Genomewide Markers for Controlling Background Variation in Association Mapping

Rex Bernardo*

Abstract

Current procedures for association mapping in plants account for population structure (Q) and kinship (K). Here I propose an association mapping procedure that uses genomewide markers (G) to account for quantitative trait loci (QTL) on background chromosomes. My objective was to determine if the G and QG models are superior to the K and QK models. I simulated mapping population sizes of N = 384, 768, and 1536 inbreds that belonged to three known subpopulations. The G and QG models showed the best adherence to the significance level (P) specified by the investigator for declaring QTL. Across different genetic models (15 or 30 QTL), population sizes, and P levels, the Q model suffered from a high number of false positives (N_{FP}). With the K and QK models, a relaxed P level led to a reasonable number of true QTL detected (N_{TQ}) with N = 384 or 768 but it led to high N_{FP} with N = 1536. Compared with the K and QK models, the G and QG models had a better balance between high N_{TQ} and low N_{FP}. The results strongly indicated that the G and QG models are superior to the K and QK models.

Association mapping enables the discovery of marker–trait associations in a collection of individuals instead of in a designed mapping population (Hästbacka et al., 1992; Risch and Merikangas, 1996). In plants, collections of inbreds that are typically used in association mapping do not constitute a single random-mating population (Remington et al., 2001; Thornsberry et al., 2001; Zhu et al., 2008) and the resulting population structure and kinship (or relatedness) among inbreds can lead to spurious marker–trait associations (Devlin and Roeder, 1999; Pritchard et al., 2000; Thornsberry et al., 2001; Yu et al., 2006; Eathington et al., 2007). Different methods to account for population structure have been proposed (Pritchard et al., 2000; Devlin et al., 2001) and the QK mixed-model method has emerged as a useful approach to account for both population structure (Q) and kinship (K) in diverse panels of inbreds (Yu et al., 2006; Stich et al., 2008).

In particular, the QK model and the related K model (kinship only) for association mapping use random markers across the genome to estimate kinship among individuals (Yu et al., 2006). The QK model and K model both include the effect of one marker being tested for significance and a polygenic background effect for each individual. Specifying kinship exploits information from relatives in estimating polygenic background effects and such polygenic effects, in turn, improve the estimate of marker–trait association for each marker being tested (Yu et al., 2006).

Published in The Plant Genome 6
doi: 10.3835/plantgenome2012.11.0028
© Crop Science Society of America
5585 Guilford Rd., Madison, WI 53711 USA
An open-access publication

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

Dep. of Agronomy and Plant Genetics, Univ. of Minnesota, 411 Borlaug Hall, 1991 Upper Buford Circle, Saint Paul, MN 55108.
Received 7 Nov. 2012.
*Corresponding author (bernardo@umn.edu).

Abbreviations: BLUP, best linear unbiased prediction; G, genomewide markers; h^2, narrow-sense heritability; K, kinship; K-S, Kolmogorov-Smirnov test statistic; L, number of QTL; N, size of mapping population; N_{FP}, number of false positives; N_M, number of markers; N_{TQ}, number of true QTL detected; Q, population structure; QTL, quantitative trait loci; RR-BLUP, ridge regression-best linear unbiased prediction; V_E, nongenetic variance; V_G, genetic variance.

An alternative to estimating polygenic background effects as specified in the QK model or K model is to use random markers not to estimate kinship but to directly estimate background marker effects. Such an approach would use methods for using large numbers of markers in genomewide selection (or genomic selection) (Meuwissen et al., 2001). Many simulation and empirical results have shown the effectiveness of genomewide selection for modeling the sum of effects of unknown quantitative trait loci (QTL) across the genome (Meuwissen et al., 2001; Bernardo and Yu, 2007; Lorenzana and Bernardo, 2009; Hayes and Goddard, 2010; Albrecht et al., 2011; Asoro et al., 2011; Heffner et al., 2011; Iwata and Jannink, 2011; Guo et al., 2012; Heslot et al., 2012; Schulz-Streeck et al., 2012). Estimates of marker effects in association mapping would be improved if background effects are better estimated via a genomewide selection framework (Meuwissen et al., 2001) than via kinship in the QK model or K model (Yu et al., 2006).

In a previous study, I have shown the usefulness of a QTL mapping approach that involves estimating genomewide background effects among recombinant inbreds developed from a single biparental cross (Bernardo, 2013). In this study, I extend such approach to association mapping in a diverse set of inbreds that have both population structure and kinship. In particular, the G model accounts for genomewide background effects whereas the QG model accounts for both population structure and genomewide background effects. My objective in this study was to determine if the G and QG models are superior to the K and QK models commonly used for association mapping in plants.

MATERIALS AND METHODS

Overview

Six association mapping procedures were compared: (i) simple model, (ii) Q model, (iii) K model, (iv) QK model, (v) G model, and (vi) QG model. The simple, Q, K, and QK models were as described by Yu et al. (2006).

The simulation procedures used in this study were largely identical to those previously described for QTL mapping (Bernardo, 2013), with the main difference being that the simulated inbreds in this study belonged to three subpopulations whereas the simulated inbreds in the Bernardo (2013) study were all derived from the same cross. For the reader's convenience, descriptions of the simulation procedures are repeated herein.

Mapping Population, Genetic Models, and Phenotypic Values

Each simulation experiment comprised a combination of genetic model, total size of mapping population (N), and significance level (P) for declaring a QTL. Each simulation experiment was repeated 400 times for N = 384, 200 times for N = 768, and 100 times for N = 1536. Each repeat differed in the location of QTL and in the genotypes, genotypic values, and phenotypic values of the inbreds. I wrote a Fortran program to conduct the simulations and data analysis.

Of the N inbreds, one-third belonged to each of three subpopulations. A simulated F_1 generation was first formed by crossing two founder inbreds and all individuals in the first subpopulation descended from this F_1 generation. The F_1 was then backcrossed to the first founder inbred to form a BC_1 generation, and all individuals in the second subpopulation descended from this BC_1 generation. Lastly, the F_1 was backcrossed to the second founder inbred to form the complementary BC_1 generation, and all individuals in the third subpopulation descended from this BC_1 generation.

The individuals within each of the three subpopulations were generated to mimic the inbred recycling that occurs in a breeding program (Bernardo, 2009). Of the N/3 individuals within each subpopulation, 10% were first-cycle inbreds, 20% were second-cycle inbreds, 30% were third-cycle inbreds, and 40% were fourth-cycle inbreds. The first-cycle inbreds were derived by seven generations of selfing from the initial generation of the subpopulation (F_1 or either of the two BC_1 crosses). The second-cycle inbreds were developed by single-seed descent from random pairs of crosses among the first-cycle inbreds. Only one second-cycle inbred was retained per cross, with this inbred being the first random inbred with a genotypic mean that exceeded the genotypic mean of at least one of its parental inbreds. The third- and fourth-cycle inbreds were developed in the same manner from pairs of crosses among inbreds in the preceding cycle. The subpopulation assignments of all N inbreds were therefore known based on pedigree.

The two founder inbreds differed at number of markers (N_M) = 768 codominant marker loci, with this number of loci being greater than the 533 single nucleotide polymorphism markers considered by Yu et al. (2006) in maize (*Zea mays* L.). The sizes of the 10 chromosomes and of the entire genome (1749 cM) corresponded to those in a published maize linkage map (Senior et al., 1996). The genome was divided into N_M bins that were 1749/N_M cM in size. A marker was located at the midpoint of each bin.

I assumed that the purpose of association mapping was to find marker–trait associations for a less-complex trait that would tend to have several QTL with large effects (Bernardo, 2008), and the trait was controlled by number of QTL (L) = 15 QTL with a narrow-sense heritability (h^2) of 0.80 or by L = 30 QTL with an h^2 of 0.70. The first founder inbred had the favorable allele at odd-numbered QTL and the second founder inbred had the favorable allele at even-numbered QTL. The L QTL were randomly located among the 10 chromosomes. The sizes of QTL effects followed a geometric series (Lande and Thompson, 1990; Bernardo and Yu, 2007). Dominance and epistasis were absent, and the genotypic value of an inbred was equal to the sum of its genotypic values across all L QTL.

The h^2 was defined relative to a population of random recombinant inbreds derived from the F_2

between the two founder inbreds. Genetic variance (V_G) in a given repeat of a simulation experiment was first calculated as the variance among genotypic values (i.e., $h^2 = 1.0$) of 400 recombinant inbreds. The $N = 384$, 768, or 1536 inbreds in the simulation experiments were assumed to have been 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 nongenetic 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) among F_2–derived recombinant inbreds. These values of V_E and V_G were used in the subsequent analyses for the K, QK, G, and QG models.

Models with Population Structure and Kinship

In the simple model, the mean performance (across environments) of the N inbreds was modeled as $\mathbf{y} = \mathbf{1}\mu + \mathbf{S}\alpha + \mathbf{e}$, in which $\mathbf{y}$ was an $N \times 1$ vector of means of the N inbreds, μ was the grand mean, α was the effect of the marker allele from the first founder inbred at the kth marker being tested, $\mathbf{e}$ was an $N \times 1$ vector of residuals, $\mathbf{1}$ was an $N \times 1$ vector of 1s, and $\mathbf{S}$ was an $N \times 1$ incidence vector that related $\boldsymbol{\alpha}$ to $\mathbf{y}$. The elements of $\mathbf{S}$ were 1 if the inbred was homozygous for the allele from the first founder inbred, –1 if the inbred was homozygous for the allele from the second founder inbred, and 0 if the inbred was heterozygous at the marker locus. The simple model therefore represented the most basic single-marker analysis and did not account for population structure or polygenic background effects. The marker effect was assumed as fixed, and the analysis for the simple model (as well as for the Q, K, and QK models described below) was repeated until each of the $N_M = 768$ markers had been tested in single-marker analysis.

The Q model, which included population structure, was $\mathbf{y} = \mathbf{1}\mu + \mathbf{S}\alpha + \mathbf{Qv} + \mathbf{e}$, in which $\mathbf{v}$ was a vector of effects for two orthogonal contrasts that, along with μ, captured the differences among the means of the three subpopulations and $\mathbf{Q}$ was an $N \times 2$ incidence matrix that related $\mathbf{v}$ to $\mathbf{y}$ (Yu et al., 2006). Effects of subpopulations were assumed as fixed. As previously mentioned, the subpopulation assignments of the N inbreds were assumed known without error in all models that involved population structure. The K model, which included polygenic background effects, was $\mathbf{y} = \mathbf{1}\mu + \mathbf{S}\alpha + \mathbf{Zu} + \mathbf{e}$, in which $\mathbf{u}$ was an $N \times 1$ vector of polygenic background effects for the inbreds and $\mathbf{Z}$ was an $N \times N$ incidence matrix that related $\mathbf{u}$ to $\mathbf{y}$. In the K model, $\mathbf{V}(\mathbf{u}) = \mathbf{K}V_G$, in which $\mathbf{K}$, the kinship matrix, had elements equal to twice the marker-based coefficient of coancestry among the N inbreds. Given that all N_M markers were polymorphic between the two founder inbreds, the elements of $\mathbf{K}$ were directly estimated as twice the simple matching coefficient, across all N_M loci, between the two marker alleles carried by one inbred and the two marker alleles carried by a second inbred. The QK model, which included population structure and polygenic background effects, was then $\mathbf{y} = \mathbf{1}\mu + \mathbf{S}\alpha + \mathbf{Qv} + \mathbf{Zu} + \mathbf{e}$.

For each simulation experiment and for each repeat, mixed-model equations were used to obtain solutions to the simple, Q, K, and QK linear models (Henderson, 1984; Yu et al., 2006; Stich et al., 2008). A t test was used for testing the effect of the kth marker at comparison-wise significance levels of $P = 0.0001$, 0.00001, and 0.000001. With $N_M = 768$ markers, these significance levels to corresponded to pre-Bonferroni, experiment-wise Type I error rates (ignoring correlations among markers) ranging from 0.000768 to 0.0768. In addition, the results suggested that a relaxed significance level may be best for the K and QK models. A significance level of $P = 0.001$ was also used for the K and QK models.

Models with Genomewide Markers

The G model combined multiple regression for a given chromosome and adjustment for genomewide marker effects at the remaining chromosomes (Bernardo, 2013). Genomewide marker effects across all $N_M = 768$ markers were first calculated by ridge regression-best linear unbiased prediction (RR-BLUP) as described by Meuwissen et al. (2001). The linear model in RR-BLUP was $\mathbf{y} = \mathbf{1}\mu + \mathbf{Mg} + \mathbf{e}$, in which $\mathbf{g}$ was an $N_M \times 1$ vector of effects of the marker allele from the first founder inbred and $\mathbf{M}$ was an incidence matrix (elements of 1, 0, or –1) that related $\mathbf{g}$ to $\mathbf{y}$. Marker effects were assumed random and the variance of each marker effect in RR-BLUP was equal to $V_M = V_G/N_M$ (Meuwissen et al., 2001).

Procedures for association mapping analysis with 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, 2013) after having corrected for genomewide marker effects across all the other chromosomes not currently being analyzed for QTL. Adjustments for population structure were not made in the G model but were made in the QG model described in the next paragraph. Suppose 106 out of the $N_M = 768$ markers were on chromosome 1. To detect QTL on chromosome 1, the phenotypic data were first adjusted for the RR-BLUP genomewide marker effects of the 768 – 106 = 662 markers found on chromosomes 2 to 10. 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 106 markers) for chromosome 1. The significance level for retaining a marker in the model was $P = 0.0001$, 0.00001, or 0.000001. 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.

In contrast to the G model, the QG model included subpopulation effects in RR-BLUP analysis of genomewide marker effects, with the RR-BLUP

linear model being $\mathbf{y} = \mathbf{1}\mu + \mathbf{Qv} + \mathbf{Mg} + \mathbf{e}$. In the per-chromosome analysis in the QG model, corrections for background effects were for genomewide marker effects across all the other chromosomes not currently being analyzed for QTL as well as for **v**. Subsequent procedures for calculating marker effects were then identical between the G and QG models.

Data Analysis and Control of Type I Error Rate

For each of the six association mapping procedures, a false positive was declared when the *k*th marker was significant 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 when a QTL had a significant left flanking marker, a significant right flanking marker, or both flanking markers as significant. The numbers of false positives, true QTL detected, and significant markers were averaged across all repeats for each simulation experiment and association mapping method. Variances across repeats were pooled across methods and were used to calculate approximate LSDs at $P = 0.05$. The lowest number of repeats (100) was used in calculating approximate LSDs.

To determine how well each of the six association mapping procedures adhered to the specified significance level, simulations were conducted for the 30 QTL, $h^2 = 0.70$ genetic model with the restrictions that (i) all 30 QTL were located on chromosome 1 and (ii) *p*-values of tests of the null hypothesis were obtained only for markers on chromosomes 2 to 10. By not obtaining *p*-values from chromosome 1, the simulations therefore satisfied the conditions that the null hypothesis of no linkage with QTL was true for all the markers considered (i.e., on chromosomes 2 to 10) and V_G was greater than zero (i.e., $V_G = 0$ if no QTL were simulated) for the K, QK, G, and QG models. The Kolmogorov-Smirnov test ($P = 0.01$) was used to determine if the empirical *p*-values followed a uniform distribution. The simulations were repeated 100 times and the frequency of nonuniform distributions of *p*-values and mean of the Kolmogorov-Smirnov test statistic across repeats was calculated.

RESULTS AND DISCUSSION

Control of Type I Error Rate

When the null hypothesis is true, *p*-values are expected to follow a uniform distribution (Murdock et al., 2008). Nonuniform *p*-values of markers would indicate that the actual Type I error rate is unequal to the significance level specified by the investigator for declaring a QTL. Simulations for which the null hypothesis of no QTL present was true indicated that, among the six association mapping procedures, the G and QG models had the best adherence to the nominal Type I error rate (Table 1). With population sizes of $N = 384$ or 768, the percentage of simulation repeats for which the *p*-values had highly significant deviations ($P = 0.01$) from a uniform distribution was 14 to 18% for the G and QG models. In contrast, the simple, Q, K, and QK models with $N = 384$ or 768 had 99 to 100% of simulation repeats with nonuniform *p*-values. When N was increased to 1536, the percentage of simulation repeats with nonuniform *p*-values decreased to 56% for the K and QK models and increased to 28 to 29% for the G and QG models.

The Kolmogorov-Smirnov test statistic (K-S), which was used to test for uniformity of *p*-values, was equal to the maximum difference between the empirical distribution of *p*-values and the expected value from a uniform distribution. Large positive values of K-S indicated an excess of small *p*-values (i.e., liberal tests of significance) whereas large negative values indicated an excess of large *p*-values (i.e., conservative tests of significance). The mean K-S was 0.32 to 0.38 with the simple model and 0.21 to 0.23 with the Q model (Table 1), indicating that these two models lead to an excess of false positives. The mean K-S with the G and QG models (0.03–0.05) was closest to the ideal value of zero. The mean K-S with the K and QK models was negative (−0.18 to −0.14) with $N = 384$ or 768 and positive (0.17) with $N = 1536$. This result indicated that tests for marker–trait associations with the K and QK models are too conservative with small N and too liberal with large N.

While Yu et al. (2006), Stich et al. (2008), and Stich and Melchinger (2009) showed plots of cumulative *p*-values, they did not conduct formal tests for deviations of the observed *p*-values from a uniform distribution and the null hypothesis in their empirical studies would not have been always true if QTL were detectable for the traits they studied. Nevertheless, the plots of cumulative *p*-values in these three previous studies did not seem entirely uniform for the following: (i) K model for *TUBB2A* gene expression in humans, Q, K, and QK models for flowering time in maize, and Q model for ear height and ear diameter in maize (Yu et al., 2006), (ii) K and QK models for grain yield in maize (Stich et al., 2008), and (iii) K and QK models for different traits in *Arabidopsis thaliana* (L.) Heynh., sugarbeet (*Beta vulgaris* L.), potato (*Solanum tuberosum* L.), and rapeseed (*Brassica napus* L.) (Stich and Melchinger, 2009).

True Quantitative Trait Loci Detected and False Positives

Across population sizes, genetic models, and significance levels for declaring QTL, the simple model had the poorest performance in terms of the number of true QTL detected (N_{TQ}) and the number of false positives (N_{FP}) declared. The simple model had N_{FP} values that were 3 to 17 times as large as the corresponding N_{TQ} values (Table 2). As previously mentioned, the simple model does not account for population structure or polygenic background effects and, as other authors have indicated (Yu et al., 2006; Stich et al., 2008; Stich and Melchinger, 2009), is not recommended.

Compared with the simple model, the Q model (which included population structure) had roughly the same N_{TQ} but had lower N_{FP} (Table 2). Such lower N_{FP} indicated that the Q model accounted for at least

a portion of the influence of population structure on marker–trait associations, and this result was consistent with the lower K-S with the Q model than with the simple model (Table 1). However, the N_{FP} values with the Q model still consistently exceeded the N_{TQ} values across simulation experiments.

The Q model attempts to account for population structure by modeling the differences in means of the subpopulations (**v**) for a given trait. To illustrate, suppose a maize mapping panel includes early-maturing flint inbreds and late-maturing dent inbreds. The difference in mean days to maturity between the flint and dent inbreds will lead to a nonzero **v** in the Q model, and the Q model will thereby account for the effect of population structure on days to maturity. In contrast, suppose that the flint and dent inbreds have equal subpopulation means for stalk lodging. In this situation, **v** in the Q model for stalk lodging will be zero and the Q model consequently becomes equal to the simple model. In other words, while marker or pedigree data may clearly indicate the presence of subpopulations (e.g., flint versus dent inbreds), the Q model cannot effectively account for the influence of subpopulations on estimates of marker–trait associations if the subpopulations have equal or similar means for the trait.

In this simulation study, the two founder inbreds complemented each other in the favorable QTL alleles they carried and therefore had similar mean performance. In terms of units of the genetic standard deviation $[(V_G)^{1/2}]$, the first founder inbred had a genotypic mean of 0.13 whereas the second founder inbred had a mean of −0.13. Discounting the effects of selection, the inbreds were then expected to have a mean of zero in the first subpopulation (F_1), 0.065 in the second subpopulation (BC_1 to the first founder inbred), and −0.065 in the third subpopulation (BC_1 to the second founder inbred). These similar subpopulation means rendered the Q model largely ineffective.

Methods have been proposed for assigning inbreds to subpopulations based on marker information, for example, STRUCTURE software (Pritchard et al., 2000) or principal components analysis (Price et al., 2006). These two marker-based approaches for assigning inbreds to subpopulations were not attempted in this study. However, it is speculated that the large differences in frequencies (i.e., 0.50, 0.75, and 0.25) of the marker allele from the first founder inbred among the three subpopulations would have led to subpopulation assignments similar to the pedigree-based assignments of the N inbreds. Furthermore, the lack of large differences among subpopulation means would likely have led to the Q model still being ineffective even if STRUCTURE software or principal components analyses were used. These points underscore that, in practice, examining the putative or known subpopulations for differences in their mean performance for a given trait would be useful before deciding to use an association mapping model that incorporates the mean effects of subpopulations (i.e., **v**).

Table 1. Control of the Type I error rate in simulations where the null hypothesis of no marker–quantitative trait loci (QTL) linkage was true for different association mapping models. Results are for different sizes of the mapping population (*N*), a genetic model with 30 QTL and a narrow-sense heritability (h^2) of 0.70, 768 markers, and a total of 100 repeats.

	N = 384		N = 768		N = 1536	
Model†	% nonuniform‡	Mean K-S§	% non-uniform	Mean K-S	% nonuniform	Mean K-S
Simple	99	0.32	100	0.34	100	0.38
Q	100	0.21	100	0.23	100	0.22
K	100	−0.14	100	−0.18	56	0.17
QK	100	−0.14	100	−0.17	56	0.17
G	18	0.03	18	0.03	29	0.05
QG	16	0.03	14	0.03	28	0.05

†Q, population structure; K, kinship; G, genomewide markers.

‡Percentage of nonuniform distributions of p-values across 100 repeats based on a Kolmogorov-Smirnov test (P = 0.01).

§K-S, Kolmogorov-Smirnov test statistic. Large positive values (mean across 100 repeats) indicate an excess of small p-values whereas large negative values indicate an excess of large p-values.

Because **v** had a minor effect, the K and QK models had similar values of N_{TQ} and of N_{FP} (Table 2). With N = 384 or 768, the K and QK models led to N_{TQ} of 0.1 to 4.8 and N_{FP} of 0.0 to 0.8 (Table 2). Increasing the population size to N = 1536 with the K and QK models led to higher N_{TQ} (5.1–9.2) but at the cost of higher N_{FP} (3.6–19.2). With N = 1536, the K and QK models had N_{FP} higher than N_{TQ} when the least stringent significance level (P = 0.0001) was used for declaring QTL. The low N_{TQ} at low N and high N_{FP} with high N were consistent with the K-S of the K and QK models at different population sizes (Table 1).

Compared with the QK model (and the largely equivalent K model), the G and QG models led to higher N_{TQ} while maintaining N_{FP} within reasonable levels (Table 2). In particular, the G and QG models performed best at the largest population size and with stringent significance levels for declaring a QTL. With N = 1536 and P = 0.00001 or 0.000001, the G and QG models had N_{TQ} of 13.1 to 18.7 and N_{FP} of 0.9 to 2.8. When the population size was reduced to N = 384, the G and QG models with P = 0.00001 or 0.000001 had lower N_{TQ} (4.5–7.5) and higher N_{FP} (1.7–4.5).

A scatterplot is useful for comparing the N_{TQ} and N_{FP} values obtained with the G, QG, and QK models (Fig. 1). The results indicated that across population sizes, genetic models, and significance levels, the G and QG models led to a better balance of maximizing N_{TQ} and minimizing N_{FP} compared with the QK model. The N_{TQ} with the QK model did not exceed 10 and these higher N_{TQ} values were accompanied by a high N_{FP}. In contrast, N_{TQ} with the G and QG models approached 20 and N_{FP} did not exceed 6. Overall, both N_{TQ} and N_{FP} were slightly higher with the G model than with the QG model (Fig. 1).

Table 2. Number of true quantitative trait loci (QTL) detected (N_{TQ}), number of false positives (N_{FP}), and number of significant markers for different association mapping models. Mapping population sizes (*N*) and significance levels for declaring QTL (*P*) varied, and 768 markers were used.

			15 QTL, $h^{2\dagger}$ = 0.80			30 QTL, h^2 = 0.70		
N	Model‡	*P*	True QTL	False positives	Significant markers	True QTL	False positives	Significant markers
384	Simple	0.0001	5.3	77.1	86.4	4.2	23.4	30.1
	Simple	0.00001	4.2	48.7	55.8	2.6	11.4	15.4
	Simple	0.000001	3.3	31.6	37.3	1.6	5.6	7.9
	Q	0.0001	4.9	20.6	28.9	3.9	9.3	15.3
	Q	0.00001	3.7	11.4	17.7	2.4	3.9	7.4
	Q	0.000001	3.0	6.8	11.7	1.4	1.8	3.8
	K	0.0001	3.0	0.8	4.7	1.0	0.2	1.3
	K	0.00001	1.9	0.3	2.7	0.4	0.0	0.5
	K	0.000001	1.2	0.1	1.6	0.1	0.0	0.1
	QK	0.0001	2.9	0.7	4.7	1.0	0.1	1.3
	QK	0.00001	1.9	0.3	2.7	0.4	0.0	0.4
	QK	0.000001	1.2	0.1	1.5	0.1	0.0	0.1
	G	0.0001	8.0	6.9	14.9	7.3	5.4	12.5
	G	0.00001	7.5	4.5	11.9	6.3	3.7	9.8
	G	0.000001	7.0	3.2	10.1	5.3	2.6	7.8
	QG	0.0001	7.8	5.2	13.0	6.9	4.1	10.7
	QG	0.00001	7.2	3.0	10.2	5.6	2.6	8.0
	QG	0.000001	6.6	2.0	8.6	4.5	1.7	6.1
		LSD(0.05)	0.3	0.3	0.3	0.3	0.2	0.3
768	Simple	0.0001	7.7	118.3	132.1	8.3	43.5	57.2
	Simple	0.00001	6.7	80.8	92.8	6.4	25.4	35.6
	Simple	0.000001	5.9	55.8	66.1	4.7	14.9	22.3
	Q	0.0001	7.5	37.5	50.9	8.4	22.2	35.7
	Q	0.00001	6.5	24.4	35.9	6.1	12.2	21.8
	Q	0.000001	5.7	17.0	27.0	4.5	6.9	13.7
	K	0.0001	4.8	0.6	6.9	3.1	0.1	3.6
	K	0.00001	3.8	0.3	5.0	1.6	0.0	1.7
	K	0.000001	3.0	0.1	3.7	0.7	0.0	0.7
	QK	0.0001	4.7	0.7	6.9	3.1	0.1	3.5
	QK	0.00001	3.8	0.3	5.0	1.6	0.0	1.7
	QK	0.000001	3.0	0.1	3.7	0.7	0.0	0.7
	G	0.0001	11.1	6.8	18.0	14.0	5.6	19.2
	G	0.00001	10.6	4.0	14.6	12.7	3.6	16.1
	G	0.000001	10.3	2.6	13.0	11.8	2.7	14.1
	QG	0.0001	10.9	5.2	16.2	13.5	4.3	17.6
	QG	0.00001	10.5	2.9	13.4	12.3	2.8	14.8
	QG	0.000001	10.1	1.8	11.9	11.2	1.9	12.8
		LSD(0.05)	0.3	0.2	0.3	0.4	0.2	0.3
1536	Simple	0.0001	9.5	165.4	183.1	12.6	70.9	93.1
	Simple	0.00001	8.9	121.6	137.9	10.7	47.3	66.0
	Simple	0.000001	8.2	89.6	104.5	9.0	31.7	47.2
	Q	0.0001	9.7	59.9	77.7	13.2	46.2	69.3
	Q	0.00001	8.8	44.5	60.5	11.3	30.9	50.5
	Q	0.000001	7.9	34.4	48.9	9.5	21.2	37.4
	K	0.0001	7.9	19.2	31.4	9.2	14.4	27.6
	K	0.00001	6.8	12.3	22.7	7.0	7.8	17.2
	K	0.000001	6.1	8.0	16.9	5.1	4.6	11.4
	QK	0.0001	7.8	15.7	27.8	9.1	12.4	25.4
	QK	0.00001	6.8	9.7	19.9	6.8	6.7	16.0
	QK	0.000001	6.1	6.2	14.9	5.1	3.6	10.4
	G	0.0001	13.6	4.6	18.6	19.8	4.5	24.1
	G	0.00001	13.4	2.2	15.9	18.7	2.8	21.3
	G	0.000001	13.2	1.4	14.8	17.8	2.0	19.6
	QG	0.0001	13.5	3.4	17.4	19.6	3.7	23.2
	QG	0.00001	13.3	1.6	15.2	18.5	2.2	20.5
	QG	0.000001	13.1	0.9	14.2	17.4	1.5	18.7
		LSD(0.05)	0.3	0.3	0.3	0.4	0.3	0.5

† h^2, narrow-sense heritability.

‡ Q, population structure; K, kinship; G, genomewide markers.

With the K and QK models, the results on control of the Type I error rate (Table 1) and the low values of both N_{TQ} and N_{FP} in some of the simulation experiments (Table 2; Fig. 1) suggested that a relaxed significance level might be useful for these two models (e.g., $P = 0.001$, which corresponded to a pre-Bonferroni, experiment-wise error rate of 0.768). The use of $P = 0.001$ in the QK model led to N_{TQ} of 2.7 to 4.4 and N_{FP} of 0.8 to 1.8 with $N = 384$ and N_{TQ} of 5.8 to 6.1 and N_{FP} of 0.7 to 1.6 with $N = 768$. When the population size was increased to $N = 1536$, the use of $P = 0.001$ in the QK model led to N_{TQ} of 9.2 to 12.0 but these higher values of N_{TQ} were accompanied by unacceptably high N_{FP} values of 25.0 to 28.2. Similar results were observed for the K model (results not shown). Such subjectivity regarding the appropriate P for a given N is a drawback of the K and QK models.

Implications and Application

The results from this study strongly suggest that a genomewide selection framework (Meuwissen et al., 2001) is useful for modeling polygenic background effects in association mapping. The G and QG models, which involve adjustment for effects of QTL at background chromosomes, was superior to the K and QK models (Yu et al., 2006) that have been routinely used in association mapping in plants (Zhu et al., 2008; Beattie et al., 2010; Ghavami et al., 2011; Gutiérrez et al., 2011; Li et al., 2011; Van Inghelandt et al., 2012).

The K and QK models use a global estimate (i.e., across all chromosomes) of kinship via the **K** matrix. Because there is no separation in kinship and in polygenic effects (**u**) between the chromosome being tested for QTL and the chromosomes not being tested for QTL, **u** could absorb some of the effects of QTL close to the marker being tested for significance. In other words, the K and QK models may have difficulty in separating the effects of QTL linked to the marker being tested from the joint effects of QTL found elsewhere in the genome. Suppose a species has only two chromosomes, and the map order of linked markers (M_i) and QTL (Q_j) is M_1 Q_1 M_2 M_3 Q_2 M_4 M_5 on chromosome 1 and M_6 Q_3 M_7 Q_4 M_8 Q_5 M_9 Q_6 M_{10} on chromosome 2. Further suppose that M_2 is the marker being tested for its significance. In the K and QK models, an effect for M_2 is fitted while attempting to simultaneously fit the joint effects of Q_2, Q_3, Q_4, Q_5, and Q_6 not via individual markers but via the fitted means of the individuals that carry different combinations of alleles at these QTL. Kinship information based on all 10 markers is used to capture information from relatives. Because markers are not used to track the effects at individual background QTL, the resolution in this example may be insufficient between effects associated with marker M_2 (mainly due to Q_1) versus the joint effects of the other QTL, especially those on the same chromosome (Q_2).

In contrast to the K and QK models, the G and QG models would attempt, in the above example, to account for the joint effects associated with all 10 markers. Adjustment for the effects of the four QTL

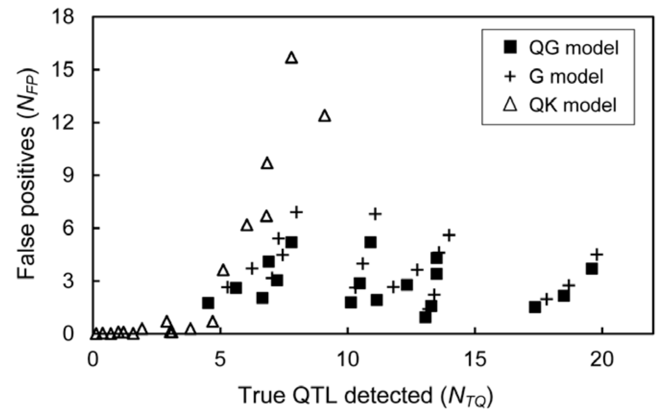

Figure 1. Number of true quantitative trait loci (QTL) detected (N_{TQ}) and number of false positives (N_{FP}) with association mapping models that use population structure and genomewide markers (QG model, solid squares), genomewide markers (G model, +), and population structure and kinship (QK model, open triangles). Results are for three significance levels ($P = 0.0001$, 0.00001, and 0.000001), two genetic models [15 QTL with narrow-sense heritability (h^2) = 0.80 and 30 QTL with $h^2 = 0.70$], and three mapping population sizes ($N = 384$, 768, and 1536).

on chromosome 2 is first made through genomewide prediction with markers M_6 to M_{10}. The effects associated with markers M_1 to M_5 on chromosome 1 are then tested by multiple regression. The results from this study suggest that finding QTL through the modeling of marker effects only (G and QG models) is superior to the modeling of the mean effect of a marker allele across all individuals (i.e., α in the K and QK linear models) along with the fitted effects of the individuals themselves (i.e., **u** in the K and QK linear models).

The foregoing illustration suggests an alternative approach for the K and QK models: instead of having a global **K** matrix, **K** matrices specific to each chromosome may be used. With 10 maize chromosomes, for example, the **K** matrix used in testing for QTL on chromosome 1 would be calculated only from the markers on chromosomes 2 to 10. This alternative approach needs to be investigated.

Strictly speaking, best linear unbiased prediction (BLUP) assumes that the genetic (V_G) and nongenetic variances (V_E) are known without error and all implementations of BLUP (including RR-BLUP with genomewide markers) are therefore approximations. The variance of marker effects in RR-BLUP is $V_M = V_G/N_M$ (Meuwissen et al., 2001) and the true values of V_G and V_E were used in the simulations in this study. In practice, both V_G and V_E would need to be estimated from the phenotypic data through an approach such as an expectation-maximization-type algorithm (Dempster et al., 1977). On the other hand, the term $V_E/V_M = V_E N_M/V_G$ in the mixed-model equations for RR-BLUP (Bernardo, 2010, p. 294) can also be expressed as $N_M(1 - h^2)/h^2$. This means that RR-BLUP actually requires information on h^2 instead of on V_G and V_E individually. Given that h^2 ranges from 0 to 1, the value of h^2 can be estimated by (i) repeatedly splitting the data set into a training set and

validation set, (ii) assuming different values of h^2 (e.g., 0.05, 0.10, 0.15,…, 0.95) in RR-BLUP, and (iii) determining which value of h^2 leads to the most accurate cross-validation predictions. This approach will circumvent the need to account for relatedness and population structure in attempting to estimate V_G and V_E by traditional approaches in an association mapping panel.

Variations of the G and QG models are possible. Multiple regression by backward elimination was conducted in this study and this approach will fail if, with large-scale genotying (Close et al., 2009; Ganal et al., 2011; Poland et al., 2012), the number of markers on a given chromosome exceeds the size of the association mapping panel. In this situation, multiple regression by forward selection may be used to prevent overparameterization. Furthermore, genomewide predictions were obtained by RR-BLUP (Meuwissen et al., 2001). Empirical results in plants have shown that genomewide predictions with RR-BLUP were as good as if not better than predictions with more complex Bayesian methods (Lorenzana and Bernardo, 2009; Heffner et al., 2009; Asoro et al., 2011; Lorenz et al., 2011; Guo et al., 2012; Schulz-Streeck et al., 2012). However, if the trait is controlled by relatively few QTL or is expressed in a binary manner, Bayesian or machine-learning approaches that constrain many markers to having null effects might be useful (Long et al., 2007; Resende et al., 2012). Software is needed for association mapping with the G and QG models as implemented in this study along with options for alternative methods for multiple regression and genomewide prediction.

References

Albrecht, T., V. Wimmer, H.J. Auinger, M. Erbe, C. Knaak, M. Ouzunova, H. Simianer, and C.C. Schön. 2011. Genome-based prediction of testcross values in maize. Theor. Appl. Genet. 123:339–350. doi:10.1007/s00122-011-1587-7

Asoro, F.G., M.A. Newell, W.D. Beavis, M.P. Scott, and J.-L. Jannink. 2011. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. Plant Gen. 4:132–144. doi:10.3835/plantgenome2011.02.0007

Beattie, A.D., M.J. Edney, G.J. Scoles, and B.G. Rossnagel. 2010. Association mapping of malting quality data from western Canadian two-row barley cooperative trials. Crop Sci. 50:1649–1663. doi:10.2135/cropsci2009.06.0334

Bernardo, R. 2004. What proportion of declared QTL in plants are false? Theor. Appl. Genet. 109:419–424. doi:10.1007/s00122-004-1639-3

Bernardo, R. 2008. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. Crop Sci. 48:1649–1664. doi:10.2135/cropsci2008.03.0131

Bernardo, R. 2009. Should maize doubled haploids be induced among F_1 or F_2 plants? Theor. Appl. Genet. 119:255–262. doi:10.1007/s00122-009-1034-1

Bernardo, R. 2010. Breeding for quantitative traits in plants. 2nd ed. Stemma Press, Woodbury, MN.

Bernardo, R. 2013. Genomewide markers as cofactors for precision mapping of quantitative trait loci. Theor. Appl. Genet. doi:10.1007/s00122-012-2032-2

Bernardo, R., and J. Yu. 2007. Prospects for genomewide selection for quantitative traits in maize. Crop Sci. 47:1082–1090. doi:10.2135/cropsci2006.11.0690

Close, T.J., P.R. Bhat, S. Lonardi, Y. Wu, N. Rostoks, L. Ramsay, A. Druka, N. Stein, J.T. Svensson, and S. Wanamaker. 2009. Development and implementation of high-throughput SNP genotyping in barley. BMC Genomics 10:582. doi:10.1186/1471-2164-10-582

Dempster, A.P., N.M. Laird, and D.B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc., B 39:1–38.

Devlin, B., and K. Roeder. 1999. Genomic control for association studies. Biometrics 55:997–1004. doi:10.1111/j.0006-341X.1999.00997.x

Devlin, B., K. Roeder, and L. Wasserman. 2001. Genomic control, a new approach to genetic-based association studies. Theor. Popul. Biol. 60:155–166. doi:10.1006/tpbi.2001.1542

Doerge, R.W., Z.-B. Zeng, and B.S. Weir. 1994. Statistical issues in the search for genes affecting quantitative traits in populations. In: Proceedings of the Symposium on Analysis of Molecular Marker Data. Joint Plant Breeding Symposium Series, Corvalis, OR. 5–6 Aug. 1994. Am. Soc. Hort. Sci. and CSSA, Madison, WI. p. 15–26.

Eathington, S.R., T.M. Crosbie, M.D. Edwards, R.S. Reiter, and J.K. Bull. 2007. Molecular markers in a commercial breeding program. Crop Sci. 47(S3):S154–S163. doi:10.2135/cropsci2007.04.0015IPBS

Ganal, M.W., G. Durstewitz, A. Polley, A. Bérard, E.S. Buckler, A. Charcosset, J.D. Clarke, E.M. Graner, M. Hansen, and J. Joets. 2011. A large maize (*Zea mays* L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS ONE 6:e28334. doi:10.1371/journal.pone.0028334

Ghavami, F., E.M. Elias, S. Mamidi, O. Ansari, M. Sargolzaei, T. Adhikari, M. Mergoum, and S.F. Kianian. 2011. Mixed model association mapping for Fusarium head blight resistance in Tunisian-derived durum wheat populations. G3: Genes, Genomes, Genet. 1:209–218.

Guo, Z., D. Tucker, J. Lu, V. Kishore, and G. Gay. 2012. Evaluation of genome-wide selection efficiency in maize nested association mapping populations. Theor. Appl. Genet. 124:261–275. doi:10.1007/s00122-011-1702-9

Gutiérrez, L., A. Cuesta-Marcos, A.J. Castro, J. von Zitzewitz, M. Schmitt, and P.M. Hayes. 2011. Association mapping of malting quality quantitative trait loci in winter barley: Positive signals from small germplasm arrays. Plant Gen. 4:256–272. doi:10.3835/plantgenome2011.07.0020

Hästbacka, J., A. de la Chapelle, I. Kaitila, P. Sistonen, W. Weaver, and E. Lander. 1992. Linkage disequilibrium mapping in isolated founder populations: Diastrophic displasia in Finland. Nat. Genet. 2:204–211. doi:10.1038/ng1192-204

Hayes, B., and M. Goddard. 2010. Genome-wide association and genomic selection in animal breeding. Genome 53:876–883. doi:10.1139/G10-076

Henderson, C.R. 1984. Applications of linear models in animal breeding. Univ. of Guelph, Guelph, ON, Canada.

Heffner, E.L., J.-L. Jannink, and M.E. Sorrells. 2011. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Gen. 4:65–75. doi:10.3835/plantgenome.2010.12.0029

Heffner, E.L., M.E. Sorrells, and J.-L. Jannink. 2009. Genomic selection for crop improvement. Crop Sci. 49:1–12. doi:10.2135/cropsci2008.08.0512

Heslot, N., H.-P. Yang, M.E. Sorrells, and J.-L. Jannink. 2012. Genomic selection in plant breeding: A comparison of models. Crop Sci. 52:146–160.

Iwata, H., and J.-L. Jannink. 2011. Accuracy of genomic selection prediction in barley breeding programs: A simulation study based on the real single nucleotide polymorphism data of barley breeding lines. Crop Sci. 51:1915–1927. doi:10.2135/cropsci2010.12.0732

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

Li, X., Y. Wei, K.J. Moore, R. Michaud, D.R. Viands, J.L. Hansen, A. Acharya, and E.C. Brummer. 2011. Association mapping of biomass yield and stem composition in a tetraploid alfalfa breeding population. Plant Gen. 4:24–35. doi:10.3835/plantgenome2010.09.0022

Long, N., D. Gianola, G.J.M. Rosa, K.A. Weigel, and S. Avendaño. 2007. Machine learning classification procedure for selecting SNPs in genomic selection: Application to early mortality in broilers. J. Anim. Breed. Genet. 124:377–389. doi:10.1111/j.1439-0388.2007.00694.x

Lorenz, A.J., S. Chao, F.G. Asoro, E.L. Heffner, T. Hayashi, H. Iwata, K.P. Smith, M.E. Sorrells, and J.-L. Jannink. 2011. Genomic selection in plant breeding: Knowledge and prospects. Adv. Agron. 113:77–123. doi:10.1016/B978-0-12-385531-2.00002-5

Lorenzana, R., and R. Bernardo. 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor. Appl. Genet. 120:151–161. doi:10.1007/s00122-009-1166-3

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Murdock, D., Y. Tsai, and J. Adcock. 2008. *P*-values are random variables. Am. Stat. 62:242–245. doi:10.1198/000313008X332421

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.-L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7:e32253. doi:10.1371/journal.pone.0032253

Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38:904–909. doi:10.1038/ng1847

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Remington, D.L., J.M. Thornsberry, Y. Matsuoka, L.M. Wilson, S.R. Whitt, J. Doebley, S. Kresovich, M.M. Goodman, and E.S. Buckler. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci. USA 98:11479–11484. doi:10.1073/pnas.201394398

Resende, M.F.R., Jr., P. Muñoz, M.D.V. Resende, D.J. Garrick, R.L. Fernando, J.M. Davis, E.J. Jokela, T.A. Martin, G.F. Peter, and M. Kirst. 2012. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). Genetics 188:695–708.

Risch, N., and K. Merikangas. 1996. The future of genetic studies of complex human diseases. Science 273:1516–1517. doi:10.1126/science.273.5281.1516

Schulz-Streeck, T., J.O. Ogutu, Z. Karaman, C. Knaak, and H.P. Piepho. 2012. Genomic selection using multiple populations. Crop Sci. 52:2453–2461. doi:10.2135/cropsci2012.03.0160

Senior, M.L., E.C.L. Chin, M. Lee, J.S.C. Smith, and C.W. Stuber. 1996. Simple sequence repeat markers developed from maize sequences found in the GENBANK database: Map construction. Crop Sci. 36:1676–1683. doi:10.2135/cropsci1996.0011183X003600060043x

Stich, B., and A.E. Melchinger. 2009. Comparison of mixed-model approaches for association mapping in rapeseed, potato, sugar beet, maize, and *Arabidopsis*. BMC Genomics 10:94. doi:10.1186/1471-2164-10-94

Stich, B., J. Möhring, H.-P. Piepho, M. Heckenberger, E.S. Buckler, and A.E. Melchinger. 2008. Comparison of mixed-model approaches for association mapping. Genetics 178:1745–1754. doi:10.1534/genetics.107.079707

Thornsberry, J.M., M.M. Goodman, J. Doebley, S. Kresovich, D. Nielsen, and E.S. Buckler, IV. 2001. *Dwarf*8 polymorphisms associate with variation in flowering time. Nat. Genet. 28:286–289. doi:10.1038/90135

Van Inghelandt, D., A.E. Melchinger, J.-P. Martinant, and B. Stich. 2012. Genome-wide association mapping of flowering time and northern corn leaf blight (*Setosphaeria turcica*) resistance in a vast commercial maize germplasm set. BMC Plant Biol. 12:56. doi:10.1186/1471-2229-12-56

Whittaker, J.C., R. Thompson, and P.M. Visscher. 1996. On the mapping of QTL by regression of phenotypes on marker-type. Heredity 77:23–32. doi:10.1038/hdy.1996.104

Yu, J., G. Pressoir, W.H. Briggs, I.V. Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38:203–208. doi:10.1038/ng1702

Zhu, C., M. Gore, E.S. Buckler, and J. Yu. 2008. Status and prospects of association mapping in plants. Plant Gen. 1:5–20. doi:10.3835/plantgenome2008.02.0089