multiple quantitative trait loci mapping in experimental crosses. Heredity 100:240–252

Yi N, Yandell BS, Churchill GA, Allison DB, Eisen EJ, Pomp D (2005) Bayesian model selection for genome-wide epistatic quantitaive trait loci analysis. Genetics 170:1333–1344

Yu J, Pressoir P, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zeng ZB (1993) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc Natl Acad Sci USA 90:10972–10976

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136:1457–1468

Zhang W, Dubcovsky J (2008) Association between allelic variation at the phytoene synthase 1 gene and yellow pigment content in the wheat grain. Theor Appl Genet 116:635–645