Theor Appl Genet (2012) 125:987–998
DOI 10.1007/s00122-012-1889-4

ORIGINAL PAPER

Comparison of biometrical approaches for QTL detection in multiple segregating families

Wenxin Liu · Jochen C. Reif · Nicolas Ranc · Giovanni Della Porta · Tobias Würschum

Received: 2 February 2012 / Accepted: 27 April 2012 / Published online: 23 May 2012
© Springer-Verlag 2012

Abstract Detection of QTL in multiple segregating families possesses many advantages over the classical QTL mapping in biparental populations. It has thus become increasingly popular, and different biometrical approaches are available to analyze such data sets. We empirically compared an approach based on linkage mapping methodology with an association mapping approach. To this end, we used a large population of 788 elite maize lines derived from six biparental families genotyped with 857 SNP markers. In addition, we constructed genetic maps with reduced marker densities to assess the dependency of the performance of both mapping approaches on the marker density. We used cross-validation and resample model averaging and found that while association mapping performed better under high marker densities, this was reversed under low marker densities. In addition to main effect QTL, we also detected epistatic interactions. Our results suggest that both approaches will profit from a further increase in marker density and that a cross-validation should be applied irrespective of the biometrical approach.

Communicated by M. Sillanpaa.

W. Liu
Crop Genetics and Breeding Department, China Agricultural University, Beijing 100193, China

W. Liu · J. C. Reif · T. Würschum (✉)
State Plant Breeding Institute, University of Hohenheim, 70599 Stuttgart, Germany
e-mail: tobias.wuerschum@uni-hohenheim.de

N. Ranc
Syngenta Seeds SAS, 12, chemin de l'Hobit, B.P. 27, 31790 Saint-Sauveur, France

G. D. Porta
Syngenta Seeds S.p.A., Via per Soresina, 26020 Casalmorano (CR), Italy

Introduction

Detection of QTL has become an indispensable tool for plant geneticists and plant breeders. QTL detection results are the basis for our understanding of the genetic architecture underlying complex traits. If QTL detection is done within breeding populations, the results are of direct relevance for applied plant breeding and facilitate the generation of novel, superior varieties by knowledge-based breeding. Classical QTL mapping relied on biparental crosses. The disadvantage of this approach is that only two alleles are considered and that QTL effects are often specific for that population and, therefore, not easily transferable (Holland 2007). To circumvent these disadvantages, QTL mapping strategies have recently shifted from biparental populations to QTL detection in multiple segregating families (Myles et al. 2009). This approach can potentially increase the QTL detection power, the accuracy of QTL localization, and provide more robust estimates of QTL effects across different genetic backgrounds.

Two conceptually different approaches are available for QTL detection in multiple segregating families: multiple-line cross QTL (MC-QTL) mapping and association mapping (Würschum 2012). The differences between both approaches are summarized in Table 1. MC-QTL mapping is an extension of linkage mapping methodology to multiple families (Blanc et al. 2006) and has recently been applied to an experimental population (Coles et al. 2010) and a breeding population in maize (Steinhoff et al. 2011). Association mapping on the other hand is based on linkage disequilibrium (LD) and can be used for QTL detection in

Table 1 Overview of the differences between the two mapping approaches

	MC-QTL mapping	Association mapping
Based on	Genetic linkage	Linkage disequilibrium
IBS/IBD usage	IBD at the parents level	IBS
Genome coverage	Markers and conditional probabilities; information every 1 cM	Only markers; information depending on density of the genetic map
Correction for family structure	Family effect, QTL effects nested within families	Inclusion of family effect in the model
Correction within families	Cofactors selected across families	Cofactors selected across families
Model	$Y = JM + X_qB_q + \sum_{c \neq q} X_cB_c + \varepsilon$	$Y = l\mu + X_fM_f + X_qb_q + \sum_{c \neq q} X_cb_c + \varepsilon$

IBS and IBD refer to identity-by-state and identity-by-descent, respectively. Details of the two models are given in "Materials and methods"

diverse panels of lines as well as in multiple families (Yu et al. 2008; Myles et al. 2009). LD is population specific and affected by different genetic factors (Flint-Garcia et al. 2003). Moreover, LD can be highly variable across the genome (Whitt et al. 2002; van Inghelandt et al. 2011; Würschum et al. 2011a) and in experiments based on multiple families is also generated by the experimental design (Verhoeven et al. 2006; Würschum et al. 2012a). The QTL detection power and the mapping resolution depend on the extent of LD. High LD between QTL and the markers facilitates the detection of QTL with medium or small effect, whereas lower LD only allows to detect QTL with large effects. A major issue for association mapping is the correction for population stratification (Sillanpää 2011), and for association mapping with multiple families, a family effect should be included in the model in addition to cofactors which correct for genetic background effects within families (Würschum et al. 2012a). Association mapping in multiple families has recently been used to detect QTL in diverse crops such as maize (e.g., Buckler et al. 2009; Lu et al. 2010; Liu et al. 2011), wheat (e.g., Reif et al. 2011a, b) or rapeseed (e.g., Würschum et al. 2012b). The major differences between the two mapping approaches are the use of identity-by-state (IBS) information in association mapping and identity-by-descent (IBD) probabilities in MC-QTL mapping. In addition, MC-QTL mapping operates with estimated conditional probabilities in the regions between the markers, whereas association mapping relies solely on the markers in the available genetic map. Therefore, association mapping can be expected to more strongly depend on the available marker density than MC-QTL mapping.

Whereas the advantages of QTL detection in multiple segregating families are obvious, it remains unclear which biometrical approach should best be applied to analyze a given population. The main goal of this study was to compare MC-QTL mapping and association mapping for QTL detection in a large elite maize population. In particular, the objectives of our study were to use cross-validation and resample model averaging to compare MC-QTL mapping and association mapping with regard to (1) the predictive power, (2) the precision of QTL estimation, and (3) their performance under different map densities. (4) In addition, both approaches were compared with regard to the detection of epistatic QTL.

Materials and methods

Plant materials and field experiments

The analyses are based on the population described in Steinhoff et al. (2011). In brief, six F_3 families, with a total of 788 individuals were obtained from a diallel cross between four dent inbreds (A, B, C, and D). The lines C and D are more closely related with each other as compared to the other two lines (Steinhoff et al. 2011). Each F_3 plant was selfed to obtain an $F_{3:4}$ family which are the bulked progenies of an individual F_3 plant. Testcross progenies were produced by mating the 788 $F_{3:4}$ families and the four parental inbreds to one elite inbred tester from the opposite heterotic pool and unrelated by pedigree. All plant materials used in this study are proprietary to Syngenta Seeds, Bad Salzuflen, Germany.

The testcross progenies were evaluated in 2007 in Italy at 10 locations with unreplicated trials. Two-row plots (8.4 m^2) were machine planted (8 plants m^{-2}) and harvested as grain trials. Data were collected for grain yield (Mg ha^{-1}), adjusted to a moisture concentration of 155 g kg^{-1}, and grain moisture (g kg^{-1}) at harvest.

Phenotypic data analyses and molecular data

In each environment, phenotypic data values were adjusted for block effects with four checks. Best linear unbiased estimates (BLUEs) across locations were determined by

assuming fixed genotypic effects for the testcross progenies and the parents. Analyses were performed using PROC GLM in the statistical software SAS (SAS Institute 2008).

For DNA extraction, each F_3 plant was represented by 15 bulked F_4 plants and genotyped with 857 SNP markers that have been checked for their adherence to the expected Mendelian segregation ratio. The consensus genetic map has a total length of 1,580 cM and an average interval length of 1.84 cM (Steinhoff et al. 2011). To assess the effect of different marker densities on the QTL mapping results of the two approaches, we constructed subsets of the genetic map. Five fixed marker complements were chosen to represent five different map densities (full marker set with 1.84, 5, 10, 15, and 20 cM) optimizing the spacing between the markers as well as possible (Fig. 3).

Linkage disequilibrium (LD) was assessed by the LD measure r^2 (Weir 1996) and significance of LD was tested with Fisher's exact tests (Hill and Robertson 1968). LD computations were performed with the software package Plabsoft (Maurer et al. 2008).

Multiple-line cross QTL mapping

QTL mapping was performed using the disconnected model suggested by Blanc et al. (2006), which models QTL effects as nested within families. Details are given in Steinhoff et al. (2011). In brief, the model used was:

$$\mathrm{Y} = \mathrm{JM} + \mathrm{X}_q\mathrm{B}_q + \sum_{c \neq q} \mathrm{X}_c\mathrm{B}_c + \varepsilon$$

where Y was a $N \times 1$ vector with the testcross BLUE values of the N progenies derived from P families. J was a $N \times P$ matrix whose elements were 0 or 1 according to whether or not individual i belonged to family p and M was a $P \times 1$ vector of family specific means. X_q (X_c) was a $N \times P$ matrix containing the expected number (ranging from 0 to 2) of allele k for each individual in family p at QTL q (cofactor c), B_q (B_c) was a $P \times 1$ vector of the expected allele substitution effects of QTL q (cofactor c) in family p, and ε was the vector of the residuals.

Cofactor selection was performed using the Schwarz (1978) Bayesian information criterion (SBC) implemented in PROC GLMSELECT implemented in the statistical software SAS (SAS Institute 2008). Testing for the presence of a putative QTL in an interval was performed with a likelihood-ratio test using statistical software R (Broman and Sen 2009). The experiment-wise type I error was determined to be $P_e < 0.10$, using 2,000 permutation runs (Churchill and Doerge 1994). Support intervals for the detected QTL were calculated based on a 1.5 LOD drop. For a better comparability with the results from the association mapping approach, a 10 cM window surrounding each detected QTL was applied. If another, colinear QTL was detected within that window, only the QTL with the higher LOD peak was kept.

The proportion of genotypic variance (p_G) explained by the detected QTL within families was calculated by a regression of the phenotypic values of the individuals minus the respective family mean (corresponds to the within family variance) on the detected QTL to obtain R^2_{adj}. The ratio $p_G = R^2_{adj}/h^2$ yielded the proportion of genotypic variance (Utz et al. 2000).

The epistasis scan for pairwise interactions was done with the model described above, which was extended by the term $\mathrm{X}_{q'}\mathrm{B}_{q'}$ for the second locus and the interaction term between the two loci q and q' $\mathrm{X}_{qq'}\mathrm{B}_{qq'}$. We used an α-level of 0.05 and followed the suggestion of Holland et al. (2002), dividing the α-level by the number of possible independent pairwise interactions between chromosome regions, assuming two separate regions per chromosome ($P < 2.6\mathrm{e}{-4}$).

Association mapping

Association mapping was done with a biometrical model which previously performed well in a comparison of different statistical approaches for association mapping in multiple families, Model B from Würschum et al. (2012a):

$$\mathrm{Y} = \mathrm{l}\mu + \mathrm{X}_f\mathrm{M}_f + \mathrm{X}_q\mathrm{b}_q + \sum_{c \neq q} \mathrm{X}_c\mathrm{b}_c + \varepsilon$$

In this model, Y is an $N \times 1$ column vector of the BLUE values of phenotypic data of N testcross progenies, coming from F families ($F = 6$); 1 is an $N \times 1$ column vector containing constant 1; μ is the intercept; X_q (X_c) is an $N \times 1$ column vector containing the marker information (coded as 0-1-2) of each individual at marker q (cofactor c); X_f is an $N \times F$ matrix whose elements are 0 or 1 according to whether an individual belongs to family f, and M_f is an $F \times 1$ vector of family effects; b_q (b_c) is the expected allele substitution effect of marker q (cofactor c); and ε is the vector of the residuals of the model.

In brief, an additive genetic model was chosen for the progenies as described by Utz et al. (2000). We applied a two-step procedure for QTL detection. In a first step, stepwise multiple linear regression was used to select a set of cofactors based on the Schwarz (1978) Bayesian criterion (SBC) with a model including a family effect and cofactors. Cofactor selection was performed using Proc GLMSELECT implemented in the statistical software SAS (SAS Institute 2008). In the second step, we calculated a P value for the association of each marker with the phenotypic value for the F test with a full model (with marker effect) against a reduced model (without marker effect) (for details see Reif et al. 2010). The applied model includes a

family effect, cofactors, and an SNP effect across families. The Bonferroni–Holm procedure (Holm 1979) was used to detect markers with significant ($P < 0.05$) main effects. The proportion of the genotypic variance explained by the detected QTL was calculated as described above. For the detection of epistasis, the model was extended and included the main effects of the two loci q and q', $X_q b_q$ and $X_{q'} b_{q'}$, and the interaction effect of the marker pair under consideration $X_{qq'} b_{qq'}$.

Cross-validation and resample model averaging

To evaluate the QTL mapping results of the two mapping methods, a fivefold cross-validation approach accounting for genotypic sampling was chosen (Utz et al. 2000; Schön et al. 2004; Liu et al. 2012). The data set (DS) was subdivided into five genotypic samples without replacement. To maintain the population structure and the relative contribution of the families to the data set, random genotypic sampling was done separately within each family and then combined across families. Four of the five genotypic samples were used as the estimation set (ES) for QTL detection, localization, and estimation of their genetic effects. The fifth genotypic sample remained as an independent sample to form the test set (TS). This TS was used to validate the QTL results from the ES and to obtain unbiased estimates of the QTL effects and the genotypic variance explained by the QTL. The randomization of genotypes in ES and TS was repeated 600 times. QTL detection was done in the DS and in the ES, whereas the TS was used to validate the results from the corresponding ES. The QTL effects estimated in the ES were used for prediction in the TS and to obtain R^2_{adj} between predicted and observed phenotypic values (Würschum et al. 2012a). The proportion of the genotypic variance of the detected QTL in the ES ($p_{G\text{-}ES}$) was compared with the proportion explained in the TS ($p_{G\text{-}TS}$) and the relative bias was calculated as $1 - (p_{G\text{-}TS}/p_{G\text{-}ES})$.

Our resample model averaging approach was similar to the subagging (80 %) described by Valdar et al. (2009). We used resampling without replacement as described for the cross-validation. In contrast to Valdar et al. (2009), we did not use forward selection to select the multiple-QTL model, but used QTL detection by association mapping or MC-QTL mapping.

Results

The linkage disequilibrium (LD) observed in the data set with 788 individuals genotyped with 857 SNP markers showed a decrease with increasing genetic map distance (Fig. 1a). For marker distances below 1 cM, the median r^2 was only around 0.2, whereas the upper quantile was around 0.5.

For the full data set (DS) including all markers, association mapping detected more QTL than MC-QTL mapping for both traits (Table 2). The proportion of genotypic variance (p_G) explained by these QTL, however, was higher for MC-QTL mapping than for association mapping. By contrast, the cross-validation approach revealed a higher proportion of explained genotypic variance in the test set (TS) for the association mapping QTL, despite the lower p_G in the DS and in the ES. This also shows in the higher relative bias in p_G estimates for MC-QTL mapping than for association mapping, i.e., the reduction of p_G from the estimation set to the test set. The unbiased estimates of p_G in the TS amounted to 12.6 and 9.9 % for

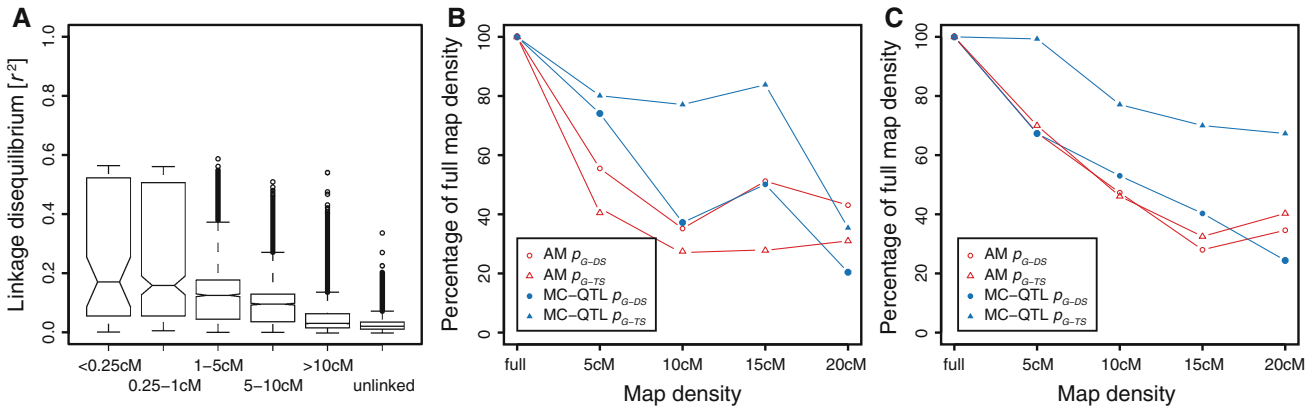

Fig. 1 Linkage disequilibrium and cross-validated performance of the two mapping approaches. **a** Distribution of linkage disequilibrium (LD) between linked marker pairs for different genetic map distances between the two markers as well as the LD observed between unlinked marker pairs (unlinked). **b**, **c** Proportion of explained genotypic variance relative to that observed with the full marker density for the QTL detected by association mapping (AM) and multiple-line cross QTL mapping (MC-QTL) for **b** grain yield and **c** grain moisture

Table 2 Comparison of the performance of association mapping (AM) and multiple-line cross QTL mapping (MCM) under different marker densities assessed by fivefold cross-validation

Number of detected QTL in the data set (QTL_{DS}) and averaged over estimation sets (QTL_{ES}). Percentage of the proportion of genotypic variance explained by the detected QTL in the data set ($p_{G\text{-}DS}$), the estimation set ($p_{G\text{-}ES}$), and in the test set ($p_{G\text{-}TS}$), as well as the relative bias ($p_{G\text{-}Bias}$)

	All markers		5 cM		10 cM		15 cM		20 cM	
	AM	MCM	AM	MCM	AM	MCM	AM	MCM	AM	MCM
Grain yield										
QTL_{DS}	9	7	5	7	4	4	4	7	5	4
$p_{G\text{-}DS}$	29.0	43.6	16.1	32.3	10.2	16.2	14.9	22.0	12.5	8.9
QTL_{ES}	7.3	8.9	5.0	6.2	2.5	5.0	2.5	4.5	2.7	2.2
$p_{G\text{-}ES}$	34.7	40.2	21.8	31.3	11.2	25.0	12.1	23.1	11.0	10.4
$p_{G\text{-}TS}$	12.6	9.9	5.1	8.0	3.1	6.8	3.5	8.3	3.9	3.5
$p_{G\text{-}Bias}$	63.7	75.4	76.6	74.4	72.3	72.8	71.1	64.1	64.5	66.3
Grain moisture										
QTL_{DS}	22	16	12	13	8	12	5	9	6	4
$p_{G\text{-}DS}$	48.6	63.2	32.8	42.5	23.0	33.5	13.6	25.5	16.8	15.4
QTL_{ES}	17.1	15.6	10.8	12.5	7.1	10.4	4.8	8.0	5.3	6.1
$p_{G\text{-}ES}$	53.5	50.0	37.6	42.8	24.7	36.2	17.4	29.5	19.2	24.0
$p_{G\text{-}TS}$	24.3	15.3	17.0	15.2	11.2	11.8	7.9	10.7	9.8	10.3
$p_{G\text{-}Bias}$	54.6	69.4	54.8	64.5	54.7	67.4	54.6	63.7	49.0	57.1

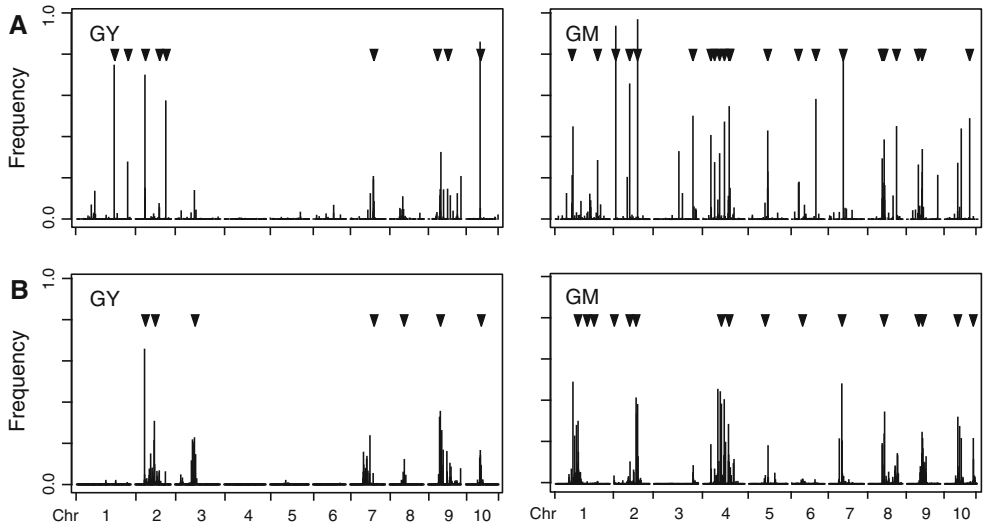

Fig. 2 Frequency distributions of QTL detected by **a** association mapping or **b** MC-QTL mapping in 600 resample model averaging (RMA) runs for grain yield (GY) and grain moisture (GM). The *arrowheads* indicate the positions of the QTL detected with the full data set

grain yield and 24.3 and 15.3 % for grain moisture, for association mapping and MC-QTL mapping, respectively.

The resample model averaging (RMA) revealed that many of the QTL positions identified by association mapping or MC-QTL mapping were identical between the two approaches (Fig. 2). There were, however, also QTL that were only detected by one of the approaches, but not by the other. Some of the QTL detected with the full data set were detected with a high frequency in the RMA runs, whereas other QTL were only detected in a small number of RMA runs. For both traits, the frequency of RMA runs in which a QTL was detected was generally lower for MC-QTL mapping than for association mapping. Another difference that became apparent in the RMA approach was that the peaks which consequently define the position of the QTL were much broader for MC-QTL mapping. With regard to the detected QTL, we found that generally the results from association mapping and from MC-QTL mapping were in good accordance, and the LOD profile and the $-\log_{10}(P$ value) bars followed the same trend (Figs. 4, 5). At positions where a peak in the LOD curve indicated a QTL, there was often also a significant P value observed in association mapping.

We compared the results obtained with the full map with results obtained with lower map densities to assess the effect of different genetic map densities. To this end, we constructed genetic maps that were subsets of the full marker map, which had average map distances of 5, 10, 15,

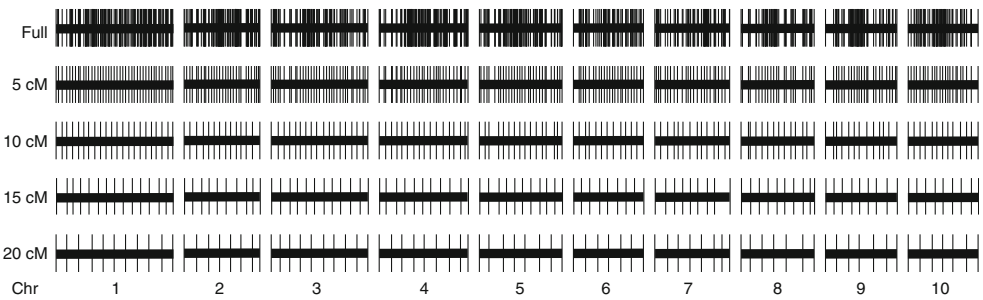

Fig. 3 Genetic maps with the different marker densities used in this study. The full map contains all 857 markers, whereas the others represent four different average map densities (5, 10, 15, and 20 cM) optimizing the spacing between the markers as well as possible

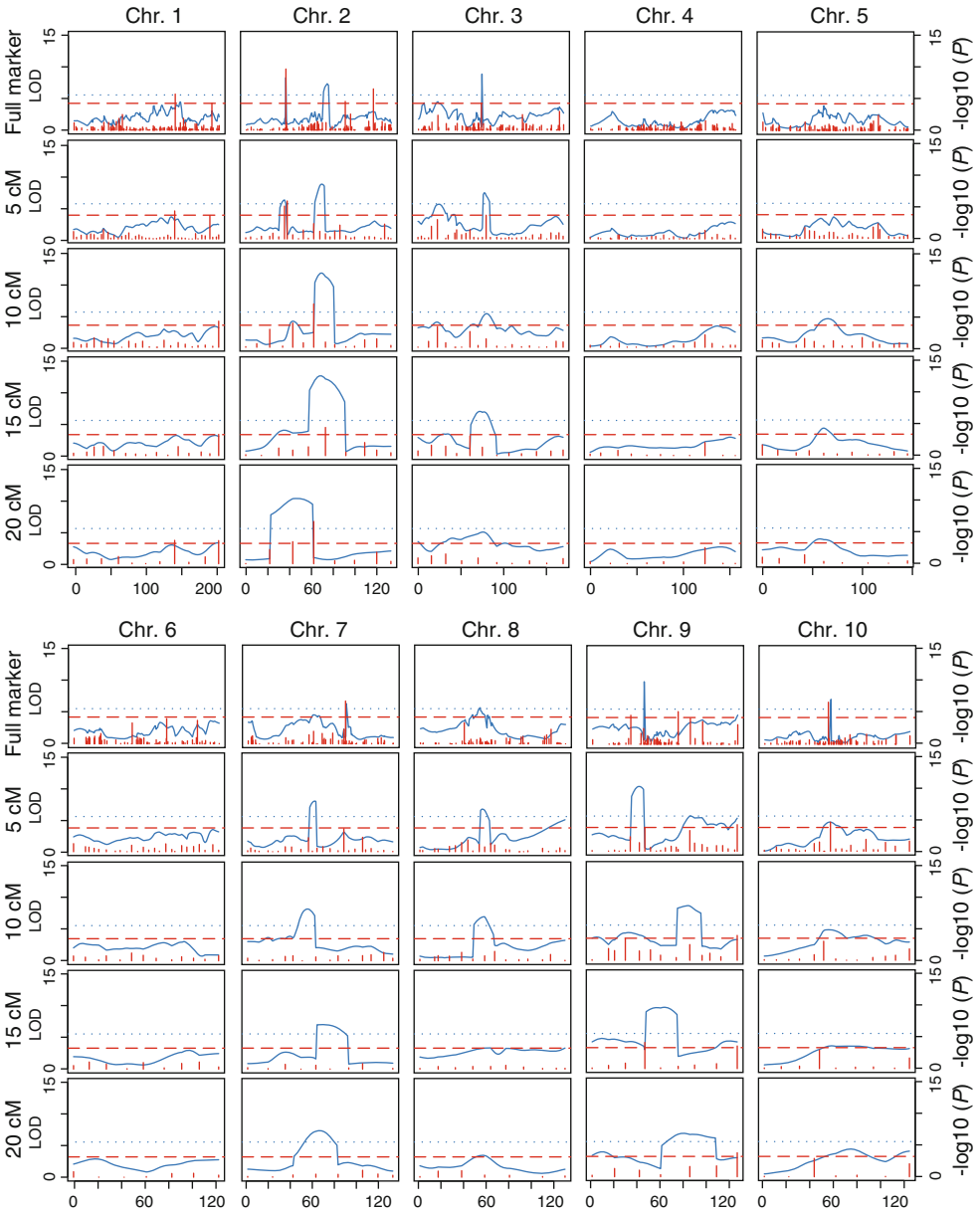

Fig. 4 Comparison of association mapping and MC-QTL mapping results for grain yield based on the different genetic maps. LOD score profiles from MC-QTL mapping (in *blue*) and $-\log_{10}(P$ values) from association mapping as *vertical bars* (in *red*) for all ten chromosomes. The *dotted line* indicates the threshold for MC-QTL mapping, which was determined by permutation tests, and the *dashed line* indicates the significance threshold for association mapping ($P < 0.05$ Bonferroni corrected). The positions on the chromosomes are given in cM (color figure online)

and 20 cM (Fig. 3). The markers in the reduced maps were chosen to optimize the spacing between them. We found that for these reduced marker maps, the number of detected QTL and the proportion of genotypic variance explained in the data set were reduced compared to the full marker set for both traits (Table 2). By contrast, the relative bias was

comparable for all map densities. In comparison to the results from the full marker set, there was an almost linear decrease of the p_G observed in the DS and in the TS with reduced map density for grain moisture, whereas for grain yield there was mainly a strong decrease from the full marker set to the 5 cM map density (Fig. 1b, c). Whereas the decrease in p_G was comparable for association mapping and MC-QTL mapping for the DS, the p_G in the TS was much less affected for MC-QTL mapping. The p_G in the TS that could be realized with the lowest map density (20 cM) was relatively low with approximately 4 % for grain yield and around 10 % for grain moisture (Table 2). Comparing the LOD profiles from MC-QTL mapping and the *P* values from association mapping for the different map densities, we found that the LOD profiles at QTL positions were rather narrow for the full marker set, but became broader with decreasing map density (e.g., grain moisture QTL on chromosome 9) (Figs. 4, 5). In addition, some LOD peaks also shifted position and became flatter or even non-significant (e.g., grain yield QTL on chromosome 2 or grain moisture QTL on chromosome 7). The RMA based on the different map densities revealed that some QTL that were detected with the full marker set were not detected any more with reduced map densities (Fig. 6). This reduction in the number of detected QTL was comparable for both approaches, association mapping and MC-QTL mapping (Table 2). The reduction in the frequency of RMA runs in which a locus was identified as QTL appeared more pronounced for MC-QTL mapping than for association mapping (Fig. 6).

We performed a full two-dimensional scan for epistatic interactions for association mapping and MC-QTL mapping with the full data set including all markers. For MC-QTL mapping, we observed five regions for grain yield and four for grain moisture, which potentially indicate an epistatic interaction (Fig. 7). By contrast, many more marker–marker interactions were detected by association mapping. For grain yield, 345 significant interactions were detected and for grain moisture, 263. There were, however, also clusters of significant interactions which are more likely to be true-positive epistatic QTL. Some loci appeared to interact with many other loci throughout the genome (e.g., locus on chromosome 1 for grain moisture). The comparison of the epistatic QTL detected with the two approaches revealed little congruency.

Discussion

Statistical properties of the two biometrical approaches

QTL detection in multiple segregating families, as opposed to single biparental populations, is becoming the method of choice for both, plant scientists and plant breeders. Different biometrical approaches for QTL detection are propagated and the aim of this study was to compare the two most prominent ones: association mapping and multiple-line cross QTL mapping (Blanc et al. 2006; Yu et al. 2008; Würschum 2012). MC-QTL mapping is a composite interval mapping (CIM) approach (Jansen and Stam 1994; Zeng 1994), in which cofactors are selected to control the genetic background, and mapping is done based on marker information as well as on conditional probabilities for the regions not covered by markers. It is an identity-by-descent approach that allows tracing the parental origin of alleles and therefore provides information beyond that of the marker status. Association mapping on the other hand is based on association mapping methodology and purely uses the identity-by-state (IBS) information contained in the marker, irrespective of which parent the allele is derived. The advantage of association mapping is that it potentially enables a higher mapping resolution and thus more accurate QTL positions for an implementation in marker-assisted selection programs. This is, however, dependent on the LD present in the population under study. Association mapping is based on LD between markers and the QTL and only high LD will facilitate the detection of medium or small effect QTL. This approach, therefore, strongly depends on the available marker density and is expected to perform better with high map densities.

The LD observed in this population was comparable to a recent study in elite maize (Van Inghelandt et al. 2011) and showed a rapid decline with genetic map distance (Fig. 1a). For closely linked markers, we observed moderate LD, which likely enables the detection of QTL with main or medium effects, whereas QTL with small effects may escape detection unless they are by chance in high LD with a marker. It has recently been shown that the choice of an appropriate biometrical model is crucial for association mapping and that a model incorporating cofactors and an effect for the segregating family performed best (Würschum et al. 2012a). A recent analysis using both MC-QTL mapping models has shown that for the data set underlying this study, the disconnected model performed much better (Steinhoff et al. 2011) and we, therefore, used that model for the comparison study.

Comparison of association mapping and MC-QTL mapping

Whereas association mapping detected more QTL, the proportion of genotypic variance explained by the QTL was higher for MC-QTL mapping (Table 2). The relative bias in p_G estimates was, however, lower for association mapping, indicating that these estimates are more robust.

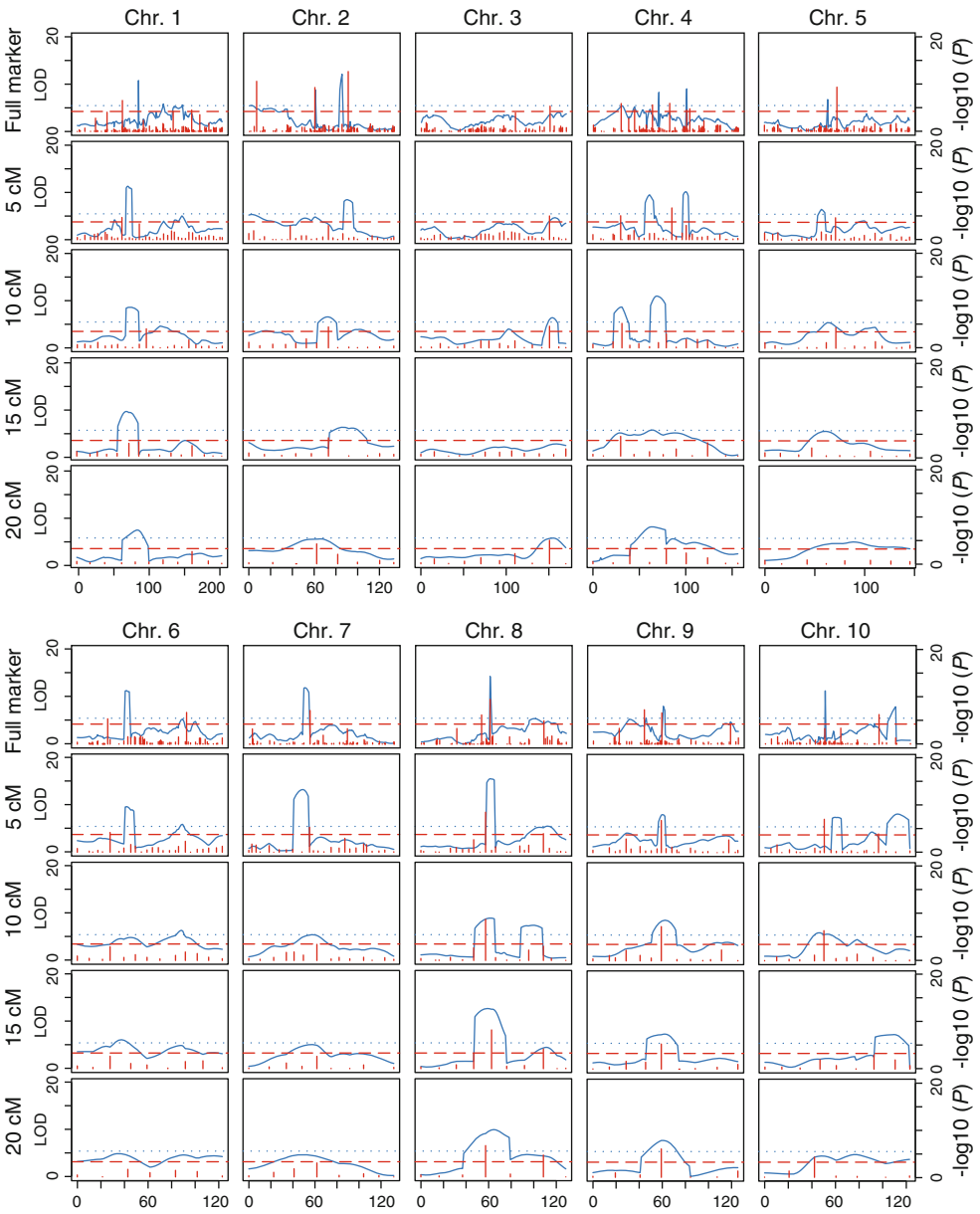

Fig. 5 Comparison of association mapping and MC-QTL mapping results for grain moisture based on the different genetic maps. LOD score profiles from MC-QTL mapping (in *blue*) and $-\log_{10}(P$ values) from association mapping as *vertical bars* (in *red*) for all ten chromosomes. The *dotted line* indicates the threshold for MC-QTL mapping, which was determined by permutation tests, and the *dashed line* indicates the significance threshold for association mapping ($P < 0.05$ Bonferroni-corrected). The positions on the chromosomes are given in cM (color figure online)

For both traits, the unbiased estimate of p_G obtained by cross-validation ($p_{G\text{-}TS}$) was higher for association mapping pointing to a higher predictive power of this approach. This difference was, however, not big and for both traits the cross-validated genotypic variance was rather low considering that almost 800 individuals were examined in this study. Schön et al. (2004) reported a much higher cross-validated genotypic variance in a biparental population of similar size. Our results with the full data set are slightly lower, but comparable to those obtained by Blanc et al. (2006) who also used six families and a total population size of 900 individuals. As all these studies were based on elite breeding material, it appears unlikely that many more QTL with large effects were segregating in the study of Schön et al. (2004). QTL effects may vary substantially between populations (Liu et al. 2011; Steinhoff et al. 2011), which is also a major drawback of QTL detection in single biparental populations, as the results are often not transferable across populations (Holland 2007). This opens another explanation for the different magnitude of validated genotypic variance between the studies. The difference between our study and that of Schön et al. (2004) is that even though both used cross-validation, this is in the former case done in multiple families in each of which the detected QTL may have different effects. Even though lower in number, these results probably reflect a more realistic estimate of the genotypic variance that can be realized by including these QTL in knowledge-based breeding.

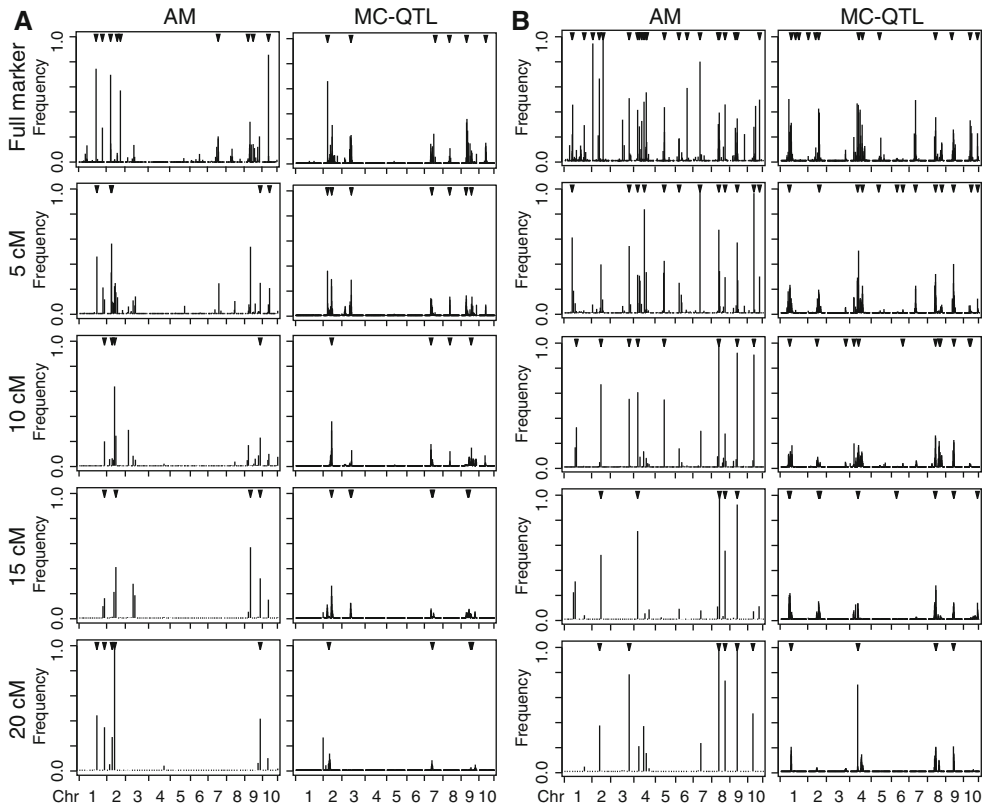

Fig. 6 Frequency distributions of QTL detected by association mapping (AM) or MC-QTL mapping in 600 resample model averaging (RMA) runs based on the different genetic maps for **a** grain yield and **b** grain moisture. The *arrowheads* indicate the positions of the QTL detected with the respective marker set with the full set of lines

The RMA approach revealed that many of the QTL positions were similar between association mapping and MC-QTL mapping (Fig. 2). As we observed also QTL that were only identified by one of the approaches, both also possess distinct properties and may be used complementarily. The QTL which were identified in a high number of RMA runs are more reliable, whereas those that were only detected with a low frequency have a higher probability of being false-positive QTL. The peaks at the QTL positions were much broader for MC-QTL mapping than for association mapping. This indicates that association mapping possesses a higher precision of QTL estimation than MC-QTL mapping. This is further supported by the plot of the LOD profiles and the $-\log_{10}(P \text{ value})$ bars, which show the same trendline, but at overlapping QTL positions the LOD peaks are generally broader (Figs. 4, 5).

QTL detection under different map densities

For many agriculturally important crops, SNP markers have become the marker type of choice and thousands or even tens of thousands of SNPs are available. Even though costs for genotyping are constantly decreasing, it still remains an issue and it is unclear how many of the available markers should be used in a customized analysis. To address this question, we performed the full analysis by association mapping and MC-QTL mapping using different marker densities (Fig. 3). We found that for both approaches, there was a comparable reduction in the number of detected QTL. For the absolute values of cross-validated genotypic variance ($p_{\text{G-TS}}$) under reduced map densities, MC-QTL mapping was better for grain yield, whereas for grain moisture the two approaches were not much different (Table 2). Considering the $p_{\text{G-TS}}$ in relation to that realized with the full marker set, it appeared that MC-QTL mapping was less susceptible to a reduction in marker density and the results remained relatively stable, except for grain yield under the lowest map density (Fig. 1b, c). This decrease in the number of detected QTL and in the predictive power was expected for association mapping, as fewer markers reduce the probability of a QTL being in LD with one of them. By contrast, MC-QTL mapping operates with conditional probabilities and as expected was found to be more robust and less affected by the reduced number of markers. A possible reason for the observed decrease of $p_{\text{G-TS}}$ under low map densities is that not every marker segregates in each of the families, such that the gaps between adjacent markers in the single families can be considerably larger

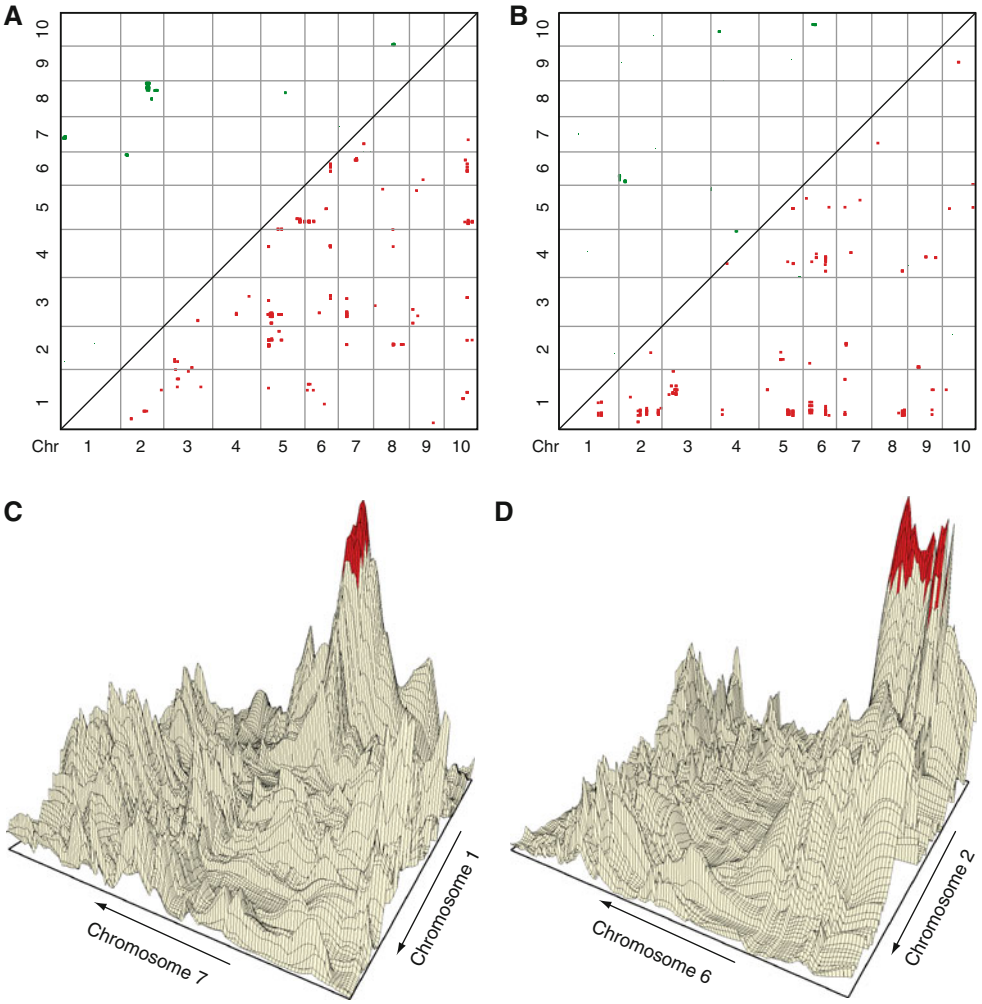

Fig. 7 Detection of epistatic QTL for **a** grain yield and **b** grain moisture by MC-QTL mapping (*above diagonal in green*) and association mapping (*below diagonal in red*). **c**, **d** Landscape of selected epistatic interactions detected by MC-QTL mapping for **c** a grain yield epistatic QTL, and **d** a grain moisture epistatic QTL. The $-\log_{10}(P$ values) are plotted and epistatic interactions below the significance threshold are shown in *red* (color figure online)

than suggested by the average map distance (Table 3). If these gaps exceed a certain size, the conditional probability estimates become less accurate resulting in a loss in QTL detection power and in the predictive power. For linkage mapping in biparental populations, 20 cM is considered to be the maximum distance between markers (e.g., Piepho 2000). Consistent with this, we observed a rapid decrease in $p_{\text{G-TS}}$ estimates for MC-QTL mapping when the marker densities in single families dropped below this threshold (compare 15 and 20 cM average consensus map density) (Fig. 1b; Table 3). This highlights that for the joint analysis of multiple families, it is the average map distance in the single families and not that in the consensus map that determines the QTL detection power.

Considering the constant decline of the cross-validated genotypic variance with marker density indicates that the predictive power could also in the full marker set be increased further by applying more markers. This is especially true for association mapping, but to a lesser extent also for MC-QTL mapping.

Detection of epistasis

Epistasis refers to interactions between two or more loci in the genome (Carlborg and Haley 2004) and epistatic QTL have recently been shown to be involved in the expression of complex agronomic traits in diverse crops (e.g., Reif et al. 2011a, b; Würschum et al. 2011b). Blanc et al. (2006), using the connected model, detected both QTL times background interactions and also QTL times QTL epistasis. A problem that generally arises with epistasis scans is the choice of an appropriate significance threshold. Here, we followed the suggestion of Holland et al. (2002) and corrected the α-level of 0.05 for the number of possible independent pairwise interactions assuming two independent regions on each chromosome (separated by the centromere). We detected epistatic QTL for grain yield and for grain moisture with both biometrical approaches (Fig. 7a, b). The major difference was in the number of detected interactions. Whereas only few were detected by MC-QTL mapping, many were detected by association mapping

Table 3 Number of polymorphic markers in each family shown for the different marker densities used in this study

Family	All markers 857 (14.2)	5 cM 289 (14.2)	10 cM 154 (17.1)	15 cM 105 (27.4)	20 cM 82 (29.3)
A × B	475 (32.5)	143 (44.0)	73 (81.6)	54 (87.8)	42 (81.6)
A × C	503 (24.5)	170 (34.9)	90 (111.2)	64 (102.9)	43 (111.2)
A × D	487 (31.1)	167 (40.9)	97 (54.3)	64 (107.6)	49 (66.8)
B × C	498 (24.0)	177 (34.6)	99 (60.3)	62 (77.2)	55 (60.7)
B × D	528 (24.0)	184 (28.3)	98 (49.1)	64 (59.8)	57 (59.8)
C × D	366 (42.4)	129 (45.8)	67 (79.8)	46 (61.9)	36 (82.8)

The size of the largest gap in cM is given in brackets

which were scattered throughout the entire genome. Interestingly, there was little congruency between the association mapping and the MC-QTL mapping results, which again highlights that despite similarities both approaches have their own specific properties.

Recent analyses have revealed the existence of epistatic master regulators, loci which are involved in a large number of interactions (Reif et al. 2011b; Würschum et al. 2011a). At least one such locus was identified here, namely the locus on chromosome 1 affecting grain moisture (Fig. 7b). This further substantiates the hypothesis that such epistatic master regulators are involved in shaping the expression of complex traits. Due to the large number of epistatic interactions identified by association mapping, we focused on the epistatic QTL detected by MC-QTL mapping. The interactions involved regions which sometimes were also identified as main effect, but mainly regions in which no main effect QTL was detected and which appear only to be involved in epistatic interactions. This is in accordance with results reported in sugar beet where the majority of the interacting regions had no main effect (Würschum et al. 2011b). The epistatic interaction landscape is illustrated by the three-dimensional plots (Fig. 7c, d).

Conclusions

We compared association mapping and linkage mapping (MC-QTL) for QTL detection in multiple segregating families. We found that with the full marker density, association mapping realized a higher cross-validated genotypic variance, whereas with lower marker densities, this picture reversed and MC-QTL mapping performed better. Both approaches, however, are likely to perform even better with higher marker densities, and the strong relative bias observed for both traits suggests that a cross-validation strategy should be employed irrespective of the applied biometrical approach.

Acknowledgments This research was conducted within the *Biometric and Bioinformatic Tools for Genomics based Plant Breeding* Project of the GABI-FUTURE initiative.

References

Blanc G, Charcosset A, Mangin B, Gallais A, Moreau L (2006) Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize. Theor Appl Genet 113:206–224

Broman KW, Sen S (2009) A guide to QTL mapping with R/qtl. Springer Science + Business Media, LLC

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ et al (2009) The genetic architecture of maize flowering time. Science 325:714–718

Carlborg Ö, Haley CS (2004) Epistasis: too often neglected in complex trait studies? Nat Rev Genet 5:618–625

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

Coles ND, McMullen MD, Balint-Kurti PJ, Pratt RC, Holland JB (2010) Genetic control of photoperiod sensitivity in maize revealed by joint multiple population analysis. Genetics 184:799–822

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Ann Rev Plant Biol 54:357–374

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–231

Holland JB (2007) Genetic architecture of complex traits in plants. Curr Opin Plant Biol 10:156–161

Holland JB, Portyanko VA, Hoffmann DL, Lee M (2002) Genomic regions controlling vernalization and photoperiod responses in oat. Theor Appl Genet 105:113–126

Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat 6:65–70

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

Liu W, Gowda M, Steinhoff J, Maurer HP, Würschum T et al (2011) Association mapping in an elite maize breeding population. Theor Appl Genet 123:847–858

Liu W, Maurer HP, Reif JC, Melchinger AE, Utz HF et al (2012) Optimum design of family structure and allocation of resources in association mapping with lines from multiple crosses (in review)

Lu Y, Zhang S, Shah T, Xie C, Hao Z, Li X et al (2010) Joint linkage–linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. Proc Natl Acad Sci USA 107:19585–19590

Maurer HP, Melchinger AE, Frisch M (2008) Population genetic simulation and data analysis with Plabsoft. Euphytica 161:133–139

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z et al (2009) Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21:2194–2202

Piepho HP (2000) Optimal marker density for interval mapping in a backcross population. Heredity 84:437–440

Reif JC, Liu W, Gowda M, Maurer HP, Möhring J et al (2010) Genetic basis of agronomically important traits in sugar beet

(*Beta vulgaris* L.) Investigated with joint linkage association mapping. Theor Appl Genet 8:1489–1495

Reif JC, Gowda M, Maurer HP, Longin CFH, Korzun V et al (2011a) Association mapping for quality traits in soft winter wheat. Theor Appl Genet 122:961–970

Reif JC, Maurer HP, Korzun V, Ebmeyer E, Miedaner T et al (2011b) Mapping QTLs with main and epistatic effects underlying grain yield and heading time in soft winter wheat. Theor Appl Genet 123:283–292

SAS Institute Inc (2008) SAS user's guide, version 9.2. SAS Institute: Cary, NC, USA

Schön CC, Utz HF, Groh S, Truberg B, Openshaw S, Melchinger AE (2004) Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. Genetics 167:485–498

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464

Sillanpää MJ (2011) Overview of techniques to account for confounding due to population stratification and cryptic relatedness in genomic data association analyses. Heredity 106:511–519

Steinhoff J, Liu W, Maurer HP, Würschum T, Longin FH et al (2011) Multiple-line cross QTL-mapping in European elite maize. Crop Sci 51:2505–2516

Utz HF, Melchinger AE, Schön CC (2000) Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. Genetics 154:1839–1849

Valdar W, Holmes CC, Mott R, Flint J (2009) Mapping in structured populations by resample model averaging. Genetics 182:1263–1277

Van Inghelandt D, Reif JC, Dhillon BS, Flament P, Melchinger AE (2011) Extent and genome-wide distribution of linkage disequilibrium in commercial maize germplasm. Theor Appl Genet 123:11–20

Verhoeven KJF, Jannink JL, Mclntyre LM (2006) Using mating designs to uncover QTL and the genetic architecture of complex traits. Heredity 96:139–149

Weir BS (1996) Genetic data analysis II, 2nd edn. Sinauer Associates, Sunderland

Whitt SR, Wilson LM, Tenaillon MD, Gaut BS, Buckler ES (2002) Genetic diversity and selection in the maize starch pathway. Proc Natl Acad Sci 99:12959–12962

Würschum T (2012) Mapping QTL for agronomic traits in breeding populations. Theor Appl Genet. doi:10.1007/s00122-012-1887-6

Würschum T, Maurer HP, Kraft T, Janssen G, Nilsson C et al (2011a) Genome-wide association mapping of agronomic traits in sugar beet. Theor Appl Genet 123:1121–1131

Würschum T, Maurer HP, Schulz B, Möhring J, Reif JC (2011b) Genome-wide association mapping reveals epistasis and genetic interaction networks in sugar beet. Theor Appl Genet 123:109–118

Würschum T, Liu W, Gowda M, Maurer HP, Fischer S et al (2012a) Comparison of biometrical models for joint linkage association mapping. Heredity 108:332–340

Würschum T, Liu W, Maurer HP, Abel S, Reif JC (2012b) Dissecting the genetic architecture of agronomic traits in multiple segregating populations in rapeseed (*Brassica napus* L.). Theor Appl Genet 124:153–161

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178:539–551

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