Theor Appl Genet (2013) 126:2563–2574
DOI 10.1007/s00122-013-2155-0

ORIGINAL PAPER

High-density genotyping: an overkill for QTL mapping? Lessons learned from a case study in maize and simulations

Michael Stange · H. Friedrich Utz · Tobias A. Schrag · Albrecht E. Melchinger · Tobias Würschum

Received: 27 March 2013 / Accepted: 5 July 2013 / Published online: 17 July 2013

Abstract High-density genotyping is extensively exploited in genome-wide association mapping studies and genomic selection in maize. By contrast, linkage mapping studies were until now mostly based on low-density genetic maps and theoretical results suggested this to be sufficient. This raises the question, if an increase in marker density would be an overkill for linkage mapping in biparental populations, or if important QTL mapping parameters would benefit from it. In this study, we addressed this question using experimental data and a simulation based on linkage maps with marker densities of 1, 2, and 5 cM. QTL mapping was performed for six diverse traits in a biparental population with 204 doubled haploid maize lines and in a simulation study with varying QTL effects and closely linked QTL for different population sizes. Our results showed that high-density maps neither improved the QTL detection power nor the predictive power for the proportion of explained genotypic variance. By contrast, the precision of QTL localization, the precision of effect estimates of detected QTL, especially for small and medium sized QTL, as well as the power to resolve closely linked QTL profited from an increase in marker density from 5 to 1 cM. In conclusion, the higher costs for high-density genotyping are compensated for by more precise estimates of parameters relevant for knowledge-based breeding, thus making an increase in marker density for linkage mapping attractive.

Communicated by J. Yan.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-013-2155-0) contains supplementary material, which is available to authorized users.

M. Stange · H. F. Utz · T. A. Schrag · A. E. Melchinger
Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70599 Stuttgart, Germany

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

Introduction

Quantitative trait locus (QTL) mapping is well established in genetic studies of plants to unravel the genetic architecture of quantitative traits and to identify QTL for knowledge-based breeding. Despite the successful implementation of association mapping for crop genetics, the classical linkage mapping in biparental populations still provides some advantages over the more recently established approaches. Especially for rare alleles, which would escape detection in association mapping approaches, linkage mapping offers a high QTL detection power due to the balanced allele frequencies in segregating populations (Würschum 2012).

Owing to the rapid advances in sequencing technologies, huge numbers of SNPs can nowadays readily be identified (Yan et al. 2009). Consequently, the marker type of choice are SNPs, which are commonly applied using high-throughput platforms. The availability of this novel source of high-density genotyping is already well exploited in genome-wide association mapping studies and genomic selection in maize (*Zea mays* L.) (Riedelsheimer et al. 2012a, b). In contrast, linkage mapping studies in maize were, until now, mostly based on low-density genetic linkage maps constructed with simple sequence repeat (SSR) markers (e.g., Ma et al. 2007; Guo et al. 2011; Martin et al. 2012). This raises the question as to whether linkage mapping in biparental populations would profit from the exploitation of the available high-density SNP arrays or if that would be a waste of resources.

High-density maps could increase the probability that at least one marker is located in each intact chromosomal segment between two recombination breakpoints in the genome (Yu et al. 2011). This could facilitate the detection of QTL within these linkage blocks. The number of recombination events and consequently the number and length of the linkage blocks depends on the population size (Yu et al. 2011). The influence of population size, number of markers, and gene effects on QTL detection power and confidence intervals of detected QTL was analyzed by Darvasi et al. (1993) in a simulation study for interval mapping based on a backcross population. They concluded that a marker density of 10–20 cM is by far sufficient for precise QTL detection and that higher marker densities have no advantages irrespective of population size and size of genetic effects. In agreement with these findings, Piepho (2000) showed in an analytical approach that a marker density below 10 cM has only negligible effects on the standard error of genetic effect estimates and the power of QTL detection.

In contrast, comparing low- and high-density QTL mapping in barley, Hori et al. (2003) concluded that higher marker densities are beneficial, since markers tightly linked to QTL are available that can be used directly for marker-assisted breeding. Further, separate detection of two tightly linked QTL was only possible with the high-density map. A higher detection power and resolution of QTL mapping with a high-density SNP map compared to a low-density RFLP plus SSR map was also found by Yu et al. (2011) in rice, especially for QTL with small genetic effects. A higher precision of QTL localization with high-density SNP maps was also reported for maize by Shi et al. (2011) and Almeida et al. (2012). The power of separating linked QTL was analyzed by Li et al. (2010) in an inclusive composite interval mapping (ICIM) simulation study of DH lines with marker densities ranging from 5 to 40 cM. However, the above-mentioned studies did not exploit the current arrays of SNPs routinely applied in maize genotyping, which offer much higher marker densities (50 k). This leads to the question if high-density genotyping, resulting in genetic linkage maps with a marker density of 1 cM, could further increase the power and precision of QTL mapping and enable a better resolution of tightly linked QTL with linkage distances below 10 cM.

To address the question, if high-density genotyping offers advantages for QTL mapping in biparental populations, we employed the same high-density linkage map in combination with both, experimental data and a simulation study. In particular, the objectives of this study were to investigate the effect of high-density versus low-density linkage maps in QTL mapping on (1) the power of QTL detection, (2) the precision of QTL localization, (3) the estimation of QTL effects, and (4) the potential to resolve tightly linked QTL.

Materials and methods

Germplasm

This study was based on 227 DH lines derived from a biparental cross between UH009 and UH007 which was described in detail earlier by Stange et al. (2013). Shortly, the parental lines UH009 and UH007 are flint inbreds developed by the maize breeding program of the University of Hohenheim. All DH lines were developed using the in vivo method described by Prigge and Melchinger (2012), where F_1 plants of each cross were pollinated by an inducer line followed by the identification of haploid seeds using an embryo color marker. Chromosome doubling was promoted by colchicine treatment to produce D_0 plants, which were then self-pollinated to produce the D_1 generation.

Field experiments

All 227 DH lines and both parental inbred lines as duplicate entries were tested together with other lines in two 10 × 20 α-designs with two replications. Experimental units were single-row plots with a length of 3 m, spaced 0.75 m apart, and comprising 20 plants. The trials were conducted in 2 years at two locations in Southwest Germany, namely Stuttgart-Hohenheim and Eckartsweier.

Fusarium graminearum Schwabe causes Giberella ear rot (GER) which leads to contamination of grain with deoxynivalenol (DON) (Pestka 2007). The resistance of DH lines to GER, DON, and additionally, days to silking (DSI) were recorded in 2008 and 2009 as described in detail by Martin et al. (2011). Briefly, artificial silk channel inoculation was performed using an aggressive isolate of *F. graminearum* on six (2008) or eight (2009) primary ears 5–6 days after silk emergence. At physiological maturity, ears were manually dehusked and visually rated for GER severity. After manual harvest of the inoculated ears and drying to an approximate moisture content of 14 %, 100 g of ground grain was taken for prediction of DON concentration on the natural log scale by near-infrared spectroscopy (NIRS) using the calibration of Bolduan et al. (2009b). Grain yield and related traits were recorded in 2009. Primary ears from five non-inoculated plants per plot were manually harvested, dried down to constant weight, and shelled to determine the grain yield (GY) in g per plant. The 100-kernel weight (HKW) in g was determined from a sample of 300 kernels and kernel number (KN) per plant was determined by counting seeds.

Phenotypic data analyses

Standard lattice analyses of the six phenotypic traits of population UH009 × UH007 were performed using software

PLABSTAT (Utz 2005). For these analyses, GER severity data were transformed using the arcsine square root function to reduce heterogeneity of variances and to meet the normality assumption. The DON concentrations, predicted on the natural log scale, were back-transformed with the natural exponential function to approximate the values in milligrams per kilogram. Heritability (h^2) on an entry-mean basis was estimated according to Hallauer et al. (2010).

Marker analyses and linkage map construction

The Illumina MaizeSNP50 Bead Chip (Illumina Inc. San Diego, USA) was used for genotyping all DH lines and their parental inbreds using bulks of up to six plants per line. For each SNP and DH line, the quality criteria as described in detail by Stange et al. (2013) were applied. In total, 7,063 SNPs and 204 DH lines met these criteria and were used for all further analyses. In addition to the SNP analyses, the DH lines had been genotyped by 106 polymorphic SSRs, as described in detail by Martin et al. (2011). Linkage blocks of the DH lines of the experimental population for all ten chromosomes are shown in Figure S1.

In a first step, a framework map was constructed to cluster the SNPs into linkage groups using software MSTMap (Wu et al. 2008). In a second step, SSRs were assigned to these linkage groups based on the linkage group information from Martin et al. (2011) and the IBM2 2008 Neighbors map accessible through the Maize Genetics and Genomics Database (Lawrence et al. 2008). Finally, chromosome-wise genetic map construction was performed for the combined set of SNPs and SSRs using software JoinMap 4.0 (Van Ooijen 2006) as described in detail by Stange et al. (2013). To assess the effect of different map densities on QTL mapping results, subsets of the full genetic map were constructed. Originating from this full genetic map with 7,169 SNP and SSR markers, we optimized the spacing between markers as well as possible (Liu et al. 2012) by choosing one marker per cM in polymorphic regions (MD = 1; 682 markers). Originating from this dense map, two more sparse maps with average marker densities in polymorphic regions of 2 (MD = 2; 439 markers) and 5 cM (MD = 5; 257 markers) were constructed.

QTL mapping

QTL mapping in the experimental population was based on adjusted entry means of DH lines averaged across environments for all six traits. For DON concentration, QTL analyses were performed with means on the natural log scale, and for GER severity, means were back-transformed to percent values. The software PlabMQTL (Utz 2012) which applies Composite Interval Mapping (CIM) with a multiple regression approach (Haley and Knott 1992), was used for the detection of QTL positions and estimations of effects for all three marker densities (MD = 1, 2, and 5 cM). The appropriate number of cofactors was chosen on basis of the smallest values of the modified Bayesian Information Criterion (mBIC; Baierl et al. 2006). Critical LOD thresholds were determined empirically with 1,000 random permutations (Churchill and Doerge 1994) separately for each marker density averaged across all traits. Around each QTL, a 1-LOD support interval was specified. We applied an additive genetic model to fit positions and effects of QTL separately for each trait. To obtain unbiased estimates of QTL effects and of the proportion of genotypic variance explained by the QTL (p_G), fivefold cross-validation (CV) was conducted with 2,000 runs (Utz et al. 2000). The proportion of genotypic variance explained by the additive QTL model was obtained as $p_G = 100 \times R^2_{adj}/h^2$, where R^2_{adj} is the adjusted coefficient of determination explained by the model and h^2 is the heritability of the trait. The p_G explained by individual QTL, expressed as normalized value in percent of p_G, was calculated according to Prigge et al. (2012).

Names were assigned to QTL following the nomenclature proposed by Schaeffer et al. (2006), which combines the letter "*q*" for QTL, an abbreviation for the name of the trait, and a number for the QTL. For example, the third QTL detected for HKW was designated as "*qhkw3*".

Simulation study

The simulation study was designed in dependence on the experimental population UH009 × UH007 and therefore, performed with the same three genetic linkage maps. We simulated five independent QTL (IQ1–IQ5) with additive genetic effects of 0.10, 0.20, 0.30, 0.50, and 0.75 on five separate chromosomes (1, 2, 4, 9, and 10) (Table S2). Each QTL was located in a polymorphic region nearest to the center of the respective chromosome. Chromosomes which best met this criterion were used to simulate the five independent QTL (IQ1–IQ5) and the remaining chromosomes were used to simulate four pairs of linked QTL, LQ6a;LQ6b to LQ9a;LQ9b on chromosomes 3, 5, 6, and 8 (Table S2). For the pairs of linked QTL, two genetic distances (5 and 10 cM) and two linkage phases (coupling and repulsion) were assumed. The pairs of linked QTL were located in polymorphic regions of the four chromosomes. The additive genetic effects of both QTL per pair were set to 0.75 with different directions to simulate the two linkage phases. No interactions between all simulated QTL were assumed and no QTL was simulated on chromosome 7 to assess the false discovery rate. The heritability of the simulated trait was set to 0.75 which approximately corresponds the mean heritability of the six phenotypic traits of population UH009 × UH007.

A base population consisting of 220,000 DH lines was generated by crossing two parental inbred lines using software PLABSIM (Frisch et al. 2000). From this base population, the mapping populations were sampled with $N = 110$ (2,000 sets), 220 (1,000 sets), and 440 (500 sets). All three population sizes in combination with the linkage maps with three marker densities (MD = 1, 2, and 5 cM) were used for composite interval mapping (CIM) with an additive model in software PlabMQTL (Utz 2012) in the same way as described before for the six traits of the experimental population UH009 × UH007. To obtain unbiased estimates of QTL effects and p_G values, CV was conducted for all sets of a given population size with 20 fivefold CV runs. Critical LOD scores were determined empirically with 20 random permutations (Churchill and Doerge 1994) separately for each set of population size samples and marker density. The number of selected cofactors, the genome-wide number of detected QTL, $p_{G\text{-}DS}$ estimated in the data set, and the cross-validated $p_{G\text{-}TS}$ were evaluated averaged across all sets of a given population size.

For the independent QTL (IQ1–IQ5) and the pairs of linked QTL (LQ6a;LQ6b to LQ9a;LQ9b), the LOD scores and for IQ1–IQ5 also the 1-LOD support intervals around each QTL were averaged across all sets of a given population size. Chromosome-wise precision of QTL detection was evaluated for all simulated QTL averaged across all sets of a given population size as the deviation between (1) the simulated QTL position and the estimated QTL positions in cM and between (2) the reference genetic effects and the estimated genetic effects. The statistical power of QTL detection was calculated for IQ1–IQ5 as the frequency with which the simulated QTL was correctly identified within a predefined interval of ±5 cM centered around the predefined QTL position among all sets of a given population size. Following Li et al. (2010), QTL detected on the chromosome, where no QTL was located (chromosome 7) or detected outside the predefined interval, were assumed to be false positive QTL. False discovery rate (FDR) was calculated as the proportion of false positives in all simulation sets for a given population size.

The power to resolve linked QTL (e.g., separate detection of LQ6a and LQ6b) was evaluated for QTL linked in coupling phase as frequency of QTL detected in one of three regions calculated across all sets of a given population size. Two regions were defined as interval ranging ±1 cM around the predefined QTL position, and the third region comprised the ghost QTL (GQ) region in between the two defined QTL intervals (Figure S2). In addition, the power to separately detect both QTL, e.g., LQ6a;LQ6b, was assessed. For QTL linked in repulsion phase two regions surrounding the QTL were defined.

Results

Experimental data: QTL mapping

Phenotypic analyses of DH lines of population UH009 × UH007 were performed in two earlier studies by Martin et al. (2012) and Stange et al. (2013). Significant ($P < 0.01$) genotypic variances were observed for all six traits. Heritabilities were highest for HKW and DSI (87 and 89 %), slightly lower for GY and KN (75 and 76 %), and lowest for DON and GER (64 and 70 %). Both parental inbred lines are from the flint heterotic pool and the full genetic linkage map with 7,169 markers based on this population showed several large monomorphic regions. We arbitrarily defined monomorphic chromosomal regions larger than 20 cM as identical-by-descent (IBD) regions. Based on this criterion approximately 27 % of the genome may be IBD in this cross with particularly large IBD regions on chromosome 3 (approximately 66 %) (Figure S3).

QTL mapping using the highest marker density (MD = 1) identified QTL for all six traits (Table S1). For each trait, the relative distribution of effect sizes of the detected QTL indicated that only QTL with medium or large effects were identified, but no QTL with small effects. Five pairs of linked QTL were detected for DSI, KN, HKW, and GY, whereof three pairs were located on chromosome 1 (for DSI, HKW, and GY), one pair on chromosome 2 (for HKW), and one pair on chromosome 10 (for KN) (Fig. 1, Table S1). The distance between linked QTL ranged from 18 to 215 cM. All pairs of linked QTL on chromosome 1 were in repulsion phase, whereas the other two were linked in coupling phase.

We next performed QTL mapping in this population based on two sets of genetic linkage maps with reduced marker density. While the full genetic linkage map had a marker density of approximately 1 cM in polymorphic regions, the two genetic maps with reduced marker density (MD = 2 and MD = 5) were generated as subsets of markers from the full map with equally spaced markers every 2 and 5 cM, respectively. Across all six traits, we observed no consistent effect of the marker density on number of selected cofactors, number of detected QTL, or proportion of genotypic variance (p_G) explained by the detected QTL in the data set and cross-validated in the test set (Table 1). In contrast, the mean length of the support intervals of the detected QTL was substantially reduced with increasing marker density for five of the six traits (Fig. 1; Table 1). The map positions of detected QTL were similar between marker densities except for the HKW QTL on chromosomes 1 and 9 (Fig. 1).

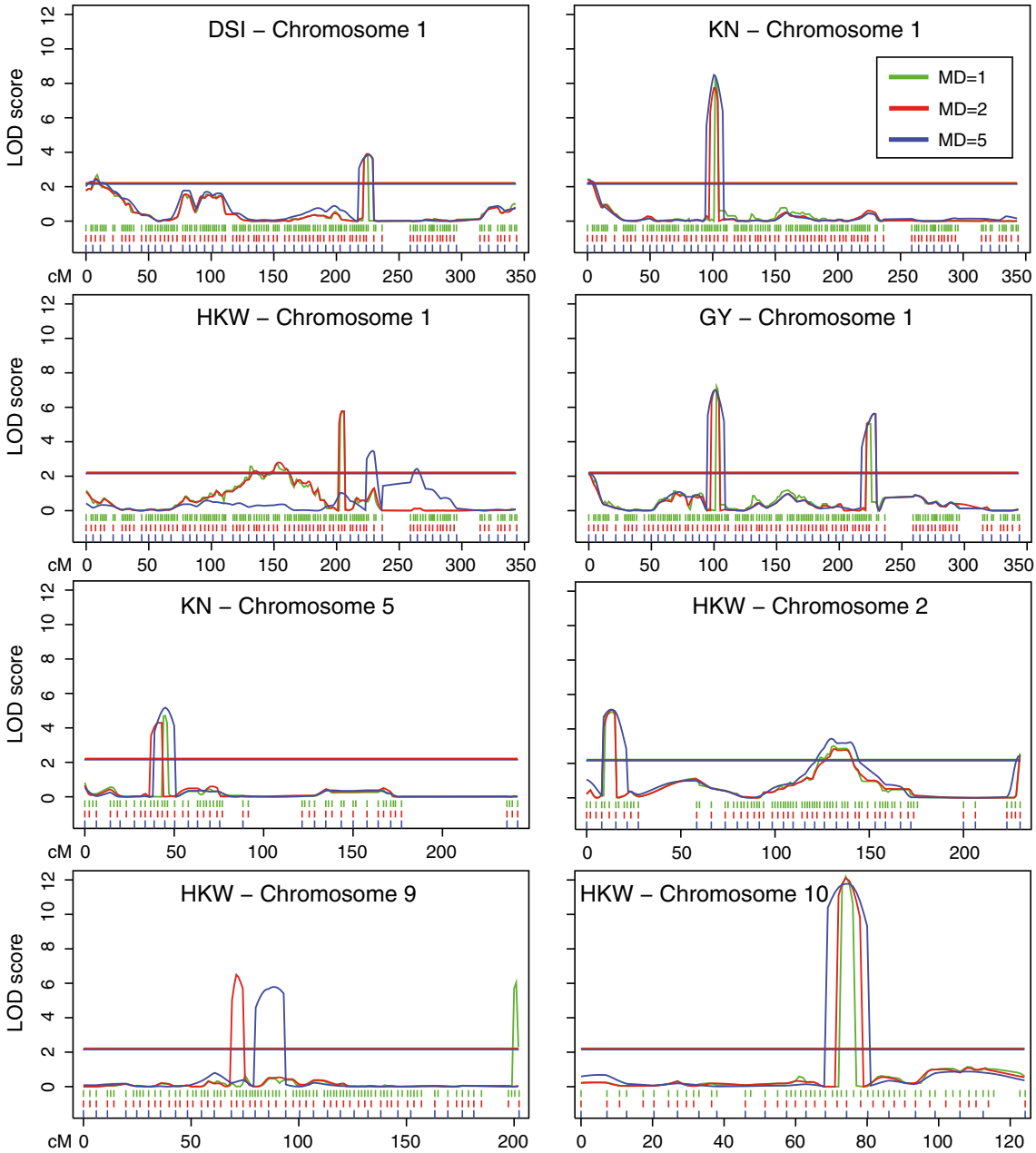

Fig. 1 Chromosome-wise LOD scores of QTL detected in the experimental population UH009 × UH007 for days to silking (DSI), kernel number (KN), 100-kernel weight (HKW), and grain yield per plant (GY) based on three map densities (MD). *Horizontal lines* indicate LOD thresholds determined empirically with 1,000 random permutations separately for each marker density, shown in the respective *color*. Marker positions for the three MD in cM are indicated by *vertical lines* in the respective *color* at the *bottom* of each plot (color figure online)

Simulation study: effect of marker density and population size on QTL mapping

To further address the effect of marker density on QTL mapping in biparental populations as well as a possible interaction between the marker density and the population size, we performed a simulation study in which we varied the marker density and population size. Whereas the LOD thresholds were slightly more stringent under the highest marker density (MD = 1), the number of selected cofactors, detected QTL, and the proportion of explained genotypic variance in the data set and in the test set were of similar magnitude for all three marker densities (Table S3). These parameters were, however, affected by population size and the number of QTL as well as the proportion of genotypic variance explained by them increased with increasing population size.

For the QTL located isolated on separate chromosomes (IQ1–IQ5), we found that the QTL detection power

Table 1 Number of selected cofactors, number of detected QTL, proportion of genotypic variance explained by all QTL per trait in the experimental population UH009 × UH007 in the data set ($\rho_{G\text{-}DS}$) and in test sets ($\rho_{G\text{-}TS}$) of fivefold cross-validation, and mean length of support intervals of detected QTL in cM

Trait	No. of cofactors			No. of QTL			$\rho_{G\text{-}DS}$			$\rho_{G\text{-}TS}$			Mean SI length		
	MD = 1	MD = 2	MD = 5	MD = 1	MD = 2	MD = 5	MD = 1	MD = 2	MD = 5	MD = 1	MD = 2	MD = 5	MD = 1	MD = 2	MD = 5
DSI	1	1	2	4	4	2	30.22	30.23	18.06	7.22	6.56	6.34	16.80	18.75	14.50
GER	1	1	1	3	3	2	28.06	27.95	20.55	6.59	7.16	4.29	11.70	13.00	18.00
DON	1	1	2	4	4	4	34.92	35.17	34.25	7.08	8.01	7.75	16.00	17.75	16.50
KN	3	3	4	6	6	7	51.23	50.18	54.74	23.17	21.06	23.79	9.00	11.50	16.86
HKW	6	6	6	8	8	7	59.41	60.23	55.30	29.31	31.85	32.65	9.90	11.50	12.43
GY	3	3	3	5	4	5	46.51	42.61	47.60	25.17	24.68	26.12	17.40	19.50	20.40

QTL detection was performed with three marker densities (MD = 1, 2, and 5 cM) for days to silking (DSI), Gibberella ear rot severity (GER), deoxynivalenol (DON) concentration, kernel number (KN), 100-kernel weight (HKW), and grain yield per plant (GY)

increased strongly with increasing QTL effect size and population size (Table 2). In contrast, marker density did not affect the QTL detection power. With regard to the estimated additive genetic effects of the QTL IQ1–IQ5, we observed that their overestimation was strongest for the smallest population ($N = 110$) and decreased with increasing population size (Table 3). For all combinations of marker density and population size, the cross-validated QTL effects were less overestimated as compared to effect estimation in the full data sets. Interestingly, we observed that the accuracy of these cross-validated effect estimates was dependent on the size of the QTL effect and on the marker density. This was most pronounced for the smallest population size, but also discernable for the larger populations. For the smallest population, only the largest QTL (IQ5) was estimated accurately under the lowest marker density (MD = 5), whereas for the highest marker density (MD = 1) medium to large effect QTL (IQ3–IQ5) were estimated accurately. The size of the support interval decreased with increasing population size and QTL effect size, but was also affected by the applied marker density (Figs. 2, 3a, b). Higher marker densities reduced the size of the support interval. By contrast, the deviation between estimated and reference QTL position and the LOD score were not affected by the marker density (Figure S4). The FDR decreased with increasing population size and slightly increased with increasing marker density which is in agreement with an increased QTL detection power with increasing marker density. However, the standard errors of FDR indicated that these slight differences between MD = 1, MD = 2, and MD = 5 were not significant (Table S4).

The simulation study also included chromosomes with two QTL to evaluate the power to resolve QTL linked in coupling or repulsion dependent on the applied marker density. For QTL in coupling, we defined an interval around each QTL as well as a ghost QTL region between them and assessed the frequency with which QTL were detected in each region (Figure S2). For QTL linked in repulsion only two regions, each surrounding a QTL, were defined. For QTL with a genetic map distance of 5 cM linked in coupling phase (LQ6a;LQ6b), a higher power to separately detect them was observed with the highest marker density (MD = 1) for all population sizes as compared to the detection of the ghost QTL (GQ6), located in the interval between them (Table 2; Fig. 3c, d). In contrast, with the lower marker densities (MD = 2 and MD = 5), the frequency of the detection of the ghost QTL increased strongly at the expense of the separate detection of the true QTL. Surprisingly, this effect was more pronounced with increasing population size. When the genetic map distance between the linked QTL was increased to 10 cM (LQ7a;LQ7b), the effect of the marker density on the power to separately detect the true QTL instead of the ghost QTL

Table 2 Power (%) of QTL detection for five simulated independent QTL (IQ1–IQ5) with additive genetic effects from 0.10 to 0.75, two pairs of linked QTL in coupling phase (LQ6a;LQ6b and LQ7a;LQ7b), and two pairs of linked QTL in repulsion phase (LQ8a;LQ8b and LQ9a;LQ9b) based on three marker densities (MD = 1, 2, and 5 cM) and three population sizes (N = 110, 220, and 440)

QTL	$N = 110$			$N = 220$			$N = 440$		
	MD = 1	MD = 2	MD = 5	MD = 1	MD = 2	MD = 5	MD = 1	MD = 2	MD = 5
Independent QTL									
IQ1 (167–177)[a]	0.2	0.5	0.3	0.7	0.6	0.5	0.8	0.8	0.6
IQ2 (110–120)	0.9	1.3	1.6	5.1	4.5	4.1	10.6	11.0	9.4
IQ3 (85–95)	5.6	4.9	4.9	19.7	18.0	17.3	54.0	51.2	48.4
IQ4 (96–106)	29.1	28.3	29.6	67.9	69.2	70.5	91.6	91.4	95.4
IQ5 (57–67)	71.0	69.4	69.4	91.9	90.5	92.5	98.6	99.0	99.4
Linked QTL with a distance of 5 cM in coupling phase									
LQ6a;LQ6b (29–31; 34–36)	1.8	0.9	0.9	1.5	0.6	0.6	28.0	1.4	1.4
LQ6a (29–31)	35.8	24.5	20.3	37.8	19.0	12.4	51.0	8.4	6.0
LQ6b (34–36)	30.5	15.4	22.9	30.1	8.2	17.9	38.6	2.8	9.4
GQ6[b] (32–33)	19.7	45.5	42.8	27.3	69.3	66.6	27.6	88.2	84.0
Linked QTL with a distance of 10 cM in coupling phase									
LQ7a;LQ7b (94–96; 104–106)	5.6	4.7	2.7	17.2	13.6	6.6	82.4	77.2	23.4
LQ7a (94–96)	27.3	18.7	10.6	30.9	17.5	7.2	67.4	57.2	13.0
LQ7b (104–106)	22.0	26.4	15.3	17.2	34.6	10.6	64.2	69.8	12.6
GQ7[b] (97–103)	40.1	42.0	61.3	38.8	45.7	77.4	34.6	30.6	77.8
Linked QTL with a distance of 5 cM in repulsion phase									
LQ8a;LQ8b (64–66; 69–71)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.2
LQ8a (64–66)	0.5	0.0	0.0	0.1	0.1	0.0	0.2	0.2	0.0
LQ8b (69–71)	0.0	0.0	0.0	0.2	0.2	0.3	0.2	0.2	0.2
Linked QTL with a distance of 10 cM in repulsion phase									
LQ9a;LQ9b (59–61; 69–71)	0.1	0.1	0.0	0.0	0.2	0.0	1.0	1.6	1.2
LQ9a (59–61)	0.3	0.3	0.1	1.0	0.8	0.2	1.6	1.8	0.8
LQ9b (69–71)	0.3	0.3	0.4	0.4	0.5	0.6	1.6	1.2	1.6

[a] Predefined interval borders used to determine the QTL detection power

[b] GQ ghost QTL detected in the interval between the linked QTL in coupling phase

was confirmed. The power to detect both true QTL simultaneously was always substantially higher for the highest marker density (MD = 1) and higher for the QTL separated by a larger genetic map distance (LQ7 as opposed to LQ6). The power to detect the QTL linked in repulsion, either separately or simultaneously, was always close to zero.

Discussion

For many crops high-density genotyping can nowadays routinely be applied. Alternatively, customized arrays with lower marker densities are available at reduced costs. Whereas high-density genotyping is certainly advantageous for association mapping or genomic selection, it may be an overkill for linkage mapping in biparental populations. The population UH009 × UH007 consists of 204 DH lines and high genotypic variances and heritabilities comparable to those reported previously were observed for all traits (Ali et al. 2005; Bolduan et al. 2009a; Buckler et al. 2009; Hallauer et al. 2010). The plants were subjected to high-density genotyping, resulting in highly saturated genetic linkage maps. Based on these genetic linkage maps, a combination of experimental data analysis and a simulation study was used to address the effect of marker density on QTL mapping in biparental populations.

Regions of IBD

Especially plants related by pedigree or coming from the same heterotic group are expected to be

Table 3 Estimated additive genetic effects in data sets (DS) and in test sets (TS) of fivefold cross-validation for five simulated independent QTL (IQ1–IQ5) with reference additive genetic effects from 0.10 to 0.75 based on three marker densities (MD = 1, 2, and 5 cM) and three population sizes ($N = 110$, 220, and 440)

QTL	MD = 1		MD = 2		MD = 5	
	DS	TS	DS	TS	DS	TS
$N = 110$						
IQ1 (0.10)[a]	0.53	0.35	0.55	0.29	0.57	0.21
IQ2 (0.20)	0.55	0.41	0.54	0.28	0.55	0.27
IQ3 (0.30)	0.56	0.34	0.57	0.35	0.58	0.37
IQ4 (0.50)	0.62	0.46	0.63	0.46	0.64	0.41
IQ5 (0.75)	0.79	0.72	0.79	0.72	0.80	0.72
$N = 220$						
IQ1 (0.10)	0.39	0.13	0.39	0.15	0.39	0.17
IQ2 (0.20)	0.40	0.24	0.41	0.25	0.41	0.23
IQ3 (0.30)	0.43	0.30	0.44	0.30	0.44	0.31
IQ4 (0.50)	0.53	0.47	0.54	0.47	0.54	0.47
IQ5 (0.75)	0.76	0.74	0.76	0.74	0.77	0.75
$N = 440$						
IQ1 (0.10)	0.30	0.15	0.32	0.21	0.32	0.25
IQ2 (0.20)	0.32	0.21	0.32	0.22	0.33	0.23
IQ3 (0.30)	0.35	0.28	0.36	0.29	0.35	0.28
IQ4 (0.50)	0.51	0.49	0.51	0.49	0.51	0.50
IQ5 (0.75)	0.75	0.74	0.76	0.74	0.76	0.75

[a] Reference additive genetic effects are given in brackets

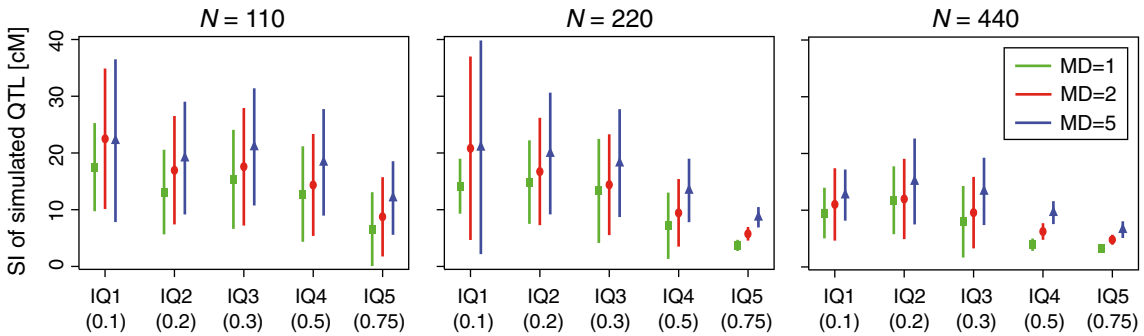

Fig. 2 Mean 1-LOD support intervals (SI; ±standard deviations as *vertical lines*) for the QTL IQ1–IQ5 with additive genetic effects from 0.10 to 0.75, respectively. Means and standard deviations are averages across all sets for a given population size based on three marker densities (MD = 1, 2, and 5 cM), respectively (color figure online)

identical-by-descent for some chromosomal regions. Consequently, no polymorphic markers will be available for these regions resulting in larger gaps in the genetic linkage maps, irrespective of the applied genotyping density. Population UH009 × UH007 showed several large monomorphic regions which were arbitrarily defined as IBD if they exceeded 20 cM (Figure S3). Based on this criterion, approximately 27 % of the entire genome may be IBD, potentially explaining the gaps observed in our linkage map. The interpretation of these gaps as IBD is substantiated by the fact that markers located in these regions are represented on the array and are known to be polymorphic in the breeding germplasm from which the two parents are derived. Another consequence of these IBD regions is that they will not contribute to the genotypic variance in that cross. Thus, in the population presented here only approximately three quarters of the genome will add to the new genetic variation that can be exploited to select superior lines.

Influence of marker density on QTL detection power

In our study, LOD thresholds were determined separately for each marker density and population size as a uniform LOD threshold cannot be assumed for diverse experimental data sets. The thresholds obtained with the highest marker density were slightly more stringent than those from the lower marker densities (Table S3), but this difference

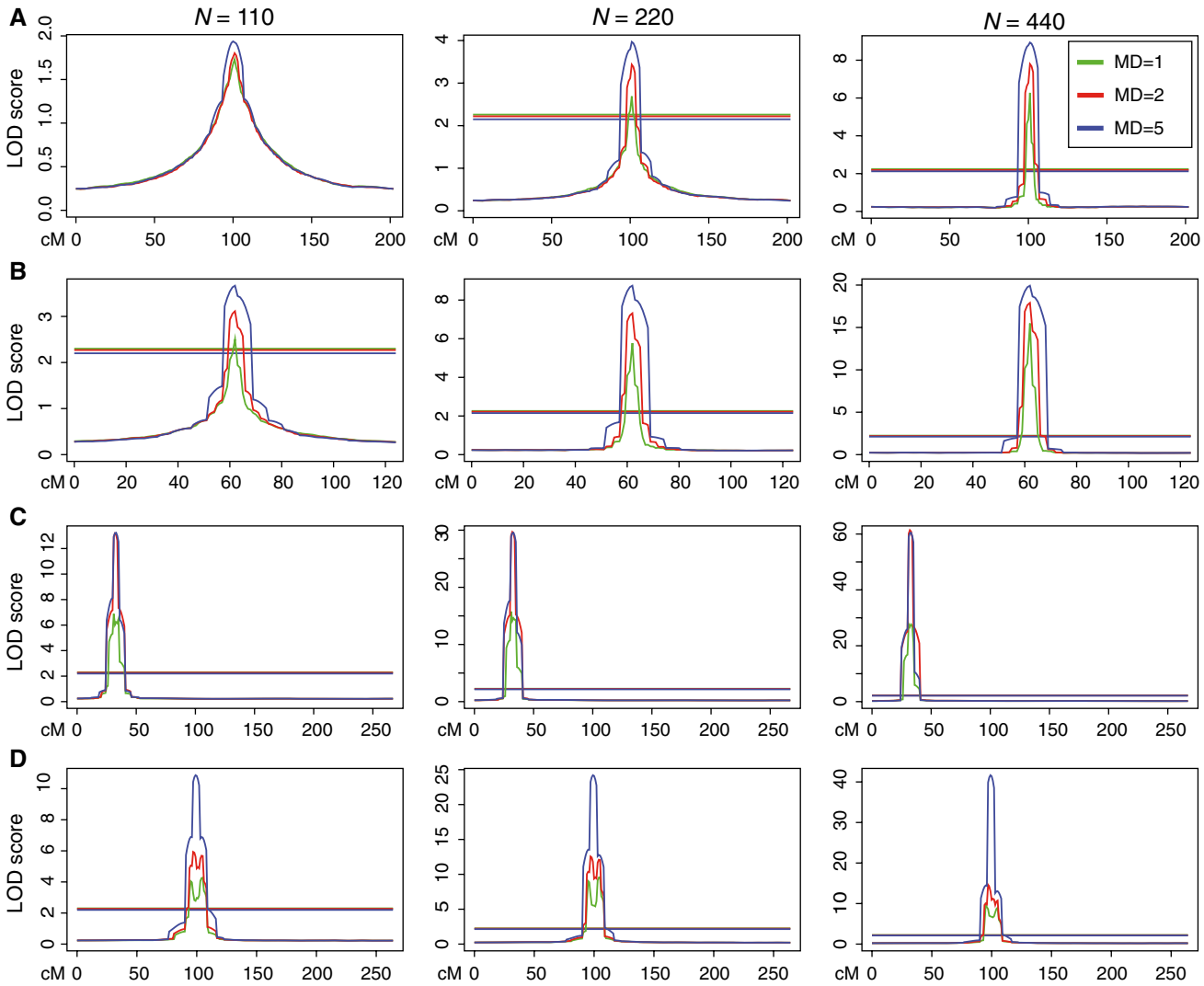

Fig. 3 Average LOD scores for independent QTL (**a**) IQ4 on chromosome 9 and (**b**) IQ5 on chromosome 10, and for linked QTL in coupling phase (**c**) LQ6a;LQ6b on chromosome 3, and (**d**) LQ7a;LQ7b on chromosome 6. LOD scores are averages across all sets for a given population size based on three marker densities (MD = 1, 2, and 5 cM), respectively. *Horizontal lines* indicate LOD thresholds determined empirically with 20 random permutations separately for each population size and MD

appears negligible and should not substantially affect the QTL detection power. Consistently, we observed no effect of the marker density on the QTL detection power neither in the experimental data nor in the simulation study with independent QTL (Tables 1, 2). This is in agreement with theoretical results from Piepho (2000) showing that the power of QTL detection for interval mapping in a backcross population changed only slightly with an increase in marker density below 10 cM. It must be noted, however, that the power to detect small simulated effect QTL (IQ1 and IQ2) was low, irrespective of the applied marker density and population size. With population sizes routinely applied in QTL mapping (100–200 individuals), only major QTL like the simulated IQ5 can be detected with sufficient power while already for the medium effect QTL the power appears insufficient for reliable detection.

The LOD scores of the detected QTL were similar between marker densities for the simulated QTL IQ1–IQ5 but concordant with the QTL detection power, increased with population size and also with an increasing genetic effect (Figure S4). This is in agreement with results reported by Li et al. (2010) for IM and ICIM. However, in contrast to our results, Li et al. (2010) observed an increase in LOD scores with increasing marker density, but these differences between marker densities diminished with decreasing population size. The differences between our results and results reported by Li et al. (2010) might be explained by different levels of marker densities. Li et al.

(2010) used MD = 40 as low density and MD = 5 as high density, whereas in our study MD = 5 was used as low density, and MD = 1 as high density. Consequently, already the lowest marker density (MD = 5) used in our study appears sufficient to reach high LOD scores, which are not further increased by applying higher marker densities.

Consistent with the unaffected QTL detection power, experimental and simulation results also showed that the increase in marker density from 5 to 1 cM did not result in a gain with regard to the proportion of genotypic variance explained by the detected QTL (Table 1; Table S3). Taken together, high-density genotyping does not increase the QTL detection power or the proportion of genotypic variance explained by the QTL which is a key parameter in marker-assisted selection.

Estimation of QTL effects

The relative distribution of QTL effects in population UH009 × UH007 indicated that for each trait only medium or large effect QTL were identified, but no QTL with small effects (Table S1). This raises the question if in this population only medium and large effect QTL are segregating for the examined traits or, if small effect QTL are present but escape detection due to insufficient QTL detection power. Alternatively, they may be detected but with overestimated effect size. To address this question, we simulated five QTL (IQ1–IQ5) with reference effects from 0.10 to 0.75 and evaluated the accuracy of effect estimates dependent on the population size and marker density. We observed that overestimation of effects was less biased after cross-validation and that the bias decreased with increasing reference effect and increasing population size (Table 3). This is in agreement with results reported by Darvasi et al. (1993), Utz et al. (2000), and Li et al. (2010). Interestingly, we observed as a general trend that the increase in marker density from 5 to 1 cM resulted in more precise effect estimates, especially for small and medium effect QTL. Together with the results on the QTL detection power, this indicates that the relative effect distribution in the experimental population is likely caused by insufficient power to detect small effect QTL, but that also some small or medium effect QTL may have been detected but with overestimated effect sizes. In applied plant breeding, population sizes are often fixed and an increase is not feasible. Our results indicate that the investment in higher marker density may pay off as more accurate QTL effect estimates can be obtained.

Precision of QTL localization

The 1-LOD support interval (SI) is often used to determine the precision of the QTL position estimates. The average length of SIs across all detected QTL in population UH009 × UH007 was short with 12.9 cM and ranged from 3 to 58 cM (Table S1). The longest SI was observed for the QTL detected for GY on chromosome 3 which is located at the end of a chromosomal region with high marker coverage followed by an approximately 30 cM long IBD region (Figure S3). The wide LOD peak of this QTL stretched over the entire IBD region until the next polymorphic region, thus explaining the large SI. In addition, four out of five SIs with a length between 20 and 30 cM showed relatively flat and broad LOD peaks which may be due to the number and positions of the selected cofactors. Even with the lowest marker density (MD = 5), the average SI length (18 cM) of QTL detected for GER was shorter compared to a previous study in this population (23 cM) based on a low-density SSR map (Martin et al. 2012). The large SI length observed by Martin et al. (2012) is in agreement with results reported by Ali et al. (2005) for Gibberella ear rot severity, where QTL localization was also rather imprecise due to the low-density map with an average marker density of only 13.8 cM.

Comparison of the average length of SIs between marker densities in population UH009 × UH007 showed a decrease of SI length with increasing marker density for five of the six traits (Table 1). This is in agreement with our simulation results where the SI length decreased with increasing marker density regardless of QTL effect and population size (Fig. 2). This effect of the marker density on the precision of QTL localization can be well seen by the progressively narrower LOD peaks with increasing marker density (Figs. 1, 3a, b). The standard deviations of the SI indicate a bias in their estimation which decreased with increasing QTL effect and population size (Fig. 2). In contrast to the clear trend of shorter SIs with increasing marker density, the accuracy of QTL localization was not affected by marker density regardless of QTL effect size and population size (Figure S4). Nevertheless, increasing marker density can improve the precision of QTL localization, especially for medium to large effect QTL and thus, might increase the efficiency of marker-assisted selection.

Resolution of linked QTL

The probability to detect two linked QTL separately is increased if they are separated by a marker interval without a QTL (Li et al. 2010). In total, five pairs of linked QTL were detected with the highest marker density in population UH009 × UH007 with large linkage distances ranging from 46 to 215 cM (Table S1). To address the question if closely linked QTL can be better separated with higher marker densities, we performed a simulation study with QTL linked at 5 (LQ6) and 10 cM (LQ7). A well-recognized problem of linked QTL is that a ghost QTL can be

detected as an artifact between the true QTL (Doerge 2002). Li et al. (2010) found that QTL linked at 10 cM distance could not be dissected with their highest marker density (MD = 5) even with large population sizes of up to 500 individuals. Only when the linkage distance was increased to 20 or 30 cM could these QTL be separated.

We observed that for all population sizes higher marker densities were beneficial as they increased the power to detect the true QTL as opposed to the ghost QTL (Table 2). The increasing power for the ghost QTL with increasing population size is likely caused by the higher error of QTL estimates with small population sizes which, seemingly increases the power to detect the true QTL. Oftentimes more frequent detection of the ghost QTL for LQ7 compared to LQ6 can be explained by the larger interval defined for the ghost QTL of LQ7 (6 versus 2 cM). Due to the small interval of GQ6, more QTL fell into the intervals of LQ6a and LQ6b resulting in a higher power compared to GQ6. The power to simultaneously detect both true QTL was lower for the more closely linked QTL LQ6 as compared to LQ7. Significantly, for both LQ6 and LQ7, this power increased substantially by increasing the marker density from 5 to 1 cM. For example, with a population size of $N = 440$, the power for simultaneous detection of both LQ7 QTL was 23.4 % with an average marker spacing of only 5 cM which more than tripled with the increase in marker density to an average spacing of 1 cM.

In order to avoid declaring a shoulder of a QTL peak as separate QTL, QTL mapping software often requires a minimum distance between two putative QTL. Owing to the previously employed low marker densities, this parameter has traditionally been set at distances (e.g., 10 cM) that are too large to enable the separate detection of QTL linked as closely as in our simulation study. An important consequence of our finding with high marker densities for separate detection of closely linked QTL (5 cM) is that for QTL mapping this threshold distance between potential QTL is reduced (in our example to 3 cM). In contrast to the QTL linked in coupling, QTL linked in repulsion phase could not be detected irrespective of the marker density. Taken together, marker density is a major factor affecting the power to resolve linked QTL which illustrates another potentially beneficial effect of high-density genotyping.

Conclusions

In this study, we investigated whether high-density maps with a marker density of 1 cM offer advantages in QTL mapping compared to low-density maps with regard to parameters relevant for plant breeding. High-density maps had no effect on the QTL detection power or the predictive power for the proportion of explained genotypic variance. By contrast, the precision of effect estimates, especially for small and medium sized QTL, the precision of QTL localization, as well as the power to resolve linked QTL profited from an increase in marker density. Thus, for QTL detection aiming at identifying QTL for marker-assisted selection, the more precise estimates of these relevant parameters may outweigh the higher costs of high-density genotyping.

Acknowledgments This research was financed by the Deutsche Forschungsgemeinschaft (DFG) research grant ME 2260/6-1 and supported by funds from DFG (Grant No. 1070, International Research Training Group "Sustainable Resource Use in North China") to M. Stange and Albrecht E. Melchinger. Part of the SNP analysis of this research was supported by the project "Cornfed" by the French National Agency for Research (ANR), the German Federal Ministry of Education and Research (BMBF), and the Spanish Ministry of Science and Innovation (MICINN). Part of this research was conducted in collaboration between the University of Hohenheim and CYMMIT under the Federal Ministry for Economic Cooperation and Development (BMZ) founded project "Abiotic stress tolerant maize for increasing income and food security among the poor in South and Southeast Asia".

Conflict of interest The authors declare that they have no conflict of interest.

Ethical standard The experiments performed within this study comply with the current laws of Germany.

References

Ali ML, Taylor JH, Jie L et al (2005) Molecular mapping of QTLs for resistance to Gibberella ear rot, in corn, caused by *Fusarium graminearum*. Genome 48:521–533

Almeida GD, Makumbi D, Magorokosho C et al (2012) QTL mapping in three tropical maize populations reveals a set of constitutive and adaptive genomic regions for drought tolerance. Theor Appl Genet 126:583–600

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

Bolduan C, Miedaner T, Schipprack W et al (2009a) Genetic variation for resistance to ear rots and mycotoxins contamination in early European maize inbred lines. Crop Sci 49:2019–2028

Bolduan C, Montes JM, Dhillon BS et al (2009b) Determination of mycotoxin concentration by ELISA and near-infrared spectroscopy in Fusarium-inoculated maize. Maize Cereal Res Commun 37:521–529

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

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

Darvasi A, Weinreb A, Minke V et al (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134:943–951

Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental populations. Nat Rev Genet 3:43–52

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longmans Green, Harlow

Frisch M, Bohn M, Melchinger AE (2000) Plabsim: software for simulation of marker-assisted backcrossing. J Hered 91:86–87

Guo J, Chen Z, Liu Z et al (2011) Identification of genetic factors affecting plant density response through QTL mapping of yield component traits in maize (*Zea mays* L.). Euphytica 182:409–422

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

Hallauer AR, Carena MJ, Miranda JB (2010) Quantitative genetics in maize breeding. Iowa State University Press, Ames

Hori K, Kobayashi T, Shimizu A et al (2003) Efficient construction of high-density linkage map and its application to QTL analysis in barley. Theor Appl Genet 107:806–813

Lawrence CJ, Harper LC, Schaeffer ML et al (2008) MaizeGDB: the maize model organism database for basic, translational, and applied research. Int J Plant Genom 2008:496957

Li H, Hearne S, Bänziger M et al (2010) Statistical properties of QTL linkage mapping in biparental genetic populations. Heredity 105:257–267

Liu W, Reif JC, Ranc N et al (2012) Comparison of biometrical approaches for QTL detection in multiple segregating families. Theor Appl Genet 125:987–998

Ma XQ, Tang JH, Teng WT et al (2007) Epistatic interaction is an important genetic basis of grain yield and its components in maize. Mol Breed 20:41–51

Martin M, Miedaner T, Dhillon BS et al (2011) Colocalization of QTL for Gibberella ear rot resistance and low mycotoxin contamination in early European maize. Crop Sci 51:1935–1945

Martin M, Miedaner T, Schwegler DD et al (2012) Comparative quantitative trait loci mapping for Gibberella ear rot resistance and reduced deoxynivalenol contamination across connected maize populations. Crop Sci 52:32–43

Pestka JJ (2007) Deoxynivalenol: toxicity, mechanisms and animal health risks. Anim Feed Sci Tech 137:283–298

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

Prigge V, Melchinger AE (2012) Production of haploids and doubled haploids in maize. Methods Mol Biol 877:161–172

Prigge V, Xu X, Li L et al (2012) New insights into the genetics of in vivo induction of maternal haploids, the backbone of doubled haploid technology in maize. Genetics 190:781–793

Riedelsheimer C, Lisea J, Czedik-Eysenberg A et al (2012a) Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. Proc Natl Acad Sci USA 109:8872–8877

Riedelsheimer C, Technow F, Melchinger AE (2012b) Comparison of whole-genome prediction models for traits with contrasting genetic architecture in a diversity panel of maize inbred lines. BMC Genomics 13:452

Schaeffer M, Byrne P, EHC J et al (2006) Consensus quantitative trait maps in maize: a database strategy. Maydica 51:357–367

Shi L, Hao Z, Weng J et al (2011) Identification of a major quantitative trait locus for resistance to maize rough dwarf virus in a Chinese maize inbred line X178 using a linkage map based on 514 gene-derived single nucleotide polymorphisms. Mol Breed 30:615–625

Stange M, Schrag TA, Utz HF, Riedelsheimer C, Bauer E, Melchinger AE (2013) High-density linkage mapping of yield and epistatic interactions in maize with doubled hapoid lines from four crosses. Mol Breed. doi:10.1007/s11032-013-9887-z

Utz HF (2005) PLABSTAT—a computer program for statistical analysis of plant breeding experiments, 3A. Universität Hohenheim, Germany

Utz HF (2012) PlabMQTL—Software for meta-QTL analysis with composite interval mapping. Version 0.5s. Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim. PlabMQTL Manual

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

Van Ooijen JW (2006) JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen

Wu Y, Bhat PR, Close TJ, Lonardi S (2008) Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. PLoS Genet 4:e1000212

Würschum T (2012) Mapping QTL for agronomic traits in breeding populations. Theor Appl Genet 125:201–210

Yan J, Yang X, Shah T et al (2009) High-throughput SNP genotyping with the GoldenGate assay in maize. Mol Breed 25:441–451

Yu H, Xie W, Wang J et al (2011) Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. PloS One 6:e17595