CrossMark

ORIGINAL ARTICLE

Choice of models for QTL mapping with multiple families and design of the training set for prediction of *Fusarium* resistance traits in maize

Sen Han[1] · H. Friedrich Utz[1] · Wenxin Liu[2] · Tobias A. Schrag[1] · Michael Stange[1,5] · Tobias Würschum[3] · Thomas Miedaner[3] · Eva Bauer[4] · Chris-Carolin Schön[4] · Albrecht E. Melchinger[1]

Abstract

Key message **QTL analysis for *Fusarium* resistance traits with multiple connected families detected more QTL than single-family analysis. Prediction accuracy was tightly associated with the kinship of the validation and training set.**

Abstract QTL mapping has recently shifted from analysis of single families to multiple, connected families and several biometric models have been suggested. Using a high-density consensus map with 2472 marker loci, we performed QTL mapping with five connected bi-parental families with 639 doubled-haploid (DH) lines in maize for ear rot resistance and analyzed traits DON, Gibberella ear rot severity (GER), and days to silking (DS). Five biometric models differing in the assumption about the number and effects of alleles at QTL were compared. Model 2 to 5 performing joint analyses across all families and using linkage and/or linkage disequilibrium (LD) information identified all and even further QTL than Model 1 (single-family analyses) and generally explained a higher proportion p_G of the genotypic variance for all three traits. QTL for DON and GER were mostly family specific, but several QTL for DS occurred in multiple families. Many QTL displayed large additive effects and most alleles increasing resistance originated from a resistant parent. Interactions between detected QTL and genetic background (family) occurred rarely and were comparatively small. Detailed analysis of three fully connected families yielded higher p_G values for Model 3 or 4 than for Model 2 and 5, irrespective of the size N_{TS} of the training set (TS). In conclusion, Model 3 and 4 can be recommended for QTL-based prediction with larger families. Including a sufficiently large number of full sibs in the TS helped to increase QTL-based prediction accuracy (r_{VS}) for various scenarios differing in the composition of the TS.

Communicated by M. Frisch.

✉ Albrecht E. Melchinger
melchinger@uni-hohenheim.de

1 Institute of Plant Breeding, Seed Science and Population Genetics (350a), University of Hohenheim, 70593 Stuttgart, Germany

2 Crop Genetics and Breeding Department, China Agricultural University, Beijing 100193, China

3 State Plant Breeding Institute (720), University of Hohenheim, 70593 Stuttgart, Germany

4 Department of Plant Breeding, Center of Life and Food Sciences Weihenstephan, Technische Universität München, 85350 Freising, Germany

5 Present Address: Strube Research GmbH and Co. KG, Hauptstraße 1, 38387 Söllingen, Germany

Introduction

Gibberella ear rot (GER) caused by *Fusarium graminearum* is a major disease of maize (*Zea mays* L.) in Europe and Canada. It reduces the yield and contaminates the grain with mycotoxins, in particular with deoxynivalenol (DON). Breeding resistant cultivars is the most effective approach for combatting the disease due to limited effect of agronomic practices and fungicides (Martin et al. 2011). For genetic improvement of GER, Martin et al. (2012) recommended marker-assisted selection, but this presupposes accurate estimates of the chromosomal location and effect of the underlying quantitative trait loci (QTL).

Linkage mapping with individual bi-parental families derived from divergent parent lines has become routine for dissecting the genetic architecture of complex traits in crops (Holland 2007), although it has certain drawbacks. (i) The mapping population originates from two parent lines and, therefore, represents only a small cross section of the breeding germplasm (Xu 1998; Liu and Zeng 2000). (ii) Mapping results from one family are often not transferable to other families (Beavis 1998; Melchinger et al. 1998), because the expression of QTL depends on its presence and in addition can be influenced by the genetic background. To overcome these limitations, multi-family QTL mapping has been proposed to detect QTL jointly from multiple bi-parental families (Jansen et al. 2003; Blanc et al. 2006; Bink et al. 2012). These populations can be either families routinely generated in practical breeding programs (Bardol et al. 2013) or created using special mating designs, e.g., the diallel design (Blanc et al. 2006), nested association mapping population (NAM, Yu et al. 2008) or multi-parent advanced generation intercross (MAGIC, Huang et al. 2015). The main difference between bi-parental and multi-family QTL mapping concerns the set of QTL that segregate in one versus several populations, which enables testing for QTL × genetic background interactions in the latter case (Blanc et al. 2006). For a given sample size, the former approach has generally higher power to detect rare QTL with large effects, segregating in only one or a small number of families, while the latter approach has higher chances to detect common QTL with small effects shared by a large number of families (Li et al. 2011; Ogut et al. 2015).

Four main categories of biometrical models have been developed for multi-family QTL mapping in plants, which differ in their assumptions about the QTL effects: (1) effects are specific to each population (e.g., FULL model in Jannink and Jansen 2001; disconnected model in Blanc et al. 2006), (2) effects of parental alleles are identical over populations (e.g., REDUCED model in Jannink and Jansen 2001; connected model in Blanc et al. 2006), (3) identical by descent (IBD) segments shared by parents have the same alleles, the effects of which are identically expressed in different populations (e.g., HaploMQM$^-$ model in Jansen et al. 2003; LDLA model in Bardol et al. 2013 and Giraud et al. 2014), (4) identical by state (IBS) segments among parents harbor identical alleles with effects consistent across genetic backgrounds (Yu et al. 2008; LDLA-1-marker model in Bardol et al. 2013; Model-B in Würschum et al. 2012). From category (1) to (4), the number of alleles at QTL decreases, resulting in a reduced number of parameters to be estimated. Thus, the power of QTL detection may increase and estimation error of QTL effects may decrease (Rebai and Goffinet 1993, 2000), if a common set of QTL can be assumed. In experimental studies,

the performance ranking of these models varied among populations of equal size and among traits (Blanc et al. 2006; Steinhoff et al. 2011; Bardol et al. 2013; Giraud et al. 2014). Therefore, further research is warranted to compare these models and provide guidance for their choice.

In recent years, genomic prediction of breeding values of untested genotypes with genome-wide markers has received considerable interest by breeders due to a dramatic reduction in the costs of genotyping (Meuwissen et al. 2001; Jannink et al. 2010). One important question in genomic prediction, and generally in marker-based prediction, is how to design the training set (TS) for achieving a high prediction accuracy. Major factors identified are the sample size and number of families in the TS and their relatedness to the validation set (VS, Riedelsheimer et al. 2013; Lehermeier et al. 2014). Multi-family QTL mapping offers the possibility to unveil the genetic basis of prediction accuracy in genomic prediction with different composition of the TS.

In our study, we compared five models of QTL mapping with multiple crosses for QTL detection and QTL-based performance prediction evaluated with cross-validation. Besides additive effects, we investigated digenic epistasis and QTL × genetic background interactions. Moreover, we examined several scenarios of composition of the TS for QTL-based prediction. Our analyses were based on a total of 639 doubled-haploid (DH) lines derived from five interconnected crosses genotyped with 56 k SNP and 363 SSR markers and phenotyped for relevant GER resistance traits (DON concentration, GER severity) and a phenological trait (days to silking) in maize.

Materials and methods

Plant material and field trials

Four flint maize inbred lines developed by the University of Hohenheim were used as parents. They represent elite breeding materials of Central Europe displaying good combining ability for grain yield in crosses with dent lines. Pedigree-based coefficients of coancestry among them range between 0.05 and 0.23 (Martin et al. 2011). Regarding resistance against *Fusarium graminearum,* parent line UH006 is highly resistant, UH007 is moderately resistant, and UH009 and D152 are highly susceptible (Bolduan et al. 2009). The four parent lines, herein denoted as R1, R2, S1 and S2, respectively, were crossed in an incomplete half-diallel design (Fig. S1) and the F$_1$ crosses were used for developing five interconnected families of DH lines ranging in size from 43 to 204 (Table 1). The DH lines were developed by applying the in vivo haploid method detailed by Prigge and Melchinger (2012).

Table 1 Family size (N), pedigree-based coefficient of coancestry among parent lines (f), family means ($\bar{X}$), and estimates of genotypic variance (σ_G^2), heritability (h^2), and genetic correlations (r_g) of deox- ynivalenol concentration (DON), Gibberella ear rot severity (GER), and days to silking (DS) for five families of doubled-haploid (DH) lines

Family	N	f	$\bar{X}$			σ_G^{2}**			h^2			r_g		
			DON	GER	DS	DON	GER	DS	DON	GER	DS	(DON, GER)	(DON, DS)	(GER, DS)
R1R2	135	0.23	3.43	39.33	81.48	0.40	102	10.08	0.77	0.77	0.88	0.96+	−0.21+	−0.32+
R1S1	204	0.22	4.84	53.06	79.81	0.25	74	7.26	0.64	0.70	0.87	1.00+	−0.53+	−0.45+
R1S2	161	0.14	4.87	59.4	81.3	0.47	102	7.76	0.80	0.79	0.92	0.98+	−0.39+	−0.40+
R2S1	96	0.19	3.7	42.97	80.99	0.56	131	12.18	0.80	0.82	0.94	0.98+	−0.64+	−0.66+
R2S2	43	0.05	4.64	56.58	81.62	0.44	115	9.21	0.74	0.83	0.89	0.99+	0.03	0.03

** Significant at $P < 0.01$ for all traits and families

[+] Absolute value of the estimate of r_g exceeds twice its standard error

All 639 DH lines and their parental lines were tested at two locations in Southwest Germany, namely Stuttgart-Hohenheim (48°43′12″ N, 9°10′48″ E) and Eckartsweier (48°31′12″ N, 7°52′12″ E) in 2 years (2008 and 2009). In each environment (year × location combination), four 10×20 α designs, each with two replicates, were grown adjacent to each other as detailed by Martin et al. (2011). The experimental units were 3 m single row plots spaced 0.75 m apart with 20 plants.

Artificial inoculation with an aggressive isolate of *F. graminearum* (IFA66) was conducted as detailed by Bolduan et al. (2009). Briefly, the inoculum (1 ml, 100,000 conidia) was injected 5–6 days after silk emergence into the silk channels of the primary ears of plants at a similar developmental stage. Six and eight plants per plot were inoculated in 2008 and 2009, respectively. At physiological maturity, the inoculated ears were manually dehusked and visually rated for GER severity from 0 to 100 %. After harvest, the ears were dried to a moisture content of approximately 14 %. DON concentration of each plot was measured with near-infrared spectroscopy (NIRS) as described in detail elsewhere (Martin et al. 2011; Miedaner et al. 2015). Moreover, the number of days to silking (DS) was recorded on a plot basis as the number of days from sowing to silk emergence of the primary ears in 50 % of the plants.

Phenotypic data analysis

Data for GER severity and DON concentration were transformed using the arcsine square root function and the natural logarithm function, respectively, to reduce heterogeneity of variances and approximate the assumption of a Gaussian distribution. After calculation of adjusted entry means in each environment, variance components across environments and entry-mean based heritabilities (h^2) were estimated for each family as detailed by Martin et al. (2011). Genotypic correlation coefficients (r_g) between traits in each family and their standard errors were calculated according to Mode and Robinson (1959). Statistical analyses of the phenotypic data were performed with software PLABSTAT (Utz 2005).

Marker screening and consensus map construction

The Illumina MaizeSNP50 array comprising 56,110 SNP markers (Ganal et al. 2011) was applied for genotyping all DH lines and the four parent lines. In each family, polymorphic SNP markers were selected, if (i) their physical map position was known and (ii) their minor-allele frequency and average call frequency exceeded 0.05 and 0.80, respectively. DH lines with more than 5 % heterozygous SNPs or an average call rate smaller than 0.80 were excluded from further analyses. In addition, the DH lines were genotyped by 123 (R1R2), 106 (R1S1), 129 (R1S2), 113 (R2S1) and 121 (R2S2) polymorphic SSRs as described in detail by Martin et al. (2011).

In each family, markers were grouped into 10 linkage groups with software MSTmap (Wu et al. 2008). Afterwards, a consensus map of each linkage group across all families was constructed with software Carthagene (de Givery et al. 2005) in the following steps. Step 1: merge the dataset of all families with the *dsmergen* command and then combine each pair of strongly correlated markers (2-point LOD score ≥3) into one locus. Step 2: build a framework map with a limited number of markers, but having a reliable order (*buildfw 10 10 {} 0*). Step 3: incorporate additional markers into the framework map using the command *buildfw keepThres AddThres {"marker order of the framework map"} 0*, where the values of *keepThres* and *AddThres* are high (≥3), but lower than in step 2. Notably, questionable markers were removed, if they caused considerable inconsistency in the marker order between the genetic and physical map or resulted in an excessive expansion of the linkage group. Step 4: repeat step 3 to add as many markers as possible while ensuring a robust marker order (*keepThres* and *AddThres* ≥3).

Clustering of parental alleles at each locus

The alleles of the parent lines at each locus of the consensus map were clustered into ancestral classes based on similarity scores between each pair of lines, which were calculated with a sliding window approach implemented in the R package "clusthaplo" (Leroux et al. 2014). Briefly, for a locus centered at a window of certain size, the similarity score between one pair of lines was computed as weighted measure of the number of IBS loci within that window, the length of the longest common genome segment of that window and, if the marker density in that window was low, the estimated genome-wide relatedness between the two lines. In this study, the first two weights were chosen from an exponential and uniform distribution, respectively. Afterwards, a Hidden Markov Model overcoming the threshold setting issue was applied to cluster the parental alleles at each locus. Clusters were firstly generated based on a set of genome-wide 35 k polymorphic SNP markers on the physical map, the positions of which were transformed into a centiMorgan scale by chromosome-wise ratios calculated from the length of the consensus map of each chromosome over its physical map length (Huang et al. 2011), and secondly extracted for the shared markers between the physical and genetic consensus map.

The window size to be used for clustering was determined in two steps. First, we investigated the LD decay along each chromosome. The LD between each pair of markers on each chromosome was calculated as r^2 (Hill and Robertson 1968) on the basis of 41 flint inbred lines including our four parents. The decay of r^2 on every chromosome was estimated according to Hill and Weir (1988). Second, to investigate how sensitive the clustering is with respect to the choice of the window size, five different window sizes ranging from 5 to 25 cM in steps of 5 cM were examined. The clustering results of each chromosome were evaluated with respect to (i) the average number of clustered ancestor alleles, (ii) the number of cluster changes defined as the change of at least one haplotype in the clustering result from locus to locus (Leroux et al. 2014), and (iii) the Pearson correlation coefficient between the modified Rogers'

distance among the four parental lines (Reif et al. 2005) and their clustering-based dissimilarity, calculated as the proportion of loci not sharing identical clusters. Finally, windows of size 20 and 10 cM were applied for chromosome 7 and the other chromosomes, respectively, on the basis of our findings and following the recommendation of Giraud et al. (2014) to choose as window size twice the genetic distance corresponding to $r^2 = 0.2$.

Detection of QTL with main effects and study of epistatic effects

Five biometric models, differing in the assumption about the number and effects of alleles at a QTL, were utilized to detect QTL with additive effects for every trait (Table 2), three based on linkage analysis (Model 1–3) and two incorporating LD and linkage information (Model 4 and 5). A detailed description of these models is given in supplementary materials.

Calculations for all five models were performed using iterative composite interval mapping (iQTLm, Charcosset et al. 2000) implemented in the MCQTL_LD software (Jourjon et al. 2005). A genome scan was performed for every marker and/or every 2 cM position, using flanking markers to infer the genotype at this position as described by Haley and Knott (1992), with multiple regression. Thresholds for declaring a putative QTL were determined by a permutation test using 1000 permutations to limit the genome-wise Type I error to 10 % for the joint analysis and 2 % for the single-family analysis to make the two comparable according to the Bonferroni correction (Blanc et al. 2006). Cofactors were selected by forward selection, restricting the distance between two adjacent cofactors to be greater than 20 cM. Support intervals of QTL positions were determined on the basis of 1-LOD unit drop. The proportion of genotypic variance (p_G) explained by each QTL (all QTL) was calculated as $p_G = R^2_{adj}/h^2$, where $R^2_{adj} = 1 - \frac{RSS_{full}/df_{full}}{RSS_{red}/df_{red}}$ and RSS_{full} and RSS_{red} refer to the residual sum of squares of the full model including the tested QTL (all QTL) and of the reduced model without the tested QTL (all QTL), respectively, and df_{full} and df_{red}

Table 2 Overview of the five biometric models used for detecting QTL with additive effects in multiple families

Model	Model name	QTL set	Number of alleles at a QTL	Reference(s)
1	Single-family model	One per family	Two per family	Blanc et al. (2006)
2	Disconnected model	Joint	Two × number of families	Blanc et al. (2006)
3	Connected model	Joint	Number of parents	Blanc et al. (2006)
4	Linkage disequilibrium and linkage model (LDLA)	Joint	Number of clustered ancestral alleles	Bardol et al. (2013), Giraud et al. (2014)
5	LDLA-1-marker model	Joint	Number of marker alleles, i.e., two (SNP) or at least two (SSR)	Würschum et al. (2012), Giraud et al. (2014)

refer to the degrees of freedom of the residual error in the full and reduced model, respectively (Giraud et al. 2014). Note that in the joint analyses, a family effect was included in both the full and reduced models and h^2 was calculated as the average heritability across all families. QTL detected with different models were considered different if their 1-LOD drop support intervals did not overlap.

QTL detected with Model 3 were further tested for digenic epistasis (Model 6) and QTL × genetic background (family) interaction (Model 7), applying the models detailed in Blanc et al. (2006) and our supplementary materials. Calculations for Model 6 were conducted with the "simple" method implemented in MCQTL_LD software (Jourjon et al. 2005). Calculations for Model 7 were conducted with a self-made program within the R environment (R Development Core Team 2008). The Type I error of both models was confined to 10 %. Estimates of R^2_{adj} and p_G were calculated for each significant interaction as described above for additive effects, except that the difference between the full and reduced model refers to the interaction term.

Cross-validation

Two cross-validation schemes, detailed below, were applied to (i) evaluate and compare Model 1–5 with two sample sizes of the TS and (ii) investigate with the connected model (Model 3), how the composition of the TS affects the prediction accuracy for the validation set (VS). Briefly, in each cross-validation run, QTL detection, localization and estimation of genetic effects were conducted in the TS and validation of the QTL results was performed in the VS as detailed by Utz et al. (2000). A Type I error of 10 % was applied to all models. Calculations were replicated 200 times with different random samples to obtain robust estimates using the R package "cvMCQTL" (Foiada et al. 2015) and our own extensions. To reduce the computation time of cross-validation, highly correlated markers from the dense genetic consensus map were removed by retaining only one marker per cM, which was polymorphic in most of the families compared with other markers within that 1 cM bin.

In Scheme 1, the same number of lines was sampled randomly without replacement from three completely connected families (R1R2, R1S1, and R2S1) to form the TS (Fig. S2). The VS was built in the same way from the remaining lines. Two TS sizes with $N_{TS} = 81$ (27 from each family) and 180 (60 from each family) and corresponding VS sizes with $N_{VS} = 48$ (16 from each family, except 15 for Model 1 in the case of R2S1 due to its small family size) and 108 (36 from each family) were employed. To enable direct comparison of prediction accuracies of the five models, the within-family prediction accuracy r_{VS} was

calculated for each family of the VS as the Pearson correlation between observed and predicted performance divided by the square root of h^2. If no QTL was identified in the TS, the prediction accuracy was set to zero. The prediction accuracy r_{VS} was averaged over all cross-validation runs and the corresponding standard deviation was determined. Moreover, the frequency of QTL detected in each 20 cM bin along the genome was recorded across cross-validation runs.

For Scheme 2, seven scenarios using TS of equal size composed of DH lines from a maximum of four families (R2S2 was excluded due to its small sample size) were investigated differing in (i) the number of families from which the DH lines were randomly sampled without replacement and (ii) the relatedness between the DH lines in the TS and VS ranging from full sib (F) and half sib (H) to unrelated (U) lines, ignoring relatedness among the four parent lines. The VS always comprised only one family. The scenarios were coded by the number of families in the TS and adding letters (F, H, U) reflecting the relationship of the genotypes in the TS to the genotypes in the VS (Table 3). If the TS included more than one family, the same number of DH lines was sampled from each family. Note that in all scenarios, both parents of the VS were parents of at least one family in the TS so that allele effects could be estimated with Model 3. The size of the TS varied from 60 to 280 in steps of 20. As with Scheme 1, r_{VS} was averaged over all cross-validation runs.

Results

Phenotypic data analysis and consensus map construction

R1R2 had the lowest family means ($\bar{X}$) of DON and GER, followed by R2S1 (Table 1). The differences in $\bar{X}$ of DS between families were small. Significant ($P < 0.01$) genotypic variances were observed for all traits in all families, with estimates being smallest for R1S1 and largest for R2S1. Heritabilities were generally high with an average of 0.78 for DON and GER and 0.90 for DS, and small differences among families. Genetic correlations were extremely tight between DON and GER ($r_g \geq 0.96$) and moderately negative ($-0.66 \leq r_g \leq -0.21$) for DS with DON or GER for all families except R2S2.

In total, 17,800 markers passed the quality check and were employed to construct the consensus map. It had a total length of 1854 cM and 2472 loci made up of 14,421 markers, out of which strongly correlated markers were combined into one locus before determining the marker order (Fig. S3). The genetic distance between adjacent loci ranged from 0 cM to a maximum of 9.4 cM across

Table 3 Composition of the training set (TS) and validation set (VS, in bold face letters) under seven scenarios and average coefficient of coancestry ($\bar{f}$) between the TS and VS

Scenario	No. of families in TS	Relatedness of TS to VS[a] and $\bar{f}$		TS population (VS family)[b]
1F	1	1F	$\bar{f} = 0.50$	R1R2(**R1R2**); R1S1(**R1S1**); R1S2(**R1S2**); R2S1(**R2S1**)
2H	2	2H	$\bar{f} = 0.25$	R1S1∪R2S1(**R1R2**); R1R2∪R2S1(**R1S1**); R1R2∪R1S1(**R2S1**)
3FH	3	1F + 2H	$\bar{f} = 0.33$	R1R2∪R1S1∪R2S1(**R1R2**); R1R2∪R1S1∪R2S1(**R1S1**); R1R2∪R1S1∪R2S1(**R2S1**); R1R2∪R1S1∪R1S2(**R1R2**); R1R2∪R1S1∪R1S2(**R1S1**); R1R2∪R1S1∪R1S2(**R1S2**)
3H	3	3H	$\bar{f} = 0.25$	R1S1∪R1S2∪R2S1(**R1R2**); R1R2∪R1S2∪R2S1(**R1S1**)
3HU	3	2H + U	$\bar{f} = 0.17s$	R1R2∪R1S1∪R1S2(**R2S1**)
4FH	4	1F + 3H	$\bar{f} = 0.31$	R1R2∪R1S1∪R1S2∪R2S1(**R1R2**); R1R2∪R1S1∪R1S2∪R2S1(**R1S1**)
4FHU	4	1F + 2H + 1U	$\bar{f} = 0.25$	R1R2∪R1S1∪R1S2∪R2S1(**R2S1**); R1R2∪R1S1∪R1S2∪R2S1(**R1S2**)

[a] F, H, and U refer to full-sib, half-sib and unrelated DH lines with $f = 0.50$, 0.25 and 0.00, respectively, ignoring relationships among the four parent lines

[b] The code inside the brackets refers to the family in the VS and the code(s) outside the bracket(s) refer(s) to family or the union (∪) of families in the TS

all chromosomes. The order of the markers on the physical map (Schnable et al. 2009), our consensus genetic map, and the SSR-based genetic map reported by Martin et al. (2012) showed collinearity across the entire genome with minor deviations in a few regions (Fig. S3). Map distances differed occasionally between the consensus map and the SSR-based genetic map, which had a total length of 2060 cM. Some regions on the physical map of each chromosome were devoid of markers. For instance, a huge segment of 75 Mb on chromosome 3, comprising the centromere, was completely lacking any markers, but the markers flanking this segment had a genetic map distance less than 5 cM, indicating extreme suppression of recombination.

Choice of window size and clustering of parental alleles

In the set of 41 flint lines used for examining the decay in LD between the markers on a chromosome, the threshold $r^2 = 0.2$ was reached at 9.6 cM for chromosome 7 and between 3.2 cM and 5.3 cM for the other chromosomes (Fig. S4). The average number of ancestral alleles at each locus obtained from clustering with five different window sizes was similar for all chromosomes (Fig. S5). The only exception was window size 5 cM, which resulted in a higher number of ancestral alleles on chromosome 6 than the other window sizes. Concerning the number of cluster changes on each chromosome, 5 cM deviated notably from the other window sizes. Nevertheless, for all window sizes, the correlation coefficients between the modified Rogers' distances among the four parent lines and their clustering-based dissimilarities were above 0.97 for all chromosomes. Therefore, we chose windows of size 20 cM for chromosome 7 and 10 cM for the other chromosomes. Clustering

of the four parental alleles at each locus to ancestral allele classes varied across loci and chromosomes (Fig. 1a, b). On average, centromeric regions had fewer clustered ancestral alleles than telomeric regions. Generally, parent lines with a higher coefficient of coancestry shared more often the same ancestral allele than others.

Detection of QTL with additive effects and epistatic QTL with different models

For DON, one to two QTL with additive effects were detected with Model 1 in each family explaining together $p_G = 14.8$–43.9 % of the genotypic variance (Table 4; Fig. 1c). The only exception was family R2S2 with the smallest sample size ($N = 43$), where no QTL was identified. Each QTL was detected in only one family, except one QTL on chromosome 2, which was shared between R1R2 and R2S1 (Table S1). In general, the favorable alleles originated from the resistant parents R1 and R2 except for one QTL on chromosome 1 in family R1S2, where the favorable allele was contributed by the susceptible parent S2. Interestingly, this QTL had only $p_G = 14.8$ %, even though the sample size in this family was fairly large ($N = 161$).

Compared with Model 1, the joint analyses (Model 2 to Model 5) of DON with all five families detected considerably more QTL (8–13) and had higher p_G values in the simultaneous fit, ranging between 34.4 and 52.9 % (Table 4, Fig. 1d). This included all QTL identified with Model 1 in all families and several new QTL, e.g., on chromosomes 4, 5, 6, 8, 9 and 10 (Table S1). Model 4 detected the largest number of QTL and had the highest p_G (52.9 %), whereas Model 5 detected besides Model 2 the least number of QTL with smallest p_G (34.4 %) and Model 3 was in between. Most of the favorable QTL alleles detected with

Fig. 1 a Number of esti-
mated ancestral alleles and **b**
clusters of the four parental
alleles at each locus along the
genome obtained with software
"clusthaplo". In **a** the position
of the centromeres is indicated
by *white darts*. In **b** each
distinct ancestral allele is given
a *single color* when it is shared
by at least two haplotypes. Note
that occurrence of the same
color at two different loci does
not imply anything about the
relatedness of the corresponding
alleles. Genome-wide values
for $-\log_{10}$ (P value of F test)
obtained from bi-parental (**c**)
and multi-family (**d**) QTL map-
ping analysis of deoxynivalenol
concentration (DON). The
location of QTL is indicated
above the curves with different
symbols for the various models.
The horizontal lines refer to the
significance thresholds of each
model
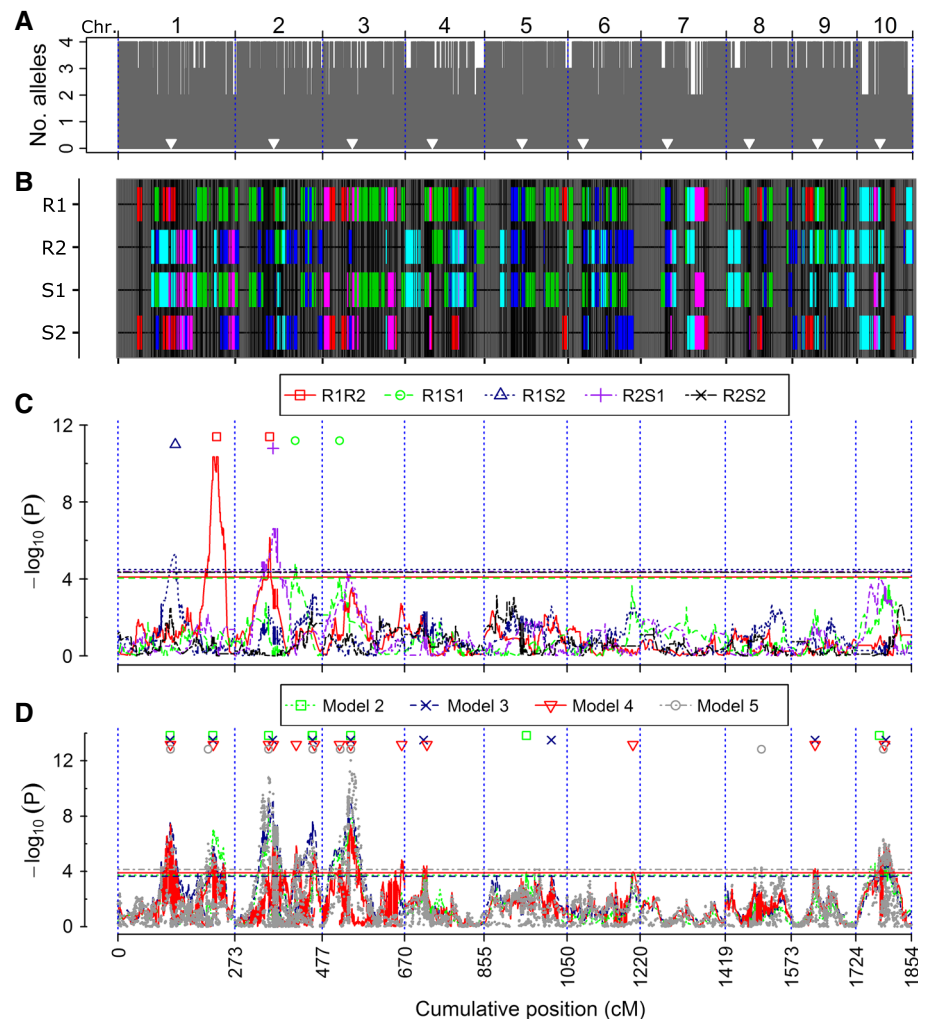
Table 4 Number of detected QTL and proportion (p_G) of the genotypic variance explained by all QTL in a simultaneous fit applying different models for deoxynivalenol concentration (DON), Gibberella ear rot severity (GER), and days to silking (DS)

Model	Family	N	DON		GER		DS	
			Number of QTL	p_G (%)[a]	Number of QTL	p_G (%)	Number of QTL	p_G (%)
1	R1R2	135	2	43.9	2	48.4	3	49.3
1	R1S1	104	2	22.2	2	23.3	2	23.2
1	R1S2	161	1	14.8	1	18.1	4	39.8
1	R2S1	96	1	30.0	1	20.1	3	61.4
1	R2S2	43	0	0	0	0	0	0
1	Total/average	–	5[b]	22.2[c]	4	22.0	8	34.7
2	All	639	8	41.2	8	39.9	11	58.3
3	All	639	9	44.0	13	51.5	12	56.3
4	All	639	13	52.9	11	39.5	11	48.3
5	All	639	8	34.4	11	22.1	13	39.6

[a] For calculation, see "Materials and methods"

[b] Number of different QTL across all families
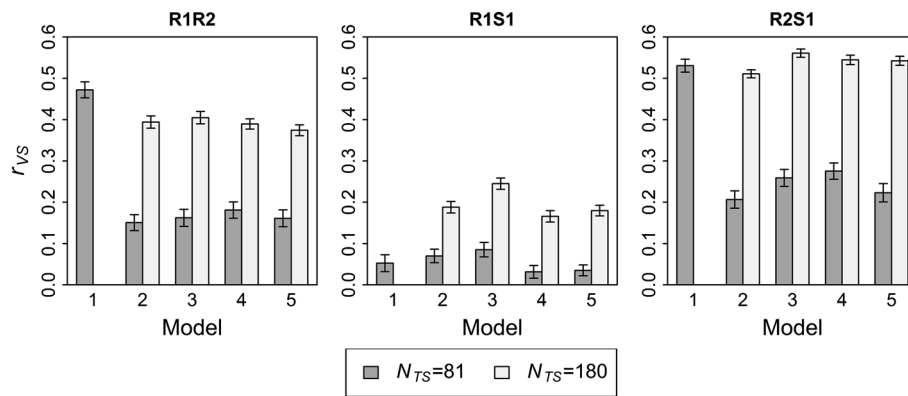
[c] Average of p_G values across all families

Fig. 2 Prediction accuracy (r_{VS}) in the validation set (*VS* indicated above each graph) composed of one family for deoxynivalenol concentration (DON) obtained by applying different models for single-family (model 1) and joint family analysis (model 2–5) of three fully connected families (R1R2, R1S1, R2S1). The height of black and gray columns refer to means across 200 cross-validation runs with $N_{TS} = 81$ and 180 and $N_{VS} = 48$ and 106, respectively, and the vertical bars show the corresponding standard deviation

Model 3, reducing DON, originated from parent line R1 with the highest resistance level (Table S2). Interestingly, the susceptible parent lines S1 and S2 also contributed resistance alleles with sizeable effects at some QTL.

For GER, with the exception of R2S2, one to two QTL displaying additive effects were detected with Model 1 in each family with p_G values from the simultaneous fit between 18.1 and 48.4 % (Table 4; Fig. S6a). Each QTL was detected in only one family except for one QTL on chromosome 2 shared between R1R2 and R1S1 (Table S1). As expected on the basis of the tight genotypic correlations between GER and DON, QTL for GER showed a high degree of co-localization and congruency of effects with QTL for DON and this applied irrespective of the model applied, but the ranking of the Model 3 and 4 in terms of the number of QTL detected and p_G differed.

For DS, Model 1 detected two to four QTL in all families except R2S2, with p_G values from the simultaneous fit ranging from 23.2 to 61.4 % (Table 4, Fig. S6c). In contrast to DON and GER, a large number of QTL identified for DS were congruent between two or three families (Table S1). For instance, two out of three QTL detected in R1R2 were also found in R1S2. For family R2S1, the QTL on chromosome 10 for DS and GER co-localized and had p_G values of 52.0 and 20.1 %, respectively. Model 2 to Model 5 detected all QTL identified with Model 1 in each family and several additional QTL (Table S1, Fig. S6d). Model 2 to Model 5 detected similar numbers of QTL (11 to 13), but Model 4 and Model 5 had smaller p_G values than Model 2 and 3.

No significant ($P < 0.05$) digenic epistasis was found for DON with Model 3. For GER, only one pair of QTL at position 67.3 cM on chromosome 9 and 62.6 cM on chromosome 10 displayed a significant ($P < 0.05$) interaction with $p_G = 5.5$ %. For DS, significant digenic epistasis was detected between the QTL at position 62.6 cM on

chromosome 2 and the QTL at 71.9 cM on chromosome 8 with $p_G = 4.4$ %. With Model 3, several significant ($P < 0.05$) QTL × genetic background (family) interactions were found, one for DON, three for GER and six for DS (Table S2), but the corresponding p_G values were consistently below 2.5 %.

Comparison of QTL mapping models via cross-validation

For Scheme 1 and DON, higher prediction accuracies r_{VS} were achieved for $N_{TS} = 81$ with Model 1 than with the joint analysis of Model 2 to 5 in each VS family except R1S1 (Fig. 2). For $N_{TS} = 180$, estimates of r_{VS} from the joint analysis increased substantially in each family and approached or even exceeded those of Model 1 with $N_{TS} = 81$. For both values of N_{TS}, there existed only minor differences between Model 2–5 in terms of r_{VS} within each family. Family R1S1, which had the lowest h^2 among the three families, had generally smaller r_{VS} values than the other two families. With Model 1 and $N_{TS} = 81$, the frequency of QTL detection in cross-validation runs was high (>0.4) for certain QTL, which were mostly specific for either R1R2 or R2S1, but generally low (<0.15) in R1S1 (data not shown). Joint analysis with Model 2–5 consistently identified QTL in those regions, where they were also detected with Model 1, but with low frequency (<0.15). Increasing the sample size to $N_{TS} = 180$ increased the QTL frequencies (>0.3) in the joint analysis considerably and all QTL identified by Model 1 with $N_{TS} = 81$ were detected.

Similar results were observed for GER and DS, except that (1) for both traits and $N_{TS} = 81$, Model 1 had in family R1S1 also a slightly higher mean r_{VS} than the joint analysis models (Fig. S7); (2) for DS and $N_{TS} = 81$, the detected QTL frequency in cross-validation runs showed

higher consistency among families than DON and GER (data not shown). For $N_{TS} = 81$, Model 3 or 4 reached generally the highest mean for r_{VS} among the joint analysis models. Both were in most cases superior to Model 2 and 5 and differences among models were less pronounced for $N_{TS} = 180$.

Effect of training set composition on prediction accuracy under different scenarios

The ranking of r_{VS} values for the different scenarios remained largely unaffected by the sample size of the TS and was almost identical for all traits (Fig. 3; Fig. S8). Prediction accuracies r_{VS} were higher under Scenario 1F than under all other scenarios for all VS families except for R1S1, where Scenario 3FH performed either equally well (DS) or better (DON, GER). In general, r_{VS} values obtained for scenarios (1F, 3FH, 4FH, 4FHU), which included different proportions of full sib DH lines in the TS, were higher than those without full sibs (2H, 3H, 3HU), irrespective of N_{TS}. Contrasting scenarios 2H with 3HU and 3FH with 4FHU showed that including unrelated lines generally reduced r_{VS} values. The increase in r_{VS} with increasing N_{TS} was generally highest for scenario 1F up to $N_{TS} = 140$, but the slope of the curves varied among the VS families.

Discussion

Historically, QTL mapping in maize started with bi-parental populations (Edwards et al. 1987). Following Lander and Botstein (1989), highly diverse parents were generally chosen to increase the chances of segregation of QTL, especially for resistance traits (Schön et al. 1993). The initial euphoria abated after it was recognized that QTL effects reported in the early studies were oftentimes highly inflated (*cf.* Schön et al. 2004) due to the so-called Beavis (1998) effect (Xu 2003), first described by Utz and Melchinger (1994). To obtain unbiased estimates of QTL effects, Utz et al. (2000) recommended to use cross-validation for separating QTL detection, corresponding to model selection, from estimation of QTL effects. Further, it was found that with small sample size of the mapping population, the power of QTL detection for quantitative traits with polygenic architecture is low (Schön et al. 2004). Different from academia, maize breeders commonly produce DH lines from several crosses, including resistant and susceptible parents, each family being only of moderate size. The five families of DH lines analyzed here are typical for this situation. The questions to be answered by our study were: (1) Should QTL mapping for marker-assisted selection under such a setting be conducted separately for each family or

Fig. 3 Prediction accuracy (r_{VS}) of QTL-based prediction across 200 cross-validation runs for deoxynivalenol concentration (DON) under different scenarios. The validation set (*VS* indicated above each graph) is composed of one family and the composition of the training set (TS) is reflected by the coding of the scenario: the number refers to the number of families included in the TS and the *letter(s)* refer(s) to their kinship with the VS, where *F, H,* and *U* denote full-sib, half sib and unrelated *lines*, respectively; for details see Table 4. QTL mapping was based on model 1 for Scenario 1F and on Model 3 for the other scenarios
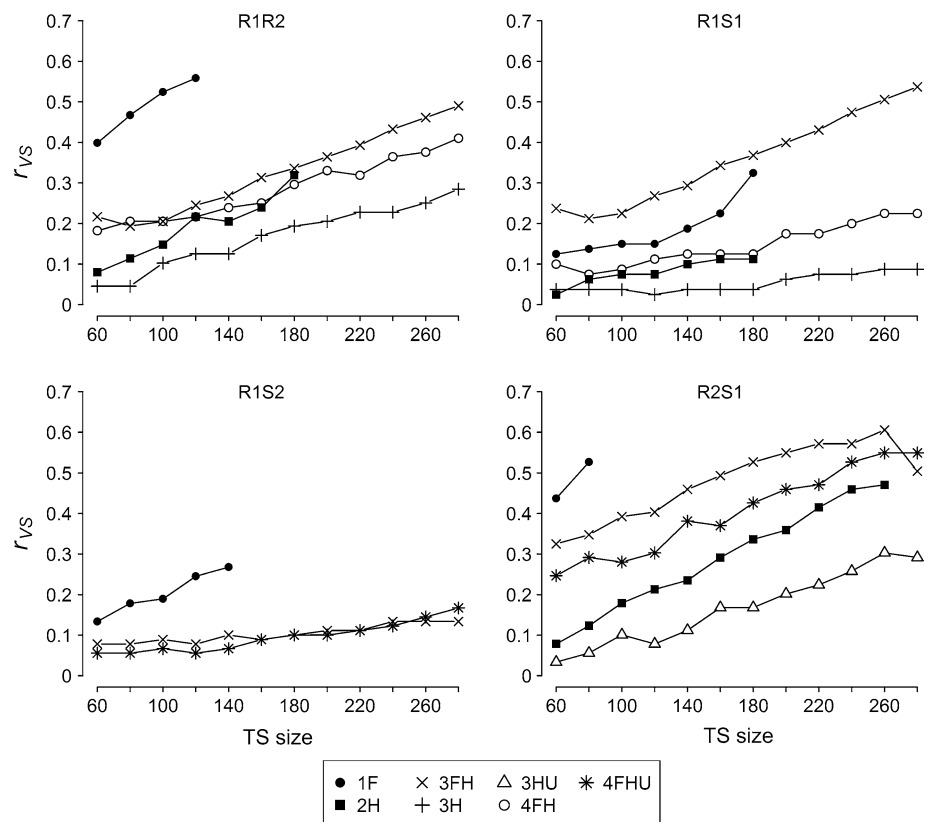
jointly across all families? (2) Which of the models proposed in the literature for joint analysis across families yield highest prediction accuracy of QTL-based prediction evaluated by cross-validation? (3) How does composition of the TS and its pedigree-relationship(s) to the VS influence the prediction accuracy?

Consensus map construction and recombination landscape

Multi-family QTL mapping requires a joint linkage map for all families included in the analysis. Construction of a consensus map can be complicated, if families differ largely in their recombination rate or even in the linear order of markers, but this is very unlikely with the interconnected families produced from related parents. Therefore, we applied *dsmergen* command in Carthagene to estimate one single recombination rate for all families and obtain consensus distance over families. Since SNP markers have only two alleles, each of these marker loci can segregate only in a subset of families of a connected design. However, owing to the high marker density provided by the MaizeSNP50 array, we found plenty of tightly linked markers segregating in different families, which enabled construction of a consensus map.

The total length of our consensus map (1854 cM) agreed well with the map lengths reported by Bauer et al. (2013) for families R1R2, R1S1 and R1S2, which ranged between 1655 and 1893 cM. Further, the linear order of markers on our consensus map was in excellent harmony with the high-density linkage map presented by Ganal et al. (2011) for the flint cross F2 × F252, but their map length was expanded due to four generations of intermating. Comparison of the consensus map and the physical map revealed strong recombination suppression in the centromeric regions of all chromosomes, most notably on chromosome 3, where a segment of 75 Mb had a map distance of 5 cM compared to a ratio of 0.07 cM per Mb averaged over the maize genome. Suppression of recombination in pericentromeric regions is in agreement with the results reported by Bauer et al. (2013) for European maize germplasm and Rodgers-Melnick et al. (2015) for US and Chinese maize germplasm.

While the consensus map should be constructed with great care, its influence on QTL mapping with multiple families depends primarily on the map density. If a high marker density is available as in our study, the recombination break points in the meiosis of the parental gamete of each DH line can be determined with high accuracy. Hence, in a genome scan with a high-density map, the genotype of the putative QTL employed for QTL mapping in the regression approach can be inferred with high fidelity from the observable genotype at tightly linked markers provided the population size is sufficiently large (Peleman et al. 2005).

Clustering parental alleles at each locus

Choice of the window size is critical for computing the similarity score between pairs of lines in "clusthaplo" (Leroux et al. 2014). Following Giraud et al. (2014), we initially chose a window size of 20 cM for chromosome 7 and 10 cM for all other chromosomes, based on the decay of LD for the same set of markers using representative lines from the same breeding pool as the four parent lines. To be on the safe side, we also varied the window size from 5 to 20 cM and observed for clustering with 5 cM a much larger average number of ancestral alleles and number of cluster changes. This is because under this setting, IBD segments greater than 5 cM are broken into pieces, which can lead to incorrect estimation of similarity score for loci at both ends of the haplotype. Chromosome 7 had the smallest average number of clustered ancestral alleles in agreement with its slow decay of LD. Centromeric regions had on average fewer clustered ancestral alleles than telomeric regions (Fig. 1a, b) in accordance with the different recombination rates along the genome mentioned above. These results differ from those of Giraud et al. (2014) who detected on average more ancestral alleles in the centromeric than in the telomeric regions, most likely because (i) we used a much higher marker density for clustering and (ii) the parent lines in our study were more closely related to each other, which facilitated accurate detection of IBD segments. Altogether, the number of ancestral alleles obtained from "clusthaplo" varied along the chromosomes and this resulted in different numbers of parameters in the LDLA model (Model 4), which caused an erratic pattern in the curves of the $-\log (P$ values) in Figs. 1d; S6b and S6d.

Detection of QTL with additive effects and epistatic QTL based on all five families

All families except R2S2 had been separately analyzed for QTL for GER and DON with low-density maps comprising between 106 and 129 SSR markers (Martin et al. 2011, 2012). We identified with Model 1 only a subset of the QTL detected previously, because different from these authors, we applied a more stringent significance level ($\alpha = 2\%$ vs. 15 %) in permutation tests to protect against a high global Type I error rate in multiple tests with several families. The QTL detected by us were always adjacent to the flanking markers reported previously, but their exact position was shifted primarily as a result of the change in the genetic map caused by the higher marker density.

In agreement with Blanc et al. (2006) and Ogut et al. (2015), the QTL detected by Model 1 in each family generally showed little congruency across families, suggesting that each family comprised a unique set of segregating QTL. The only congruent QTL for DON was found for

families R1R2 and R2S1 on chromosome 2, explaining a high percentage of the genotypic variance (p_G = 20.3 and 30.0 %, respectively), and the favorable allele originated in both families from the common parent R2. This is in accordance with the findings of Blanc et al. (2006) that congruent QTL among families often have large effects and originate from a common parent. In contrast, several QTL for DS detected with Model 1 were consistent across two or three families, suggesting that the level of congruency of QTL across families depends strongly on the trait. While most QTL for all three traits were family specific, we detected with Model 3 only few significant QTL × genetic background (family) interactions. Either the family sizes in our study were too small to warrant sufficient power for detecting this type of epistasis, or epistatic effects are small for the investigated traits. Results from studies with the US NAM panel with 200 recombinant inbred lines from each of 25 families on flowering date (Buckler et al. 2009) and a genome-wide association mapping study of *Fusarium verticillioides* with 1687 lines from the USDA gene bank (Zila et al. 2013) support the latter explanation.

For all traits, Model 2 detected all the QTL identified with Model 1 and additional QTL even though both models assume that allele effects of QTL are nested within families. Thus, joint analysis of several families with Model 2 can benefit from more replicates of QTL genotypes, which leads to a higher power of QTL detection, especially if common QTL are shared between families. This finding is somewhat different from the results by Blanc et al. (2006), where the QTL detected with these two models displayed greater discrepancies. This may be due to different genetic architecture of the traits and/or the higher marker density in our study which increased the power of QTL detection for both models. Model 3 detected more QTL than Model 2 for DON and GER. This is in line with Blanc et al. (2006) and can be explained by (i) a smaller number of parameters to be estimated in Model 3 than in Model 2, which leads to a higher power of QTL detection (Rebai and Goffinet 2000), and (ii) the low importance of epistasis observed for these traits. Contrary to expectation, Model 4 did not outperform the other joint analysis models for GER and DS. This may be attributable to the small number of parents in our study so that the gain in power for Model 4, expected from reducing the number of parameters by clustering the parental alleles, was limited. This is different from the study of Bardol et al. (2013), which involved more parents so that clustering the parental alleles resulted in a substantial reduction of the parameters in the model. Although the number of QTL detected for GER and DS was not smallest for Model 5, it yielded the lowest values of p_G for all three traits. This implies that either some of the QTL detected by Model 5 were false positives or estimates of the allele effects at the detected QTL were inaccurate, as expected if the numbers

of alleles at QTL exceed those at adjacent markers. The latter explanation is consistent with Lu et al. (2012) and Bardol et al. (2013), who observed that multi-allelic models capture a greater proportion of the genetic variance than bi-allelic models.

Comparison of models via cross-validation

For marker-assisted selection, breeders are interested in the prediction accuracy of genotypes on the basis of the detected QTL. To warrant a fair comparison of the different models for QTL detection, unbiased estimates of the prediction accuracy were determined by cross-validation using Scheme 1 with the following features: (i) Three completely interconnected families (R1R2, R1S1, and R2S1) were analyzed so that every pair of parental alleles could be contrasted with greatest power (Wu and Jannink 2004). (ii) The same number of DH lines was sampled from each family for composition of the TS and VS so that each family contributed equally to QTL detection, estimation of parameters and prediction. (iii) All models were compared with the same sample size for the TS (N_{TS} = 81 or 180) and VS (N_{VS} = 48 or 108). (iv) The same Type I error of 10 % was applied for all models.

In contrast to Ogut et al. (2015), who found that joint analysis generally had higher prediction abilities than single-family analysis, our results showed that prediction accuracies r_{VS} for individual families determined with cross-validation were for most traits and families with N_{TS} = 81 lower for the joint analysis models than for Model 1 (Fig. 2; Fig. S7). Obviously, the superiority of Model 1 for small sample sizes depends strongly on the genetic architecture of the trait across all families. If specific QTL with large effects prevail in each family, as applies to DON and GER, the power of detecting these QTL is lower for Model 2–5 than for Model 1, because only a subset of genotypes (one-third under Scheme 1) will segregate in the TS used by these models. In contrast, if a QTL with a small effect segregates in one family, but has a large effect in the other families, joint analysis will most likely detect this QTL and using it for prediction can help to increase r_{VS}. If the number of DH lines from each family in the TS is larger so that N_{TS} = 180 for the joint analysis, then Model 3 reached generally similar values for r_{VS} as Model 1 for N_{TS} = 81. Thus, the superiority of Model 1 over Model 3 seems to be strongly dependent on the number of individuals from the family to be predicted that are included in the TS besides the genetic architecture of the trait and the congruency of QTL across families.

Depending on the trait, Ogut et al. (2015) reported generally poor consistency of the QTL detected by Model 1 and 2. We found that QTL detected by Model 1 with high frequency were all identified by the joint analysis models

with low ($N_{TS} = 81$) or high ($N_{TS} = 180$) frequencies (data not shown). Furthermore, for Scheme 1 and $N_{TS} = 81$, Model 2 generally had lower r_{VS} than Model 3 and Model 4 (Fig. 2; Fig. S7). Although the difference was not very big, this finding still suggests that Model 2 is most likely not the best choice in the experiments involving families of small size ($N \leq 27$). Moreover, since Model 2 assumes QTL effects are specific to families (Table 2), full sibs must be included in the TS to estimate the QTL effects in the TS and predict other full sibs in the VS. This feature imposes considerable restrictions on possible composition of the TS, and, therefore, makes Model 2 less attractive than Model 3 and 4. For Scheme 1 and $N_{TS} = 180$, no substantial differences in r_{VS} among the four joint analysis models were observed for all traits, contrary to the findings on p_G based on the full data set (Table 4). This could be explained by the composition of the TS with different sampling of genotypes from the individual families and different assumptions about the number of QTL alleles. For all traits, Model 5 generally reached for $N_{TS} = 81$ slightly lower r_{VS} values than Model 3 and Model 4. This finding is consistent with that of p_G for the full data set. Thus, Model 5 most likely explains a smaller proportion of the genotypic variance than the other models allowing for multi-allelic QTL, even though it had a similar power of QTL detection, as reflected by the number detected QTL (Table 4). In conclusion, multi-family QTL analysis is superior to single-family analysis only if each family is represented by an adequate sample size (generally >60) in the TS and common QTL do exist. Model 3 or 4 exceeded Model 2 and 5, when evaluated with cross-validation, but differences among these models were generally small.

It must be noted that the p_G values presented in Table 4 are mostly likely inflated, because they were determined without cross-validation. To obtain an idea about the upward bias and its dependency on the model, we determined for Scheme 1 in addition to the p_G values for the VS, which correspond to the square of the r_{VS} values, also the p_G values for the TS, using the same method as described for the full data set. Compared to the TS, the p_G values in the VS averaged for Model 2–5 only about 45 % for $N_{TS} = 81$ and about 70 % for $N_{TS} = 180$ (data not shown). Moreover, the relative size of the upward bias in the p_G values of the TS [(p_G in TS/p_G in VS) − 1] was almost the same for Model 2–5 so that choice of the best model could be determined from analyses of the full data set without cross-validation.

Design of the training set for QTL-based prediction

Riedelsheimer et al. (2013) examined the prediction accuracy r_{VS} for genomic prediction with GBLUP using the same five families and the phenotypic and genotypic data (excluding SSR markers) as used in this study. Regarding the composition of the TS, they found that r_{VS} was highest for scenario 1F and much lower for scenarios 2H and 3H with a further minor reduction for 3HU. We found in most cases the same ranking for these scenarios (Fig. 3; Fig. S8). Scenarios (1F, 3FH, 4FH, 4FHU), where the TS included various proportions of full sibs to the genotypes in the VS, had generally higher r_{VS} in our study than scenarios (2H, 3H, 3HU) with only half sibs or with half sibs and unrelated lines. This is in agreement with experimental result on genomic prediction with full-sib and half-sib families by Lehermeier et al. (2014) and Foiada et al. (2015). Moreover, r_{VS} increased in most cases linearly with increasing N_{TS} and the slope was generally steeper for the scenarios with full sibs in the TS. Thus, the average kinship $\bar{f}$ between the TS and VS was the main factor determining r_{VS} for a given sample size. The only exception was family R1S1, where r_{VS} for DON and GER was higher for scenario 3FH with $\bar{f} = 0.33$ than for 1F with $\bar{f} = 0.50$. Possible explanations, why prediction including full sibs in the TS generally achieved higher accuracy than all other scenarios, could be that related families share more QTL than less related ones. These QTL could be rare QTL, which segregate and have significant effects in only one or a limited number of families as observed in our study for DON and GER. The number of families included in the TS hardly affected r_{VS}, once both parents of the VP were parents of at least one population in the TS. However, the kinship between the TS and VS seems less influential, when both related and less related families comprise a large number of common QTL, as demonstrated for R1R2 for DS, where r_{VS} was similar for all scenarios (Fig. S8b). In addition, we found that the number of QTL and the size of QTL effects for all traits depended strongly on the family. The deviation observed for R1S1 can be explained by the observation that R1S1 comprised both rare and common QTL for all the three traits, and the average kinship $\bar{f}$ between the TS and VS did not accurately reflect the resemblance at the QTL for scenarios 3FH and 1F. Altogether, our results suggest a major influence of $\bar{f}$ on r_{VS}, but the association seems to depend on the trait. While kinship measurements between genotypes, either based on pedigree or genome-wide markers, basically estimate the genome-wide resemblance, they may fail to reflect the specific resemblance with respect to the QTL influencing the trait of interest (Würschum and Kraft 2015). Nevertheless, identical linkage phases between marker and QTL may be more persistent between related materials than less related ones, as discussed by Riedelsheimer et al. (2013) and supported by our analysis of ancestral alleles with "clusthaplo".

In conclusion, our results strongly emphasize to apply QTL-based prediction only, if the TS includes at least 60 genotypes being full sibs to the VS. If no full sibs are

available, both parents of the VS should be included as parents in half-sib families in the TS. Inclusion of unrelated DH lines seems of questionable value. Without a sufficient number of full sibs in the TS, the risk of having a very low prediction accuracy is high even with the most advanced methods of multi-family QTL mapping. Thus, breeders are advised to ascertain a high degree of connectedness among the families, which they want to use for QTL-based prediction in marker-assisted selection or for genomic selection.

Author contribution statement AEM and TM designed the experiments; TM and AEM conducted the experiments with the help of others mentioned in Acknowledgements; EB and CCS generated the SNP data used in this study; MS conducted the phenotypic data analysis, conducted the marker quality check and compiled the linkage maps for each family; SH constructed the consensus map and performed all QTL analyses with the help of HFU and WL; SH and AEM drafted the manuscript which was edited by all authors.

Compliance with ethical standards

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

Ethical standard The experiments reported in this study comply with the current laws of Germany.

References

Bardol N, Ventelon M, Mangin B et al (2013) Combined linkage and linkage disequilibrium QTL mapping in multiple families of maize (Zea mays L.) line crosses highlights complementarities between models based on parental haplotype and single locus polymorphism. Theor Appl Genet 126:2717–2736. doi:10.1007/s00122-013-2167-9

Bauer E, Falque M, Walter H et al (2013) Intraspecific variation of recombination rate in maize. Genome Biol 14:R103. doi:10.1186/gb-2013-14-9-r103

Beavis WD (1998) QTL analyses: power, precision, and accuracy. In: Paterson AH (ed) Molecular dissection of complex traits. CRC press, New York, pp 145–162

Bink MCAM, Totir LR, ter Braak CJF et al (2012) QTL linkage analysis of connected populations using ancestral marker and pedigree information. Theor Appl Genet 124:1097–1113. doi:10.1007/s00122-011-1772-8

Blanc G, Charcosset A, Mangin B et al (2006) Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize. Theor Appl Genet 113:206–224. doi:10.1007/s00122-006-0287-1

Bolduan C, Miedaner T, Schipprack W et al (2009) Genetic variation for resistance to ear rots and mycotoxins contamination in early European maize inbred lines. Crop Sci 49:2019–2028. doi:10.2135/cropsci2008.12.0701

Buckler ES, Holland JB, Bradbury PJ et al (2009) The genetic architecture of maize flowering time. Science 325:714–718. doi:10.1126/science.1174276

Charcosset A, Mangin B, Moreau L, Combes L, Jourjon MF et al (2000) Heterosis in maize investigated using connected RIL populations. In: Quantitative genetics and breeding methods: the way ahead. INRA, Paris, pp 89–98

de Givry S, Bouchez M, Chabrier P et al (2005) CARTHA GENE: multipopulation integrated genetic and radiation hybrid mapping. Bioinformatics 21:1703–1704. doi:10.1093/bioinformatics/bti222

Edwards MD, Stuber CW, Wendel JF (1987) Molecular-marker-facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116:113–125

Foiada F, Westermeier P, Kessel B et al (2015) Improving resistance to the European corn borer: a comprehensive study in elite maize using QTL mapping and genome-wide prediction. Theor Appl Genet 128:875–891. doi:10.1007/s00122-015-2477-1

Ganal MW, Durstewitz G, Polley A et al (2011) A large maize (zea mays L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS one. doi:10.1371/journal.pone.0028334

Giraud H, Lehermeier C, Bauer E et al (2014) Linkage disequilibrium with linkage analysis of multiline crosses reveals different multiallelic QTL for hybrid performance in the flint and dent heterotic groups of maize. Genetics 198:1717–1734. doi:10.1534/genetics.114.169367

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

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

Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. Theor Popul Biol 33:54–78. doi:10.1016/0040-5809(88)90004-4

Holland JB (2007) Genetic architecture of complex traits in plants. Curr Opin Plant Biol 10:156–161. doi:10.1016/j.pbi.2007.01.003

Huang X, Paulo MJ, Boer M, Effgen S, Keizer P, Koornneef M, van Eeuwijk FA (2011) Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. Proc Natl Acad Sci 108:4488–4493

Huang BE, Verbyla KL, Verbyla AP et al (2015) MAGIC populations in crops: current status and future prospects. Theor Appl Genet 128:999–1017

Jannink JL, Jansen R (2001) Mapping epistatic quantitative trait loci with one-dimensional genome searches. Genetics 157:445–454

Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Brief Funct Genom 9:166–177. doi:10.1093/bfgp/elq001

Jansen RC, Jannink JL, Beavis WD (2003) Mapping quantitative trait loci in plant breeding populations. Crop Sci 43:829. doi:10.2135/cropsci2003.0829

Jourjon MF, Jasson S, Marcel J et al (2005) MCQTL: multi-allelic QTL mapping in multi-cross design. Bioinformatics 21:128–130. doi:10.1093/bioinformatics/bth481

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps (published erratum appears in Genetics 1994 Feb; 136(2):705). Genetics 121:185–199

Lehermeier C, Krämer N, Bauer E et al (2014) Usefulness of multiparental populations of maize (*Zea mays* L.) for genome-based prediction. Genetics 198:3–16. doi:10.1534/genetics.114.161943

Leroux D, Rahmani A, Jasson S et al (2014) Clusthaplo: a plugin for MCQTL to enhance QTL detection using ancestral alleles in multi-cross design. Theor Appl Genet 127:921–933. doi:10.1007/s00122-014-2267-1

Li H, Bradbury P, Ersoz E et al (2011) Joint QTL linkage mapping for multiple-cross mating design sharing one common parent. PLoS One. doi:10.1371/journal.pone.0017573

Liu Y, Zeng ZB (2000) A general mixture model approach for mapping quantitative trait loci from diverse cross designs involving multiple inbred lines. Genet Res 75:345–355. doi:10.1017/S0016672300004493

Lu Y, Xu J, Yuan Z et al (2012) Comparative LD mapping using single SNPs and haplotypes identifies QTL for plant height and biomass as secondary traits of drought tolerance in maize. Mol Breed 30:407–418. doi:10.1007/s11032-011-9631-5

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. doi:10.2135/cropsci2010.11.0664

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. doi:10.2135/cropsci2011.04.0214

Melchinger AE, Utz HF, Schön CC (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics 149:383–403. doi:10.1016/1369-5266(88)80015-3

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

Miedaner T, Han S, Kessel B, et al (2015) Prediction of deoxynivalenol and zearalenone concentrations in *Fusarium graminearum* inoculated backcross populations of maize by symptom rating and near-infrared spectroscopy. Plant Breed 009:n/a–n/a. doi:10.1111/pbr.12297

Mode CJ, Robinson HF (1959) Pleitropism and the genetic variance and covariance. Biometrics 15:518–537. doi:10.2307/2527650

Ogut F, Bian Y, Bradbury PJ, Holland JB (2015) Joint-multiple family linkage analysis predicts within-family variation better than single-family analysis of the maize nested association mapping population. Hered (Edinb) 114:552–563. doi:10.1038/hdy.2014.123

Peleman JD, Wye C, Zethof J, Sorensen AP, Verbakel H, van Oeveren J, Gerats T, van der Voort JR (2005) Quantitative trait locus (QTL) isogenic recombinant analysis: a method for high-resolution mapping of QTL within a single population. Genetics 171(3):1341–1352. doi:10.1534/genetics.105.045963

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

Rebai A, Goffinet B (1993) Power of tests for QTL detection using replicated progenies derived from a diallel cross. Theor Appl Genet 86:1014–1022. doi:10.1007/BF00211055

Rebai A, Goffinet B (2000) More about quantitative trait locus mapping with diallel designs. Genet Res 75:243–247

Reif JC, Melchinger AE, Frisch M (2005) Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. Crop Sci 45:1–7. doi:10.2135/cropsci2005.0001

Riedelsheimer C, Endelman JB, Stange M et al (2013) Genomic predictability of interconnected biparental maize populations. Genetics 194:493–503. doi:10.1534/genetics.113.150227

Rodgers-Melnick E, Bradbury PJ, Elshire RJ et al (2015) Recombination in diverse maize is stable, predictable, and associated with genetic load. Proc Natl Acad Sci 112:201413864. doi:10.1073/pnas.1413864112

Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115. doi:10.1126/science.1178534

Schön CC, Lee M, Melchinger AE et al (1993) Mapping and characterization of quantitative trait loci affecting resistance against second-generation European corn borer in maize with the aid of RFLPs. Hered (Edinb) 70:648–659. doi:10.1038/hdy.1993.93

Schön CC, Utz HF, Groh S et al (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. doi:10.1534/genetics.167.1.485

Steinhoff J, Liu W, Maurer HP et al (2011) Multiple-line cross quantitative trait locus mapping in European elite maize. Crop Sci 51:2505. doi:10.2135/cropsci2011.03.0181

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

Utz HF (2005) PLABSTAT: a computer program for the statistical analysis of plant breeding experiments. University of Hohenheim, Germany

Utz HF, Melchinger AE (1994) Comparison of different approaches to interval mapping of quantitative trait loci. In: Ooijen JW van, Jansen J (ed), Biometrics plant Breed Appl Mol markers Wageningen: the Netherlands, 6–8 July 1994. 1994, 195–204 ST

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

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. Heredity 93(1):77–78

Wu XL, Jannink JL (2004) Optimal sampling of a population to determine QTL location, variance, and allelic number. Theor Appl Genet 108:1434–1442. doi:10.1007/s00122-003-1569-5

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. doi:10.1371/journal.pgen.1000212

Würschum T, Kraft T (2015) Evaluation of multi-locus models for genome-wide association studies: a case study in sugar beet. Heredity 114:281–290

Würschum T, Liu W, Gowda M et al (2012) Comparison of biometrical models for joint linkage association mapping. Hered (Edinb) 108:332–340. doi:10.1038/hdy.2011.78

Xu S (1998) Mapping quantitative trait loci using multiple families of line crosses. Genetics 148:517–524

Xu S (2003) Theoretical basis of the Beavis effect. Genetics 165:2259–2268

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178:539–551. doi:10.1534/genetics.107.074245

Zila CT, Samayoa LF, Santiago R et al (2013) A Genome-Wide association study reveals genes associated with Fusarium ear rot resistance in a maize core diversity panel. G3: Genes|Genomes|Genet. doi:10.1534/g3.113.007328