Theor Appl Genet (2013) 126:999–1009
DOI 10.1007/s00122-012-2032-2

ORIGINAL PAPER

Genomewide markers as cofactors for precision mapping of quantitative trait loci

R. Bernardo

Received: 28 March 2012 / Accepted: 6 December 2012 / Published online: 28 December 2012

Abstract In composite interval mapping of quantitative trait loci (QTL), subsets of background markers are used to account for the effects of QTL outside the marker interval being tested. Here, I propose a QTL mapping approach (called G model) that utilizes genomewide markers as cofactors. The G model involves backward elimination on a given chromosome after correcting for genomewide marker effects, calculated under a random effects model, at all the other chromosomes. I simulated a trait controlled by 15 or 30 QTL, mapping populations of $N = 96$, 192, and 384 recombinant inbreds, and $N_M = 192$ and 384 evenly spaced markers. In the C model, which utilized subsets of background markers, the number of QTL detected and the number of false positives depended on the number of cofactors used, with five cofactors being too few with $N = 384$ and 20–40 cofactors being too many with $N = 96$. A window size of 0 cM for excluding cofactors maintained the number of true QTL detected while decreasing the number of false positives. The number of true QTL detected was generally higher with the G model than with the C model, and the G model led to good control of the type I error rate in simulations where the null hypothesis of no marker–QTL linkage was true. Overall, the results indicated that the G model is useful in QTL mapping because it is less subjective and has equal, if not better, performance when compared with the traditional approach of using subsets of markers to account for background QTL.

Communicated by F. van Eeuwijk.

R. Bernardo (✉)
Department of Agronomy and Plant Genetics, University of Minnesota, 411 Borlaug Hall, 1991 Upper Buford Circle, Saint Paul, MN 55108, USA
e-mail: bernardo@umn.edu

Introduction

Composite interval mapping (CIM; Zeng 1993, 1994; Jansen 1992, 1993) is a very common method, if not the most common method, used for mapping quantitative trait loci (QTL) in plants. Although newer QTL mapping methods, such as inclusive CIM (Li et al. 2007), multiple interval mapping (Kao et al. 1999), and Bayesian LASSO (Yi and Xu 2008) have been proposed, CIM remains a popular and standard method because of its implementation in software, such as QTL Cartographer (Wang et al. 2006), PLABQTL (Utz and Melchinger 1996), R/qtl (Broman et al. 2003), and GenStat (http//:www.vsni.co.uk). In simple interval mapping, the presence of a QTL at specific positions between two adjacent markers is evaluated by maximum likelihood or by regression-based tests (Lander and Botstein 1989; Haley and Knott 1992). In CIM, simple interval mapping is combined with multiple regression given a set of background markers. These background markers help account for the effects of QTL outside the marker interval being tested. Results have shown that such use of background markers as cofactors in CIM leads to more precise estimates of QTL location compared with simple interval mapping (Jansen and Stam 1994; Zeng 1994; Liu 1998; Cornforth and Long 2003).

No definitive rules, however, have been established for determining the appropriate number of background markers to use in CIM. Having too few markers as cofactors in CIM could lead to an insufficient amount of background effects being captured, whereas having too many markers could lead to model overfitting. Jansen and Stam (1994)

proposed that the number of cofactors should not exceed twice the square root of the size of the mapping population. The standard method (Method 6) for selecting cofactors in QTL Cartographer (Wang et al. 2006) first ranks all markers according to their effects by stepwise regression. The user then chooses a specific number of background markers to include and a window size for excluding cofactors. To illustrate, QTL Cartographer has a default number of background markers equal to five and a window size of 10 cM. If these default values are used, the software then chooses as cofactors the five most important markers from stepwise regression, with the condition that each of the five background markers is at least 10 cM away from the two markers that flank the interval being tested. Different numbers of background markers and window sizes for excluding cofactors could lead to different results.

Developments since the 1990s have led to methods for genomewide prediction of performance for quantitative traits (Meuwissen et al. 2001). Unlike regression-based procedures that predict performance via subsets of markers with fixed effects, genomewide prediction does not involve significance tests, but instead uses all markers in a random effects model to predict performance. Genomewide prediction would make the unresolved issue of the appropriate number of background markers in CIM irrelevant. The use of genomewide markers to correct for background QTL effects, via partial least squares, has been previously proposed by van Eeuwijk et al. (2000) and applied to empirical data sets by van Eeuwijk et al. (2002), Bjørnstad et al. (2004), and Vargas et al. (2006). However, rigorous comparisons of genomewide markers versus subsets of significant markers to correct for background QTL effects are still lacking.

In this paper, I propose and evaluate a mapping procedure that utilizes genomewide markers as random cofactors for finding QTL and estimating QTL effects. Background marker effects are obtained by ridge-regression best linear unbiased prediction (RR-BLUP), which has emerged as a simple, fast, and effective procedure for obtaining genomewide predictions in plants (Lorenzana and Bernardo 2009; Heffner et al. 2009; Lorenz et al. 2011; Asoro et al. 2011; Guo et al. 2012). I then present simulation results comparing this new procedure with the traditional procedure of using subsets of background markers in QTL mapping. For ease of comparison, QTL detection in the simulations relied on determining whether or not a QTL was adjacent to a marker declared as significant, rather than testing for a QTL at different points between two markers in an interval mapping approach. For brevity, controlling background effects through genomewide markers was called the G model, whereas controlling background effects through subsets of markers was called the C (for composite) model.

Materials and methods

Mapping population, genetic models, and phenotypic values

The genetic models in this study were largely based on those in previous studies (Bernardo 2004; Bernardo and Yu 2007). Each simulation experiment comprised a combination of genetic model, size of mapping population, and number of markers. The simulation experiments were repeated 1,000 times. Each repeat differed in the location of QTL and in the genotypes, genotypic values, and phenotypic values of the individuals. I wrote a Fortran program to conduct the simulations and data analysis.

A simulated F_1 generation, formed by crossing two parental inbreds, was selfed eight times to form an F_2-derived mapping population of $N = 96$, 192, or 384 recombinant inbreds. The two parental inbreds differed at $N_M = 192$ or 384 codominant marker loci. The sizes of the ten chromosomes (ranging from 128 to 241 cM) and of the entire genome (1,749 cM) corresponded to those in a published maize (*Zea mays* L.) linkage map (Senior et al. 1996). The genome was divided into N_M bins that were $1{,}749/N_M$ cM in size. A marker was located at the midpoint of each bin.

The premise was to map major QTL for a less complex trait that would tend to have several QTL with large effects (Bernardo 2008), and the trait was controlled by $L = 15$ QTL with an entry-mean heritability of $h^2 = 0.80$ or by $L = 30$ QTL with an h^2 of 0.70. The first parent had the favorable allele at odd-numbered QTL and the less favorable allele at even-numbered QTL. The L QTL were randomly located among the ten chromosomes without considering the position of any marker, but with no two QTL being located in the same position. The complementary QTL alleles from the two parents and the random locations of the QTL resulted in random linkage phases (coupling or linkage) between any linked pair of QTL. The sizes of QTL effects followed a geometric series (Lande and Thompson 1990). At the kth QTL, the genotypic values were a^k for the favorable homozygote, 0 for the heterozygote, and $-a^k$ for the less favorable homozygote, where $a = (1 - L)/(1 + L)$ as specified by Lande and Thompson (1990). Dominance was therefore absent and epistasis was likewise absent. The genotypic value of a recombinant inbred was equal to the sum of its genotypic values across all L QTL.

Phenotypic values were simulated as follows. First, genetic variance (V_G) in a given repeat of a simulation experiment was calculated as the variance among genotypic values (i.e., $h^2 = 1.0$) of 5,000 recombinant inbreds. Phenotypic values were then simulated for the $N = 96$, 192, or 384 recombinant inbreds evaluated in eight

environments with one replication in each environment. Phenotypic values were obtained by adding a random nongenetic effect to the genotypic value of each recombinant inbred in each environment. The nongenetic effects were normally and independently distributed with a mean of zero and a variance of V_E. The V_E was scaled to achieve a target entry-mean h^2 of 0.70 (for $L = 30$ QTL) or 0.80 (for $L = 15$ QTL). Estimates of V_E and V_G were then obtained from an analysis of variance of the simulated phenotypic data. These variance component estimates were used in subsequent G model analysis. The phenotypic data for QTL analysis comprised the means across environments of the N recombinant inbreds.

C model

In the C model, the number of background markers used as cofactors was $N_{CF} = 5$, 10, 20, and 40 and the window sizes were 0, 5, and 10 cM. For convenience, different C models were denoted by the number of background markers used, e.g., C-5 refers to a C model with $N_{CF} = 5$. In accordance with a procedure implemented in QTL Cartographer, stepwise regression by forward selection was used to rank the top 45 markers, which was the maximum number of markers needed with $N_{CF} = 40$ and the densest marker distribution ($N_M = 384$), and if the most important markers are all adjacent but need to be excluded according to the maximum window size of 10 cM. Assume these 45 markers are denoted by R1, R2, R3 … R45, where the number after R indicates the ranking of the marker according to priority as a cofactor. The marker with the lowest p value based on the single marker linear regression was first identified and designated as marker R1. Next, the marker with the lowest p value, given that marker R1 was also in the model, was identified and designated as marker R2. Marker R3 was the marker with the lowest p value given that markers R1 and R2 were in the multiple regression model. The procedure continued until R45 was identified.

The N_{CF} markers were then chosen from markers R1 to R45, excluding any of the R markers within the window size from the marker being tested for its association with a QTL. Suppose the QTL Cartographer defaults of $N_{CF} = 5$ markers and a window size of 10 cM were used. Marker M_1, on chromosome 1 at the 5 cM position, was being tested for its association with a QTL. Further suppose that marker R1 was at the 30 cM position on chromosome 1, marker R2 was at the 10 cM position on chromosome 1, and markers R3–R6 were on other chromosomes and were unlinked to M_1. In this situation, markers R1, R3, R4, R5, and R6 were chosen as the $N_{CF} = 5$ cofactors for marker M_1, with marker R2 being excluded because it was less than 10 cM from marker M_1.

For the kth out of $N_M = 192$ or 384 markers, multiple regression was conducted with the kth marker and its set of N_{CF} cofactors in the model. Significance of the kth marker was tested at both $p = 0.0001$ and 0.00001 based on a t test of the regression coefficient for the kth marker. A false positive was declared whenever the kth marker had a significant regression coefficient, but no QTL was present in either of the marker's adjacent intervals (Doerge et al. 1994; Whittaker et al. 1996). A true QTL was declared to have been detected whenever a QTL had a significant left flanking marker, a significant right flanking marker, or had both flanking markers as significant. The number of markers with significant effects was recorded.

G model

Genomewide marker effects were calculated by RR-BLUP as described by Meuwissen et al. (2001) and Bernardo and Yu (2007). Marker effects were assumed random and the variance of each marker effect in RR-BLUP was equal to the estimate of V_G (i.e., from analysis of variance of the simulated phenotypic data) divided by N_M.

Procedures for QTL analysis in the G model comprised two steps. In the first step, multiple regression by backward elimination was performed on a chromosome by chromosome basis (Bernardo 2004), after having corrected for genomewide marker effects across all the other chromosomes not currently being analyzed for QTL. Suppose 53 out of $N_M = 384$ markers were found on chromosome 1. To detect QTL on chromosome 1, the phenotypic data were first adjusted for the RR-BLUP genomewide marker effects of the $(384 - 53) = 331$ markers found on chromosomes 2–9. With such per-chromosome adjusted data as the dependent variable, backward elimination was used to allow the examination of the full model (i.e., all 53 markers) for chromosome 1. The significance level for retaining a marker in the model was $p = 0.0001$ or 0.00001. These procedures were then repeated for each chromosome. In the second step (which was for obtaining final estimates of marker effects but not for retesting the significance of effects), multiple regression coefficients were obtained by jointly analyzing all the markers found significant in the per-chromosome analysis. Unadjusted phenotypic data were used as the dependent variable in this second step.

The G model therefore combined backward elimination for a given chromosome and adjustment for genomewide marker effects at the remaining chromosomes. To facilitate comparisons with the C model, which did not involve backward elimination, a G model involving single-marker analysis (denoted by G-SM) was also used. For the kth marker, the G-SM analysis involved adjustment of the phenotypic data for the genomewide effects at the $(N_M - 1)$ markers other than the kth marker, and single-marker linear

regression with the adjusted data as the dependent variable. Unlike the G model, the G-SM model therefore involved adjustment of the phenotypic data for all other markers on the same chromosome as the marker being tested for its significance. A final estimation of marker effects (but not retesting their significance) was done by multiple regression with all the markers found significant in the single-marker analysis. For the G model and G-SM model, the numbers of false positives, true QTL detected, and significant markers were determined in the same manner as for the C model.

Data analysis

For each experiment and for each QTL analysis model, the numbers of false positives, true QTL detected, and significant markers were averaged across the 1,000 repeats. Variances across the 1,000 repeats of an experiment were used in z tests of the significance ($p = 0.05$) of differences among the means for the C models, G model, and G-SM model. To have a rough but global estimate of precision, approximate least significant differences (LSD) were obtained using the mean variance across the C, G, and G-SM models.

Control of type I error rate

Simulations were conducted to determine the extent to which the G, G-SM, and C models controlled the type I error rate. The simulations were conducted for the 15 QTL, $h^2 = 0.80$ genetic model with the restrictions that (1) all 15 QTL were located on chromosome 1 and (2) p values associated with tests of the null hypothesis in the G, G-SM, and C models were obtained for markers on chromosomes 2–10. Such simulations therefore satisfied the conditions that the null hypothesis of no linkage with QTL was true for all the markers on chromosomes 2–10, and that V_G was greater than zero (i.e., $V_G = 0$ if no QTL were simulated) for RR-BLUP in the G and G-SM models. Simulations were conducted for the same values of N, N_M, and N_{CF} as previously indicated. Given that the p values are expected to follow a uniform distribution if the null hypothesis is true (Murdock et al. 2008), the Kolmogorov–Smirnov test ($p = 0.01$) was used to determine if the empirical p values followed a uniform distribution. The simulations were repeated 200 times and the frequency of non-uniform distributions of p values was determined.

Results

In the C models, which utilized subsets of N_{CF} background markers, the ratio between the number of true QTL detected (N_{TQ}) and the number of false positives (N_{FP}) was highly dependent on the window size used. Consider a trait controlled by $L = 15$ QTL with an entry-mean heritability of $h^2 = 0.80$, a number of markers of $N_M = 384$, a population size of $N = 384$, and a significance level of $p = 0.00001$ for declaring a significant QTL. When the window size was 10 cM, N_{FP} (which ranged from 16.9 to 18.1) exceeded N_{TQ} (which ranged from 8.9 to 10.9) across N_{CF} values of 5–40 (results not shown). When the window size was reduced to 5 cM, N_{TQ} ranged from 8.5 to 10.5, whereas N_{FP} was drastically reduced, ranging from 3.6 to 7.6. When the window size was 0 cM (i.e., background markers were included as cofactors regardless of their proximity to the marker being tested for its association with a QTL), N_{TQ} ranged from 8.4 to 9.1, whereas N_{FP} was further reduced to 0.4–5.5. The same general trend was observed for other population sizes, for a less stringent significance level of $p = 0.0001$, and for a trait controlled by $L = 30$ QTL and with $h^2 = 0.70$. Given the inferiority of the C models with window sizes greater than 0 cM, only the results from the C models with a window size of 0 cM are presented in the rest of this article (including Tables 1, 2; Fig. 1).

In the C models, the appropriate N_{CF} depended on the size of the mapping population, with fewer background markers being more suitable for smaller N and more background markers being more suitable for larger N. When the mapping population was small ($N = 96$), the C-40 model (i.e., with $N_{CF} = 40$) led to N_{FP} values that exceeded N_{TQ} regardless of N_M and of the number of QTL controlling the trait (Table 1). With $N = 96$ and $N_M = 384$, the C-20 model led to N_{FP} values that often approached or exceeded N_{TQ}. In contrast, when the mapping population was large ($N = 384$), the C-5 model led to N_{FP} values significantly higher than those with the C-10, C-20, and C-40 models, but with no gain in N_{TQ}.

For a given simulation experiment, a less stringent significance level ($p = 0.0001$ instead of 0.00001) led to higher values of N_{TQ}, N_{FP}, and number of significant markers (Table 1). The number of QTL controlling the trait had little effect on both N_{TQ} and N_{FP}: for a given combination of N, N_M, and model for QTL analysis, differences in L led to only small differences in both N_{TQ} and N_{FP}. Consider a moderate population size of $N = 192$ and a moderate number of markers of $N_M = 192$. With the G model and a significance level of $p = 0.00001$, N_{TQ} was 8.2 with $L = 15$ QTL controlling the trait and was 8.0 with $L = 30$ QTL (Table 1). Likewise, N_{FP} was 1.4 with $L = 15$ and 0.9 with $L = 30$. Given the consistent effects of the significance level used and the minimal effects of the number of QTL on both N_{TQ} and N_{FP}, the results for each of the C, G, and G-SM models were pooled across significance levels and L, and are presented in Fig. 1.

Table 1 Numbers of true QTL detected (N_{TQ}), false positives (N_{FP}), and significant markers for the G, G-SM, and C models (with a window size of 0 cM) for detecting QTL

Model	Significance level	15 QTL, $h^2 = 0.80$			30 QTL, $h^2 = 0.70$		
		True QTL	False positives	Significant markers	True QTL	False positives	Significant markers
$N = 96$, $N_M = 192$							
G	0.0001	6.5	3.5	9.7	5.2	1.6	6.3
G	0.00001	5.6	1.8	7.2	3.8	0.8	4.2
G-SM	0.0001	4.9	1.8	7.5	2.1	0.2	2.2
G-SM	0.00001	3.6	0.7	4.7	1.0	0.1	1.0
C-5	0.0001	3.7	0.3	3.8	2.4	0.4	2.5
C-5	0.00001	2.9	0.2	2.9	1.5	0.2	1.5
C-10	0.0001	4.6	0.7	5.1	3.4	0.8	3.8
C-10	0.00001	3.7	0.3	3.9	2.4	0.4	2.5
C-20	0.0001	5.7	2.6	8.1	5.2	2.6	7.3
C-20	0.00001	4.8	1.4	6.0	3.9	1.5	5.0
C-40	0.0001	7.3	12.4	19.7	8.6	10.8	19.0
C-40	0.00001	6.6	9.1	15.6	7.2	8.0	14.8
	LSD (0.05)	0.14	0.17	0.23	0.16	0.14	0.21
$N = 192$, $N_M = 192$							
G	0.0001	9.1	2.8	11.9	9.7	1.7	10.5
G	0.00001	8.2	1.4	9.5	8.0	0.9	8.1
G-SM	0.0001	8.1	3.3	13.1	6.3	0.6	6.7
G-SM	0.00001	6.9	1.4	9.5	4.2	0.2	4.1
C-5	0.0001	6.1	0.5	6.6	4.8	0.4	4.7
C-5	0.00001	5.3	0.2	5.3	3.7	0.2	3.4
C-10	0.0001	7.0	0.3	7.0	6.2	0.5	6.0
C-10	0.00001	6.1	0.1	5.9	4.7	0.2	4.4
C-20	0.0001	7.3	0.7	7.8	7.1	0.8	7.2
C-20	0.00001	6.2	0.3	6.3	5.5	0.4	5.3
C-40	0.0001	7.4	2.6	9.9	7.8	2.3	9.4
C-40	0.00001	6.3	1.1	7.2	6.0	1.0	6.4
	LSD (0.05)	0.13	0.11	0.19	0.19	0.08	0.18
$N = 384$, $N_M = 192$							
G	0.0001	11.1	1.3	12.9	14.2	1.2	14.3
G	0.00001	10.4	0.6	11.1	12.4	0.6	12.0
G-SM	0.0001	10.5	4.1	17.1	12.2	1.1	13.6
G-SM	0.00001	9.5	1.8	13.2	9.9	0.4	10.1
C-5	0.0001	8.8	2.0	12.2	9.2	1.0	10.4
C-5	0.00001	7.8	0.9	9.5	7.4	0.4	7.4
C-10	0.0001	9.5	0.3	9.8	10.3	0.4	9.8
C-10	0.00001	8.7	0.1	8.6	8.9	0.2	8.2
C-20	0.0001	9.6	0.3	9.9	11.3	0.4	10.6
C-20	0.00001	8.5	0.1	8.4	9.4	0.2	8.6
C-40	0.0001	9.0	0.6	9.6	10.5	0.6	10.2
C-40	0.00001	7.8	0.2	7.9	8.4	0.2	7.7
	LSD (0.05)	0.13	0.11	0.19	0.19	0.07	0.19
$N = 96$, $N_M = 384$							
G	0.0001	5.5	5.3	10.7	4.2	3.2	7.1
G	0.00001	4.8	3.2	7.9	3.1	1.8	4.7
G-SM	0.0001	2.6	1.0	4.4	0.6	0.1	0.7

Table 1 continued

Model	Significance level	15 QTL, $h^2 = 0.80$			30 QTL, $h^2 = 0.70$		
		True QTL	False positives	Significant markers	True QTL	False positives	Significant markers
G-SM	0.00001	1.6	0.4	2.4	0.2	0.0	0.2
C-5	0.0001	3.5	1.0	4.5	2.1	1.0	3.0
C-5	0.00001	2.9	0.6	3.4	1.5	0.6	1.9
C-10	0.0001	4.5	1.6	6.0	3.1	2.0	4.9
C-10	0.00001	3.9	1.0	4.8	2.4	1.2	3.4
C-20	0.0001	5.5	5.0	10.4	4.5	5.7	10.0
C-20	0.00001	5.0	3.3	8.2	3.8	3.9	7.5
C-40	0.0001	6.8	19.8	26.8	7.2	19.6	26.8
C-40	0.00001	6.5	16.6	23.2	6.5	16.6	23.1
	LSD (0.05)	0.14	0.18	0.21	0.14	0.17	0.19
$N = 192$, $N_M = 384$							
G	0.0001	8.5	5.1	13.5	8.8	3.7	12.0
G	0.00001	7.9	2.9	10.7	7.4	2.3	9.3
G-SM	0.0001	5.7	2.3	10.2	2.6	0.3	3.0
G-SM	0.00001	4.6	1.1	7.1	1.3	0.1	1.4
C-5	0.0001	6.3	2.3	9.3	4.5	1.3	5.7
C-5	0.00001	5.4	1.0	6.7	3.5	0.7	4.0
C-10	0.0001	7.2	0.9	7.9	5.8	1.5	6.9
C-10	0.00001	6.4	0.6	6.8	4.7	0.9	5.3
C-20	0.0001	7.7	1.8	9.3	7.2	2.5	9.3
C-20	0.00001	6.7	0.9	7.5	6.0	1.4	7.0
C-40	0.0001	8.1	6.8	14.8	8.5	6.6	14.7
C-40	0.00001	7.2	3.8	10.9	7.1	3.8	10.6
	LSD (0.05)	0.13	0.16	0.21	0.17	0.13	0.19
$N = 384$, $N_M = 384$							
G	0.0001	11.0	3.6	14.8	13.6	3.2	16.3
G	0.00001	10.5	1.9	12.5	12.2	2.0	13.7
G-SM	0.0001	9.2	4.5	17.5	7.7	0.9	9.8
G-SM	0.00001	8.1	2.2	13.3	5.6	0.3	6.3
C-5	0.0001	9.3	9.6	21.7	9.4	5.7	16.9
C-5	0.00001	8.4	5.5	15.9	7.5	2.4	10.6
C-10	0.0001	9.9	1.1	11.1	10.0	1.7	11.5
C-10	0.00001	9.1	0.5	9.5	8.7	1.0	9.3
C-20	0.0001	10.0	0.9	10.8	11.4	1.4	12.2
C-20	0.00001	9.0	0.4	9.3	9.9	0.8	10.2
C-40	0.0001	9.7	2.1	11.8	11.4	2.3	13.2
C-40	0.00001	8.6	0.8	9.4	9.5	1.0	10.1
	LSD (0.05)	0.13	0.20	0.29	0.18	0.15	0.26

In the C models, the numeral (e.g., C-5) indicates the number of background markers used as cofactors

In Fig. 1, the results for C-40 with $N = 96$ were excluded because, as indicated earlier, the C-40 model was inappropriate with a small mapping population. The results for C-5 with $N = 384$, $N_M = 384$, $p = 0.0001$, and $L = 5$ led to an N_{FP} value (9.6, Table 1) that exceeded the N_{TQ} value (9.3) and the results for this simulation experiment were likewise excluded from Fig. 1.

The N_{TQ} values increased as N increased (Fig. 1). At a given N, an increase in N_{TQ} was generally accompanied by an increase in N_{FP}. Increasing the number of markers from

Table 2 Control of the type I error rate in simulations where the null hypothesis of no marker-QTL linkage was true for the G, G-SM, and C models (with a window size of 0 cM) for detecting QTL

N and N_M	Percentage of non-uniform distributions of p values[a]					
	G	G-SM	C-5	C-10	C-20	C-40
$N = 96$, $N_M = 192$	2	88	4	10	11	68
$N = 192$, $N_M = 192$	2	100	7	5	17	18
$N = 384$, $N_M = 192$	0	100	7	5	16	20
$N = 96$, $N_M = 384$	42	15	23	31	41	13
$N = 192$, $N_M = 384$	3	41	27	29	59	65
$N = 384$, $N_M = 384$	2	99	22	25	49	86

In the C models, the numeral (e.g., C-5) indicates the number of background markers used as cofactors. The results are for a 15 QTL, $h^2 = 0.80$ model and for a total of 200 repeats

[a] Based on a Kolmogorov–Smirnov test for a uniform distribution ($p = 0.01$)

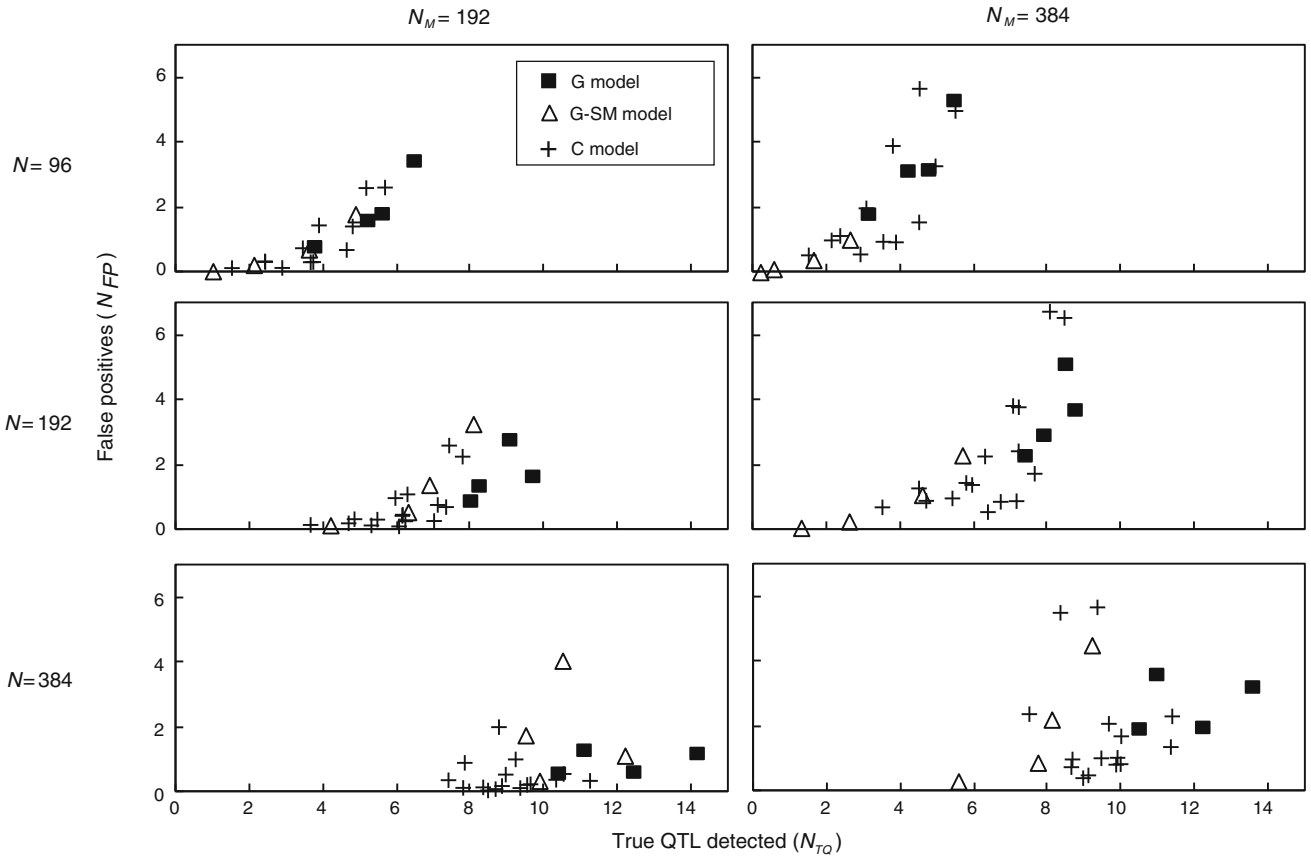

Fig. 1 Numbers of true QTL detected and false positives with different C models (*plus sign*), the G model (*solid squares*), and the G-SM model (*open triangles*). The results with the G model and G-SM model are for different significance levels ($p = 0.0001$ and 0.00001) and genetic models (15 QTL with $h^2 = 0.80$, and 30 QTL with $h^2 = 0.70$). The results with the C model (window size of 0 cM only) are for the two significance levels, two genetic models, and four numbers of cofactors ($N_{CF} = 5$, 10, 20, and 40). Specific data points that were not plotted are noted in the text

$N_M = 192$ to $N_M = 384$ led to an increase in N_{FP}. As indicated later in the "Discussion", this result needs to be interpreted with caution given how a true QTL and a false positive were defined in this study. The N_{TQ} values were generally highest with the G model. The N_{TQ} values were generally lower with the G-SM model than with the C models when $N_M = 384$ markers were used, but not when $N_M = 192$ markers were used.

There were instances wherein the use of a less stringent significance level in the C models ($p = 0.0001$) and a more

stringent significance level in the G model ($p = 0.00001$) led to equal N_{FP} values but unequal N_{TQ} values. Consider the simulation experiment with $N = 96$, $N_M = 192$, and $L = 30$. The N_{FP} was 0.8 when a significance level of $p = 0.00001$ was used in the G model or when a significance level of $p = 0.0001$ was used in the C-10 model (Table 1). Although the N_{FP} values were equal, the corresponding N_{TQ} was slightly but significantly higher in the G model (3.8) than in the C-10 model (3.4). Likewise, with $N = 384$, $N_M = 192$, and both numbers of QTL studied, the G model with $p = 0.00001$ led to the same N_{FP} (0.6) but significantly higher N_{TQ} when compared with the C-40 model with $p = 0.0001$ (Table 1).

The N_{FP} values were generally lowest with the G-SM model. Furthermore, the N_{FP} values in Fig. 1 were generally lower with the C models than with the G model, with the following notable exceptions: C-40 model with $N = 192$, $N_M = 384$, $p = 0.0001$, and $L = 15$; C-40 model with $N = 192$, $N_M = 384$, $p = 0.0001$, and $L = 30$; C-5 model with $N = 384$, $N_M = 384$, $p = 0.00001$, and $L = 15$; and C-5 model with $N = 384$, $N_M = 384$, $p = 0.0001$, and $L = 30$.

The variances of N_{TQ} and N_{FP} were of comparable magnitude across the C, G, and G-SM models (results not shown). For the simulation experiments with the fewest QTL ($L = 15$), fewest markers ($N_M = 192$), and fewest recombinant inbreds ($N = 96$), the variances of N_{TQ} across the 1,000 repeats ranged from 1.6 to 3.3 for the C models, 2.8–3.1 for the G model, and 2.9–3.5 for the G-SM model. The corresponding variances of N_{FP} ranged from 0.2 to 13.3 for the C models, 2.6–5.6 for the G model, and 1.2–3.3 for the G-SM model.

Control of the type I error rate, as evidenced by a uniform distribution of empirical p values when the null hypothesis was true, was best with the G model, intermediate with the C models, and poorest with the G-SM model. With the G model, less than 3 % of the repeats of the simulation experiments had non-uniform p values ($p = 0.01$), except when a small mapping population ($N = 96$) and a large number of markers ($N_M = 384$) were used (Table 2). In the latter situation, the frequency of non-uniform distributions of p values increased to 42 %. The frequency of non-uniform distributions of p values ranged from 15 to 100 % with the G-SM model and from 4 to 86 % with the C models. With the C models, the frequency of non-uniform distributions of p values was lower with $N_M = 192$ markers than with $N_M = 384$ markers.

Discussion

This study compared the use of genomewide markers to account for background effects (G model) versus the traditional procedure of using subsets of markers to account for background effects (C model) in mapping QTL. The G model is less subjective than the C model, and the results indicated that the G model is also at least as powerful, if not more powerful, than the C model.

The results underscored the dependence of C models, such as CIM (Zeng 1993, 1994; Jansen 1992, 1993) on the investigator's choice of number of cofactors (N_{CF}) and window size for excluding cofactors. According to the rule proposed by Jansen and Stam (1994), the N_{CF} values should be $N_{CF} \leq 19$ for a mapping population of $N = 96$ recombinant inbreds, $N_{CF} \leq 27$ for $N = 192$, and $N_{CF} \leq 39$ for $N = 384$. The simulation results indeed indicated that the use of $N_{CF} = 20$ or 40 cofactors when the population size was $N = 96$ led to the number of false positives exceeding the number of true QTL detected. On the other hand, the results also indicated that the QTL Cartographer (Wang et al. 2006) default of $N_{CF} = 5$ was too small when the mapping population was large ($N = 384$) and given the $L = 15$ or 30 QTL controlling the trait in this study.

Furthermore, the results did not support using the default window size of 10 cM in QTL Cartographer. The number of true QTL detected was maintained and the number of false positives decreased when the window size was reduced to 0 cM. This result was consistent with recommendations by Zeng (1994) and by Lynch and Walsh (1998, p. 465) to use a window size of 0 cM in CIM. Consider four adjacent markers in the following order: M_1–M_2–M_3–M_4. In testing for the presence of a QTL in the M_2–M_3 interval, including M_1 as a possible cofactor by specifying a window size of 0 cM in CIM is expected to account for QTL to the left of M_1. Likewise, including M_4 as a possible cofactor by specifying a window size of 0 cM is expected to account for QTL to the right of M_4. However, a QTL in the M_1–M_2 interval or in the M_3–M_4 interval would not be isolated (Whittaker et al. 1996) and the effects of QTL in these two flanking intervals cannot be cleanly separated from the effect of QTL in the M_2–M_3 interval being tested (Zeng 1994; Whittaker et al. 1996).

In contrast to the C model, the G model circumvents the need to choose subsets of background markers as cofactors and the need to choose a window size for excluding markers as cofactors. As shown in Fig. 1, the number of true QTL detected was generally higher with the G model than with the C models across different population sizes and numbers of markers. This slight increase in power to detect QTL was accompanied by a slight increase in the number of false positives, particularly when the mapping population was small ($N = 96$). These results suggest that it might be possible to constrain the G model and the C models to have similar false-positive rates by specifying different significance levels for the two models. However, the larger issue is that if the use of the same significance

level in the C models and the G model leads to different numbers of false positives for the same data set, the two approaches must then differ in their adherence to the significance level specified by the investigator.

In particular, when the null hypothesis is true, the resulting p values are expected to follow a uniform distribution (Murdock et al. 2008). A significant deviation of the empirical p values from a uniform distribution then indicates a failure of the statistical test to adhere to the specified significance level. Simulations that constrained the null hypothesis of no marker-QTL linkage to be true indicated that the G model generally led to the best control of the type I error rate (Table 2). However, the G model failed to adhere to the specified significance level when the mapping population was small ($N = 96$) and the number of markers was large ($N_M = 384$). The accuracy of genome-wide predictions increases with larger N (Lorenzana and Bernardo 2009; Heffner et al. 2011; Guo et al. 2012) and, conversely, poor prediction of background effects due to a small mapping population is expected to lead to less reliable results. When many markers are present on each chromosome due to a large N_M, a small mapping population would lead to few degrees of freedom in per-chromosome backwards elimination in the G model. The high frequency (42 %) of non-uniform distributions of p values with the G model for $N = 96$ and $N_M = 384$ was therefore not unexpected.

As expected, a more stringent significance level ($p = 0.00001$ instead of $p = 0.0001$) reduced both the number of true QTL detected and the number of false positives regardless of the model for detecting QTL. Although the statistical properties of CIM have been well studied (Zeng 1993, 1994; Jansen 1992, 1993), the simulations indicated that CIM may not always adhere well to the specified type I error rate if all the QTL are found only on one chromosome. This result was due to the joint effects of (1) specifying a fixed number of cofactors rather than a significance threshold for their inclusion in CIM and (2) having only 1–2 recombination events per chromosome in a recombinant inbred (Smith et al. 2008). Suppose a mapping population of $N = 96$ has been genotyped with $N_M = 192$ markers, and all 15 QTL are found on chromosome 1. In this situation, only a few markers on chromosome 1 should suffice as cofactors in CIM, and the frequency of non-uniform distributions of p values was only 4 % with the C-5 model for $N = 96$ and $N_M = 192$ (Table 2). However, increasing the number of cofactors would lead to choosing some cofactors unlinked to QTL from the other chromosomes, or to a high level of collinearity among the multiple cofactors on chromosome 1. Either situation would contribute to a high frequency of non-uniform distributions of p values (e.g., 68 % with the C-40 model for $N = 96$ and $N_M = 384$; Table 2).

Extensions of CIM such as inclusive CIM (Li et al. 2007) and multiple interval mapping (Kao et al. 1999) have been proposed. Inclusive CIM does not require prior specification of N_{CF}, but instead requires that a significance threshold be specified for including and removing a marker as a cofactor. Although simulation results indicated that inclusive CIM is robust with regards to the significance threshold for cofactors (Li et al. 2007), no definitive rules exist for specifying such significance threshold. Multiple interval mapping builds a multiple-QTL model by fitting putative QTL as cofactors. As Li et al. (2007) indicated, multiple interval mapping avoids pitfalls associated with CIM but introduces other pitfalls due to complexities in the model selection procedure. Li et al. (2007) found that different model selection methods in multiple interval mapping led to different numbers of QTL detected.

The power to detect QTL was lower with the G-SM model than with the G model. As previously indicated, the G model involved backward elimination on a given chromosome after adjusting for genomewide marker effects on the other chromosomes. In contrast, the G-SM model involved single-marker analysis after correcting for genomewide marker effects at the other ($N_M - 1$) markers. The reduced power in the G-SM model may be attributed to linked markers capturing a portion of the effect of the marker being tested for its association with a QTL. Furthermore, the high frequencies of non-uniform distributions of p values (Table 2) indicated that the G-SM model did not adhere well to the specified type I error rate. The use of the G-SM model over the G model is, therefore, generally not recommended and, at best, the G-SM model should be used only as an exploratory method. In particular, if the number of markers on a chromosome approaches or exceeds N, the per-chromosome backward elimination procedure in the G model will fail because of a lack of degrees of freedom. In this situation, a possible approach is to use the G-SM model at a relaxed significance level in preliminary screening for marker–QTL associations along a chromosome. A reduced set of markers on the chromosome can then be subjected to G model analysis.

The definitions of a true QTL detected (i.e., a QTL with one or both flanking markers declared significant) and of a false positive (i.e., a significant marker not flanked on either side by a QTL) were the same as those in a previous study (Bernardo 2004). Given this definition, having more markers may seem to lead to fewer true QTL detected or to more false positives (Fig. 1). Suppose marker M_1 is at the 1 cM position, a QTL is at the 6 cM position, and marker M_4 is at the 12 cM position. Both M_1 and M_4 are then found significant. In this situation, a true QTL is detected and neither M_1 nor M_4 is a false positive. Now suppose the number of markers is increased and, in addition to M_1 and M_4, marker M_2 is at the 4 cM position and marker M_3 is at the 8 cM position. Further suppose that M_1, M_2, and M_3 are

found significant. In this situation, M_1 is now declared as a false positive because it does not immediately flank the QTL. Or, suppose M_1 and M_4 are found significant whereas M_2 and M_3 are nonsignificant. In this situation, which could arise if not enough recombinations occurred between M_2 and M_3, both M_1 and M_4 are declared as false positives and the QTL is declared undetected. These examples show that while the definitions of a true QTL detected and of a false positive led to a straightforward means for comparing the results with the C, G, and G-SM models, comparisons are best made under the same number of markers.

Furthermore, the simulation experiments in this study did not involve interval mapping, but instead relied on tests of the significance of effects of the markers themselves. As illustrated in the examples in the preceding paragraph, this approach led to a straightforward means of interpreting the results. In addition, arguments have been raised against the continued need for interval mapping procedures (F. van Eeuwijk 2011, personal communication). When interval mapping and CIM were proposed in the 1990s, marker technologies available at that time (e.g., restriction fragment length polymorphisms and simple sequence repeats) enabled mapping populations to be genotyped with markers that were typically spaced 10–15 cM apart (Bernardo 2008). However, developments since the 2000s in high-throughput genotyping with single-nucleotide polymorphisms have led to the availability of closely spaced markers (Bhattramakki and Rafalski 2002; Hyten et al. 2008). If markers are spaced, say, 3 cM apart, simple tests for QTL effects could be done at the marker positions (i.e., every 3 cM) instead of interval mapping at putative QTL positions every 1–2 cM.

Lastly, genomewide marker effects were calculated by RR-BLUP (Meuwissen et al. 2001; Bernardo and Yu 2007) in this study. Empirical results for several plant species have consistently shown that genomewide predictions with RR-BLUP were as good, if not better, than predictions with more complex Bayesian methods (Lorenzana and Bernardo 2009; Heffner et al. 2009; Lorenz et al. 2011; Asoro et al. 2011; Guo et al. 2012). If the trait is controlled by relatively few QTL, it may be advantageous to use a Bayesian genomewide prediction approach that constrains many markers to having effects of zero or a partial least squares approach (van Eeuwijk et al. 2000, 2002) that extracts latent factors that account for as much background variation as possible. Such topic needs to be studied.

References

Asoro FG, Newell MA, Beavis WD, Scott MP, Jannink J (2011) Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. Plant Genome 4:132–144

Bernardo R (2004) What proportion of declared QTL in plants are false? Theor Appl Genet 109:419–424

Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci 48:1649–1664

Bernardo R, Yu J (2007) Prospects for genomewide selection for quantitative traits in maize. Crop Sci 47:1082–1090

Bhattramakki D, Rafalski A (2002) Discovery and application of single nucleotide polymorphism markers in plants. In: Henry RJ (ed) Plant genotyping: the DNA fingerprinting of plants. CABI Publishing, Oxford, pp 179–191

Bjørnstad Å, Grønnerød S, Mac Key J, Tekauz A, Crossa J, Martens H (2004) Resistance to barley scald (*Rhynchosporium secalis*) in the Ethiopian donor lines 'Steudelli' and 'Jet', analyzed by partial least squares regression and interval mapping. Hereditas 141:166–179

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

Cornforth TW, Long AD (2003) Inferences regarding the numbers and locations of QTLs under multiple-QTL models using interval mapping and composite interval mapping. Genet Res 82:139–149

Doerge RW, Zeng ZB, Weir BS (1994) Statistical issues in the search for genes affecting quantitative traits in populations. In: Analysis of molecular marker data (supplement). Joint Plant Breed Symp Ser, Am Soc Hort Sci, Crop Sci Soc Am, Madison, p 15–26

Guo Z, Tucker D, Lu J, Kishore V, Gay G (2012) Evaluation of genome-wide selection efficiency in maize nested association mapping populations. Theor Appl Genet 124:261–275

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324

Heffner EL, Sorrells ME, Jannink J (2009) Genomic selection for crop improvement. Crop Sci 49:1–12

Heffner EL, Jannink J, Sorrells ME (2011) Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Genome 4:65–75

Hyten DL, Song Q, Choi I-Y, Specht JE, Matukumalli LK, Nelson RL, Shoemaker RC, Young ND, Cregan PB (2008) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor Appl Genet 116:945–952

Jansen RC (1992) A general mixture model for mapping quantitative trait loci by using molecular markers. Theor Appl Genet 85:252–260

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

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

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

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

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

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

Liu B-H (1998) Computational tools for study of complex traits. In: Paterson AH (ed) Molecular dissection of complex traits. CRC Press, Boca Raton, pp 43–79

Lorenz AJ, Chao S, Asoro FG, Heffner EL, Hayashi T, Iwata H, Smith KP, Sorrells ME, Jannink J (2011) Genomic selection in

plant breeding: knowledge and prospects. Adv Agron 113:77–123

Lorenzana R, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor Appl Genet 120:151–161

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinnauer Associates Inc., Sunderland

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Murdock D, Tsai Y, Adcock J (2008) *P*-values are random variables. Am Stat 62:242–245

Senior ML, Chin ECL, Lee M, Smith JSC, Stuber CW (1996) Simple sequence repeat markers developed from maize sequences found in the GENBANK database: map construction. Crop Sci 36:1676–1683

Smith JSC, Hussain T, Jones ES, Graham G, Podlich D, Wall S, Williams M (2008) Use of doubled haploids in maize breeding: implications for intellectual property protection and genetic diversity in hybrid crops. Mol Breed 22:51–59

Utz HF, Melchinger AE (1996) PLABQTL: a program for composite interval mapping of QTL. J Agric Genomics 2:1–5

van Eeuwijk FA, Crossa J, Vargas M, Ribaut JM (2000) Variants of factorial regression for analysing QTL by environment interaction. In: Gallais A, Dillmann C, Goldringer I (eds) Quantitative genetics and breeding methods: the way ahead, vol 96. INRA editions, Versailles, Les colloques, p 107–116

van Eeuwijk FA, Crossa J, Vargas M, Ribaut JM (2002) Analysing QTL by environment interaction by factorial regression, with an application to the CIMMYT drought and low nitrogen stress programme in maize. In: Kang MS (ed) Quantitative genetics, genomics and plant breeding. CAB International, Wallingford, pp 245–256

Vargas M, van Eeuwijk FA, Crossa J, Ribaut JM (2006) Mapping QTLs and QTL × environment interaction for CIMMYT maize drought stress program using factorial regression and partial least squares methods. Theor Appl Genet 112:1009–1023

Wang S, Basten C, Zeng Z-B (2006) Windows QTL Cartographer. North Carolina State University, Raleigh

Whittaker JC, Thompson R, Visscher PM (1996) On the mapping of QTL by regression of phenotypes on marker-type. Heredity 77:23–32

Yi N, Xu S (2008) Bayesian LASSO for quantitative trait loci mapping. Genetics 179:1045–1055

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

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