Theor Appl Genet (2009) 118:741–751
DOI 10.1007/s00122-008-0934-9

ORIGINAL PAPER

Molecular marker-based prediction of hybrid performance in maize using unbalanced data from multiple experiments with factorial crosses

Tobias A. Schrag · Jens Möhring · Hans Peter Maurer · Baldev S. Dhillon · Albrecht E. Melchinger · Hans-Peter Piepho · Anker P. Sørensen · Matthias Frisch

Received: 14 June 2008 / Accepted: 9 November 2008 / Published online: 2 December 2008

Abstract In hybrid breeding, the prediction of hybrid performance (HP) is extremely important as it is difficult to evaluate inbred lines in numerous cross combinations. Recent developments such as doubled haploid production and molecular marker technologies have enhanced the prospects of marker-based HP prediction to accelerate the breeding process. Our objectives were to (1) predict HP using a combined analysis of hybrids and parental lines from a breeding program, (2) evaluate the use of molecular markers in addition to phenotypic and pedigree data, (3) evaluate the combination of line per se data with marker-based estimates, (4) study the effect of the number of tested parents, and (5) assess the advantage of haplotype blocks. An unbalanced dataset of 400 hybrids from 9 factorial crosses tested in different experiments and data of 79 inbred parents were subjected to combined analyses with a mixed linear model. Marker data of the inbreds were obtained with 20 AFLP primer–enzyme combinations. Cross-validation was used to assess the performance prediction of hybrids of which no or only one parental line was testcross evaluated. For HP prediction, the highest proportion of explained variance (R^2), 46% for grain yield (GY) and 70% for grain dry matter content (GDMC), was obtained from line per se best linear unbiased prediction (BLUP) estimates plus marker effects associated with mid-parent heterosis (TEAM-LM). Our study demonstrated that HP was efficiently predicted using molecular markers even for GY when testcross data of both parents are not available. This can help in improving greatly the efficiency of commercial hybrid breeding programs.

Communicated by H. C. Becker.

T. A. Schrag · H. P. Maurer · B. S. Dhillon · A. E. Melchinger (✉)
Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany
e-mail: melchinger@uni-hohenheim.de

J. Möhring · H.-P. Piepho
Bioinformatics Unit of the Institute for Crop Production and Grassland Research, University of Hohenheim, 70599 Stuttgart, Germany

A. P. Sørensen
Keygene, P.O. Box 216, 6700 AE Wageningen, The Netherlands

M. Frisch
Institute of Agronomy and Plant Breeding II, Justus-Liebig-University, 35392 Giessen, Germany

Introduction

Hybrid maize breeders develop a large number of inbred lines and estimate their value in cross combinations (Hallauer 1990). The number of crosses increases rapidly with the number of inbreds and their field evaluation requires large resources. The generation of inbred lines has been facilitated and accelerated in recent years by the doubled haploid technology, which is being used increasingly in commercial hybrid breeding programs (Schmidt 2004; Seitz 2005). This has a strong influence on the allocation of resources in hybrid maize breeding (Longin et al. 2007).

Depending upon the resources, only a small proportion of all possible experimental hybrids can be evaluated in field trials. In this situation, prediction of hybrid performance (HP) utilizing field trial data available from related crosses has been attempted to identify promising inter-group hybrids (i.e., crosses between lines from different heterotic groups) without having them tested in the field. Bernardo (1994) presented a best linear unbiased

prediction (BLUP) approach based on phenotypic trait data and coancestry coefficients estimated from pedigree records or marker data. The approach was extended by including marker data directly into the model to account for quantitative trait loci (QTL), which however resulted only in marginal improvements (Bernardo 1998, 1999). Charcosset et al. (1998) evaluated the prediction of HP, comparing different marker-based approaches to account for specific combining ability (SCA). Their results for inter-group crosses indicated higher prediction efficiencies with BLUP and factorial regression models compared with a genetic distance model.

Vuylsteke et al. (2000) presented a linear regression approach to predict HP and SCA for grain yield (GY) using marker-based estimates for the genotypic value of inter-group crosses. This approach was extended and validated for the prediction of GY and grain dry matter content (GDMC) in factorial crosses (Schrag et al. 2006) and was further improved to account for multiple testing and correlation among marker loci by the identification of haplotype blocks, in which stretches of closely linked markers were regarded as one unit (Schrag et al. 2007). In both studies, four experiments on factorial crosses were analyzed separately. Therefore, the performance of crosses between inbreds evaluated in different factorials remained unpredicted. This situation is commonly encountered in plant breeding programs and consequently is of great interest to plant breeders.

Conventional procedures of line development by recurrent selfing enable field testing of inbred lines and rejection of poor ones during inbreeding. In contrast, the doubled haploid technology generates homozygous lines representing a random sample of all possible inbred lines from the parental cross. Assessing their breeding value only by conventional field tests is prohibitive considering the resources required. Genotyping these lines with molecular markers before producing their seed, conducting testcross (TC) evaluation in field experiments and predicting HP by using estimates of marker effects obtained from crosses between already evaluated lines in the same breeding program would greatly enhance the efficiency of hybrid breeding programs with doubled haploids. However, to our knowledge prediction of the performance of hybrids between inbreds, for which no TC evaluation data are available, has not been addressed so far.

Mid-parent heterosis (MH), defined as the difference between the performance of a hybrid and the mean performance of its parents, accounts for up to 76% of the GY of maize hybrids (Hallauer and Miranda Filho 1988). It is therefore of great importance in hybrid maize breeding. While the performance of parental lines per se has not been of much value in predicting HP due to masking dominance effects (Smith 1986; Hallauer 1990), prediction of MH based on marker effects combined with inbred line per se performance seems to be a promising approach, which has not been evaluated earlier.

Our objectives were to (1) predict the GY and GDMC performance of untested maize hybrids using a combined analysis of hybrids and their parental inbred lines across several field experiments from an applied breeding program, (2) compare predicted performance based on molecular markers with that based exclusively on phenotypic and pedigree data, (3) test the improvement in the prediction of HP by combining observed line per se data with marker-based MH estimates, (4) compare the performance prediction of hybrids of which no versus one parental line was evaluated for testcross performance, and (5) study the benefit of combining correlated markers into haplotype blocks for prediction.

Materials and methods

Molecular data

Seventy-nine parental inbred lines (47 from the dent pool and 32 from the flint pool) were genotyped with 20 amplified fragment length polymorphism (AFLP) primer–enzyme combinations (Vos et al. 1995), as described in detail by Schrag et al. (2006). Positions of 910 mapped AFLP bands were obtained from an integrated AFLP map (Vuylsteke et al. 1999). Each marker was analyzed for polymorphism and the proportion of missing marker observations among the evaluated inbreds.

Phenotypic and pedigree data

Nine dent × flint factorial mating experiments were conducted within the maize breeding program of the University of Hohenheim. Each experiment was carried out within 1 year at 5–7 locations in Germany under diverse agroecological conditions. In total, there were 54 experiment × location combinations (trials) across 6 years and 11 locations. Across all experiments, 400 inter-group crosses between 47 dent and 32 flint inbred parents were evaluated along with 20 commercial hybrids, which were used as checks. By combining these data, an unbalanced 47 × 32 factorial dataset comprising 400 tested and 1,104 untested hybrids was generated (Fig. 1). The 79 inbred parents were evaluated for their per se performance in 11 trials across 3 years and 3–5 locations. Eight year × location combinations were in common for the hybrid trials and inbred trials. All trials were conducted using adjacent α-designs with two-row plots and two to three replications. GY was recorded in Mg ha^{-1} adjusted to 155 g kg^{-1} grain moisture, while GDMC was recorded in

percent. Pedigree data for all inbred lines were gathered from breeding history tracing back at least two ancestral generations.

Biometrical analysis of phenotypic data

Data for GY and GDMC from all experiments were subjected to combined analyses with a mixed linear model. Main effects for years, locations, and check varieties were treated as fixed, which allowed to account for performance differences between experiments. Genotypic effects, all interactions, and block effects for trials, replications within trials, and incomplete blocks within replications were treated as random. Genotypic effects of the crosses were partitioned into general combining ability (GCA) effects of the parental dent and flint inbreds, and SCA effects of the crosses (Gardner and Eberhart 1966). It was assumed that GCA effects were normally distributed with variances $A_1 \times \sigma^2_{\text{dent}}$ and $A_2 \times \sigma^2_{\text{flint}}$ and SCA effects with variance $D \times \sigma^2_{\text{sca}}$, where the matrices, A_1, A_2 and D, were computed from coefficients of coancestry among the inbred lines (Bernardo 2002). Coefficients of coancestry were calculated using SAS procedure INBREED (SAS Institute Inc. 2000). Covariance between the per se performance of an inbred and its corresponding GCA effect as a parent of crosses was considered in the model. Consequently, estimation of GCA effects may benefit from line per se data and vice versa, line per se effects may benefit from HP data if available. The same covariance structure was used for the genotype × location interaction, except that effects for the SCA × location interaction were assumed independent. The residual error was assumed to be specific for each trial. All other block variances were assumed to be homogeneous. Mixed linear model analyses were performed with ASReml (Gilmour et al. 2002).

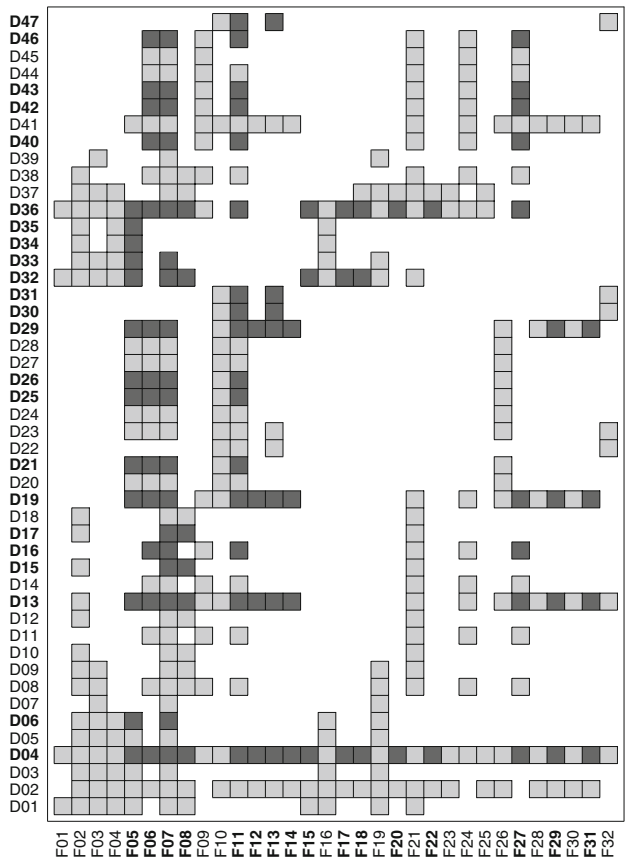

Fig. 1 Unbalanced factorial crosses showing one possible random subdivision for the estimation set and test set in cross-validation of hybrid performance prediction. The available dataset was based on 400 crosses (indicated by cells). In each of 300 cross-validation runs, 23 of the 47 dent lines (D01–D47) and 16 of the 32 flint lines (F01–F32) were randomly assigned as testcross evaluated lines (*bold letters*). The crosses between the testcross evaluated lines formed the estimation set (indicated by *dark gray cells*) and all remaining crosses formed the test set (indicated by *light gray cells*)

Evaluation of the efficiency of prediction methods

Cross-validation was performed to evaluate the efficiency of the investigated prediction methods. For each cross-validation run, the entire 47 × 32 factorial dataset was divided into an estimation set and a test set (Fig. 1). In order to mimic the situation in commercial hybrid breeding programs, in each cross-validation run 23 dent and 16 flint lines were randomly assigned as 'TC evaluated', which corresponds to half of the parental lines from each heterotic group. The remaining inbreds were designated as 'TC unevaluated'. The crosses between TC-evaluated lines formed the estimation set, which by expectation covered 24.5% of all crosses. The remaining 75.5% crosses formed the test set, such that 25.5% of all crosses were termed as 'Type 0' meaning that none of the parents were evaluated for TC performance, and 50% of all crosses were of 'Type 1' meaning that one parent was evaluated for TC performance. In a complete 47 × 32 factorial (1,504 crosses), this procedure would assign 368 crosses to the estimation set and 1,136 crosses to the test set. However, since only 400 out of 1,504 crosses were studied, the numbers of crosses in the estimation and test sets were accordingly lower (Fig. 1).

For each cross-validation run, phenotype data collected in field experiments were analyzed separately on the basis of the specific estimation set, using the mixed linear model. Initially, re-estimating variance components in each cross-validation run was considered, however, occasional lack of convergence was observed. For this reason, cross-validation was performed using variance estimates based on the whole dataset. For each prediction method under investigation, the BLUPs obtained from the estimation set were then used as the basis for performance prediction of the

hybrids in the test set. Prediction efficiency was assessed by two statistics, namely the proportion of explained variance (R^2) and the square root of the mean square deviation (RMSD) between predicted and observed HP values of all hybrids in the test set as well as separately for Type 0 and Type 1 hybrids. Higher estimate of R^2 and lower estimate of RMSD indicated better efficiency. Cross-validation was carried out with 300 randomized replications. Boxplots were based on Tukey's five number summary (Tukey 1977).

Prediction methods

PP-GS

Based on phenotypic and pedigree data only ('PP-'), the performance of test-set hybrids was predicted by BLUPs for GCA and SCA ('-GS') of the corresponding dent and flint lines and their crosses, estimated by the mixed linear model analysis from the hybrids in the estimation set. The GCA estimates for TC-unevaluated parental lines were obtained from data of related TC-evaluated inbred lines using coefficients of coancestry in the mixed linear model analysis. Likewise, the SCA estimates of hybrids in the test set were obtained from related hybrids in the estimation set.

PP-L

The mid-parent performance based on BLUPs of parent line ('-L') per se performance was considered the predictor of HP.

MLR-H

The prediction of HP in the test set was regarded as a multiple linear regression (MLR) problem, as described in detail by Schrag et al. (2007). Briefly, the HP effects ('-H') of the genotypic classes among the hybrids in the estimation set were computed at each AFLP marker locus. Markers were added to the MLR model by forward selection using the Schwarz Bayesian criterion (Schwarz 1978). With the resulting model, performance of the test-set hybrids was predicted. The MLR analysis was restricted to markers with no missing observations.

TEAM-H

The genotypic value of each hybrid was estimated by the sum of marker class effects across all AFLP markers that were significantly associated with the investigated trait (Schrag et al. 2007). This sum was termed total effects of associated markers (TEAM). Values of HP ('-H') were regressed on the TEAM values across all hybrids in the estimation set. The regression parameters thus obtained were used to predict the performance of hybrids in the test set on the basis of their TEAM values.

MLR-LM

Another approach for HP prediction combined line per se performance with MH ('-LM'), the latter being predicted using AFLP markers. The BLUPs for performance of the parental lines per se were estimated with the mixed linear model analysis. For each estimation-set hybrid, MH was calculated as the difference between HP and the mean performance of the two parental lines per se, and used to calibrate the MLR model based on AFLP marker data. Subsequently, HP of the test-set hybrids was predicted by adding (1) MH predicted with MLR and (2) the mid-parent performance estimated from the mean of BLUPs of the corresponding parental lines per se performance.

TEAM-LM

Analogous to MLR-LM, the HP of the test-set hybrids was predicted by adding (1) MH predicted with TEAM and (2) the mid-parent performance estimated from the mean of BLUPs of the corresponding parental lines per se performance.

Analysis of haplotype block structure

Adjacent AFLP marker loci were combined into haplotype blocks, considering marker data of all 79 inbred lines and using the HB2 and HB3 approaches described in detail by Schrag et al. (2007). Briefly, for the HB2 approach, the maximum block length was four markers, and the markers affected by missing observations were considered as blocks with the length of one. In the HB3 approach, a maximum of 15 markers displaying strong linkage disequilibrium were combined into a haplotype block. All marker-based prediction methods were applied to the HB2 and HB3 haplotype block data analogously to the use of single AFLP marker data described above.

Results

Molecular and phenotypic data

For all analyses, a set of 732 AFLP markers was chosen that was polymorphic and had less than 30% missing observations in the 79 inbred lines. Of these markers, a proportion of 19% had no missing marker observation among all lines. For the haplotype blocks, a proportion of

16% of the 705 HB2 blocks and 12% of the 424 HB3 blocks had no missing observation. For GY, 40–59% of HP was attributable to heterosis, whereas for GDMC, contribution of heterosis ranged from −8 to 5% (Table 1). With both traits the estimate of the GCA variance component was larger for dent than for flint lines. The ratio of SCA variance to GCA variance (average of dent and flint GCA variance components) was 0.07 for GY and 0.04 for GDMC. Correlation coefficients between mid-parent performance and HP were 0.44 for GY and 0.81 for GDMC, both estimates being highly significant ($P < 0.001$).

Efficiency of prediction methods

For GY, the number of hybrids in the estimation set ranged from 60 to 145 across all cross-validation runs and the average was 98.3, which corresponded to 24.6% of the 400 crosses studied. For GDMC, the number of hybrids in the estimation set was 43–140 with an average of 97.2, which was 24.3% of the 400 hybrids.

The prediction methods explained 16.5–46.0% of the observed variance for GY and 16.1–69.9% for GDMC (Table 2). For GY, the TEAM methods had by far the highest R^2, followed by the MLR and PP-GS methods. The poorest prediction efficiency was observed for PP-L, which had the lowest R^2 and exceptionally high RMSD. For GDMC, the highest R^2 and the lowest RMSD were obtained by the application of the TEAM-LM method. For R^2, it was followed by PP-L and MLR-LM, indicating that the methods that considered inbred line per se estimates (TEAM-LM, PP-L, MLR-LM) explained consistently higher proportions of the GDMC variance in contrast to TEAM-H, PP-GS, and MLR-H. Within both groups of methods (considering inbred line per se estimates versus others) the ranking was identical for R^2 (TEAM > PP > MLR) and RMSD (TEAM < PP < MLR) of GDMC. For both traits, the haplotype block analyses applied to TEAM-H and MLR-H showed only small differences in R^2 among single AFLP marker data, HB2, and HB3 (Fig. 2).

Prediction of performance of Type 0 and Type 1 hybrids

For GY, the two TEAM methods showed the highest R^2 for the performance prediction of Type 0 hybrids, the R^2 being 0.37 for TEAM-H and TEAM-LM (Fig. 3). The difference in R^2 (ΔR^2) between Type 0 and Type 1 hybrids was only 0.12–0.14 for the TEAM methods. For GDMC, TEAM-LM and PP-L had high R^2 for Type 0 hybrids (0.66 for TEAM-LM and 0.64 for PP-L). Both these methods showed small ΔR^2 between Type 0 and Type 1 hybrids (0.06 for TEAM-LM and 0.01 for PP-L). Further, for GDMC, all methods based on line per se values had smaller ΔR^2 than the counterpart methods. The MLR-H and MLR-LM methods generally resulted in lower R^2 than the respective TEAM methods and their ΔR^2 between Type 0 and Type 1 hybrids was larger than for TEAM in all cases. The PP methods showed extreme results for ΔR^2 with practically no reduction in PP-L (0.01 for both traits) and the largest in PP-GS (0.33 for GY and 0.38 for GDMC).

Table 1 Biometrical analyses of phenotypic data of 400 crosses and their 79 parental inbred lines: mean and range of hybrid performance (HP), mid-parent performance, mid-parent heterosis (MH), ratio MH/HP as well as the estimates of variance components of general combining ability (GCA) and specific combining ability (SCA) and their standard errors (SEs)

Criterion	Grain yield (Mg ha^{-1})		Grain dry matter content (%)	
	Mean	Range	Mean	Range
Performance-related measure				
HP	10.76	9.44–12.38	70.01	65.12 to 74.80
Mid-parent performance	5.44	4.45–6.43	71.18	62.28 to 77.73
MH	5.32	4.01–6.60	−1.16	−5.93 to 3.58
MH/HP	0.49	0.40–0.59	−0.02	−0.08 to 0.05
Criterion	Grain yield (Mg ha^{-1})		Grain dry matter content (%)	
	Estimate	SE	Estimate	SE
Variance component				
GCA (dent)	0.221	±0.064	3.741	±0.893
GCA (flint)	0.151	±0.056	1.347	±0.479
SCA	0.013	±0.009	0.111	±0.050

Table 2 Efficiency of various methods for prediction of grain yield and grain dry matter content of hybrids of which no or only one parental line was evaluated for testcross performance

Method	Grain yield (Mg ha^{-1})		Grain dry matter content (%)	
	Median R^2	Median RMSD	Median R^2	Median RMSD
PP-GS	0.240	0.519	0.281	1.624
PP-L	0.165	5.441	0.648	1.933
MLR-H	0.229	0.615	0.161	2.226
MLR-LM	0.241	0.646	0.488	1.937
TEAM-H	0.451	0.415	0.339	1.517
TEAM-LM	0.460	0.433	0.699	1.379

Discussion

Biometrical analysis of field data

The materials evaluated showed a high degree of heterosis for GY (mean MH/HP ratio 49%) whereas heterosis was of little importance for GDMC (mean MH/HP ratio −2%). This enabled the efficiency evaluation of prediction methods for two traits having extreme expression of heterosis.

The present study was based on combined analyses of nine experiments, four of which were investigated previously in separate analyses by Schrag et al. (2006). As compared with the mean of these previous analyses, the present estimate of GCA variance component was comparable for GY of flint lines, smaller for GY of dent lines and GDMC of flint lines, and larger for GDMC of dent lines. The present estimates of SCA variance component as well as the SCA:GCA ratios for both traits were smaller. The fact that GCA and SCA variances depend on genetic materials and test environments may explain the slight variation in these estimates compared with the corresponding values reported by Schrag et al. (2006). Variance components for GDMC in the present study were in good agreement with those reported by Parisseaux and Bernardo (2004) based on 22,774 single crosses belonging to nine heterotic patterns. Furthermore, Fischer et al. (2008) reported similarly low SCA:GCA ratios for GY. Overall, this result indicates the minor importance of SCA in the germplasm used in the maize breeding program at the University of Hohenheim, emphasizes the predominant role of GCA for GY and GDMC, and consequently suggests that variation in HP for these traits may be explained largely by GCA estimates.

With the mixed linear model, it was possible to analyze an unbalanced dataset from separate experiments comprising different years and locations. Additionally, with the combined analysis of experiments on crosses and inbred parents, the BLUP estimates were obtained for HP, parental line per se performance, MH, GCA and SCA. This approach facilitates the utilization of the large amount of phenotypic data that are generated in commercial hybrid breeding programs, and allows the performance prediction of crosses of which one or even both parents have not been evaluated.

Efficiency of prediction methods

Heterosis has a major influence on maize GY, which (1) reduces the correlation between parental line per se performance and HP and (2) leads to large differences between the average performance levels of inbred lines and hybrids. It can be regarded as a major cause for the low R^2 and the exceptionally high RMSD value for the PP-L method. Keeping in view the complex inheritance of GY, it may be stated that the differences between hybrids and inbreds due to MH were modeled successfully with the marker-based TEAM methods ($R^2 \geq 45\%$; RMSD ≤ 0.43). In TEAM-H, markers were used for prediction of HP, whereas in TEAM-LM, markers were used to predict MH instead, which was combined with line per se BLUP estimates. This slightly increased R^2 over that of TEAM-H. In summary, the TEAM approaches provided by far the best predictions for GY, indicating that marker data can advantageously be used for this purpose. In contrast, the line per se (PP-L) approach failed to predict the performance for a heterotic trait like GY, as was expected from the literature (Smith 1986; Hallauer 1990). However, combining line per se performance estimates with a good marker-based approach for MH prediction (such as TEAM) provided equally reliable HP predictions.

The major proportion of HP variance in GDMC is contributed by the performance of the parental inbreds per se. This may explain the generally higher GDMC prediction efficiency of the methods that included per se estimates (PP-L, MLR-LM, TEAM-LM) in comparison with the other methods. Heterosis was very low for GDMC. However, the general performance levels of hybrids and inbreds still differed. This systematic deviation is a source of RMSD, which may explain the higher RMSD for PP-L than PP-GS. Since GCA and SCA values were estimated from hybrids, PP-GS better reflected the general performance level of hybrids and resulted in a lower systematic deviation. The marker-based MH prediction by TEAM-LM accounted for this systematic deviation between hybrids and inbreds in addition to a small proportion of the explained HP variance. Consequently, the TEAM-LM method in comparison with PP-L further improved the efficiency of GDMC prediction, as reflected by an increased R^2 and reduced RMSD. For the non-heterotic trait GDMC, methods using line per se performance were

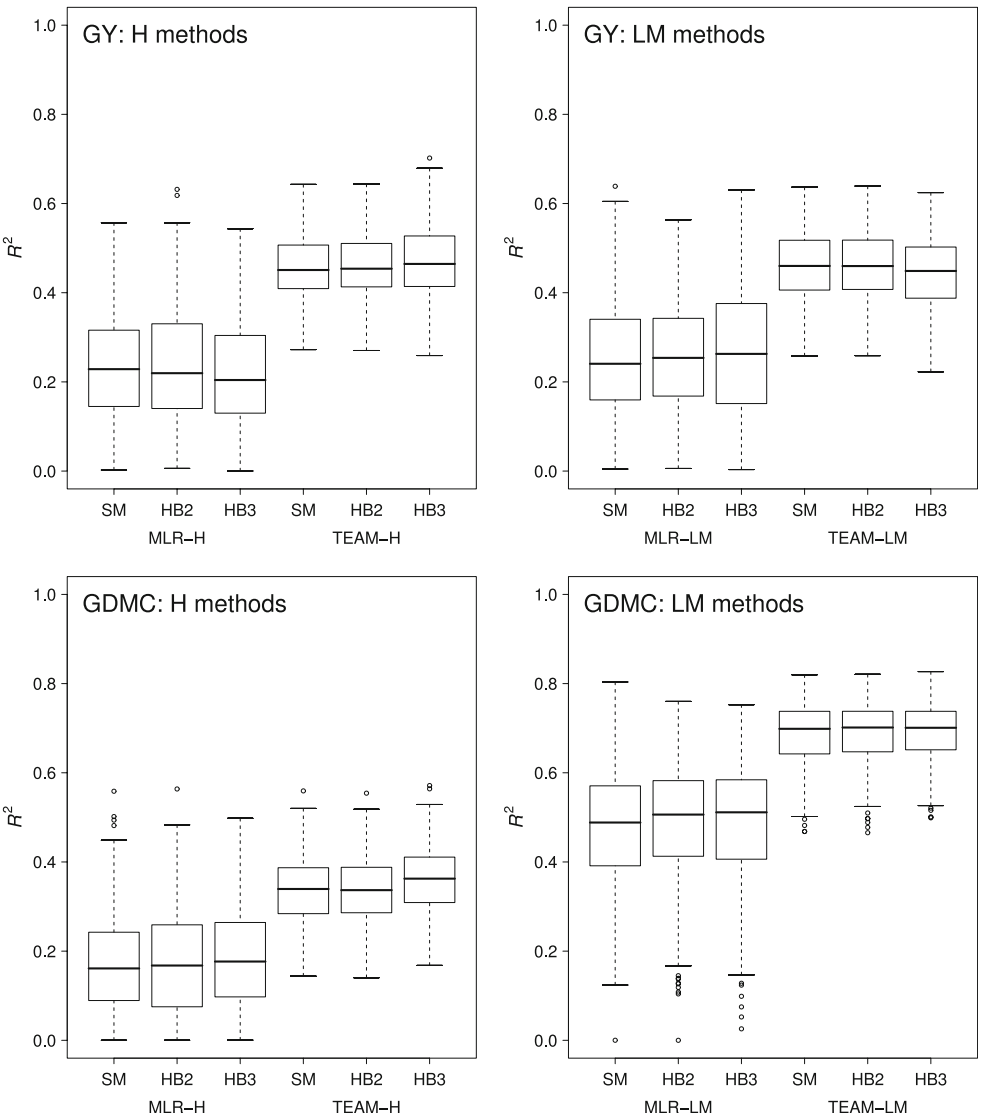

Fig. 2 Efficiency (R^2) of marker-based methods (MLR-H, MLR-LM, TEAM-H, TEAM-LM) applied to single AFLP marker data (SM) and haplotype blocks (HB2, HB3) for prediction of grain yield (GY) and grain dry matter content (GDMC) of hybrids of which no (Type 0) or only one (Type 1) parental line was evaluated for testcross performance

advantageous and the usefulness of marker-based approaches was limited. MLR failed and marker-based predictions with TEAM based on HP directly were poor. Only when combining line per se performance estimates with marker-based heterosis (TEAM-LM), predictions could be improved slightly by the inclusion of marker data. In the present study, the efficiency of various methods varied with the degree of heterosis but at the same time TEAM-LM was the most efficient method to predict HP for both GY and GDMC.

Haplotype blocks

The TEAM methods do not account for correlation structure among the AFLP marker loci. Therefore, the use of haplotype blocks was expected to further improve the prediction efficiency of TEAM. Such approaches were presented by Schrag et al. (2007), where the haplotype block approaches HB2 and HB3 were slightly advantageous for prediction with TEAM and MLR methods. In the current study, prediction efficiencies based on 732 single AFLP markers and 705 HB2 haplotype blocks were similar (Fig. 2), which can be explained by the fact that the haplotype structure obtained with the HB2 algorithm differed only marginally from the single AFLP marker data. This was due to the property of the HB2 procedure that adjacent markers could only be combined into a haplotype block if they were completely unaffected by missing marker observations. Application of the HB3 algorithm to the 732 AFLP markers resulted in a distinct block structure comprising 424 haplotype blocks, of which a very low number of 51 blocks were completely unaffected by missing marker observations. Since the MLR method could utilize only these 51 blocks as its database, this may have impaired the prediction efficiency of MLR based on HB3 data for GY. In contrast, the TEAM

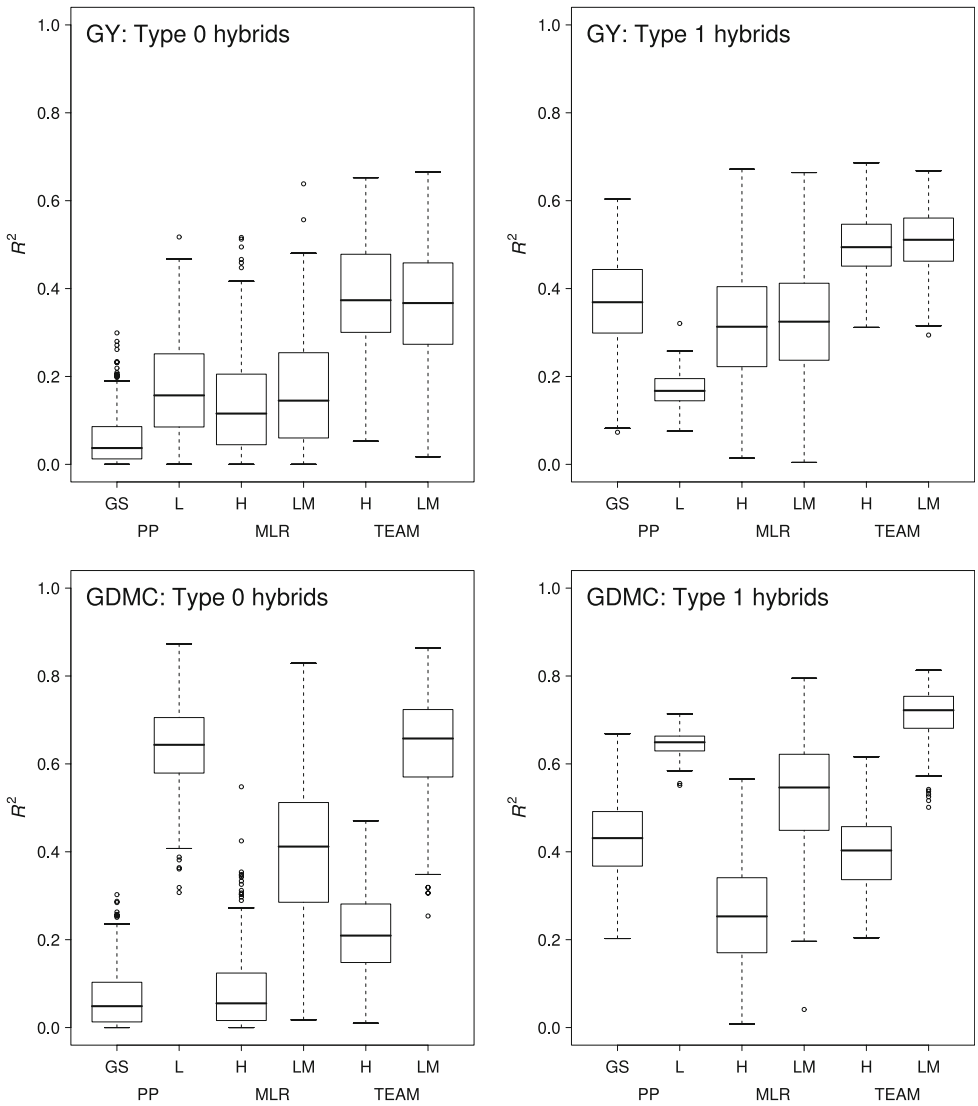

Fig. 3 Efficiency (R^2) of various methods for prediction of grain yield (GY) and grain dry matter content (GDMC) of hybrids of which no (Type 0) or only one (Type 1) parental line was evaluated for testcross performance

methods (1) had less adverse affect due to missing marker observations and (2) should benefit from addressing the correlation structure of markers by the use of haplotype blocks. Therefore, improved prediction efficiency with the TEAM methods was expected. However, a small increase in R^2 was observed for TEAM-H only. In general, the use of haplotype blocks did not increase substantially the prediction efficiency in our data. However, on higher marker densities, e.g., with SNP chips, haplotype blocks are expected to be more advantageous. Missing marker data represent a major problem for haplotype block analysis, which among other reasons may explain the poor improvement of prediction efficiencies. Missing data of markers can be predicted from observed genotypes of tightly linked markers (Balding 2006). Such data imputation promises to be a simple and effective solution for improving the HP prediction especially if based on haplotype blocks and also for the MLR approach.

Prediction of performance of Type 0 and Type 1 hybrids

Performance prediction of hybrids having untested parental lines is of major interest to plant breeders. If hybrids are regarded as the sampling unit, a very high proportion of parental lines is left in the estimation set due to the structure of the factorial mating design. The extreme case for such a scenario is a hybrid-based leave-one-out strategy (Vuylsteke et al. 2000; Schrag et al. 2006), where large numbers of testcrosses of each parental line are included in each cross-validation run. Such a strategy may result in very high prediction efficiency but it does not address the demand of plant breeders who are interested in predicting the performance of hybrids involving parental lines that have not been TC evaluated. Thus, the efficiency of the prediction methods was studied for Type 0 hybrids (TC-unevaluated × TC-unevaluated parents) and Type 1

hybrids (TC-evaluated × TC-unevaluated parents). This was achieved by treating the parental inbred lines instead of crosses as sampling unit in cross-validation.

Prediction efficiency of PP-L differed only marginally between Type 0 and Type 1 hybrids. The per se BLUP value was mainly based on the performance of parental lines per se, and to a limited extent on TC evaluation as correlation between parent per se performance and GCA was considered. This may explain the marginal differences between Type 0 and Type 1 hybrids with the application of PP-L in the present study. In contrast to PP-L, prediction efficiency of PP-GS differed considerably between Type 0 and Type 1 hybrids. The PP-GS method was based on GCA and SCA estimates obtained from the phenotypic performance of the crosses. When the parental lines have not been evaluated in cross combinations, these estimates were obtained from phenotypic data of the crosses involving related inbred parents and their coefficients of coancestry in the mixed linear model analysis. However, these coefficients are only expected values of genome distribution and, therefore, may not reflect the actual relatedness of the inbred lines investigated, which is influenced by selection and genetic drift during inbreeding (Melchinger et al. 1991; Bernardo et al. 1996). Moreover, the coefficients indicate overall expectations for the whole genome and do not refer to specific genomic regions that may be relevant for the predicted trait. This limitation can be overcome by approaches such as TEAM or MLR, which are based on molecular markers associated with the trait under investigation.

Size of estimation and test sets

Subdivision of data into estimation and test sets for cross-validation was done considering the situation across multiple generations of a commercial plant breeding program. We sampled 23 dent and 16 flint lines as TC-evaluated parents in the cross-validation runs, leading to a balanced subdivision into the estimation set as well as Type 0 and Type 1 hybrids in the test set. An expansion of the estimation set to 38 dent × 26 flint lines resulted in a strong downsizing in the Type 0 component of test set and increased variance of prediction efficiency, which reflects the sampling effect due to the smaller test sets. In comparison with the 23 × 16 scenario, the database of the 38 × 26 estimation set was substantially larger and resulted in higher GY prediction efficiency for both PP-GS and TEAM-LM methods (data not shown). Increases of similar magnitude were also reported by Bernardo (1994) and Charcosset et al. (1998).

Replacing line per se evaluation by additional hybrids

In this study, all parental inbred lines were evaluated for their line per se performance. However, given a fixed budget, additional hybrids could have been evaluated for their performance instead of these parental lines. Such hybrids would enhance the database for effect estimation and could improve the connectivity among the several field experiments within the analysis. The question is open, whether under such a scenario, the GCA/SCA-based prediction benefits to such an extent that it reaches the efficiency of the line per se-based approaches.

The line per se-based approaches predicted GDMC very well, both based on phenotypic data only (PP-L) and combined with the marker TEAM approach (TEAM-LM). In general, the low heterosis of GDMC and the low relevance of SCA for this trait support the GCA/SCA-based prediction. However, for hybrids, of which only one or no parental line was TC evaluated, the prediction efficiency of GCA/SCA is much lower than for the line per se-based approaches that even with additional hybrids the GCA/SCA-based prediction can hardly be improved up to the level of line per se-based prediction approaches.

For GY, the line per se-based approach is not generally superior to the GCA/SCA approach. Even the combination of line per se performance with markers (TEAM-LM) was equal but not superior to the marker-based TEAM-H approach, which not generally requires line per se data. Thus, for GY it could be interesting to study whether the prediction efficiency of PP-GS but also of TEAM-H can further be increased by replacing the line per se evaluations with additional hybrids, which may improve the connectivity between the experiments. Thus, the experiments could be augmented with hybrids from previous experiments. A simple and common method for improving connectivity is to include several check varieties in each trial, as was done in the current work.

Application in plant breeding programs

Inbred lines in large numbers are continuously produced in commercial maize breeding programs but by far not all possible cross combinations are evaluated because of resource constraints. On the other hand, our study indicates that on the basis of marker assayed inbred lines, the performance of maize hybrids can be predicted efficiently for the heterotic trait GY as well as the non-heterotic trait GDMC for crosses of which one or even both inbred parents have not been TC evaluated. Relative response to indirect selection (say selection based on markers to improve GY) versus direct selection depends, in addition to other factors, on relative selection intensities and the proportion of additive-genetic variance of GY accounted for by the markers versus the heritability of GY (Dudley 1993; Falconer and Mackay 1996). With the development of high-throughput marker platforms, molecular markers have become available in large numbers and the cost per

marker data point has decreased considerably (Eathington et al. 2007). Further, the cost of genotyping is expected to decrease further compared with that for phenotyping in field trials. Thus, more intense selection can be carried out based on markers as compared to direct selection for GY. Also, selection based on increased marker density is expected to enhance the proportion of variance accounted for by markers. These factors positively affect the advantage of marker-assisted selection with large impact on allocation of resources in commercial plant breeding programs.

Marker-assisted selection among inbred lines and prediction of HP will particularly enhance the utility of doubled haploid technology, which generates a vast number of homozygous lines in short time. Marker assays may be conducted immediately after the development of doubled haploids. Promising doubled haploids may be identified using marker effects estimated in the previous breeding cycle, and selected for seed multiplication and evaluation of their per se performance. The HP may be predicted using molecular markers (e.g., TEAM-H) or using molecular markers combined with per se performance data (e.g., TEAM-LM). The hybrids having high predicted performance and parents with high per se performance required for commercial seed production may be identified, their seed produced and field trials conducted on their performance. The data generated in each new cycle of breeding may be analyzed with a mixed linear model approach and used to re-estimate marker-QTL associations for continuous refinement of the estimates of marker effects. This approach will not only enable maize breeders to enhance selection intensity but will also positively impact the selection response per unit of time by reducing the number of seasons required to develop a commercial hybrid.

Acknowledgments This project was supported by the German Research Foundation DFG in the framework program 'Heterosis in Plants' (research grants FR 1615/3-1 and 3-3). The staff at the Plant Breeding Research Station at Eckartsweier and Hohenheim is gratefully acknowledged for conducting the field experiments. The authors appreciate the editorial work of J. Muminović and thank H.F. Utz for his valuable comments and suggestions.

References

Balding DJ (2006) A tutorial on statistical methods for population association studies. Nat Rev Genet 7:781–791

Bernardo R (1994) Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Sci 34:20–25

Bernardo R (1998) Predicting the performance of untested single crosses: trait and marker data. In: Lamkey KR, Staub JE (eds) Concepts and breeding of heterosis in crop plants. Crop Science Society of America, Madison, pp 117–127

Bernardo R (1999) Marker-assisted best linear unbiased prediction of single-cross performance. Crop Sci 39:1277–1282

Bernardo R (2002) Breeding for quantitative traits in plants. Stemma Press, Woodbury

Bernardo R, Murigneux A, Karaman Z (1996) Marker-based estimates of identity by descent and alikeness in state among maize inbreds. Theor Appl Genet 93:262–267

Charcosset A, Bonnisseau B, Touchebeuf O, Burstin J, Dubreuil P, Barriere Y, Gallais A, Denis JB (1998) Prediction of maize hybrid silage performance using marker data: comparison of several models for specific combining ability. Crop Sci 38:38–44

Dudley JW (1993) Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. Crop Sci 33:660–668

Eathington SR, Crosbie TM, Edwards MD, Reiter RS, Bull JK (2007) Molecular markers in a commercial breeding program. Crop Sci 47:S154–S163

Falconer DS, Mackay TF (1996) Introduction to quantitative genetics. Longman Group, Essex

Fischer S, Möhring J, Schön CC, Piepho H-P, Klein D, Schipprack W, Utz HF, Melchinger AE, Reif JC (2008) Trends in genetic variance components during 30 years of hybrid maize breeding at the University of Hohenheim. Plant Breed 127:446–451

Gardner CO, Eberhart SA (1966) Analysis and interpretation of the variety cross diallel and related populations. Biometrics 22:439–452

Gilmour AR, Cullis BR, Welham SJ, Thompson R (2002) ASReml reference manual. Release 1.0. VSN International, Hemel Hempstead

Hallauer AR (1990) Methods used in developing maize inbreds. Maydica 35:1–16

Hallauer AR, Miranda Filho JB (1988) Quantitative genetics in maize breeding. Iowa State University Press, Ames

Longin CFH, Utz H, Melchinger A, Reif JC (2007) Hybrid maize breeding with doubled haploids. II. Optimum type and number of testers in two-stage selection for general combining ability. Theor Appl Genet 114:393–402

Melchinger AE, Messmer MM, Lee M, Woodman WL, Lamkey KR (1991) Diversity and relationships among U.S. maize inbreds revealed by restriction fragment length polymorphisms. Crop Sci 31:669–678

Parisseaux B, Bernardo R (2004) In silico mapping of quantitative trait loci in maize. Theor Appl Genet 109:508–514

SAS Institute Inc. (2000) SAS System 8. SAS Institute Inc., Cary

Schmidt W (2004) Hybridmaiszüchtung bei der KWS SAAT AG. Bericht über die 54. Tagung der Vereinigung der Pflanzenzüchter und Saatgutkaufleute Österreichs 2003, Gumpenstein, pp 1–6 (in German)

Schrag TA, Melchinger AE, Sørensen AP, Frisch M (2006) Prediction of single-cross hybrid performance for grain yield and grain dry matter content in maize using AFLP markers associated with QTL. Theor Appl Genet 113:1037–1047

Schrag TA, Maurer HP, Melchinger AE, Piepho H-P, Peleman J, Frisch M (2007) Prediction of single-cross hybrid performance in maize using haplotype blocks associated with QTL for grain yield. Theor Appl Genet 114:1345–1355

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

Seitz G (2005) The use of doubled haploids in corn breeding. In: Proceedings of the 41st annual Illinois corn Breeders' School 2005, Urbana-Champaign, pp 1–7

Smith OS (1986) Covariance between line per se and testcross performance. Crop Sci 26:540–543

Tukey JW (1977) Exploratory data analysis. Addison-Wesley, Reading

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP—a new technique for DNA-fingerprinting. Nucleic Acids Res 23:4407–4414

Vuylsteke M, Mank R, Antonise R, Bastiaans E, Senior ML, Stuber CW, Melchinger AE, Lübberstedt T, Xia XC, Stam P, Zabeau M, Kuiper M (1999) Two high-density AFLP® linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. Theor Appl Genet 99:921–935

Vuylsteke M, Kuiper M, Stam P (2000) Chromosomal regions involved in hybrid performance and heterosis: their AFLP®-based identification and practical use in prediction models. Heredity 85:208–218