Theor Appl Genet (2012) 125:707–713
DOI 10.1007/s00122-012-1862-2

ORIGINAL PAPER

Impact of selective genotyping in the training population on accuracy and bias of genomic selection

Yusheng Zhao · Manje Gowda · Friedrich H. Longin · Tobias Würschum · Nicolas Ranc · Jochen C. Reif

Received: 12 October 2011 / Accepted: 22 March 2012 / Published online: 6 April 2012

Abstract Estimating marker effects based on routinely generated phenotypic data of breeding programs is a cost-effective strategy to implement genomic selection. Truncation selection in breeding populations, however, could have a strong impact on the accuracy to predict genomic breeding values. The main objective of our study was to investigate the influence of phenotypic selection on the accuracy and bias of genomic selection. We used experimental data of 788 testcross progenies from an elite maize breeding program. The testcross progenies were evaluated in unreplicated field trials in ten environments and fingerprinted with 857 SNP markers. Random regression best linear unbiased prediction method was used in combination with fivefold cross-validation based on genotypic sampling. We observed a substantial loss in the accuracy to predict genomic breeding values in unidirectional selected populations. In contrast, estimating marker effects based on bidirectional selected populations led to only a marginal decrease in the prediction accuracy of genomic breeding values. We concluded that bidirectional selection is a valuable approach to efficiently implement genomic selection in applied plant breeding programs.

Communicated by J. Wang.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-012-1862-2) contains supplementary material, which is available to authorized users.

Y. Zhao · M. Gowda · F. H. Longin · T. Würschum · J. C. Reif (✉)
State Plant Breeding Institute, University of Hohenheim, 70599 Stuttgart, Germany
e-mail: jochreif@uni-hohenheim.de

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

Introduction

Genomic selection promises to substantially increase selection gain in crop and livestock breeding programs (Meuwissen et al. 2001; Heffner et al. 2010). It differs from previous strategies such as marker-assisted selection by using a large number of molecular markers ideally covering the whole genome to estimate genomic breeding values (Jannink et al. 2010). As the first step in genomic selection, marker effects are estimated in a training population. In the second step, individuals related or unrelated to the training population that have been genotyped but not phenotyped are selected based on their marker profiles.

First empirical findings for dairy cattle supported the potential of genomic selection for livestock breeding (e.g., Hayes et al. 2009; Luan et al. 2009). For plant breeding, empirical evaluation of the accuracy of genomic selection has been done using cross-validation in biparental families of maize, barley and Arabidopsis (Lorenzana and Bernardo 2009). Crossa et al. (2010, 2011) examined empirically the prospects of genomic selection in diverse panels of wheat and maize lines. Albrecht et al. (2011) and Zhao et al. (2012a, b) used cross-validation based on testcross progenies of inbred lines from applied maize breeding programs to estimate the accuracy of genomic selection. The findings of the above-mentioned studies suggest that genomic selection can be an effective strategy in plant breeding with high accuracies in predicting genomic breeding values.

Using routinely generated phenotypic data of applied plant breeding programs is an economic approach to implement genomic selection. Advanced breeding

populations are selected by applying multi-stage selection: at early stages several families with large population size are commonly evaluated in a small number of environments and at later stages few families with a small number of individuals are phenotyped in a large number of environments (Longin et al. 2006, 2007). A training population consisting of routinely phenotyped individuals at early stages is suboptimal as genotypic values are estimated only with a moderate heritability. While heritability is high for lines tested at later stages, truncation selection may lead to a severe bias in the genetic makeup of the training population. To the best of our knowledge, no study was published on the role of truncation selection in training populations on a bias of genomic selection, although it is of high relevance for applied plant breeding populations.

In the context of linkage mapping, selective genotyping was proposed to improve the efficiency of QTL detection under a restricted budget (Lebowitz et al. 1987; Lander and Botstein 1989; Sun et al. 2010). For selective genotyping, DNA markers were only assayed for the most informative progenies, those with high or low phenotypic values of the target trait. This bidirectional selection showed to be much more powerful than random sampling (Navabi et al. 2009) but power and type I error were still better in the full data set compared to bidirectional selective genotyping (Sun et al. 2010). The possibility of selective genotyping in the context of genomic selection has to the best of our knowledge not yet been investigated.

The main goal of our study was to investigate the influence of phenotypic selection on the accuracy and bias of genomic selection using large multi-parental populations from a commercial maize breeding program. Our objectives were to (1) study the prediction accuracy of genomic breeding values and bias in unidirectional or bidirectional selected populations; (2) assess how the balanced and unbalanced family sizes affect the accuracy to predict genomic breeding values; and (3) evaluate how to efficiently implement genomic selection in applied breeding programs.

Materials and methods

Genotypic and phenotypic data

The field experiments were described in detail by Steinhoff et al. (2011). Briefly, six segregating families, with a total of 788 individuals were obtained from diallel crosses between four elite dent inbreds. Testcross progenies were evaluated in 2007 in Italy at 10 locations with unreplicated trials (Supplementary Table S1). The number of progenies in each family varied from 104 to 143. Data were recorded for grain yield (Mg ha^{-1}) at an adjusted moisture concentration of 155 g kg^{-1}. Each of the segregating families was evaluated in separate but adjacent field trials connected with common checks. Genotypic variances of the total population (σ^2_G) were estimated with following regression model: $y_{Env} = 1_n\mu + ZU_{Env} + GU_G + e$, where y_{Env} refers to the values of phenotypes of single environments adjusted using check varieties (Steinhoff et al. 2011), 1_n is a vector with the length of the number of individuals g times the number of environments l, μ denotes the overall mean, Z is a design matrix assigning fixed environment effects to the phenotypes, U_{Env} is a vector of environments effects, G is a design matrix for random genotypic effects, U_G refers to vector of the effects of the genotypes and here e is a residual term comprising genotype–environment interaction and error associated with y_{Env}. Moreover, best linear unbiased estimates (BLUEs) of means for testcross progenies and parents across environments were determined by assuming fixed genotypic effects. The 788 individuals have been fingerprinted with 857 high-quality SNP markers used for the further analyses (details of the fingerprinting have been described previously by Steinhoff et al. 2011).

Data analysis

Genomic selection was carried out similarly as described earlier by Zhao et al. (2012a). In brief, breeding values were estimated by the following model $y = 1_g\mu + Xa + e$, where y is a $g \times 1$ vector of BLUEs of genotype means estimated across environments based on the above defined model, $X = (X_1, X_2, \ldots, X_m)$ is a $g \times m$ matrix and m refers to the number of markers, where $X_j = (X_{1j}, X_{2j}, \ldots, X_{gj})^T$ is a $g \times 1$ vector denoting the genotype of the individuals for marker j. $X_{ij} = 0$ if individual i is homozygous for the first allele at locus j, $X_{ij} = 1/\sqrt{(2 - F_{inb})p_j(1 - p_j)}$ if heterozygous, $X_{ij} = 2/\sqrt{(2 - F_{inb})p_j(1 - p_j)}$ if individual i is homozygous for the second allele at locus j, where F_{inb} denotes the inbreeding coefficient of individual i and p_j refers to the allele frequency at marker j, $a = (a_1, a_2, \ldots, a_m)^T$ is a $m \times 1$ vector, and a_j is the effect of the jth marker. Meuwissen et al. (2001) proposed that the regression coefficients are independent and random draws from a normal distribution and, consequently, the variance of a_j is assumed to be $\sigma^2_{Genotype}/m \times \sigma^2_{Genotype}$ was chosen to reflect the genetic variance of the underlying population. The estimates of a_j were obtained from mixed-model equations (Henderson 1984). We used an estimate of the error variance of the BLUEs across environments, i.e., σ^2_{Error} (which comprises the genotype times environment interaction variance and the error variance) divided by the number of environments l. Consequently, the shrinkage parameter λ^2 was defined as $(\sigma^2_{Error}/l)/(\sigma^2_{Genotype}/m)$ (Meuwissen et al. 2001). Given the

estimates of a and μ as $\hat{a}$ and $\hat{\mu}$, genomic breeding values were predicted as $PV = 1_g\hat{\mu} + X\hat{a}$.

Cross-validation

We applied fivefold cross-validation based on genotypic sampling to evaluate the accuracy of genomic selection focusing on selected training populations. Here, the entire data set was randomly split into five subsets holding the proportion of individuals per population constant. Four subsets were combined and formed the training set for estimating genetic effects. The remaining subset formed the prediction set, in which predictions derived from the training set are tested for their predictive ability by estimating the Pearson correlation coefficient (r_{MP}) (Albrecht et al. 2011). The accuracy of genomic selection was expressed as $r_{GS} = r_{MP}/h$ (Dekkers 2007; Albrecht et al. 2011), where h was the square root of heritability. The regression of the observed phenotypes on the predicted phenotypes was used as a measure of the bias for the estimated genomic breeding values, where a slope of the regression of $r = 1$ denotes no bias, $r < 1$ implies an underestimation of the observed phenotypes on average r times, and a regression coefficient of $r > 1$ implies an overestimation of the observed phenotypes on average r times. In addition, we investigated the accuracy of the estimates of the marker effects by correlating marker effects estimated for the training population with those estimated for the whole population. The sampling of training and prediction set was repeated 5,000 times.

We estimated the marker effects and predicted the genomic breeding values for varying training and prediction sets using mass selection based on BLUEs across environments as follows:

1. Random—*Training set:* We first selected randomly 80 % of the individuals of the total data set. Then we further randomly selected 10–90 % of the individuals. *Prediction set:* The remaining 20 % of the individuals of the total data set were used to evaluate the prediction accuracy.
2. Unidirectional—*Training set:* We first selected randomly 80 % of the individuals of the total data set. Then we applied unidirectional selection to choose 10–90 % of the individuals with highest genotypic values. *Prediction set:* The remaining 20 % of the individuals of the total data set were used to evaluate the prediction accuracy.
3. Unidirectional-mod—*Training set*: We first selected randomly 80 % of the individuals of the total data set. Then we applied unidirectional selection to choose 30–50 % individuals with highest genotypic values. *Prediction set*: For the remaining 20 % of the individuals of the total data set unidirectional selection was applied to choose 30–50 % individuals with highest genotypic values. Prediction accuracy was evaluated using these selected individuals.
4. Bidirectional—*Training set*: We first randomly selected 80 % of the individuals of the total data set. Then we applied bidirectional selection to choose 10–90 % of the individuals with extreme (lowest or highest) genotypic values. *Prediction set*: The remaining 20 % of the individuals of the total data set were used to evaluate the prediction accuracy.
5. Unbalanced bidirectional—*Training set*: We first selected randomly 80 % of the individuals of the total data set. Then we applied unbalanced bidirectional selection with varying proportions of inferior individuals to choose 30–50 % of the individuals with extreme genotypic values. *Prediction set*: The remaining 20 % of the individuals of the total data set were used to evaluate the prediction accuracy.

Furthermore, we estimated the marker effects and predicted the genomic breeding values using family selection. For this scenario, we selected an unequal number of individuals per biparental family depending on the mean performance of the family. Sampling within biparental families was done randomly. This scenario reflects the situation that family sizes may vary according to the parental mean.

Results

With bidirectional selection and a high selection intensity of 10 %, prediction accuracy of genomic breeding values was superior compared to random selection and explains 74 % of the level observed for the full data set (Fig. 1). In contrast, with unidirectional selection and high selection intensities prediction accuracy of genomic breeding values was low with values of r_{GS} <0.20, i.e., 29 % of the level observed for the full data set. Prediction accuracies of genomic breeding values were always lower for unidirectional compared to random selection. Interestingly, prediction accuracy of genomic breeding values was higher when unidirectional selection was applied only for the training population compared to a scenario where unidirectional selection was applied for both the training and prediction set (Fig. 2). This higher accuracy was more pronounced with increasing number of selected individuals.

For bidirectional selection, we observed no bias in predicting genomic breeding values (Fig. 3). In contrast, with unidirectional selection prediction of genomic breeding values led to an underestimation of the observed

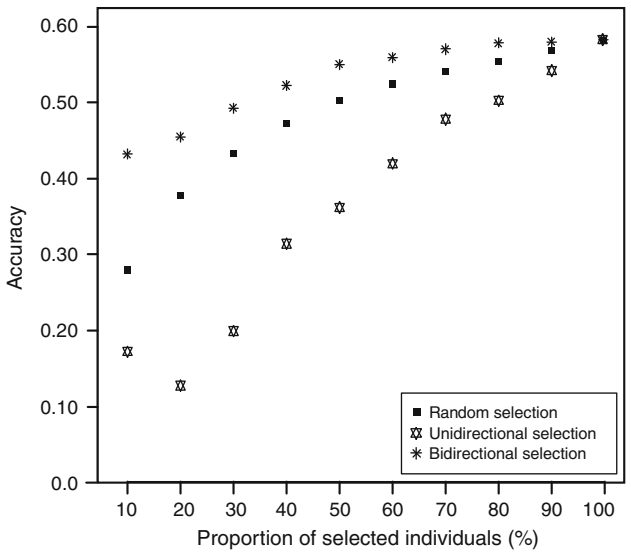

Fig. 1 Prediction accuracy of genomic breeding values for grain yield in association with different proportion of selected individuals of the total population applying (1) random selection, (2) unidirectional selection, and (3) bidirectional selection

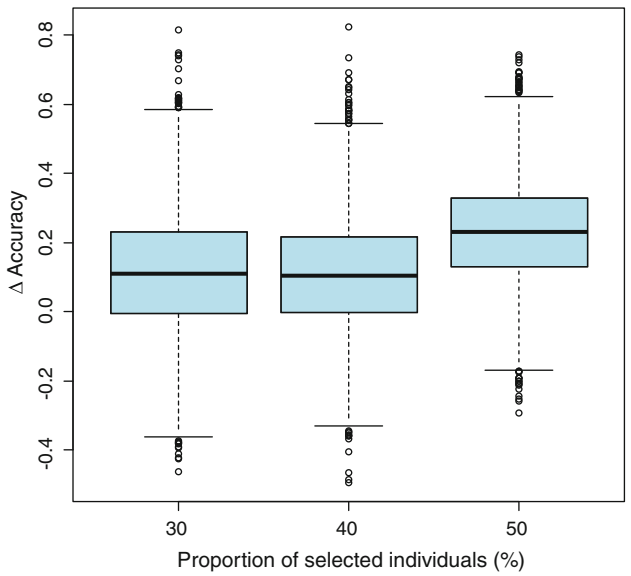

Fig. 2 Box-whisker plots for the difference in prediction accuracy (Δ accuracy) between unidirectional versus unidirectional-mod selection. For unidirectional selection, selection was applied only for the training population. In contrast, for unidirectional-mod selection, selection was applied for the training and the prediction population

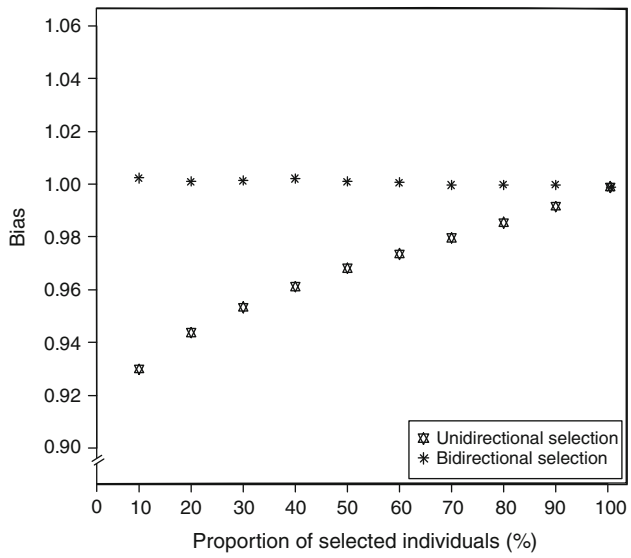

Fig. 3 Association between the bias to predict genomic breeding values for grain yield and different proportions of selected individuals of the total population applying (1) unidirectional selection and (2) bidirectional selection

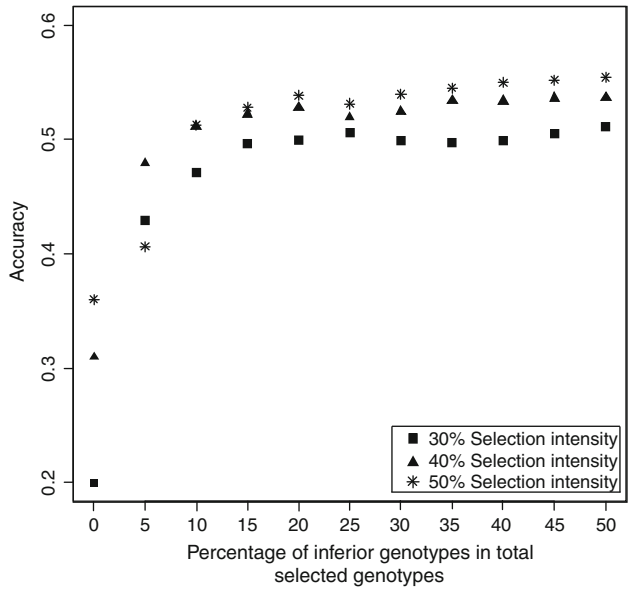

Fig. 4 Prediction accuracy of genomic breeding values for grain yield for unbalanced bidirectional selection with varying proportions of inferior versus superior genotypes for three selection intensities

breeding value by up to 8 %. With decreasing selection intensities this bias was monotonically decreasing.

For unbalanced bidirectional selection with a proportion of 10 % of inferior genotypes, prediction accuracy of genomic breeding values amounted to 81 % ($r_{GS} = 0.47$) of the accuracy reached with the full data set (Fig. 4). Selection intensity of 30 % resulted in substantially lower accuracies compared to selection intensity of 40 and 50 %. Moreover, with unbalanced bidirectional selection, estimated genomic breeding values underestimated the observed breeding values by up to 4 % (Fig. 5). As the proportion of inferior genotypes in training population reached 30 %, the observed bias was negligible.

Promising families can efficiently be pre-selected based on midparent performance and, therefore, family sizes often increase in applied plant breeding programs with increasing midparent performance. We investigated the

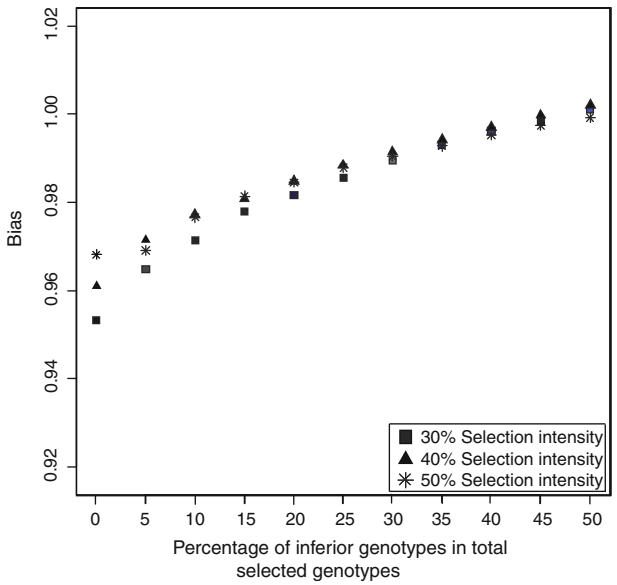

Fig. 5 Bias in prediction of genomic breeding values for grain yield for unbalanced bidirectional selection with varying proportions of inferior versus superior genotypes with three selection intensities

Table 1 Decrease in prediction accuracy (Δ) for two scenarios of individuals selection within families according to their average grain yield performance (A-GY in Mg ha^{-1}) (N_1 and N_2) versus balanced sampling of an equal number of individuals within families

Family	A-GY	N_1	Δ	N_2	Δ
Pop-A × B	13.69	50	−0.04	20	−0.11
Pop-A × C	13.89	90	0.01	80	−0.05
Pop-A × D	13.82	70	0.00	50	−0.05
Pop-B × C	13.88	80	0.02	65	0.03
Pop-B × D	13.78	60	−0.04	35	−0.08
Pop-C × D	14.02	100	0.04	95	0.09
Total	13.84	450	0.00	345	−0.03

influence of this unbalanced sampling and observed a negligible change in prediction accuracy of genomic breeding values for balanced versus unbalanced sampling when total population size amounted to 450 individuals (Table 1). Nevertheless, with a total population size of 345 individuals unbalanced sampling led to a reduction in prediction accuracy of genomic breeding values.

Discussion

Using routinely generated phenotypic data of applied plant breeding programs represents a cost-effective strategy to implement genomic selection. Nevertheless, breeding populations are based on multi-stage selection, and therefore, individuals phenotyped at a large number of environments often belong to highly selected populations. Truncation selection leads to a severe bias in the genetic makeup of the training population, which could have a strong impact on the accuracy of genomic selection. This stimulated us to investigate the influence of selection in the training population on the accuracy of genomic selection using experimental data from an elite maize breeding program.

Bias and prediction accuracy for unidirectional selection

We observed an underestimation of genomic breeding values of up to 8 % for scenarios where unidirectional selection was applied with high selection intensities (Fig. 3). The bias is relevant if estimated genomic breeding values are compared to phenotypic data of other varieties. This is for instance relevant for maize breeding programs based on doubled haploids, where uniform seed production is a major concern for the first cycles of selection. In this scenario, performance of lines with low amount of seeds in early selection stages can only be predicted using genomic breeding values while those with sufficient seeds are phenotypically evaluated. In such a situation bias will severely hamper a fair comparison of those lines where only genomic predicted breeding values are available versus lines which were also evaluated in field trials.

Because of limited economic resources, plant breeders often focus on the most promising crosses. One frequently used criterion for the choice of promising crosses is the average performance of the parents, which is a good predictor of the mean performance of a cross (e.g., Melchinger et al. 1998; Utz et al. 2001; Miedaner et al. 2006). We mimicked unbalanced population size based on the expected average testcross performance and observed a reduction in the accuracy to predict genomic breeding values only for scenarios with very unbalanced population sizes (Table 1). This is in line with results of the study of Zhao et al. (2012b) who also reported a low impact of unbalanced population sizes on the accuracy to predict genomic breeding values for grain yield in maize. Consequently, unbalanced population sizes due to selection based on average performance of the parents do not substantially hamper the use of routinely generated phenotypic data for genomic selection.

Estimating marker effects in populations where unidirectional selection was applied with high selection intensity led to substantially lower accuracies to predict genomic breeding values compared to the situation where marker effects were estimated in unselected populations of the same size (Fig. 1). We observed the same trend also on the level of the marker effects (Supplementary Fig. S1). Our findings can be explained by a substantial shift in allele

frequencies due to selection and reduced genetic variance in selected training populations. On the other hand, accuracy to predict genomic breeding values was further decreased if unidirectional selection was applied not only in the estimation set, but also in the prediction set (Fig. 2). This can be due to a high selection pressure within the prediction set causing a substantial decrease in the genetic variance but with no impact on the variance of the error prediction.

After an evaluation of the per se performance of inbred lines, first tests for general combining ability (GCA) are often conducted in commercial maize breeding programs in a limited set of around 3–4 environments (Longin et al. 2007). As traits measured for line per se performance are not tightly correlated with testcross performance for grain yield (e.g., Mihaljevic et al. 2005), populations evaluated in the early GCA tests might be considered as nearly unselected. We investigated a scenario reflecting the use of phenotypic data of the early GCA test where marker effects were estimated in plants phenotyped at low intensity (three locations) to predict genotypic values of plants phenotyped with high intensity (10 locations) using fivefold cross-validation and observed an accuracy of $r_{GS} = 0.37$ (Supplementary Fig. S2). This accuracy is substantially higher compared to scenarios of unidirectional selection (Fig. 1) with high selection intensity and phenotypic evaluation at 10 locations. Our finding can be interpreted as an indicator that use of unselected early GCA test data should be preferred compared to using unidirectional selected germplasm of later cycles of selection for estimating the marker effects.

Prediction accuracy and bias based on bidirectional selection

A cost-effective strategy to estimate marker effects with high precision could be to assay also genotypes with low yield potential besides the high-performing individuals in later selection cycles. Our results clearly underlined the high potential of this bidirectional selection with only moderate decrease in accuracies to predict genomic breeding values compared to the full data set (Fig. 1). The bias in estimating genomic breeding values was also negligible, which is a further advantage of bidirectional selection. As a constant update of the prediction model required for genomic selection (for review see Albrecht et al. 2011) ideally conducted with genotypic values determined with high heritability, bidirectional selection is a promising approach to implement genomic selection in applied breeding programs.

Our finding of a negligible bias for bidirectional compared to unidirectional selection (Fig. 3) is in contrast to the results of Lander and Botstein (1989) and Melchinger et al. (2012) where they reported an overestimation of QTL effects with bidirectional selection. This difference can be explained by the assumption of fixed QTL effects applying multiple linear regression QTL mapping approaches as opposed to random marker effects for genomic selection based on random regression best linear unbiased prediction. Assuming random marker effects leads to shrinkage of the marker effects towards the mean, for bidirectional selection the mean does not change resulting in marker effects estimated with a negligible bias. In contrast, mean changes for unidirectional selection and consequently shrinkage of marker effects causes a bias in genomic selection.

With the aim to further decrease the need to test low-performing individuals at many locations, we studied whether accuracy can be substantially increased compared to random selection if only a small proportion of inferior genotypes are used in bidirectional selection. Our findings clearly suggested that even with a proportion of 10–15 % of inferior genotypes high accuracies to predict genomic breeding values can be reached (Fig. 4). The bias in this scenario was also negligible (Fig. 5). Consequently, our findings showed that an update of estimation of marker effects can be cost-efficiently realized in the framework of routine breeding programs by implementing bidirectional selection with 10–15 % of inferior genotypes in the phenotypic evaluation of breeding germplasm in later cycles of selection.

Conclusions and outlook

Our findings are based on six biparental populations each consisting of at least 104 individuals. Therefore, the results on impact of selection on accuracy and bias for genomic selection are relevant for scenarios focusing on the genetic variation within biparental populations. The genetic variation within biparental population is of utmost importance, because across population variation can be predicted very precisely based on GCA effects of the parents (Melchinger et al. 1998). Precise estimates for GCA effects are commonly available for second cycle maize breeding programs. The six populations were established using a restricted number of four parental lines, which questions the relevance for a situation of a more diverse population as reported for instance by Albrecht et al. (2011). Unbalanced sampling in our study revealed that accuracy still amounted to 0.35 in small populations consisting of only 20 individuals (Table 1). This can be interpreted as an indicator that besides linkage information of within-biparental population, linkage disequilibrium across the six biparental populations is also exploited, and, therefore, our results are also relevant for genetically broader populations. Nevertheless, it is of high interest to validate our findings with populations established using a larger number of parents.

The results of our re-sampling procedure revealed a substantial loss in the accuracy to predict genomic breeding values in populations with unidirectional selection. Consequently, estimation of marker effects using routinely generated phenotypic data of applied plant breeding programs in advanced cycles of selection cannot be recommended. In contrast, use of populations with bidirectional selection is an interesting alternative for a cost-efficient implementation of genomic selection. Using an unbalanced bidirectional selection strategy based on only a small proportion of inferior genotypes further improves the efficiency to implement genomic selection in applied plant breeding programs.

In the context of linkage mapping Bayesian approaches have been suggested for QTL detection from single-tail sample of phenotype distributions (e.g., Sillanpää and Hoti 2007). It is of utmost interest to extend these models for genomic selection with the aim to estimate marker effects in unidirectional selected breeding populations. This need is apparent considering the drastic reduction in accuracy of genomic selection when estimating marker effects in unidirectional selected populations.

Acknowledgments This research was conducted within the Biometric and Bioinformatic Tools for Genomics based Plant Breeding project supported by the German Federal Ministry of Education and Research (BMBF) within the framework of GABI–FUTURE initiative. Y. Zhao and M. Gowda were supported by BMBF within the HYWHEAT project (Grant ID: FKZ0315945D).

References

Albrecht T, Wimmer V, Auinger H-J, Erbe M, Knaak C, Ouzunova M, Simianer H, Schön CC (2011) Genome-based prediction of testcross values in maize. Theor Appl Genet 123:339–350

Crossa J, Campos G, Pérez P, Gianola D, Burgueño J, Araus JL, Makumbi D, Singh RP, Dreisigacker S, Yan J, Arief V, Banziger M, Braun H-J (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186:713–724

Crossa J, Pérez P, Campos G, Mahuku G, Dreisigacker S, Magorokosho C (2011) Genomic selection and prediction in plant breeding. J Crop Improv 25:239–261

Dekkers JCM (2007) Prediction of response to marker-assisted and genomic selection using selection index theory. J Animal Breed Genet 124:331–341

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: genomic selection in dairy cattle: progress and challenges. J Dairy Sci 92:433–443

Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010) Plant breeding with genomic selection: gain per unit time and cost. Crop Sci 50:1681–1690

Henderson CR (1984) Application of linear models in animal breeding. University of Guelph, Guelph

Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Brief Funct Genomics 9:166–177

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Lebowitz RL, Soller M, Beckmann JS (1987) Trait-based analysis for the detection of linkage between marker loci and quantitative trait loci in cross between inbred lines. Theor Appl Genet 73:556–562

Longin CFH, Utz HF, Reif JC, Schipprack W, Melchinger AE (2006) Hybrid maize breeding with doubled haploids: I. One-stage versus two-stage selection for testcross performance. Theor Appl Genet 112:903–912

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

Lorenzana R, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor Appl Genet 120:151–161

Luan T, Wooliams JA, Lien S, Kent M, Svendsen M, Meuwissen THE (2009) The accuracy of genomic selection in Norwegian red cattle assessed by cross validation. Genetics 183:1119–1126

Melchinger AE, Gumber RK, Leipert RB, Vuylsteke M, Kuiper M (1998) Prediction of test cross means and variances among F_3 progenies of F_1 crosses from test cross means and genetic distances of their parents in maize. Theor Appl Genet 96:503–512

Melchinger AE, Orsini E, Schön CC (2012) QTL mapping under truncation selection in homozygous lines derived from biparental crosses. Theor Appl Genet 124:543–553

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

Miedaner T, Schneider B, Oettler G (2006) Means and variances for *Fusarium* head blight resistance of F2-derived bulks from winter triticale and winter wheat crosses. Euphytica 152:405–411

Mihaljevic R, Schön CC, Utz HF, Melchinger AE (2005) Correlations and QTL correspondence between line per se and testcross performance for agronomic traits in four populations of European maize. Crop Sci 45:114–122

Navabi A, Mather DE, Bernier J, Spaner DM, Atlin AN (2009) QTL detection with bidirectional and unidirectional selective genotyping: marker-based and trait-based analyses. Theor Appl Genet 118:347–358

Sillanpää MJ, Hoti F (2007) Mapping quantitative trait loci from a single-tail sample of the phenotype distribution including survival data. Genetics 177:2361–2377

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

Sun Y, Wang J, Crouch JH, Xu Y (2010) Efficiency of selective genotyping for genetic analysis of complex traits and potential applications in crop improvement. Mol Breed 26:451–493

Utz HF, Bohn M, Melchinger AE (2001) Predicting progeny means and variances of winter wheat crosses from phenotypic values of their parents. Crop Sci 41:1470–1478

Zhao Y, Gowda M, Liu W, Würschum T, Maurer HP, Longin CFH, Ranc N, Reif JC (2012a) Accuracy of genomic selection in European maize elite breeding populations. Theor Appl Genet 124:769–776

Zhao Y, Gowda M, Liu W, Würschum T, Maurer HP, Longin CFH, Ranc N, Reif JC (2012b) Choice of shrinkage parameter and prediction of genomic breeding values in maize elite breeding populations. Theor Appl Genet (in review)