Available online at www.sciencedirect.com

ScienceDirect



Use of genomic selection and breeding simulation in cross prediction for improvement of yield and quality in wheat (*Triticum aestivum* L.)

Ji Yao, Dehui Zhao, Xinmin Chen, Yong Zhang*, Jiankang Wang*

The National Key Facility for Crop Gene Resources and Genetic Improvement, National Wheat Improvement Center, Institute of Crop Sciences, Academy of Agricultural Sciences, Beijing 100081, China

ARTICLE INFO

Article history:

Received 19 March 2018

Received in revised form 5 May 2018

Accepted 10 May 2018

Available online 4 June 2018

Keywords:

Breeding simulation

Cross prediction

Genomic selection

Parental selection

Usefulness

ABSTRACT

In wheat breeding, it is a difficult task to select the most suitable parents for making crosses aimed at the improvement of both grain yield and grain quality. By quantitative genetics theory, the best cross should have high progeny mean and large genetic variance, and ideally yield and quality should be less negatively or positively correlated. Usefulness is built on population mean and genetic variance, which can be used to select the best crosses or populations to achieve the breeding objective. In this study, we first compared five models (RR-BLUP, Bayes A, Bayes B, Bayes ridge regression, and Bayes LASSO) for genomic selection (GS) with respect to prediction of usefulness of a biparental cross and two criteria for parental selection, using simulation. The two parental selection criteria were usefulness and midparent genomic estimated breeding value (GEBV). Marginal differences were observed among GS models. Parental selection with usefulness resulted in higher genetic gain than midparent GEBV. In a population of 57 wheat fixed lines genotyped with 7588 selected markers, usefulness of each biparental cross was calculated to evaluate the cross performance, a key target of breeding programs aimed at developing pure lines. It was observed that progeny mean was a major determinant of usefulness, but the usefulness ratings of quality traits were more influenced by their genetic variances in the progeny population. Near-zero or positive correlations between yield and major quality traits were found in some crosses, although they were negatively correlated in the population of parents. A selection index incorporating yield, extensibility, and maximum resistance was formed as a new trait and its usefulness for selecting the crosses with the best potential to improve yield and quality simultaneously was calculated. It was shown that applying the selection index improved both yield and quality while retaining more genetic variance in the selected progenies than the individual trait selection. It was concluded that combining genomic selection with simulation allows the prediction of cross performance in simulated progenies and thereby identifies candidate parents before crosses are made in the field for pure-line breeding programs.

© 2018 Crop Science Society of China and Institute of Crop Science, CAAS. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

* Corresponding authors.

E-mail addresses: zhangyong05@caas.cn, (Y. Zhang), wangjiankang@caas.cn. (J. Wang).

Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.

1. Introduction

Breeding techniques have evolved rapidly in the last two decades. Historically, breeders used only phenotypic traits to make key progeny advance decisions, but the efficiency of this practice is reduced by genotype-by-environment ($G \times E$) interaction, errors of measurement, and limitations of phenotyping methods. For example, wheat quality traits are normally tested in later breeding generations owing to (1) the lack of sufficient seeds in early generations and (2) the expense and time required by quality testing in the lab. For this reason, marker-assisted selection (MAS) was introduced in early generations to enable selection on such traits and increase selection efficiency [1–3]. MAS involves two major steps. First, a biparental population is developed to identify quantitative trait loci (QTL) for traits of interest; then, markers linked to the identified QTL are used in selection [3, 4]. Nowadays, QTL mapping for detecting genes for breeding traits in major crops is routine [3], and functional markers have been developed using cloned-gene information [5]. However, the practical use of MAS in breeding programs is limited for various reasons, including genetic background interactions of identified QTL, large $G \times E$ interaction effects, and low power to detect minor QTL [3, 4].

To overcome the disadvantages of MAS, genomic selection (GS) using genomewide markers to estimate trait breeding values of individuals was proposed [6, 7]. GS involves establishing a prediction model in a training population using both phenotypic and genotypic data, and predicting trait genomic estimated breeding value (GEBV) for individuals with only genotypic information [6]. Compared to MAS and prediction methods based only on pedigree information, GS can often result in higher genetic gain and better prediction accuracy in recurrent selection and hybrid performance prediction [8–12]. GS has been intensively studied in major crops [8, 13, 14] and for key traits such as yield, disease resistance, and quality [15, 16].

In pure-line (or, equivalently, fixed-line) breeding such as in wheat, one of the critical challenges for breeders is to identify suitable parents for creating genetic variation to maximize selection response in subsequent breeding cycles [17, 18]. As in hybrid breeding programs, the number of possible crosses is usually far greater than the number that breeders can make in the field. Desirable crosses should have high means and large genetic variances in progeny population, affording breeders higher chances to select superior progenies. To evaluate crosses among a large number of parental lines by accounting for both population mean and genetic variance, the usefulness concept was proposed [19]. By definition, usefulness is a function of progeny mean, genetic variance, and heritability. A modification of usefulness, called superior progeny value, was proposed by Zhong and Jannink [20]. It focuses on mean and genetic variance of the progenies by ignoring heritability, which is equivalent to usefulness when heritability is equal to 1.

Progeny mean of a biparental cross can be well predicted by the midparent value from the two parents. However, predicting the genetic variance of in a progeny population is not easy. Phenotypic distance, pedigree-based parentage

distance, and genetic distance have been proposed to predict the genetic variance of progeny [19, 21]. One problem with these predictors is that they provide the same variance estimate for all traits, a practice that does not reflect reality [22]. Some approaches have used GS to predict progeny genetic variance [22–26]. However, different GS models performed inconsistently well on different traits [27–29]. Few studies have been conducted on the prediction of usefulness under different GS models and the consequent genetic gain after progeny selection.

In wheat breeding, although grain yield is the primary focus, breeders want to improve several quality traits at the same time. Normally, quality traits such as grain protein content are negatively correlated with yield and highly influenced by $G \times E$ interaction. Breeding programs would be more efficient if the best crosses and their parents could be predicted and identified. Parental-selection strategy, based on the predicted progeny variance either by assuming linkage equilibrium or by accounting for linkage disequilibrium between parents, has been evaluated for improving grain yield, grain protein, loaf volume, and mixing time individually [24]. There has been no report on parental selection to improve both yield and grain quality. Few studies have focused on other important quality traits such as extensibility, extension area, maximum resistance, and stability time, which reflect protein quality. Two previous studies showed that the application of GS rapidly reduced genetic diversity [30, 31]. It is not clear how to maintain genetic diversity when using GS for cross prediction.

The objectives of this study were (1) to evaluate the ability of different GS models to predict usefulness under different trait architectures, (2) to evaluate genetic gain using usefulness in parental selection, in comparison with midparent GEBV, and (3) to evaluate different parental selection criteria to arrive at the parents best able to improve yield and grain quality together while maintaining genetic diversity. A simulation experiment was conducted for objectives (1) and (2) and a population of 57 wheat fixed lines was used for objective (3).

2. Materials and methods

2.1. Plant materials, field experiment, and phenotyping

A wheat population was constructed from 48 high quality and three high-yielding cultivars or advanced lines bred in China and six high-quality cultivars from the U.S and Australia. Names and origins of the 57 wheat fixed lines are presented in Supplementary Table S1. For phenotyping, the 57 wheat fixed lines were evaluated in two locations: Jining, Shandong province, and Gaoyi, Hebei province in two seasons from 2013 to 2015, using randomized complete block designs. Each trial had two replications, each of 6 rows 4 m in length and 0.2 m in width for each plot. Yield and thousand-kernel weight were evaluated in both seasons. Other yield components and agronomic traits including spike number, kernel number, heading date, and plant height were evaluated in Jining and Gaoyi only in the 2013–2014 season [32].

Grain quality parameters were evaluated in the 2013–2014 season. Firmness was assessed by the single-kernel characterization system (SKCS 4100 Perten Instruments AB, Sweden). Grain protein content was assessed by near-infrared spectroscopy (Foss 1241, Sweden) and reported on a 14% grain moisture basis. Grain samples were tempered and milled with a Buhler Quadrumat mill (MLU 220, Uzvil, Switzerland) using method AACCC 26-21A. Dough quality parameters were assessed by mixograph, farinograph, and extensograph using methods AACCC54-40A, 54-21, and 54-10, respectively. The direct extraction rate was set at 65%.

The 11 traits studied included yield, thousand-kernel weight (TKW), spike number (SN), kernel number (KN), heading date (HD), plant height (PH), protein content (PC), extensibility (EXT), extension area (EA), maximum resistance (MR), and stability time (ST).

2.2. DNA extraction and genotyping

Wheat leaf tissues were sampled and DNA was extracted by the modified CTAB method [33]. Genotypic data for cultivars was obtained with the 90 K iSelect SNP chip [34] featuring 81,587 SNPs. SNP genotyping was performed with Genome Studio. A total of 38,833 SNPs remained for the 57 genotypes after one round of quality control. Further quality control measures included the removal of monomorphic markers, SNPs with minor-allele frequencies lower than 0.05, and SNPs with heterozygosity greater than 20%. After SNP filtering, 7588 SNPs with known genetic map positions on the map of Wang et al. [34] were selected. Distribution of the selected SNPs on the 21 wheat chromosomes is presented in Fig. 1. SNP genotype at each locus was coded as (1, 0, -1), where 1 indicated the homozygous genotype of the major allele, -1 the homozygous genotype of the minor allele, and 0 the heterozygous genotype. The 57 wheat lines together with their phenotypic traits and SNP genotyping data formed the training population to be used for evaluating prediction accuracy of GS models and parental selection strategies for improving yield and quality.

2.3. Phenotypic data analysis

For each phenotypic trait, the following linear model was used to conduct analysis of variance (ANOVA) and test the significance of $G \times E$ interactions [35].

$$Y_{ijk} = \mu + G_i + E_j + R_{k(j)} + GE_{ij} + e_{ijk} \quad (1)$$

In Eq. (1), Y_{ijk} represents a phenotypic observation, μ the overall mean, G_i the effect of genotype i , E_j the effect of location–year combination (environment) j , $R_{k(j)}$ the effect of replication k nested in environment j , GE_{ij} the $G \times E$ interaction between genotype i and environment j , and e_{ijk} the residual effect associated with genotype i in environment j and replication k . Trait heritability per plot and per mean was calculated by Eq. (2).

$$H_{\text{per plot}}^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{ge}^2 + \sigma_{\text{err}}^2}, H_{\text{per mean}}^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{m} + \frac{\sigma_{\text{err}}^2}{mr}} \quad (2)$$

In Eq. (2), σ_g^2 represents genotypic variance, σ_{ge}^2 the variance of $G \times E$ interaction, σ_{err}^2 the residual variance, m the number of location–year combinations, and r the number of replications in each environment [35].

Best linear unbiased predictions (BLUPs) were obtained for each trait using the model described in Eq. (1) except that the genotypic effect was defined as random, assuming $G \sim N(0, \sigma_g^2)$. ANOVA was performed with QTL IciMapping software [36] and BLUPs were calculated with the *lme4* package in R [37]. Pearson correlations among traits were calculated using BLUPs of phenotypic traits.

A selection index incorporating yield, maximum resistance, and stability time was constructed for improving the three traits simultaneously. Economic weights were set at 0.6, 0.2, and 0.2 for yield, maximum resistance, and extensibility, respectively. The index was calculated as $0.6 \times \text{yield} + 0.2 \times \text{maximum resistance} + 0.2 \times \text{extensibility}$. This index reflected two major wheat breeding objectives, grain yield and industrial quality. Yield was given higher weight than the

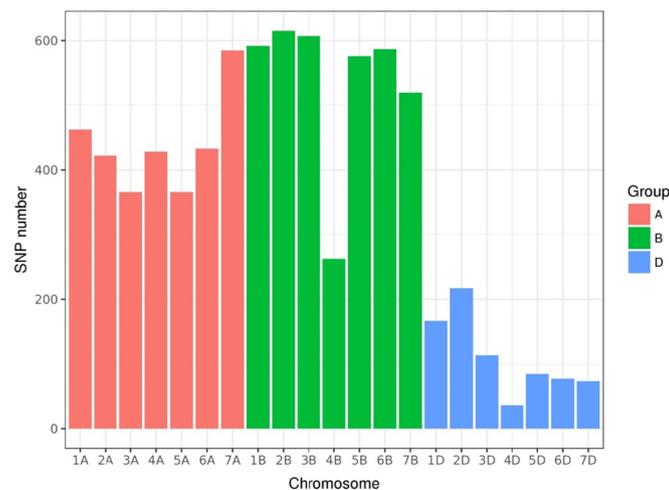


Fig. 1 – Distribution of 7588 selected SNPs on the 21 wheat chromosomes.

other two quality traits, as it is more important than quality in wheat breeding. The other two quality traits were considered equally important. Predicted genotypic values (BLUPs), which account for genetic correlations between traits [38], were used to calculate the index.

2.4. Simulation experiment

A simulation study was conducted to assess the prediction accuracy and genetic gain afforded by different GS models, using 57 parents. The genotyping data of the parents at the 7588 SNPs was the same as that of the training population. To assess the impact of genetic architectures of different traits, the number of QTL was set at three levels: 30, 50, and 100. QTL effects were generated from the geometric distribution [1] as shown in Eq. (3).

$$QTL_k = \left(\frac{L-1}{L+1} \right)^k \quad (3)$$

In Eq. (3), QTL_k is additive effect of the k^{th} QTL effect and L is the number of QTL. SNPs corresponding to the three QTL numbers were randomly sampled from the 7588 SNPs and used as QTL in each genetic architecture. The remaining SNPs were used as markers in prediction and simulation. Favorability or unfavorability of alleles at loci in the parents depended on their SNP genotypes.

The phenotype of each of the 57 parents was simulated according to the linear model $y_i = l_i + e_i$, where y_i is the simulated phenotypic value of parent i , l_i is the genotypic value obtained by summarizing all QTL effects in parent i , and e_i is the random error of parent i from the normal distribution $e_i \sim N(0, \sigma_e^2)$. Three levels of trait heritability: 0.3, 0.5, and 0.8, were used to represent low, medium, and high heritability, respectively. Error variance was calculated as $\sigma_e^2 = \frac{\sigma_g^2}{H^2} - \sigma_g^2$ for each heritability level. Phenotypes were simulated for a total number of nine genetic architectures and repeated 10 times, resulting in 90 simulated data sets.

The usefulness (U) of a cross or the progeny population from the cross is defined as $U = \mu + ih\sigma_g$, where μ is mean of the progeny population derived from the cross, i is the standardized selection intensity based on the selected proportion, h is the square root of the trait heritability, and σ_g is the standard deviation of genetic variance of the progeny population. In this study, we followed Zhong and Jannink [20] by ignoring h in the formula. The U of each biparental cross was estimated by the following steps. First, five GS models were applied to each simulated phenotypic data set to obtain marker effects. The ridge regression BLUP model and four Bayesian models (Bayes A, Bayes B, Bayes ridge regression, and Bayes LASSO) were implemented with the R packages *rrBLUP* and *BGLR* [39, 40]. Parameters of all Bayesian models were set according to Perez and de los Campos [40]. The iteration number of the Bayesian models was set at 12,000 and the first 3000 iterations were discarded as burn-in [40].

Second, a half-diallel mating design was used for the 57 parents, resulting in 1596 biparental crosses. For each cross, 5000 recombinant inbred lines (RIL) were simulated with the *R/qtl* package [41] using the genetic map of Wang et al. [34]. GEBVs of the 5000 RILs were calculated from the estimated

marker effects and simulated RIL genotypes. Finally, a progeny mean was obtained by averaging the GEBVs of the 5000 RILs and genetic variances of the progeny were calculated accordingly. The estimated progeny mean and genetic variance were used to calculate usefulness at three proportions of selection: 0.10, 0.05, and 0.01. The true usefulness and genetic variance of each cross were calculated from QTL information in the corresponding genetic architecture.

Prediction accuracy for progeny mean, progeny genetic variance, and usefulness were calculated as the Pearson coefficient of correlation between true and predicted values of the 1596 crosses. Correlation between midparent GEBV and true usefulness was also calculated for evaluating whether midparent GEBV was adequate for predicting usefulness. For each simulated data set, genetic gains from different selection proportions and prediction models were calculated as follows. (1) The 100 crosses with highest usefulness were selected. For each selected cross, the simulated RILs were selected according to the corresponding selection proportion. (2) Mean of the true genotypic values of the selected progeny was denoted as TVP, and mean of genotypic values in each simulated data set was denoted as TVS. Genetic gain was calculated as $(TVP - TVS) / \sigma_{TVS}$, where σ_{TVS} is the standard deviation of TVS. For comparison with the selection by usefulness, midparent GEBV was used to select the top 100 crosses and genetic gain was calculated similarly.

2.5. Parental selection schemes in training population

Fivefold cross validation was performed to assess the prediction accuracies of the GS models for all traits in the training population. Missing marker data were imputed using the EM algorithm implemented in the *A.mat* function in the *rrBLUP* R package [39]. The full training data set was divided into five subsets of equal size. For each validation, four of the five subsets were used as a training set and the remaining subset as a validation set. The cross validation was repeated 10 times. Genomic prediction accuracy was calculated as the correlation between $GEBV_{pre}$ and OBV, where $GEBV_{pre}$ was GEBV predicted by GS models, and OBV was BLUP obtained from the phenotypic analysis. The model with the highest prediction accuracy was intended to be used in subsequent analyses. However, no significant difference among models in usefulness prediction in the simulation experiment or the training population was found. The best crosses predicted by the five models were largely the same. In cross validation, Bayes B performed slightly better and was accordingly used for estimating marker effects and performing subsequent analyses in the training population.

Usefulness at a selection proportion of 0.05 was calculated for all traits in the training population, following the same approach as applied in the simulation experiment with all 7588 SNPs. Four schemes were considered to select parents most suitable for improving yield and quality: (1) select parents that produce high-yield usefulness crosses, then select 10% of their simulated progeny that have highest EXT, and finally select from these progeny the 50% with highest MR; (2) select parents producing high-MR usefulness crosses, then select 10% of their simulated progeny with highest EXT, and finally select from these the 50% with highest yield; (3)

select parents that producing high-EXT usefulness crosses, then select the 10% of their simulated progeny with highest MR, and finally select from these the 50% with highest yield; (4) using the selection index described above in the phenotype data analysis section, select parents based on the usefulness of the index in their derived crosses, and then select 5% of their simulated progeny with highest index value. For each scheme, the top 100 crosses were selected and the mean and genetic variance of the selected progenies was calculated accordingly. The selection response of each trait for each scheme was obtained by subtracting the mean of trait BLUPs in the training population from the mean of the selected progeny.

3. Results

3.1. Analysis of variance and correlation between phenotypic traits

The results of ANOVA for the 11 phenotypic traits are summarized in Table 1. All traits showed highly significant differences in genotypic effect among the 57 wheat lines. $G \times E$ interaction was significant for all traits except spike number. Coefficient of variation (CV) ranged from 0.07 to 0.58. Protein content had the smallest CV (0.07) and stability time had the largest (0.58). Trait heritability at the plot level ranged from 0.45 to 0.93. Heritability at the level of means across environments and replications ranged from 0.57 to 0.99 (Table 1). Heritability per mean is always higher than heritability per plot because less variance from $G \times E$ interaction and random error is included in the variance of phenotypic mean (Eq. (2)). Heading date had the lowest heritability at 0.57, owing to a large amount of missing data. Only one location was included in the ANOVA for heading date, due to the large amount of missing data in one other location.

Trait correlations are presented in Table 2. Correlations between quality traits ranged from 0.12 to 0.94. MR and ST were significantly correlated with each other, but not with PC and EXT. EXT had the highest negative correlation with yield. Yield and TKW both were negatively correlated with quality traits, but correlations between yield and MR and between yield and ST were close to zero and not significant. In contrast

to the other quality traits, MR and ST both showed positive but nonsignificant correlations with other yield components and agronomic traits except for plant height.

3.2. Accuracy of genomic selection models in the simulation experiment

Heritability was the main factor that influenced the accuracy of progeny variance prediction (Fig. 2). The accuracy ranged from 0.17 to 0.22 at the low heritability of 0.3, from 0.25 to 0.32 at the moderate heritability of 0.5, and from 0.37 to 0.42 at the high heritability of 0.8. The prediction accuracy was slightly affected by QTL number. The accuracy was increased by 0.04 from 30 to 100 QTL at the heritability of 0.3, increased by 0.05 from 30 to 100 QTL at the heritability of 0.5, and increased by 0.03 from 30 to 100 QTL at the heritability of 0.8. A minor difference in prediction accuracy was observed among different GS models (Fig. 2). Bayesian LASSO performed slightly better for QTL numbers 30 and 50, but slightly underperformed for QTL number 100. This may be due to that Bayesian LASSO can estimate a few large QTL effects accurately.

The prediction accuracy of usefulness obtained in the simulation experiment showed a trend similar to that of progeny genetic variance (Figs. 2 and 3). Heritability exerted a strong influence on the prediction accuracy of usefulness. Accuracy ranged from 0.25 to 0.43 at the low heritability of 0.30, from 0.36 to 0.53 at the moderate heritability of 0.50, and from 0.40 to 0.61 at the high heritability of 0.8. With the decrease in selection proportion from 0.1 to 0.01, prediction accuracy decreased from 0.48 to 0.41 on average. A similar trend was observed across different heritabilities and QTL numbers. Because different GS models provided similar predictions of midparent GEBV, only midparent GEBV from π BLUP is presented here. Prediction accuracies obtained from midparent GEBV were consistently lower than those obtained from usefulness calculated by GS models (Fig. 3). When heritability increased from 0.3 to 0.8, on average accuracy from usefulness was 0.01 to 0.05 higher than that from midparent GEBV. In contrast, accuracy from usefulness increased from 0.02 to 0.04, compared with midparent GEBV. This trend was also observed with the increasing QTL number. GS models showed marginal differences in all combinations of QTL

Table 1 – Analysis of variance of yield, yield components, and agronomic and quality traits.

Trait	Mean \pm SD	V_G	V_{GE}	CV	H^2 (per plot)	H^2 (per mean)
Yield (kg ha ⁻¹)	7549.20 \pm 1182.45	3314.00 ^{***}	606.97 ^{***}	0.16	0.66	0.96
Thousand-kernel weight (g)	41.75 \pm 6.04	24.56 ^{***}	2.73 ^{***}	0.14	0.79	0.98
Spike number	40.29 \pm 7.85	41.96 ^{***}	2.89 ^{ns}	0.19	0.53	0.83
Kernel number	33.08 \pm 5.80	11.16 ^{***}	1.78 [*]	0.18	0.45	0.79
Heading date (d) ^a	180.98 \pm 3.39	27.76 ^{***}	NA	0.28	0.57	0.57
Plant height (cm)	77.69 \pm 9.90	89.96 ^{***}	3.30 ^{***}	0.13	0.91	0.99
Protein content (%)	11.40 \pm 0.82	0.53 ^{***}	0.02 ^{***}	0.07	0.88	0.98
Extension area (cm ²)	105.83 \pm 34.70	1148.51 ^{***}	28.29 ^{***}	0.33	0.93	0.99
Extensibility (mm)	169.64 \pm 20.93	354.47 ^{***}	11.08 ^{***}	0.12	0.86	0.97
Maximum resistance (BU)	481.71 \pm 154.50	23,100.02 ^{***}	532.58 ^{***}	0.32	0.93	0.99
Stability time (min)	14.18 \pm 8.27	56.16 ^{***}	10.65 ^{***}	0.58	0.78	0.98

CV, coefficient of variation; V_G , estimated variance of genotypic effects; V_{GE} , estimated variance of $G \times E$ interactions; NA, data not available. * and ***, significant at the 0.05 and 0.001 probability levels, respectively; ns, non-significant.

^a Heading date was analyzed only in Jining owing to large numbers of missing data in Gaoyi.

Table 2 – Pearson coefficients of correlation between traits in training population, calculated with BLUPs.

Trait	Yield	TKW	SN	KN	HD	PH	PC	EA	EXT	MR	ST
Yield		*	**	ns	ns	***	***	ns	***	ns	ns
TKW	0.33		***	ns	ns	ns	ns	***	***	**	***
SN	-0.40	-0.61		***	ns	**	*	**	***	ns	ns
KN	0.25	0.10	-0.68		ns	**	ns	ns	*	ns	ns
HD	0.24	-0.16	0.21	-0.07		ns	ns	ns	ns	ns	ns
PH	-0.59	-0.24	0.37	-0.35	0.05		***	ns	***	ns	ns
PC	-0.56	-0.15	0.32	-0.23	0.06	0.46		**	***	ns	ns
EA	-0.23	-0.49	0.40	-0.07	0.02	0.12	0.36		***	***	***
EXT	-0.68	-0.46	0.50	-0.27	-0.25	0.47	0.52	0.46		ns	ns
MR	-0.02	-0.37	0.25	0.02	0.09	-0.05	0.22	0.94	0.15		***
ST	-0.07	-0.43	0.20	0.07	0.13	-0.07	0.15	0.73	0.12	0.77	

TKW, thousand-kernel weight; SP, spike number; KN, kernel number; HD, heading date; PH, plant height; PC, protein content; EA, extension area; EXT, extensibility; MR, maximum resistance; ST, stability time; *, **, and ***, significant at the 0.05, 0.01, and 0.001 probability levels, respectively; ns, non-significant.

number, heritability, and selection proportion (Fig. 3). No single model performed best in all genetic architectures.

3.3. Genetic gains from genetic selection models and parental selection

Fig. 4 shows the genetic gains obtained from different GS models and selection proportions. Heritability had a major influence on genetic gain. Higher heritability resulted in higher gain across all models and parental selection methods. Genetic gains from usefulness were consistently higher than those from midparent GEBV across GS models and heritabilities. Gain from usefulness was 0.06 higher than that from midparent GEBV at a heritability of 0.30, 0.08 higher at a heritability of 0.50, and 0.13

higher at a heritability of 0.80. When selection proportion decreased, the difference in genetic gain between usefulness and midparent GEBV increased. Gain from usefulness was 0.01 to 0.03 higher for selection proportion 0.10, 0.01 to 0.05 higher for selected proportion 0.05, and 0.04 to 0.09 higher for selection proportion 0.01. QTL number also affected genetic gain at different heritabilities. The trend of the effect of QTL number was similar to those of heritability and selection proportion. For selection proportions 0.10, 0.05, and 0.01, the respective gains from usefulness were 0.01, 0.02, and 0.05 higher for QTL number 30, 0.02, 0.04, and 0.05 higher for QTL number 50, and 0.03, 0.04, and 0.07 higher for QTL number 100. Genetic gains for different GS models showed negligible differences across genetic architectures.

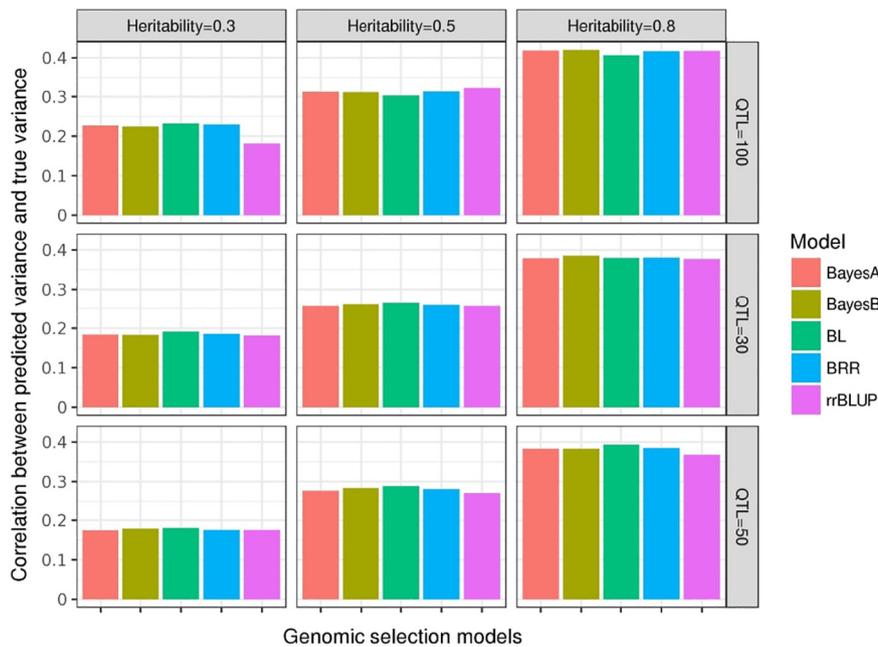


Fig. 2 – Prediction accuracy for progeny genetic variance among genomic selection models, genetic architectures, and selection proportions. BL, Bayes LASSO; BRR, Bayes ridge regression.

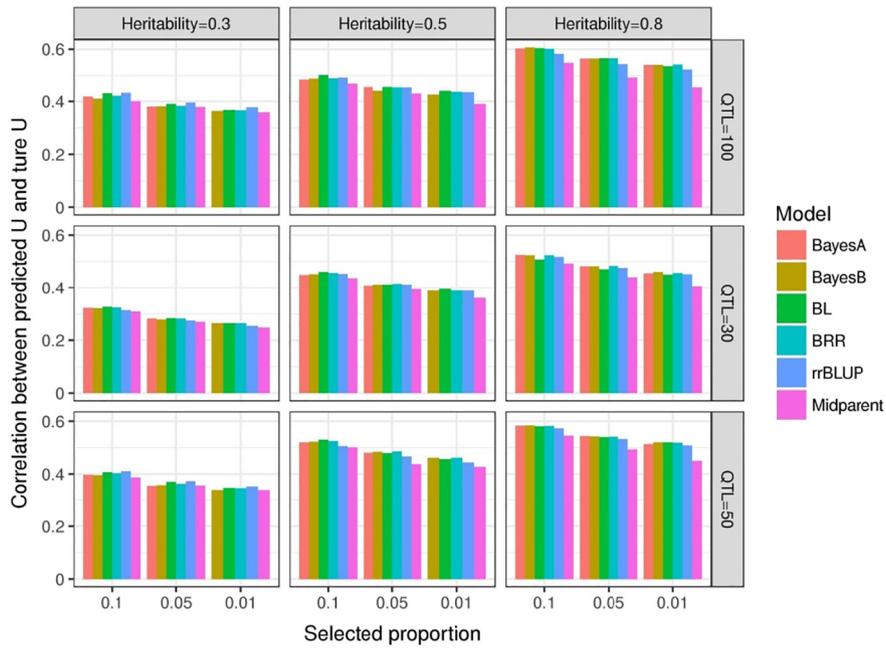


Fig. 3 – Prediction accuracy of usefulness among genomic selection models, genetic architectures, and selection proportions. BL, Bayes LASSO; BRR, Bayes ridge regression; Midparent, midparent GEBV; U, usefulness.

3.4. Cross prediction and parental selection in the training population

The accuracies of the five GS models for each trait in training population are presented in Fig. 5. MR showed the lowest prediction accuracy ($r = 0.340$ on average across models). Yield and plant height showed the highest prediction

accuracy ($r = 0.811$ and $r = 0.807$ on average across models). In general, traits with high heritability showed high accuracy, except for MR, EA, ST, and HD. Bayes B showed the highest prediction accuracy for EA, HD, MR, PH, and TKW, Bayes ridge regression showed the highest prediction accuracy for EXT, KN, and SN, and rrBLUP showed the highest prediction accuracy for ST, PC, Yield, and HD. Prediction accuracy from

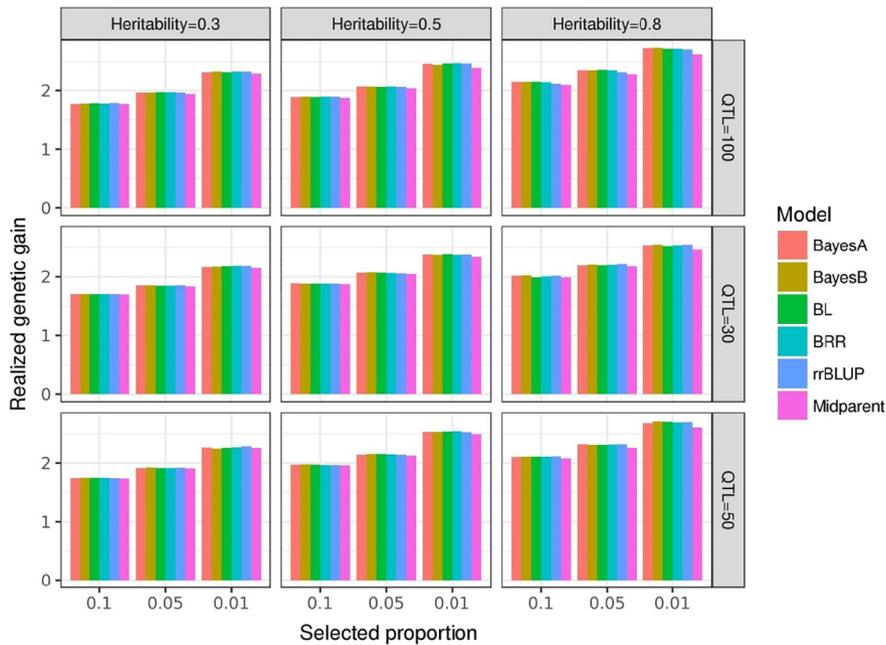


Fig. 4 – Genetic gains from different parental selection criteria and genomic selection models. BL, Bayes LASSO; BRR, Bayes ridge regression; Midparent, midparent GEBV.

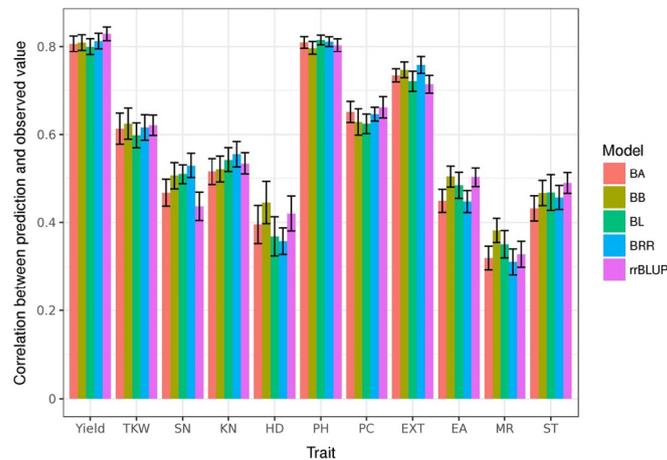


Fig. 5 – Prediction accuracy for different traits in the training population. TKW, thousand-kernel weight; SP, spike number; KN, kernel number; HD, heading date; PH, plant height; PC, protein content; EA, extension area; EXT, extensibility; MR, maximum resistance; ST, stability time; BL, Bayes LASSO; BRR, Bayes ridge regression.

cross validation differed among traits and models (Fig. 5), but simulation results showed little difference among models in predicting cross usefulness and genetic gain (Figs. 3 and 4). Bayes B gave the highest accuracy in predicting usefulness for most traits and was accordingly chosen for further analyses in the training population.

The triangular relationship shown in Fig. 6 was observed between progeny mean and genetic variance of simulated progeny for all traits. For yield, most crosses gave high progeny means and low variances (Fig. 6-A). In contrast, quality traits, for example MR, showed more complete triangles (Fig. 6-B). In Fig. 6, the top 100 crosses predicted by usefulness are highlighted in blue and located in the high progeny mean area. The top 100 crosses predicted by usefulness and midparent GEBV were largely common to all traits. The number of crosses common to the five quality traits ranged from 83 to 89, and for other traits from 85 to 92. In general, there were fewer common crosses between usefulness and midparent GEBV for quality traits.

Trait correlations changed from parental to progeny populations. As an example, Fig. 7 shows distribution of

correlations between two quality traits and yield. In the training population, the correlation was -0.02 between MR and yield, and -0.68 between EXT and yield (Table 2). However, the correlations ranged from -0.61 to 0.63 for MR (Fig. 7-A) with a mean of -0.008 , and from -0.88 to 0.27 for EXT (Fig. 7-B) with a mean of -0.51 . Thus, even though a quality trait showed a strong negative correlation with yield in training population, it was still possible to see a positive correlation in the progeny population derived from two parents selected from the training population.

Table 3 presents selection responses and variances of the selected progenies from four parental selection schemes. Yield, EXT and MR were assumed to be the three targeted traits. Scheme 1 generated the largest response for yield (892.50), but the responses for EXT and MR were negative (-15.19 and -48.73). For scheme 2, the response was near zero for yield (4.35), but positive for EXT and MR (5.65 and 163.59). The response for MR in scheme 2 was the largest among all schemes. In scheme 3, the response for yield was negative (-1029.15), but responses for EXT and MR were positive (25.97 and 3.29). The response for EXT in scheme 3 was the largest

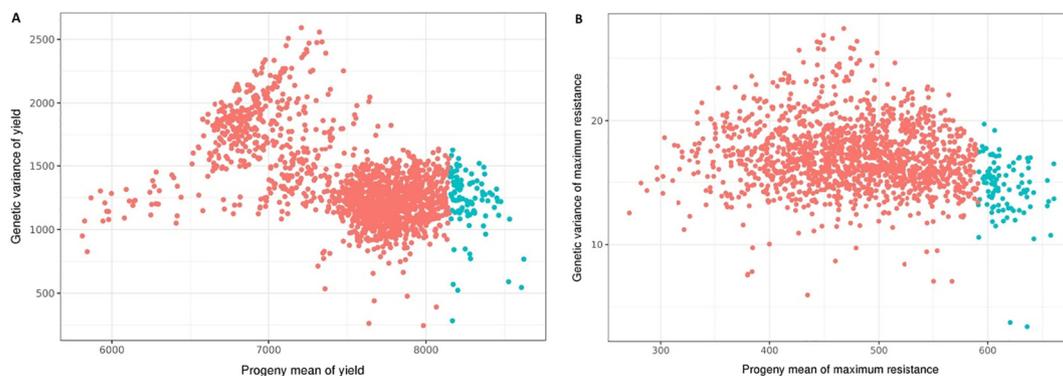


Fig. 6 – Relationship between midparent GEBV and variance in each biparental cross from the 57 fixed lines in the training population. The 100 crosses selected for highest trait usefulness are highlighted in blue. A for yield; B for maximum resistance.

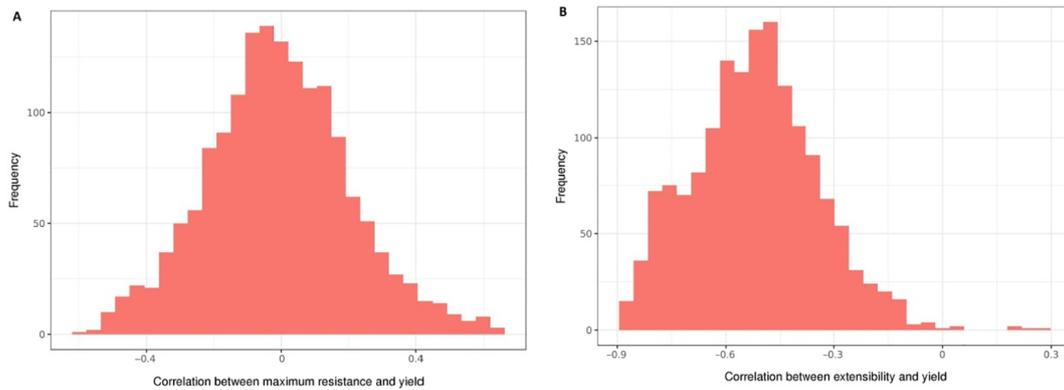


Fig. 7 – Frequency distribution of correlations between yield and maximum resistance (A), and between yield and extensibility (B) in the 1596 biparental crosses from 57 parents in the training population.

among all schemes. For scheme 4, the responses for yield and MR were both positive (643.20 and 76.86), and the response for EXT was negative (–3.70) but was higher than that for scheme 1. Responses for yield and MR from scheme 4 were the second largest among the four schemes. For the other quality traits, ST and EA, small positive responses (1.63 and 12.12) were obtained from scheme 4 owing to the positive correlation among quality traits (Table 2). For other yield components and agronomic traits, near-zero responses were observed except for plant height. Plant height showed a negative response in scheme 4 (–3.76), which would be desirable when breeders want to reduce the trait.

All parental selection schemes reduced genetic variance significantly. In general, the larger the response observed for a trait, the greater was the observed reduction in genetic variance (Table 3). Genetic variance in selected progeny was the smallest for yield in scheme 1 (647.55) but was the largest in scheme 3 (8204.10). Because crosses in scheme 2 were selected by usefulness on MR, genetic variance of MR in the selected progeny (303.88) was smallest. Similarly, scheme 3 gave the smallest variance for EXT (15.87). Scheme 4 gave the

largest genetic variance in selected progenies for MR and EXT (4157.07 and 36.64), and second largest genetic variance for yield (206.09). For traits not included in the index, scheme 4 gave the highest genetic variances for EA and ST (231.73 and 8.92), second highest for PC (0.10), and second lowest for TKW and SN (3.66 and 2.94).

Scheme 4 was the only one that gave positive responses for yield and quality, and large genetic variances in the selected progeny. Accordingly, it was used to select the parents for yield and quality improvement. Table 4 presents the 10 crosses with the highest usefulness on selection index, together with GEBVs of the two parents, midparent GEBV, and progeny mean at a selection proportion of 0.05 for each cross. Progeny mean greater than midparent GEBVs was highlighted. ‘Fengdecunmai 5’ appeared most frequently in the top ten crosses, indicating that this wheat cultivar can be used as a good candidate parent for yield and quality improvement. Progeny means of yield and MR were either higher than or close to corresponding midparent GEBVs. For EXT, progeny means in two crosses, ‘12CA29’ × ‘Fengdecunmai 5’ and ‘Zhengmai 129’ × ‘Fengdecunmai 5’,

Table 3 – Selection response and genetic variance of the selected progenies from four parental selection schemes (S1 to S4). The top 5% progeny in each of the 100 selected crosses, for a total of 25,000 simulated progeny, were used to calculate mean and variance. For comparison, the mean and variance of training population (TP) are also presented.

Trait	Selection response					Genetic variance of the selected progenies				
	S1	S2	S3	S4	TP	S1	S2	S3	S4	TP
Yield	892.50	4.35	–1029.15	628.20	7549.20	647.55	7353.60	8204.10	3091.35	41,962.35
EXT	–15.19	5.65	25.97	–3.70	169.63	23.30	35.88	15.87	36.64	321.86
MR	–48.73	163.59	3.29	76.86	481.71	2356.74	303.88	4036.00	4157.07	21,746.57
ST	–2.13	5.16	0.67	1.63	14.18	3.50	3.48	5.05	8.92	36.1
TKW	0.93	–2.38	–3.73	–0.31	41.75	2.24	5.83	4.19	3.66	22.31
SN	–0.87	1.95	3.98	–0.25	40.29	1.58	7.41	5.92	2.94	30.14
KN	0.00	–0.28	–1.22	0.37	33.08	0.73	1.50	1.27	1.38	6.22
EA	–18.35	34.72	15.48	12.12	105.83	110.97	38.32	182.97	231.73	1078.54
PC	–0.51	0.26	0.42	–0.24	11.40	0.04	0.06	0.47	0.10	0.48
HD	0.33	0.23	–0.86	0.05	191.83	0.16	0.27	0.89	0.18	1.58
PH	–0.75	–1.07	8.30	–3.76	77.69	6.60	26.60	14.26	6.85	82.31

TKW, thousand-kernel weight; SP, spike number; KN, kernel number; HD, heading date; PH, plant height; PC, protein content; EA, extension area; EXT, extensibility; MR, maximum resistance; ST, stability time.

Table 4 – Top 10 candidate crosses selected by usefulness of selection index.

Biparental cross		Yield				EXT				MR			
P1	P2	Progeny	P1	P2	MP	Progeny	P1	P2	MP	Progeny	P1	P2	MP
12CA29	Fengdecunmai 5	8302.5 ^a	8140.5	8341.5	8241.0	168.7 ^a	163.4	162.1	162.7	639.2	685.5	623.3	654.4
CA0998	Fengdecunmai 5	8640.0	9012.0	8341.5	8677.5	158.9 ^a	140.0	162.1	151.1	517.6 ^a	380.9	623.3	502.1
13CA47	Fengdecunmai 5	8292.0	8257.5	8341.5	8299.5	170.2 ^a	176.2	162.1	169.1	587.1 ^a	510.5	623.3	566.9
13CA48	Fengdecunmai 5	8259.0 ^a	8056.5	8341.5	8199.0	169.8 ^a	172.7	162.1	167.4	599.0	590.0	623.3	606.6
Jishi021	Fengdecunmai 5	8091.0 ^a	7671.0	8341.5	8005.5	174.1 ^a	181.0	162.1	171.6	663.0	715.1	623.3	669.2
Zhongyou 255	Fengdecunmai 5	8140.5 ^a	7890.0	8341.5	8115.0	174.7 ^a	180.0	162.1	171.0	632.0 ^a	639.0	623.3	631.1
12CA29	CA0998	8548.5	8140.5	9013.5	8575.5	161.3 ^a	163.4	140.0	151.7	552.2 ^a	685.5	380.9	533.2
Zhengmai 129	Fengdecunmai 5	8454.0	8569.5	8341.5	8455.5	166.2 ^a	162.7	162.1	162.4	527.5	376.5	623.3	499.9
GY12023	Fengdecunmai 5	8230.5 ^a	7921.5	8341.5	8131.5	169.3 ^a	172.8	162.1	167.4	622.0	628.7	623.3	626.0
CA0996	Fengdecunmai 5	8530.5 ^a	8599.5	8341.5	8470.5	161.6 ^a	156.6	162.1	159.3	514.9 ^a	328.8	623.3	476.0

EXT, extensibility; MR, maximum resistance; P1, parent 1; P2, parent 2; RILs was selected using a proportion of 0.05, and used in calculating the progeny mean; MP, midparent GEBV.

^a Progeny means over midparent GEBVs.

were higher than those of both parents. For yield and MR, none of the 10 crosses gave a progeny mean higher than those of both parents. However, two crosses, 'Zhongyou 255' × 'Fengdecunmai 5' and 'CA0996' × 'Fengdecunmai 5', gave positive responses for all three traits. The 10 crosses in Table 4 gave positive responses for EXT and at least one other trait among yield and MR, and could be used to improve yield and at least one quality trait.

4. Discussion

4.1. Factors affecting prediction of genetic variance and usefulness

Genetic variance of the progeny population of a biparental cross can be predicted by simulation. Heritability was the most important factor that affected the accuracy of prediction of genetic variance and usefulness. It has been shown that, for a given number of markers, larger training population and higher heritability increase the estimation accuracy of marker effects [42]. Accurate estimation of marker effects resulted in higher prediction accuracy for the progeny mean. Accordingly, high accuracy of prediction of genetic variance and usefulness was achieved when heritability was high (Figs. 2 and 3). An increase in accuracy of prediction of genetic variance with higher heritability was also observed by Tiede et al. [25]. Compared with heritability, QTL number had a much smaller effect on accuracy of prediction of genetic variance and usefulness.

When a GS model was used to predict individual GEBV, prediction accuracy was influenced by trait genetic architecture [27, 43–45]. In some studies, the investigated models had similar prediction accuracy, but other studies showed significant differences in different traits [42, 46, 47]. In this study, different accuracies were observed using cross validation for five GS models (Fig. 5). However, when these models were used to predict usefulness and progeny genetic variance of a biparental cross, similar accuracies were observed. Lado et al. [24] also reported that Bayes LASSO and rBLUP yielded similar marker effects and prediction accuracy in two populations from CIMMYT's wheat breeding program.

4.2. Inclusion of genetic variance in cross performance prediction

The midparent value is a commonly used predictor for cross prediction and parental selection. However, it does not account for progeny genetic variance that affects genetic gain after selection. Our simulation experiment demonstrated that parental selection by usefulness resulted in higher genetic gain. The difference in genetic gain between usefulness and midparent GEBV was larger under higher heritability, owing to the higher accuracy in estimating genetic variance. Zhong and Jannink [20] reported that influence of progeny genetic variance on usefulness decreased with an increase in QTL number. In our study, three QTL numbers were used, but a decreasing trend was not observed. The result in this study was in agreement with that reported by Lehermeier et al. [48].

The relative gain between usefulness and midparent GEBV changed with selection proportion. The smaller the selection proportion, the larger was the relative gain (Fig. 4). This observation is understandable when genetic variance is considered in cross prediction. When the selection proportion is large, genetic variance has less weight on usefulness. Thus, parents selected by usefulness become similar to those selected by midparent GEBV, and the genetic gains from the two criteria become similar. It has been reported [48, 49] that genetic variance was underestimated in a simulation approach. Thus, the effect of including genetic variance in cross prediction could be larger than that observed in the simulation experiment. In our simulation, 5000 RILs were simulated from each biparental cross to enable precise estimation of genetic gain under a small selection proportion. We understand that the number of RILs could be different in practical breeding. If early-generation selection is not considered, we assume that the number of advanced lines from each cross could be in the thousands.

4.3. Realization of parental selection through cross prediction for improving multiple traits

When usefulness and midparent GEBV each was used for cross prediction in the training population, a large number of common crosses were found. This result shows that midparent GEBV

dominated in cross prediction. In the training population, because the ratio between variance of mean and variance of genetic standard deviation was high for all traits (data not shown), the effect of accounting for genetic variance in parental selection becomes limited, according to Zhong and Jannink [20]. The finding that quality traits had slightly fewer common crosses than yield, yield components, and agronomic traits suggests that the usefulness of quality traits is more strongly influenced by genetic variance than that of other traits. Similar results were reported for silking date and protein concentration in maize [23], and protein content and loaf volume in wheat [24].

In the training population, triangular relationships were found between predicted progeny mean and genetic variance for all traits, as previously reported by Bernardo and Lado et al. [23, 24]. Progeny generated by parents with extreme trait values have low variance and low progeny mean for low \times low crosses and high progeny mean for high \times high crosses. When two parents have a large phenotypic difference, they generate progenies with a medium mean and relatively high variance. Thus, a triangular relationship is formed (Fig. 6). More crosses with low variance and high progeny mean were observed for yield than for quality traits, for example MR. This finding differed from that of Lado et al. [24], who found that the triangular relationship was stronger for quality traits than for yield. The difference may be associated with population structure, as each population has its own gene frequency, genotypic frequency, trait heritability, and genetic variance and correlation.

Improvement of quality and yield simultaneously is a challenge in wheat breeding. Owing to the negative correlation between yield and quality traits, a parental selection scheme based on yield usefulness resulted in a negative response for quality traits, and vice versa (Table 3). However, individual genotypes may carry different favorable alleles, and genetic correlations between traits in a progeny population can be much different from that observed in the parental population. Even when two traits were negatively correlated in the parental population, positive correlations could be observed in some progeny populations (Fig. 7). To improve yield and quality simultaneously, parents that could generate breeding populations with no or a positive correlation between these target traits are desired. Trait correlation in the parental population is also relevant. If traits are strongly negatively correlated in a parental population, there is a low possibility of identifying desirable parents. The structure of the training population has been reported [50, 51] to affect significantly the accuracy of GS, and it will likewise influence parental selection for improving multiple traits.

When multiple traits are targeted in breeding, an index combining trait values and economic weights can be constructed and then used in selection [52, 53]. In addition to single-trait selection, a selection index constructed from yield and two quality traits was used in this study. Parental selection with the selection index resulted in positive responses for both yield and MR, and identified candidate parents for improving yield and quality traits simultaneously (Tables 3 and 4). The tradeoff effect on traits of interest could be overcome by incorporating different weights in the index. On the other hand, maintenance of genetic diversity is important to achieving genetic gain in the next cycle of

breeding [30]. A high level of genetic diversity will also help breeders to deal with potential challenges from unfavorable environments, such as drought and other abiotic stresses [54]. In this study, we found that parental selection based on usefulness of a selection index could result in higher genetic variance in newly generated progenies than could other parental selection schemes. Application of the selection index for parental selection not only led to genetic gains for both yield and quality, but also retained substantial genetic variance for subsequent cycles of breeding.

4.4. Justifications for assumptions used in this study

In this study, genetic variance was estimated by simulation and then used to calculate usefulness. There are other methods for estimating genetic variance analytically [48, 49]. However, simulation is able to predict trait correlations in newly generated progeny populations. In addition, no analytic approach could predict genetic gain and genetic diversity after parental selection. One other advantage is that more complicated genetic models, such as epistasis, could be considered in simulation as well [55, 56].

To simulate progenies and calculate usefulness from progeny GEBV, marker effects were needed. The five genomic selection models used in this study can estimate marker effects. In genomic selection, there are some other methods, such as genomic best linear unbiased prediction (GBLUP), reproducing kernel Hilbert space (RKHS) and machine learning. However, those methods cannot give estimates of marker effects and accordingly were not considered. The other requirement for progeny simulation is a genetic linkage map of markers and genes. During the SNP filtering process, 7588 SNPs on a consensus genetic map were finally chosen. Considering the relatively small population size used in this study, the SNP filtering process reduced redundancy of SNPs. Lado et al. [24] employed two wheat populations, one of 1465 fixed lines genotyped with 3884 SNPs and the other of 5984 lines genotyped with 1164 SNPs. In comparison, we used more SNPs with a smaller size of population.

We assumed an additive genetic model and found no major differences among GS models in genetic variance prediction. Non-additive effects (such as epistasis) may influence the accuracy of genetic variance prediction. Incorporating epistasis may improve prediction accuracy; however, modeling epistasis effects for a large number of markers is still a challenge [57]. Incorporating epistasis variance in cross prediction awaits further investigation. $G \times E$ interaction was also not taken into account. Marker effects can be obtained by including $G \times E$ interaction [58, 59], and then used to estimate genetic variance for each environment. Like environment-specific breeding values, optimal parents may be different across environments. The value of incorporating $G \times E$ interaction in GS and cross prediction remains an open question.

In this study we simulated a large number of RILs for predicting usefulness of a biparental cross, an operation that is applicable to pure-line breeding programs such as for wheat, soybean, and conventional rice. For hybrid breeding programs such as for maize, breeders select suitable parents to make superior F_1 hybrids. However, in hybrid breeding, breeders also need to cross two inbred parents to produce the

next generation of inbreds. For this reason, cross prediction using the progeny simulation approach employed in this study will also be useful in inbred development for hybrid breeding programs.

5. Conclusions

Minor differences were observed among GS models in cross prediction. Use of the usefulness criterion for parental selection resulted in higher genetic gain than use of midparent GEBV. Selecting suitable parents to make crosses using a simulation approach can disrupt the negative correlation between yield and quality traits. Use of a selection index is one option for selecting parents with the aim of improving both yield and quality traits and maintaining genetic diversity for long-term breeding objectives.

Supplementary data for this article can be found online at <https://doi.org/10.1016/j.cj.2018.05.003>.

Acknowledgments

This work was supported by the National Key Basic Research Program of China (2014CB138105) and the National Natural Science Foundation of China (31371623).

REFERENCES

- [1] R. Lande, R. Thompson, Efficiency of marker-assisted selection in the improvement of quantitative traits, *Genetics* 124 (1990) 743–756.
- [2] B. Servin, O.C. Martin, M. Mezard, F. Hospital, Toward a theory of marker-assisted gene pyramiding, *Genetics* 168 (2004) 513–523.
- [3] R. Bernardo, Molecular markers and selection for complex traits in plants: learning from the last 20 years, *Crop Sci.* 48 (2008) 1649–1664.
- [4] Y. Xu, J.H. Crouch, Marker-assisted selection in plant breeding: from publications to practice, *Crop Sci.* 48 (2008) 391–407.
- [5] Y. Liu, Z. He, R. Appels, X. Xia, Functional markers in wheat: current status and future prospects, *Theor. Appl. Genet.* 125 (2012) 1–10.
- [6] T.H.E. Meuwissen, B.J. Hayes, M.E. Goddard, Prediction of total genetic value using genome-wide dense marker maps, *Genetics* 157 (2001) 1819–1829.
- [7] R. Bernardo, J. Yu, Prospects for genomewide selection for quantitative traits in maize, *Crop Sci.* 47 (2007) 1082–1090.
- [8] T. Albrecht, V. Wimmer, H.J. Auinger, M. Erbe, C. Knaak, M. Ouzunova, H. Simianer, C.C. Schon, Genome-based prediction of testcross values in maize, *Theor. Appl. Genet.* 123 (2011) 339–350.
- [9] M. Denis, J.M. Bouvet, Efficiency of genomic selection with models including dominance effect in the context of eucalyptus breeding, *Tree Genet. Genomes* 9 (2012) 37–51.
- [10] J.M. Massman, A. Gordillo, R.E. Lorenzana, R. Bernardo, Genomewide predictions from maize single-cross data, *Theor. Appl. Genet.* 126 (2013) 13–22.
- [11] T. Guo, H. Li, J. Yan, J. Tang, J. Li, Z. Zhang, L. Zhang, J. Wang, Performance prediction of F₁ hybrids between recombinant inbred lines derived from two elite maize inbred lines, *Theor. Appl. Genet.* 126 (2013) 189–201.
- [12] S. Xu, D. Zhu, Q. Zhang, Predicting hybrid performance in rice using genomic best linear unbiased prediction, *Proc. Natl. Acad. Sci. U. S. A.* 111 (2014) 12456–12461.
- [13] Z. Guo, D.M. Tucker, J. Lu, V. Kishore, G. Gay, Evaluation of genome-wide selection efficiency in maize nested association mapping populations, *Theor. Appl. Genet.* 124 (2012) 261–275.
- [14] Y. Beyene, K. Semagn, S. Mugo, A. Tarekegne, R. Babu, B. Meisel, P. Sehabiague, D. Makumbi, C. Magorokosho, S. Oikeh, J. Gakunga, M. Vargas, M. Olsen, B.M. Prasanna, M. Banziger, J. Crossa, Genetic gains in grain yield through genomic selection in eight biparental maize populations under drought stress, *Crop Sci.* 55 (2015) 154–163.
- [15] E.L. Heffner, J.L. Jannink, H. Iwata, E. Souza, M.E. Sorrells, Genomic selection accuracy for grain quality traits in biparental wheat populations, *Crop Sci.* 51 (2011) 2597–2606.
- [16] S.D. Battenfield, C. Guzman, R.C. Gaynor, R.P. Singh, R.J. Pena, S. Dreisigacker, A.K. Fritz, J.A. Poland, Genomic selection for processing and end-use quality traits in the CIMMYT spring bread wheat breeding program, *Plant Genome* 9 (2016) <https://doi.org/10.3835/plantgenome2016.01.0005>.
- [17] R. Bernardo, Parental selection, number of breeding populations, and size of each population in inbred development, *Theor. Appl. Genet.* 107 (2003) 1252–1256.
- [18] J.R. Witcombe, S. Gyawali, M. Subedi, D.S. Virk, K.D. Joshi, Plant breeding can be made more efficient by having fewer, better crosses, *BMC Plant Biol.* 13 (2013) 22.
- [19] H.F. Utz, M. Bohn, A.E. Melchinger, Predicting progeny means and variances of winter wheat crosses from phenotypic values of their parents, *Crop Sci.* 41 (2001) 1470–1478.
- [20] S. Zhong, J.L. Jannink, Using quantitative trait loci results to discriminate among crosses on the basis of their progeny mean and variance, *Genetics* 177 (2007) 567–576.
- [21] E. Souza, M.E. Sorrells, Prediction of progeny variation in oat from parental genetic relationships, *BMC Plant Biol.* 82 (1991) 233–241.
- [22] M. Mohammadi, T. Tiede, K.P. Smith, Popvar: a genome-wide procedure for predicting genetic variance and correlated response in biparental breeding populations, *Crop Sci.* 55 (2015) 2068–2077.
- [23] R. Bernardo, Genomewide selection of parental inbreds: classes of loci and virtual biparental populations, *Crop Sci.* 54 (2014) 2586–2595.
- [24] B. Lado, S. Battenfield, C. Guzman, M. Quincke, R.P. Singh, S. Dreisigacker, R.J. Pena, A. Fritz, P. Silva, J. Poland, L. Gutierrez, Strategies for selecting crosses using genomic prediction in two wheat breeding programs, *Plant Genome* 10 (2017) 1–12.
- [25] T. Tiede, L. Kumar, M. Mohammadi, K.P. Smith, Predicting genetic variance in biparental breeding populations is more accurate when explicitly modeling the segregation of informative genomewide markers, *Mol. Breed.* 35 (2015) 199.
- [26] C. Lehermeier, G. de los Campos, V. Wimmer, C.C. Schon, Genomic variance estimates: with or without disequilibrium covariances? *J. Anim. Breed. Genet.* 134 (2017) 232–241.
- [27] S. Zhong, J.C. Dekkers, R.L. Fernando, J.L. Jannink, Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study, *Genetics* 182 (2009) 355–364.
- [28] N. Heslot, H. Yang, M.E. Sorrells, J.L. Jannink, Genomic selection in plant breeding: a comparison of models, *Crop Sci.* 52 (2012) 146–160.
- [29] P. Perez-Rodriguez, D. Gianola, J.M. Gonzalez-Camacho, J. Crossa, Y. Manes, S. Dreisigacker, Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat, *G3-Genes Genomes Genet.* 2 (2012) 1595–1605.
- [30] A. Jacobson, L. Lian, S.Q. Zhong, R. Bernardo, Minimal loss of genetic diversity after genomewide selection within biparental maize populations, *Crop Sci.* 55 (2015) 783–789.

- [31] J. Rutkoski, R.P. Singh, J. Huerta-Espino, S. Bhavani, J. Poland, J.L. Jannink, M.E. Sorrells, Genetic gain from phenotypic and genomic selection for quantitative resistance to stem rust of wheat, *Plant Genome* 8 (2015) 1–10.
- [32] X. Kong, Y. Zhang, D. Zhao, X. Xia, C. Wang, Z. He, Noodle quality evaluation of new wheat cultivars from northern China winter wheat regions, *Acta Agron. Sin.* 8 (2016) 1143–1159 (in Chinese with English abstract).
- [33] M.G. Murray, W.F. Thompson, Rapid isolation of high molecular weight plant DNA, *Nucleic Acids Res.* 8 (1980) 4321–4325.
- [34] S. Wang, D. Wong, K. Forrest, A. Allen, S. Chao, B. Huang, M. Maccaferri, S. Salvi, S.G. Milner, L. Cattivelli, A.M. Mastrangelo, A. Whan, S. Stephen, G. Barker, R. Wieseke, J. Plieske, M. Lillemo, D. Mather, R. Appels, R. Dolferus, G. Brown-Guedira, A. Korol, A.R. Akhunova, C. Feuillet, J. Salse, M. Morgante, C. Pozniak, M. Luo, J. Dvorak, M. Morell, J. Dubcovsky, M. Ganal, R. Tuberosa, C. Lawley, I. Mikoulitch, C. Cavanagh, K.J. Edwards, M. Hayden, E. Akhunov, International wheat genome sequencing, characterization of polyploid wheat genomic diversity using a high-density 90000 single nucleotide polymorphism array, *Plant Biotechnol. J.* 12 (2014) 787–796.
- [35] J. Wang, *Quantitative Genetics*, Science Press, Beijing, China, 2017 (in Chinese).
- [36] L. Meng, H. Li, L.Y. Zhang, J. Wang, QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations, *Crop J.* 3 (2015) 269–283.
- [37] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014.
- [38] J.J. Ceron-Rojas, J. Crossa, V.N. Arief, K. Basford, J. Rutkoski, D. Jarquin, G. Alvarado, Y. Beyene, K. Semagn, I. DeLacy, A genomic selection index applied to simulated and real data, *G3-Genes Genomes Genet.* 5 (2015) 2155–2164.
- [39] J.B. Endelman, Ridge regression and other kernels for genomic selection with R package rrBLUP, *Plant Genome* 4 (2011) 250–255.
- [40] P. Perez, G. de los Campos, Genome-wide regression and prediction with the BGLR statistical package, *Genetics* 198 (2014) 483–495.
- [41] K.W. Broman, H. Wu, S. Sen, G.A. Churchill, R/qtl: QTL mapping in experimental crosses, *Bioinformatics* 19 (2003) 889–890.
- [42] V. Wimmer, C. Lehermeier, T. Albrecht, H.J. Auinger, Y. Wang, C.C. Schon, Genome-wide prediction of traits with different genetic architecture through efficient variable selection, *Genetics* 195 (2013) 573–587.
- [43] H.D. Daetwyler, R. Pong-Wong, B. Villanueva, J.A. Woolliams, The impact of genetic architecture on genome-wide evaluation methods, *Genetics* 185 (2010) 1021–1031.
- [44] B.J. Hayes, J. Pryce, A.J. Chamberlain, P.J. Bowman, M.E. Goddard, Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits, *PLoS Genet.* 6 (2010), e1001139.
- [45] C. Riedelsheimer, F. Technow, A.E. Melchinger, Comparison of whole-genome prediction models for traits with contrasting genetic architecture in a diversity panel of maize inbred lines, *BMC Genomics* 13 (2012) 452.
- [46] Z.A. Desta, R. Ortiz, Genomic selection: genome-wide prediction in plant improvement, *Trends Plant Sci.* 19 (2014) 592–601.
- [47] R. Howard, A.L. Carriquiry, W.D. Beavis, Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures, *Trends Plant Sci.* 4 (2014) 1027–1046.
- [48] C. Lehermeier, S. Teyssedre, C.C. Schon, Genetic gain increases by applying the usefulness criterion with improved variance prediction in selection of crosses, *Genetics* 207 (2017) 1651–1661.
- [49] L. Lian, A. Jacobson, S.Q. Zhong, R. Bernardo, Prediction of genetic variance in biparental maize populations: genomewide marker effects versus mean genetic variance in prior populations, *Crop Sci.* 55 (2015) 1181–1188.
- [50] D. Akdemir, J.I. Sanchez, J.L. Jannink, Optimization of genomic selection training populations with a genetic algorithm, *Genet. Sel. Evol.* 47 (2015) 38.
- [51] J. Isidro, J.L. Jannink, D. Akdemir, J. Poland, N. Heslot, M.E. Sorrells, Training set optimization under population structure in genomic selection, *Theor. Appl. Genet.* 128 (2015) 145–158.
- [52] C. Lin, Index selection for genetic improvement of quantitative characters, *Theor. Appl. Genet.* 52 (1978) 49–56.
- [53] A.W. Schulthess, Y. Wang, T. Miedaner, P. Wilde, J.C. Reif, Y. Zhao, Multiple trait and selection indices genomic predictions for grain yield and protein content in rye for feeding purposes, *Theor. Appl. Genet.* 129 (2015) 273–287.
- [54] Y. Fu, Understanding crop genetic diversity under modern plant breeding, *Theor. Appl. Genet.* 128 (2015) 2131–2142.
- [55] J. Wang, M. van Ginkel, D. Podlich, G. Ye, R. Trethowan, W. Pfeiffer, I.H. DeLacy, M. Cooper, S. Rajaram, Comparison of two breeding strategies by computer simulation, *Crop Sci.* 43 (2003) 1764–1773.
- [56] J. Wang, M. van Ginkel, R. Trethowan, G. Ye, I. DeLacy, D. Podlich, M. Cooper, Simulating the effects of dominance and epistasis on selection response in the CIMMYT wheat breeding program using QuCim, *Crop Sci.* 44 (2004) 2006–2018.
- [57] Y. Jiang, J.C. Reif, Modeling epistasis in genomic selection, *Genetics* 201 (2015) 759–768.
- [58] N. Heslot, D. Akdemir, M.E. Sorrells, J.L. Jannink, Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions, *Theor. Appl. Genet.* 127 (2014) 463–480.
- [59] M. Lopez-Cruz, J. Crossa, D. Bonnett, S. Dreisigacker, J. Poland, J.L. Jannink, R.P. Singh, E. Auriague, G. de los Campos, Increased prediction accuracy in wheat breeding trials using a marker \times environment interaction genomic selection model, *G3-Genes Genomes Genet.* 5 (2015) 569–582.