

ORIGINAL ARTICLE

Optimum design of family structure and allocation of resources in association mapping with lines from multiple crosses

W Liu^{1,2}, HP Maurer², JC Reif², AE Melchinger³, HF Utz³, MR Tucker⁴, N Ranc⁵, G Della Porta⁶ and T Würschum²

Family mapping is based on multiple segregating families and is becoming increasingly popular because of its advantages over population mapping. Although much progress has been made recently, the optimum design and allocation of resources for family mapping remains unclear. Here, we addressed these issues using a simulation study, resample model averaging and cross-validation approaches. Our results show that in family mapping, the predictive power and the accuracy of quantitative trait loci (QTL) detection depend greatly on the population size and phenotyping intensity. With small population sizes or few test environments, QTL results become unreliable and are hampered by a large bias in the estimation of the proportion of genotypic variance explained by the detected QTL. In addition, we observed that even though good results can be achieved with low marker densities, no plateau is reached with our full marker complement. This suggests that higher quality results could be achieved with greater marker densities or sequence data, which will be available in the near future for many species.

Heredity (2013) **110**, 71–79; doi:10.1038/hdy.2012.63; published online 10 October 2012

INTRODUCTION

Family mapping refers to association mapping in lines derived from multiple biparental crosses (Myles *et al.*, 2009) and is a powerful genomic tool to dissect the genetic architecture of quantitative traits (Yu *et al.*, 2008; Buckler *et al.*, 2009). It is based on the analysis of multiple segregating families, typically from connected crosses. Compared with association mapping in a diverse panel of lines (population mapping), the more balanced allele frequencies in family mapping can facilitate a higher quantitative trait loci (QTL) detection power and reduce the confounding effects of genetic relatedness (McMullen *et al.*, 2009). In addition, family mapping holds the promise that the identified QTL are more stable than the QTL detected in biparental populations, which are often not transferable from one population to another (Holland, 2007). Family mapping has been applied not only to natural populations (Wu *et al.*, 2002; Kover *et al.*, 2009) but also to populations from plant breeding programs in different crop species (for example, Reif *et al.*, 2010; Liu *et al.*, 2011).

A recent comparison between biometric models available for family mapping indicated that they differ substantially with regard to their QTL detection power, correction for family structure and their use of the available linkage disequilibrium (LD) (Würschum *et al.*, 2012). Even though different family mapping studies are available, it remains unclear which model is most appropriate for the analysis of family mapping data sets.

For MAS programs to be superior to field evaluation and classical phenotypic selection, QTL positions and effects must be estimated with high precision and the proportion of genotypic variance explained by the QTL must be high. In linkage mapping, it has been shown through simulation studies (Beavis 1998) and experimental data (Utz *et al.*, 2000; Schön *et al.*, 2004), that the QTL effects and the proportion of explained genotypic variance are often overestimated, leading to an excessively optimistic assessment of MAS. This effect is strongly influenced by the size of the mapping population, with smaller populations showing a much larger reduction in the amount of genotypic variance explained by the QTL in an independent test set (TS). The reasons for this inflated estimation of explained genotypic variance include epistatic and environmental (G × E) interactions, and that QTL detection and estimation of their effects are performed in the same population. Different approaches have been suggested to obtain unbiased estimates of QTL effects. The cross-validation approach, applied by Schön *et al.* (2004), also appears appropriate for family mapping to obtain more realistic estimates of the genotypic variance explained by the detected QTL (Würschum *et al.*, 2012).

To date, however, no study has evaluated the bias in the estimation of QTL effects and the precision of QTL prediction in family mapping approaches. In addition, fundamental questions remain when planning family mapping experiments, and these relate to the optimum design of family structure and the allocation of resources. In this study, we used complementary simulation and experimental

¹Crop Genetics and Breeding Department, China Agricultural University, Beijing, China; ²State Plant Breeding Institute, University of Hohenheim, Stuttgart, Germany; ³Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Stuttgart, Germany; ⁴ARC Centre of Excellence for Plant Cell Walls, University of Adelaide, Urrbrae, South Australia, Australia; ⁵Syngenta Seeds SAS, chemin de l'Hobit, Saint-Sauveur, France and ⁶Syngenta Seeds SpA, Casalmorano, Italy
Correspondence: Dr T Würschum, State Plant Breeding Institute, University of Hohenheim, Fruwirthstrasse 21, Stuttgart 70599, Germany.
E-mail: tobias.wuerschum@uni-hohenheim.de

approaches to address these questions. In detail, our objectives were to (1) investigate the optimum allocation of resources in family mapping by computer simulations, (2) evaluate the precision of QTL position estimates and the predictive power of QTL identified by family mapping based on experimental data of a large maize population (930 lines evaluated in six environments), (3) investigate the optimum allocation of resources by using an array of genotypic and environmental subpopulations derived from the full data set, (4) assess the performance of data sets that are either balanced or unbalanced with regard to family size or the parental contribution to the mapping population and (5) analyze the effect of the applied marker density on family mapping.

MATERIALS AND METHODS

Simulation study

For the simulation study considering the optimum allocation of resources, a random-mating population of infinite size was simulated using the software package Plabsoft (Maurer *et al.*, 2008). This population was used to generate an infinite population of doubled haploid (DH) lines from which the parents were randomly sampled. Three different scenarios concerning LD in the base population were simulated. Adjacent loci were either simulated to be in linkage equilibrium (no LD) or in high or low LD (the decay of LD with genetic map distance for the high and low LD scenario is shown in Supplementary Figure S1). Loci on different chromosomes were in linkage equilibrium in all three scenarios. Each genotype possessed two chromosomes with 150 cM length that had a marker every cM with allele frequencies of 0.5. In the middle of one of these chromosomes, a QTL was simulated that in the population of DH parental lines explained 5% of the genotypic variance, whereas the remaining genotypic variance was equally contributed by QTL on 100 background chromosomes with allele frequencies of 0.5.

A total budget of 10 000 plot equivalents was assumed. This restricted budget was used for (1) developing and genotyping of DH lines (three plot equivalents per DH) and (2) field testing (one plot per DH and test environment). The number of replications per environment was set to one. In studies without an upper limit for test environments, this has recently been shown to maximize the heritability in plant breeding programs (Longin *et al.*, 2007). The costs for each scenario were thus: $N \times (3 \times \text{plots} + \text{Env} \times \text{plots})$, where N refers to the number of lines and Env to the number of test environments. Genotypes were randomly sampled from the base population and converted into DH lines to generate the parents. The parents (3–15) were crossed in two popular mating designs, a diallel (DIA) design or a single round robin (SRR) design, to generate the DH mapping populations. Family sizes were determined by the total budget and ranged from 666 (3 parents, 2 environments) to 39 (15 parents, 14 environments) for SRR and from 666 (3 parents, 2 environments) to 5 (15 parents, 14 environments) for DIA. The ratio of the variance components for genotype, genotype by location and residual was assumed to be 1:2:2. The threshold for QTL detection was $P < 0.001$ and the second chromosome without QTL was used to estimate the false-positive rate. A total of 300 runs, corresponding to 300 data sets, were conducted for each combination of parents and locations (test environments).

Plant materials, field experiments and molecular markers

The analyses are based on the population described in Liu *et al.* (2011). Nine elite inbred lines originating from the stiff stalk heterotic group were used as parents and crossed in an incomplete DIA design. In total, 11 segregating families were generated by single seed descent until the F_3 generation or *in vivo* DH induction (Supplementary Figure S2). The 930 genotypes (292 F_3 and 638 DH lines) were crossed to a tester, which was an elite inbred line from the opposite European heterotic pool. All plant materials used in this study are proprietary to Syngenta Seeds SAS, chemin de l'Hobit, Saint-Sauveur, France.

The 930 testcross progenies were evaluated in unreplicated trials in six environments in 2007. Two-row plots (8.2 m²) were machine planted (5.5–7.0 plants m⁻²). Data were recorded for grain yield (Mgha⁻¹, adjusted to 155 g kg⁻¹ of grain moisture) and for grain moisture (content in g kg⁻¹ at harvest stage).

The 930 genotypes were fingerprinted following standard protocols with 425 single-nucleotide polymorphism (SNP) markers. These markers were randomly distributed across the genome with an average marker distance of 2.8 cM and a maximum gap between adjacent markers of 23 cM. Map positions of all markers were based on the integrated linkage map of Liu *et al.* (2011) with a total map length of 1207 cM.

Phenotypic data analyses

All quantitative genetic parameters were estimated on the basis of the testcross performance of the 930 maize lines. An analysis of variance was performed based on the model: $y_{ij} = \mu + G_i + E_j + e_{ij}$, where y_{ij} is the adjusted entry mean of the i th maize line at the j th environment, μ the intercept term, G_i is the genetic effect of the i th maize line modeled as fixed effect, E_j the effect of the j th environment and e_{ij} the error term including the genotype-by-environment interaction effect. Both G_i and E_j were modeled as random effects to obtain estimates of variance due to genotype (σ^2_G) and variance of the residuals (σ^2_e) (confounded with genotype-by-environment interaction variance, $\sigma^2_{G \times E}$) as described by Searle (1971, p. 475). Heritability (h^2) on an entry-mean basis was estimated as the ratio of genotypic to phenotypic variance according to Hallauer and Miranda (1981): $h^2 = \sigma^2_G / (\sigma^2_G + \sigma^2_{e/E})$, where E refers to the number of environments. In addition, best linear unbiased estimates across the six environments were calculated for the 930 testcross progenies for both traits with the model described above, except that G_i was modeled as fixed effect.

Association mapping

For QTL mapping, an additive genetic model was chosen for the testcross progenies as described by Melchinger *et al.*, (1998). The estimated best linear unbiased estimates across locations were used for the family mapping analysis. We used a multiple-regression approach that has previously been shown to perform well in a model comparison for family mapping (Model A and Model B from Würschum *et al.* (2012)). Briefly, the applied models included cofactors (Model A) or cofactors and an additional family effect (Model B), and the approach is based on a two-step procedure for QTL detection. In the first step, cofactors were selected by stepwise multiple linear regression based on the Schwarz Bayesian criterion (Schwarz, 1978). In the second step, we calculated a P -value for the F-test with a full model (including SNP effect) versus a reduced model (without SNP effect). Cofactor selection was performed using Proc GLMSELECT implemented in the statistical software SAS (SAS Institute, 2008).

For the detection of main-effect QTL, a genome-wide scan for marker–trait associations was conducted. We tested for significance with $P < 0.05$ and controlled for multiple testing by applying the Bonferroni–Holm procedure (Holm, 1979). The total proportion of genotypic variance (p_G) explained by the detected QTL was calculated by fitting all QTL simultaneously in a linear model to obtain R^2_{adj} . The ratio $p_G = R^2_{\text{adj}}/h^2$ yielded the proportion of genotypic variance (Utz *et al.*, 2000).

Cross-validation, resampling and RMA

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

Our resample model averaging (RMA) approach to reveal the QTL frequency distributions (Utz *et al.*, 2000) was similar to the subbagging (80%) described by Valdar *et al.* (2009). We used resampling without replacement as described for cross-validation. In contrast to the study by Valdar *et al.* (2009), we did not use forward selection to select the multiple QTL model, but used QTL detection by Model B.

To analyze the effect of the marker density on family mapping, four different sets of markers were generated: 100% (all 425 SNP markers), 80% (340 SNPs), 60% (255 SNPs), 40% (170 SNPs) and 20% (85 SNPs). These markers were either sampled randomly (unbalanced) or to maintain the relative contribution of each of the chromosomes (balanced).

For the optimum allocation of resources in family mapping, subsamples from the full population ($N=930$, $E=6$) were drawn without replacement representing (a) genotypic subsamples and (b) environmental subsamples. Genotypic subsamples were always sampled with an equal percent contribution of all families to ensure a similar population structure in the subsample as in the reference population. All 12 possible combinations derived from subsamples with $N=930$, 660, 440 and 220 individuals and $E=6$, 4 and 2 environments were analyzed. The subsampling of genotypes and environments was repeated 600 times to result in 600 different data sets per $N-E$ combination, except for the full data set (930, 6), where only one data set exists. For each data set, the heritability, the localization of QTL and the proportion of genotypic variance explained by these QTL were estimated, and cross-validation was carried out as described above.

The balanced data set ($n=440$) was obtained by genotypic subsampling. The balanced_{Family} data set consisted of 11 families, each of which contained 40 individuals, resulting in a total population size of 440. The balanced_{Parents} data set balanced the contribution of the parents to the population as much as possible and also had a total population size of 440. In all, 600 runs of random sampling of individuals from the families according to these criteria were carried out to result in 600 balanced data sets each.

RESULTS

We performed a simulation study to evaluate the optimum allocation of resources in a theoretical framework, varying the number of parents from 3 to 15, and the number of test environments from 2 to 14. In addition, we compared two recently described biometrical models for family mapping, Model A and Model B (Würschum *et al.*, 2012). The models differ in that both incorporate cofactors, but in addition, Model B includes an effect for the segregating family. For the DIA design, the optimum with regard to the QTL detection power was very similar for both biometrical models, with 7 or 8 parents and two test environments (Figure 1). For SRR, this optimum shifted towards a higher number of parents with 10 for Model A and 13 for Model B. The optimum number of test environments was always two, except for SRR Model B, where the optimum was found for four test environments. For both designs, a higher power was observed for Model B. The QTL detection power at the optima was 0.72 (Model A, DIA), 0.85 (Model B, DIA), 0.74 (Model A, SRR) and 0.82 (Model B, SRR).

A marked difference between the two biometrical models was observed for the false-positive rate. Whereas Model B showed a very low false-positive rate for all tested combinations of parental number and test environments, Model A exhibited a high false-positive rate of up to 0.8 in the combination optimizing the QTL detection power. In addition, the simulation study revealed a dependency of the false-positive rate on the extent of LD. Low LD in the base population, that is, LD that decays after a short genetic map distance, increased the false-positive rate as compared with long-ranging LD.

We also tested the two biometrical models on experimental data by using the full data set with 930 individuals evaluated in six test environments and genotyped with 425 SNP markers. The number of detected QTL and the genotypic variance explained by these QTL was comparable between the two models for grain yield, whereas for grain

moisture, Model A detected more QTL, which explained a much higher proportion of the genotypic variance (p_G) (Table 1). It must be noted here that the two traits differ substantially with regard to the ratio of within-family and among-family variance (Zhao *et al.*, 2012). Whereas the among-family variance for grain yield is negligible, it is twice as high compared with the within-family variance for grain moisture. When the detected QTL were analyzed separately within the four largest families ($A \times D$, $A \times E$, $A \times F$, $E \times B$), we observed for both traits that the two biometrical models were comparable or that Model B was even slightly better, indicating that much of the p_G of Model A was due to among-family variance. We used a fivefold cross-validation approach in which the effects of the QTL detected with the full data set were estimated in the ES (80% of genotypes from each family) to subsequently carry out a prediction in the TS (remaining 20% of genotypes). We observed that the relative bias (the relative reduction in p_G from the ES compared with the TS) was slightly higher for Model A compared with Model B. Regarding the two traits, the relative bias was higher for grain yield as compared with grain moisture, which may be due to the lower heritability of the former (Supplementary Table S1). This is in accordance with results from our simulation study showing that Model A has a much higher false-positive rate than Model B (Figure 1). Taken together, Model A appears to possess an enhanced risk of detecting QTL due to among-family variance and potentially a higher false-positive rate. For further analyses using experimental data, we therefore focused on Model B.

The analysis of the experimental data from 930 genotypes evaluated at six locations yielded estimates of heritability of 0.51 for grain yield and 0.72 for grain moisture (Supplementary Table S1). We tested different combinations of genotypes and/or test environments, and, as might be expected according to quantitative genetics theory, observed that the heritability was not affected by the size of the population, but was strongly affected by the number of test environments.

To assess the precision and the reliability of the detected QTL, we used a recently described RMA approach (Valdar *et al.*, 2009). Most of the QTL detected with the full data set were identified in more than 40% of the RMA runs (Figure 2). The RMA, however, also revealed that some QTL detected with the full data set were only selected as QTL in <20% of the runs (for example, grain moisture QTL on chromosome 4). By contrast, the RMA also identified some QTL positions that were not detected with the full data set, but were identified in a high number of RMA runs (for example, grain yield QTL on chromosome 7). We observed strong effects on the precision of QTL detection for the 12 combinations of sample size and phenotyping intensity. A reduction in either of the two parameters led to a severely lower frequency of RMA runs in which the QTL was detected. This effect was more pronounced for a reduction in sample size than for a reduced number of test environments. For the combination of 220 genotypes evaluated in two environments, the frequency of RMA runs in which a QTL was identified was generally below 10%, even for those QTL that were identified in the majority of runs in the full data set.

The resampling approach revealed that for grain yield and for grain moisture, the number of QTL detected in the ES decreased with both a reduced number of genotypes and a reduction in test environments (Supplementary Table S2). In addition, the proportion of genotypic variance explained by the detected QTL in the ES and in the TS decreased with a reduced number of genotypes (Figure 3 and Supplementary Figure S3). The reduction in the number of test environments led to a reduction in p_G for grain yield but not for grain moisture, whereas the bias in the estimation of p_G was comparable for all 12 combinations. The variation in p_G estimates in the ES and in

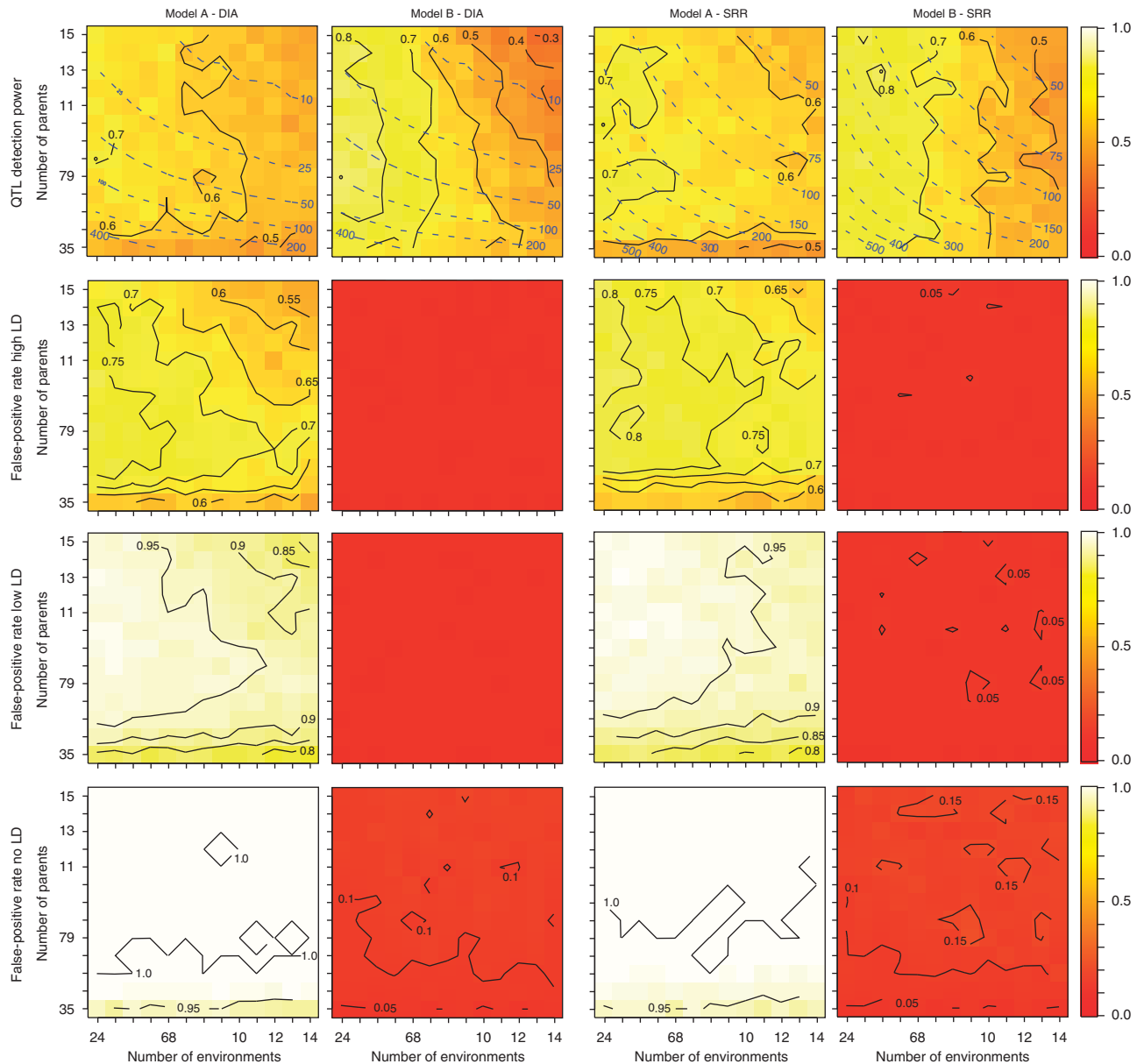


Figure 1 QTL detection power and false-positive rate of two biometric models for family mapping, Model A and Model B, assessed in a simulation study. The simulation is based on a fixed budget, with varying numbers of parents and test environments. Results are shown for two mating designs, DIA and SRR, and the false-positive rate for different levels (high, low, no) of LD. The dashed blue lines indicate the family size.

the TS increased when the number of genotypes or the number of test environments were reduced (Supplementary Figure S4). This effect was more pronounced for a reduction in the number of test environments than for the reduced sample sizes. A similar result was observed for the bias in p_G estimates.

We next compared populations with 440 genotypes that were either unbalanced, that is, each family size is decreased proportionally relative to the full data set, or balanced. The balanced data sets either had similar family sizes ($\text{balanced}_{\text{Family}}$) or the population was more balanced with regard to the contribution of the 11 parents ($\text{balanced}_{\text{Parents}}$). This comparison revealed that for both traits, the unbalanced data set performed slightly better than the balanced data sets with regard to the number of detected QTL (Supplementary Table S2) as well as the p_G (Figure 4).

We analyzed the full data set with reduced numbers of markers that were either sampled randomly throughout the entire genome (unbalanced) or sampled to maintain the representation of the chromosomes as with the full marker complement (balanced). We observed a reduction in the number of detected QTL with the reduction in marker density (Table 2). With only 20% of the markers, we still detected approximately half of the QTL identified with the full marker set. No difference was observed between balanced and unbalanced sampling of markers. Regarding the proportion of genotypic variance explained by the detected QTL under different marker densities, we observed an almost linear reduction of p_G with reduced marker density (Figure 5). As for the number of detected QTL, however, the p_G obtained with the lowest marker density (20%) was still considerably high and in the TS amounted 84% and 69% of

Table 1 Fivefold cross-validation in family mapping with the full data set (930 individuals)

Parameter	Grain yield		Grain moisture	
	Model A	Model B	Model A	Model B
QTL _{DS}	6	6	17	10
ρ_{G-DS}	40.8	44.1	71.9	27.4
$\rho_{G-A \times D}$	35.0	45.3	63.0	61.6
$\rho_{G-A \times E}$	31.1	30.8	33.1	33.7
$\rho_{G-A \times F}$	35.3	35.0	23.9	38.7
$\rho_{G-E \times B}$	8.8	8.8	14.0	22.0
ρ_{G-ES}	48.4	34.7	68.7	27.1
ρ_{G-TS}	35.9	26.2	57.1	23.2
Bias	12.5	8.5	11.7	4.0
Relative bias	25.8	24.5	16.9	14.4

Number of detected QTL (QTL_{DS}) from two biometric models, proportion of genotypic variance (in percent) explained by the detected QTL across all families in the data set (ρ_{G-DS}) and within the four largest single families ($\rho_{G-A \times D}$, $\rho_{G-A \times E}$, $\rho_{G-A \times F}$, $\rho_{G-E \times B}$). Proportion of genotypic variance averaged over estimation sets (ρ_{G-ES}) and averaged over test sets (ρ_{G-TS}). Bias and relative bias (%) in the estimation of ρ_G .

that of the full marker density for grain yield and grain moisture, respectively. We observed no strong differences in the variance of ρ_G for the different marker densities and no difference between the balanced versus the unbalanced marker sets.

DISCUSSION

Simulation study

Simulation studies have recently been used to address questions with regard to the optimum design of family mapping studies (Verhoeven *et al.*, 2006; Stich *et al.*, 2007; Stich 2009). In our study, the DIA and SRR designs were comparable with regard to the QTL detection power. An optimum family mapping design should maximize the number of parents included in the study to sample a high allelic diversity and to enable a high mapping resolution due to a lower LD generated by the sampling of the parents. These considerations support the finding of Verhoeven *et al.* (2006) and suggest SRR as a promising design for family mapping.

The quality of family mapping experiments is not solely determined by the QTL detection power, but also by the false-positive rate. Yu *et al.* (2008) observed that for two traits of different complexity, the QTL detection power and the false-positive rate did not reach a plateau even for a total population size of 5000 individuals in the maize nested association mapping design. As expected, both parameters also greatly depended on the heritability, and thus on the complexity of the trait as well as on the phenotyping intensity. Stich (2009) did not control for population structure in his analyses, but this correction has been shown to be essential to control false-positive QTL, even in family mapping (Verhoeven *et al.*, 2006; Würschum *et al.*, 2012). Our simulation study confirmed this and clearly showed that the model without correction for population structure (Model A) exhibited a much higher false-positive rate (Figure 1). The interpretation of the observed false-positive rate (approximately 0.8) is that in ~80% of the runs, a QTL is detected on a chromosome where no QTL is located. In addition, we observed a strong effect of the LD in the base population from which the parents are sampled on the false-positive rate. The false-positive rate increased with decreasing LD, that is, LD that decays within a shorter distance. A possible explanation for this observation is that with low LD, there are more independent tests for associations with the trait than with high LD. In

the most extreme case of no LD, each marker represents an independent test, and consequently the probability of one of them being falsely detected as QTL increases. We thus note that low LD is desirable as it enables a high mapping resolution, but it can also increase the rate of false-positive QTL.

Stich *et al.* (2010) also investigated the optimum allocation of resources with a fixed budget and found that the QTL detection power reached a plateau with 4–7 test environments. By contrast, our results mostly show a plateau for the optima for QTL detection power in 2–4 test environments (Figure 1), which corresponds to the phenotyping intensity commonly applied in breeding programs for first GCA tests (Longin *et al.*, 2007). This illustrates that with the assumptions underlying our simulation study, a sufficient heritability can be achieved with few test environments and it is more advantageous to direct resources towards increasing population size, rather than towards additional test environments.

For Model B, the optimum number of parents was found to be 7 and 13 for the DIA and the SRR design, respectively. Interestingly, both optima were found for a combination with family sizes of around 100 individuals, even though the model is not nested and family size should consequently be of no relevance. The reason may be that with a certain number of parents, an optimum is reached with regard to the probability of a QTL segregating in families, and thus of informative individuals in the population. The rather low number of parental lines observed here as optimum is likely due to the fact that we simulated only two QTL alleles and a high probability of the QTL segregating in some families is reached already with fewer parents. It must be noted that this is likely dependent on the QTL allele frequency, as for rare alleles this optimum shifts towards a higher number of parents (unpublished results).

The examples mentioned above clearly show that simulation studies are a valuable tool, but have certain restrictions in that the assumed scenarios underlying these studies are difficult to choose and do not always reflect reality. This is especially true when considering that experimental data are often unbalanced and do not strictly adhere to the theoretically defined mating designs evaluated in simulation studies (Würschum, 2012). Another simplified assumption of simulation studies is that QTL effects are similar in different families. Thus, simulation approaches must be complemented by the analysis of experimental data.

Predictive power and accuracy of QTL

We used experimental data from 930 genotypes evaluated in six test environments and applied a fivefold cross-validation to assess the predictive power of the detected QTL. Applying Model B, we observed that the relative bias was 21.8% for grain yield and 11.8% for grain moisture (Table 1). Thus, family mapping appears to possess a good predictive power and is not hampered by a large bias in ρ_G estimates. The unbiased estimate of the proportion of genotypic variance explained by the detected QTL (ρ_{G-TS}) was approximately 25% for both traits, which is promising for such complex traits and comparable to other studies in maize (Schön *et al.*, 2004). It must be noted, however, that the QTL due to among-family variance detected for grain moisture with Model A were verified by the cross-validation approach. Thus, these QTL, which are prone to an enhanced false-positive rate, can only be avoided by the choice of an appropriate biometrical model, for example, Model B.

As a measure of the precision of the QTL position estimates, we performed an RMA approach (Valdar *et al.*, 2009). Most of the QTL detected with the full data set were supported by their identification as QTL in a high number of the RMA runs (Figure 2). Some QTL

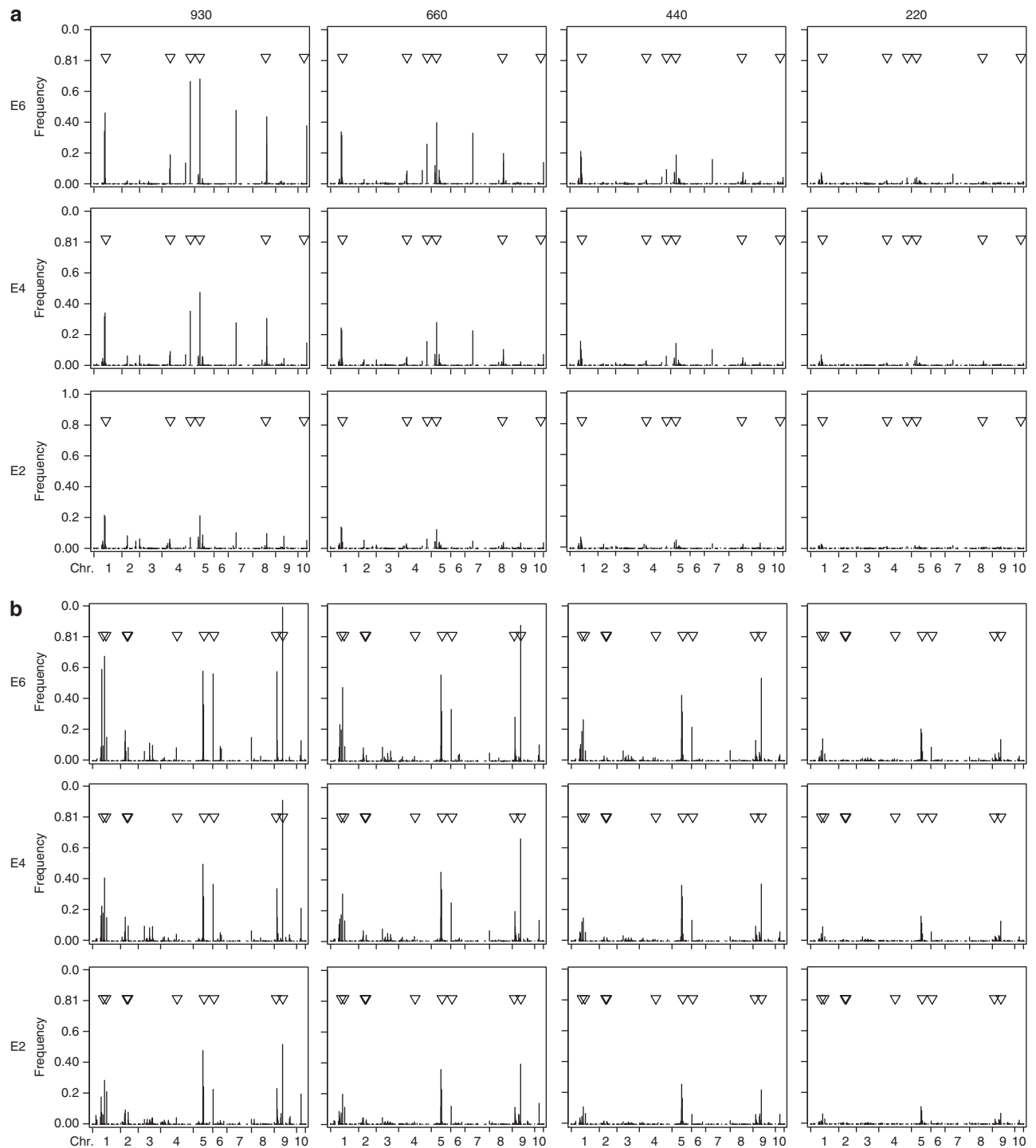


Figure 2 Frequency distributions of QTL detected with Model B in experimental data in 600 RMA runs and 12 combinations of sample size (930, 660, 440, 220) and test environments (E6, E4, E2) for (a) grain yield and (b) grain moisture. The arrowheads indicate the positions of the QTL detected with the full data set. Markers are shown at their position on the chromosomes.

were, however, only identified in few RMA runs and are therefore less reliable, whereas other QTL not detected with the full data set could be identified by this approach. The QTL positions were well defined, confirming the high mapping resolution of association mapping in multiple families.

For the 12 combinations of sample size and test environments, the RMA approach revealed a strong effect of both parameters. The reduction in sample size resulted in particularly low accuracy of QTL

detection (Figure 2). For 220 genotypes, no QTL was consistently detected for grain yield. The same trend was observed for grain moisture even though less pronounced, which may be attributed to the higher heritability of this trait or the underlying genetic architecture. Thus, despite the fact that QTL were detected in these data sets, these results appear unreliable. In conclusion, for family mapping to produce reliable results, a high heritability must be achieved by using a substantial number of test environments and the

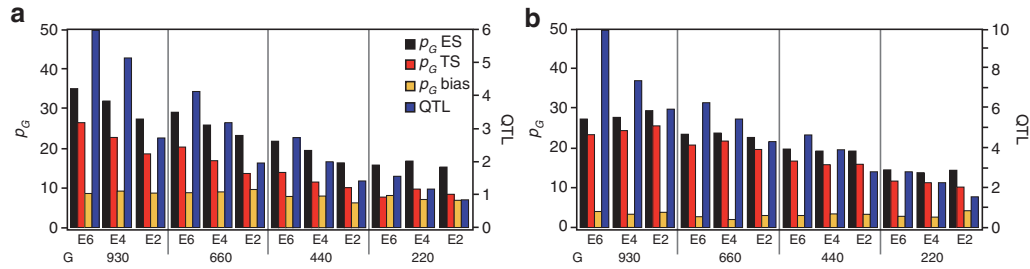


Figure 3 Proportion of genotypic variance (ρ_G) explained by the QTL detected with Model B in the ES (ρ_G ES), in the TS (ρ_G TS) and bias (ρ_G bias), and the number of detected QTL shown for 12 combinations of sample size (930, 660, 440, 220) and test environments (E6, E4, E2) for (a) grain yield and (b) grain moisture.

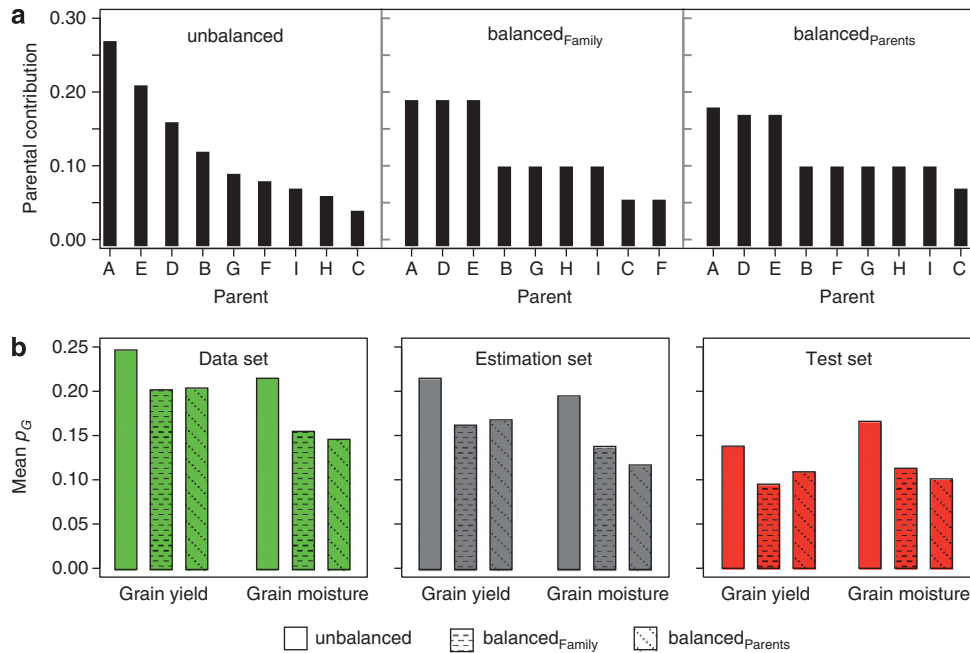


Figure 4 Effect of unbalanced versus balanced population sizes. (a) Contribution of the parents (based on pedigree) to the three different population sets each with 440 individuals. Unbalanced has relative family sizes composed as in the full data set, balanced_{Family} has equal family sizes of 40 and balanced_{Parents} balances the parental contribution as well as possible. (b) Mean proportion of genotypic variance (ρ_G) explained by the detected QTL in the data set, the ES and in the TS for grain yield and grain moisture.

Table 2 Average number of QTL (90% quantiles within parenthesis) detected in the full data set (930 genotypes) with different marker densities for grain yield and grain moisture

	Marker density								
	20%	20%*	40%	40%*	60%	60%*	80%	80%*	100%
<i>Grain yield</i>									
QTL	4.0	4.0	4.6	4.5	5.2	5.2	5.7	5.8	6
	(2; 6)	(3; 6)	(3; 6)	(3; 6)	(3; 7)	(3; 7)	(4; 7)	(4; 7)	(—)
<i>Grain moisture</i>									
QTL	5.3	5.4	6.5	6.5	6.9	7.0	7.4	7.4	10
	(3; 8)	(3; 8)	(4; 9)	(4; 9)	(5; 10)	(5; 10)	(5; 11)	(5; 11)	(—)

The asterisk indicates the balanced sampling of markers.

population must have a certain minimum size. The RMA approach can serve as a good indicator for the quality of the obtained mapping results.

Effect of sample size and phenotyping intensity

In contrast to simulation studies, research projects and breeding programs are faced with a fixed total budget and the goal is to invest that budget optimally. Whereas the complexity of the trait is given, the following parameters can be varied: (1) population size; (2) phenotyping intensity; and (3) genotyping intensity. To assess the former two parameters, we subdivided the full data set into 12 numerical combinations of genotypes and test environments. We observed that the heritability was strongly affected by the number of test environments, but not by the population size (Supplementary Table S1). According to quantitative genetics theory and previous

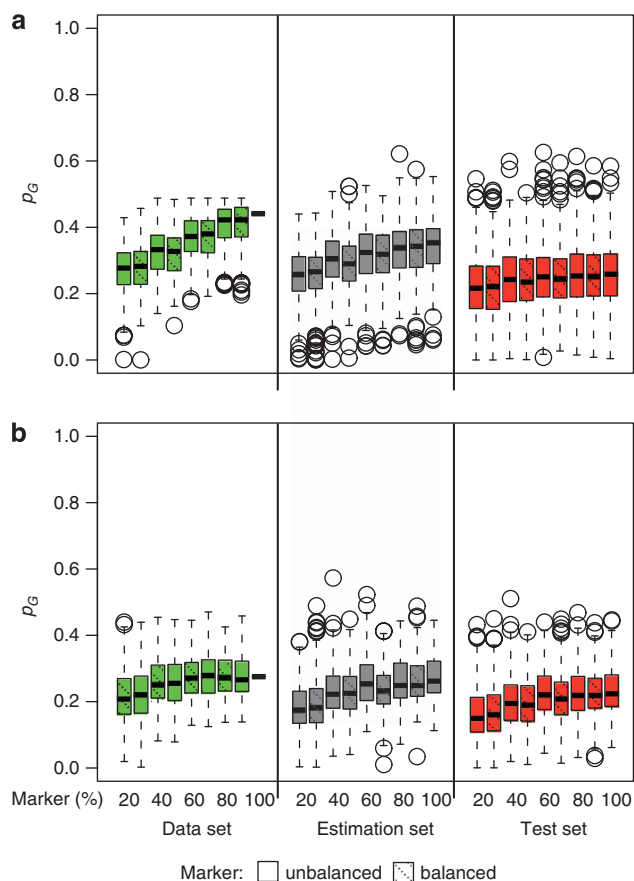


Figure 5 Effect of different marker densities in family mapping. Boxplots show the variation in the proportion of genotypic variance (p_G) explained by the detected QTL in the full data set, the ES and in the test set for five different marker densities (100%, 80%, 60%, 40% and 20%) for (a) grain yield and (b) grain moisture. Markers were either sampled randomly (unbalanced) or to maintain the proportion of markers derived from each chromosome as with the full marker set.

studies (Beavis, 1994; Utz and Melchinger, 1994; Falconer and Mackay, 1996), a higher heritability will warrant a higher QTL detection power. In accordance with these expectations, we observed that the number of detected QTL decreased strongly with decreasing heritability (Supplementary Table S2). In addition, we also detected fewer QTL with decreasing population sizes even though heritability was similar. This indicates that besides heritability, population size is a critical parameter affecting QTL detection power and that small mapping populations can maintain a high heritability, but nevertheless possess a reduced QTL detection power. This is in line with theoretical considerations suggesting that in linkage mapping, QTL detection power can be improved by allocating resources to more genotypes as compared with fewer genotypes with more replications (Knapp and Bridges, 1990).

The estimation of p_G in the ES and in the TS in 600 cross-validation runs for the 12 combinations of sample size and number of test environments revealed a decrease in the predictive power for smaller population sizes (Figure 3). For grain yield, p_G also decreased with less test environments, whereas this had no effect on grain moisture. This may be due to a smaller genotype-by-environment interaction component for grain moisture or to the different genetic architecture of the two traits. For grain moisture, QTL with strong effect may be detected irrespective of the number of test environments

(Figure 2), maintaining a high p_G . Notably, when considering current costs for phenotyping and genotyping (Supplementary Figure S3), a reduction in the number of test environments affected p_G for grain yield to a similar extent as the reduction in population size (Figure 3). For both traits, the variation in p_G estimates in the ES and in the TS, as well as the bias, were more strongly affected by the reduction in the number of test environments than by the sample size (Supplementary Figure S4). The much larger variation in p_G estimates for fewer environments is also due to the variation in heritability estimates. As the bias can be compared for all combinations, the relative bias was stronger for the combinations with less genotypes or less environments. Taken together, our results show that the considerations proposed from the analysis of biparental populations (Schön *et al.*, 2004) are also valid for populations based on multiple families. Population size and robust phenotyping are key factors for successful family mapping studies.

Composition of the family mapping population

We also investigated the effect of balanced and unbalanced populations based on a family mapping population with 440 genotypes. These analyses showed that, with regard to the cross-validated proportion of genotypic variance, the unbalanced data set performed slightly better than either the balanced_{Family} or the balanced_{Parents} data sets. This was surprising, as intuitively we expected the balanced data sets to perform better. This did not appear to be caused by different heritability levels, as these were comparable for the three data sets despite differences in the heritability of the individual families (Liu *et al.*, 2011). We employed the RMA approach to assess which QTL were frequently detected in the unbalanced and in the balanced_{Family} data sets (Supplementary Figure S5). This revealed differences in the frequency distributions for some of the major QTL compared with the full data set. The unbalanced set was sampled to maintain the relative contribution of each family as in the full data set. Consequently, large families contribute more to the unbalanced set than the balanced set, and thus provide a higher QTL detection power for QTL segregating in these large families. Consistent with this, we observed that QTL more frequently detected in the unbalanced data set did indeed segregate in the large families (Supplementary Table S3). By contrast, QTL that were more frequently detected in the balanced data set did not segregate in the large families, but mainly in the smaller families. These will be under-represented in the unbalanced data set, whereas their contribution to the population is increased in the balanced data set. This effect is likely also responsible for the slight differences between the balanced_{Family} and the balanced_{Parents} data sets observed for both traits. Taken together, the QTL detection power in family mapping can be improved by increasing the allele frequencies, which is achieved by maximizing the number of informative individuals, that is, individuals from segregating families. It may thus be advantageous to compile an unbalanced family mapping population if this increases the proportion of families segregating for candidate QTL. This, however, requires prior knowledge that can be taken into consideration when QTL mapping experiments are planned. Without such prior knowledge, the best strategy is to balance the family mapping population with regard to the parental allele frequencies.

Influence of marker density

Association mapping in multiple segregating families is based on LD, and a high marker density is therefore expected to be crucial for the QTL detection power (Liu *et al.*, 2012). We varied the marker density from 100% (425 SNPs) to 20% (85 SNPs), which corresponds to an

average marker density of 2.8 and 14.2 cM, respectively. For both traits, the number of detected QTL decreased with reduced marker density, but not as strongly as might have been expected (Table 2). This may be due to the LD present in this population, which will allow detection of QTL with strong effect even for lower marker densities. Regarding p_G , we observed a constant increase with marker density and no plateau was reached even for the highest marker density available in this study (Figure 5). No effect was observed between the balanced and the unbalanced sampling of markers. Marker density is certainly a critical issue for family mapping and higher marker densities or even sequence data will increase the QTL detection power. Depending on the LD present in the data set (Hamblin *et al.*, 2011), however, a few hundred markers will already be sufficient for the detection of QTL with large or medium effects, and thus for a satisfactory result in family mapping. The extent of LD in the mapping population also determines whether linked QTL can be separated or not. This in turn can affect the bias in QTL estimation as linkage in coupling can result in overestimation and linkage in repulsion in underestimation of QTL effects.

CONCLUSIONS

We have used a simulation study and the analysis of experimental data to assess the optimum allocation of resources in family mapping studies. Our results highlight the importance of a high heritability achieved by robust phenotyping of the individuals. The family mapping population should balance the contribution of parental alleles, and results can likely be improved by higher marker densities that will be available in the near future. A promising strategy appears to be to use cross-validation and RMA to assess the quality of the obtained mapping results.

DATA ARCHIVING

Data have been deposited at Dryad: doi:10.5061/dryad.m6079.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This research was supported by two projects from the German Federal Ministry for Education and Research (BMBF): the Biometric and Bioinformatic Tools for Genomics based Plant Breeding project (GABI-FUTURE initiative) and CornFed (FK 0315461A).

- Holland JB (2007). Genetic architecture of complex traits in plants. *Curr Opin Plant Biol* **10**: 156–161.
- Holm S (1979). A simple sequentially rejective bonferroni test procedure. *Scand J Stat* **6**: 65–70.
- Knapp SJ, Bridges WC (1990). Using molecular markers to estimate quantitative trait locus parameters: power and genetic variances for unreplicated and replicated progeny. *Genetics* **126**: 769–777.
- Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD *et al.* (2009). A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet* **5**: e1000551.
- Lande R, Thompson R (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* **124**: 743–756.
- Liu W, Gowda M, Steinhoff J, Maurer HP, Würschum T, Longin CF *et al.* (2011). Association mapping in an elite maize breeding population. *Theor Appl Genet* **123**: 847–858.
- Liu W, Reif JC, Ranc N, Della Porta G, Würschum T (2012). Comparison of biometrical approaches for QTL detection in multiple segregating families. *Theor Appl Genet* **125**: 987–998.
- Longin CFH, Utz HF, Reif JC, Wegenast T, Schipprack W, Melchinger AE (2007). Hybrid maize breeding with doubled haploids: III. Efficiency of early testing prior to doubled haploid production in two-stage selection for testcross performance. *Theor Appl Genet* **115**: 519–527.
- Maurer HP, Melchinger AE, Frisch M (2008). Population genetic simulation and data analysis with Plabsoft. *Euphytica* **161**: 133–139.
- McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li HH, Sun Q (2009). Genetic properties of the maize nested association mapping population. *Science* **325**: 737–740.
- 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.
- Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE *et al.* (2009). Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell* **21**: 2194–2202.
- Reif JC, Liu W, Gowda M, Maurer HP, Möhring J, Fischer S *et al.* (2010). Genetic basis of agronomically important traits in sugar beet (*Beta vulgaris* L.) investigated with joint linkage association mapping. *Theor Appl Genet* **121**: 1489–1499.
- SAS Institute. Inc (2008). *SAS User's Guide, Version 9.2*. SAS Institute: Cary, NC.
- Schön CC, Utz HF, Groh S, Truberg B, Openshaw S, Melchinger AE (2004). Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics* **167**: 485–498.
- Schwarz G (1978). Estimating the dimension of a model. *Ann Stat* **6**: 461–464.
- Searle SR (1971). *Linear Models*. Wiley: New York.
- Stich B, Yu J, Melchinger AE, Piepho HP, Utz HF, Maurer HP *et al.* (2007). Power to detect higher-order epistatic interactions in a metabolic pathway using a new mapping strategy. *Genetics* **176**: 563–570.
- Stich B (2009). Comparison of mating designs for establishing nested association mapping populations in maize and *Arabidopsis thaliana*. *Genetics* **183**: 1525–1534.
- Stich B, Utz HF, Piepho HP, Maurer HP, Melchinger AE (2010). Optimum allocation of resources for QTL detection using a nested association mapping strategy in maize. *Theor Appl Genet* **120**: 553–561.
- Utz HF, Melchinger AE (1994). Comparison of different approaches to interval mapping of quantitative trait loci. pp 195–204, in *Biometrics in plant breeding: applications of molecular markers. Proceedings of the Ninth Meeting of the EUCARPIA Section Biometrics in Plant Breeding*, edited by Van Ooijen JW, Jansen J (CPRO-DLO: Wageningen, Netherlands).
- Utz HF, Melchinger AE, Schön CC (2000). Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. *Genetics* **154**: 1839–1849.
- Valdar W, Holmes CC, Mott R, Flint J (2009). Mapping in structured populations by resample model averaging. *Genetics* **182**: 1263–1277.
- Verhoeven KJF, Jannink JL, McIntyre LM (2006). Using mating designs to uncover QTL and the genetic architecture of complex traits. *Heredity* **96**: 139–149.
- Würschum T (2012). Mapping QTL for agronomic traits in breeding populations. *Theor Appl Genet* **125**: 201–210.
- Würschum T, Liu W, Gowda M, Maurer HP, Fischer S, Schechert A *et al.* (2012). Comparison of biometrical models for joint linkage association mapping. *Heredity* **108**: 332–340.
- Wu R, Ma CX, Casella G (2002). Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations. *Genetics* **160**: 779–792.
- Yu J, Holland JB, McMullen MD, Buckler ES (2008). Genetic design and statistical power of nested association mapping in maize. *Genetics* **178**: 539–551.
- Zhao Y, Gowda M, Liu W, Würschum T, Maurer HP, Longin FH *et al.* (2012). Accuracy of genomic selection in European maize elite breeding populations. *Theor Appl Genet* **124**: 769–776.

- Beavis WD (1994). The power and deceit of QTL experiments: lessons from comparative QTL studies. *49th Annual Corn and Sorghum Industry Research Conference*. ASTA: Washington, DC, pp 250–266.
- Beavis WD (1998). QTL analyses: power, precision and accuracy. *Molecular Dissection of Complex Traits*, edited by Paterson. AHCRC Press Boca Raton FL pp 145–162.
- Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C *et al.* (2009). The genetic architecture of maize flowering time. *Science* **325**: 714–718.
- Falconer DS, Mackay TFC (1996). *Introduction to Quantitative Genetics*, 4th edn. Longman House: London.
- Hallauer AR, Miranda JB (1981). *Quantitative Genetics in Maize Breeding*. Iowa State University Press: Ames, IA.
- Hamblin MT, Buckler ES, Jannink JL (2011). Population genetics of genomics-based crop improvement methods. *Trends Genet* **27**: 98–106.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)