The Crop Journal 11 (2023) 1884–1892

Contents lists available at ScienceDirect

The Crop Journal

journal homepage: www.keaipublishing.com/en/journals/the-crop-journal/

Genomic prediction of yield performance among single-cross maize hybrids using a partial diallel cross design

Ping Luo [a,b,c,1], Houwen Wang [a,1], Zhiyong Ni [c,1], Ruisi Yang [a], Fei Wang [a], Hongjun Yong [a], Lin Zhang [d], Zhiqiang Zhou [a], Wei Song [e], Mingshun Li [a], Jie Yang [f], Jianfeng Weng [g], Zhaodong Meng [g], Degui Zhang [a], Jienan Han [a], Yong Chen [a], Runze Zhang [a], Liwei Wang [e], Meng Zhao [g], Wenwei Gao [c], Xiaoyu Chen [a], Wenjie Li [a], Zhuanfang Hao [a,*], Junjie Fu [a,*], Xuecai Zhang [b,*], Xinhai Li [a,*]

[a] State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China
[b] International Maize and Wheat Improvement Center (CIMMYT), Texcoco 56237, Mexico
[c] College of Agronomy, Xinjiang Agricultural University, Urumqi 830091, Xinjiang, China
[d] College of Agronomy, Northeast Agricultural University, Harbin 150030, Heilongjiang, China
[e] Institute of Cereal and Oil Crops, Hebei Academy of Agriculture and Forestry Sciences, Shijiazhuang 050035, Hebei, China
[f] Food Crops Research Institute, Xinjiang Academy of Agricultural Science, Urumqi 830091, Xinjiang, China
[g] Maize Research Institute of Shandong Academy of Agricultural Sciences, Jinan 250100, Shandong, China

A R T I C L E   I N F O

Article history:
Received 3 July 2023
Revised 27 September 2023
Accepted 28 September 2023
Available online 29 October 2023

Keywords:
Maize
Genomic prediction
Prediction model
Genetic effects
Hybrid performance

A B S T R A C T

Genomic prediction (GP) in plant breeding has the potential to predict and identify the best-performing hybrids based on the genotypes of their parental lines. In a GP experiment, 34 elite inbred lines were selected to make 285 single-cross hybrids in a partial-diallel cross design. These lines represented a mini-core collection of Chinese maize germplasm and comprised 18 inbred lines from the Stiff Stalk heterotic group and 16 inbred lines from the Non-Stiff Stalk heterotic group. The parents were genotyped by sequencing and the 285 hybrids were phenotyped for nine yield and yield-related traits at two locations in the summer sowing area (SUS) and three locations in the spring sowing area (SPS) in the main maize-producing regions of China. Multiple GP models were employed to assess the accuracy of trait prediction in the hybrids. By ten-fold cross-validation, the prediction accuracies of yield performance of the hybrids estimated by the genomic best linear unbiased prediction (GBLUP) model in SUS and SPS were 0.51 and 0.46, respectively. The prediction accuracies of the remaining yield-related traits estimated with GBLUP ranged from 0.49 to 0.86 and from 0.53 to 0.89 in SUS and SPS, respectively. When additive, dominance, epistasis effects, genotype-by-environment interaction, and multi-trait effects were incorporated into the prediction model, the prediction accuracy of hybrid yield performance was improved. The ratio of training to testing population and size of training population optimal for yield prediction were determined. Multiple prediction models can improve prediction accuracy in hybrid breeding.

1. Introduction

To keep up with population increase, the output of staple cereals including maize (*Zea mays* L.), rice (*Oryza sativa* L.), and wheat (*Triticum aestivum* L.) must double over the next two decades [1]. Maize provides food and feedstocks wordwide [2]. Maize breeding aims to identify hybrids with high yield potential in multiple environments. Such trials employ partial-diallel designs featuring cross-mating between heterotic groups, to identify the best hybrids among a few elite breeding lines. They also permit accurately estimating the general and specific combining abilities (GCA and SCA) of the parental lines. Owing to time and resource constraints, only few lines can be evaluated in such studies.

Genomic prediction (GP) permits increasing the efficiency of multiple-location studies. In GP, the effects of all markers are estimated simultaneously from a training population that has been both phenotyped and genotyped, and then the genomic breeding

* Corresponding authors.
E-mail addresses: lixinhai@caas.cn (X. Li), XC.Zhang@cgiar.org (X. Zhang), fujunjie@caas.cn (J. Fu), haozhuanfang@163.com (Z. Hao).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.cj.2023.09.009

values of the untested genotyped lines are estimated as the sum of all marker effects [3]. Genomic prediction permits early selection before phenotypes of traits are collected. Because the genotypes of hybrids are inferred from the genotypes of their parents rather than being determined by sequencing, GP lowers the cost of hybrid breeding [4,5]. In maize, GP has been implemented to predict and select target traits in breeding lines [6,7] and to predict the performance of hybrids [8,9].

The most common maize cultivar is a hybrid, and prediction of its performance is essential [10–12]. Hybrid performance is influenced by both additive and non-additive effects, including dominance and all types of epistatic interactions [5,13–15]. In hybrid breeding experiments, prediction models incorporating non-additive effects have been employed extensively [16]. Modeling non-additive effects in genomic hybrid prediction presents both advantages [17–21] and drawbacks [5,16,22–25]. Adding non-additive effects into a genomic best linear unbiased prediction (GBLUP) model did not improve plant height in a maize $F_{2:3}$ population, but did increase yield [21]. For predicting traits with low heritability, such as yield, prediction accuracy can be improved by joint modeling. Its advantage is that genotype × environment interactions and correlations among traits are appropriately accounted for. This feature increases statistical power and parameter estimate precision, increasing prediction accuracy and reducing the bias of genomic selection [26–29]. The best GP models must be identified for each trait and population.

We wished to identify GP models suitable for diverse growing regions based on the two major maize producing regions in China: The Yellow–Huai Valley (summer sowing area, SUS) and Northeast China (spring sowing area, SPS). The objectives of the present study were to: (1) assess the prediction accuracy of GBLUP models for maize hybrid yield and yield-related traits in SUS and SPS; (2) evaluate the prediction accuracy of maize hybrids using various genetic-effects models, genotype × environment interaction model, and multi-trait effects model; and (3) evaluate the prediction accuracy of yield under several ratios of training to testing population (TRN: TST) and population sizes.

2. Materials and methods

2.1. Plant materials, experimental design, and phenotypic data collection

Thirty-four elite inbred lines, representing a mini-core collection of Chinese maize germplasm, were selected to make single-cross hybrids. Among the 34, 18 belonged to the Stiff Stalk (SS) heterotic group and the other 16 belonged to the Non-Stiff Stalk (NSS, Sipingtou/Iodent) group. A partial-diallel cross design was applied to generate 285 single-cross hybrids using parental lines from differing groups. Three hybrids: HS06 × HS27, HS12 × HS23, and HS17 × HS33, were not made, owing to the differences in flowering times of the parents.

All hybrids were planted in trials during the spring and summer of 2022 in five locations. In each location, all hybrids were planted in two-row plots using an incomplete block design with two replications. The row length was 4.0 m and the row spacing was 0.6 m. Field managements followed best local practices. The two SUS trials were planted in Hebei (HB) and Henan (HN), and the three SPS trials in Heilongjiang (HLJ), Jilin (JL), and Liaoning (LN) (Table S1). In each location, measurements of nine agronomic traits were recorded (Table S2): grain yield (GY), ear weight (EarW), ear length (EarL), ear diameter (EarD), tip length (TipL), kernel row number (RowN), kernel number (KerN), 100-kernel weight (100KW), and shelling percentage (SP).

2.2. Phenotypic data analysis

The lme4 package in R [30] was used to fit a linear mixed model to the nine traits over all trials to provide the following best linear unbiased predictor (BLUP) value:

$$Y_{ij} = \mu + \mathrm{Line}_i + \mathrm{Env}_j + (\mathrm{Line} \times \mathrm{Env})_{ij} + (\mathrm{Env} \times \mathrm{Rep})_{ij} + \mathrm{error}_{ijn}$$

where μ is the general mean, Line_i is the genotype effect of the i th inbred, Env_j is the effect of the j th environment, $(\mathrm{Line} \times \mathrm{Env})_{ij}$ is genotype–environment interaction effect, $(\mathrm{Env} \times \mathrm{Rep})_{ij}$ is environment-by-replication interaction, and error_{ijn} is the error of the j th environment and the n th replication.

Generalized heritability was calculated based on the variance components:

$$H^2 = \left(\delta_g^2\right) / \left(\delta_g^2 + \delta_g e^2 / y + \delta_e^2 / yr\right)$$

where y and r represent the numbers of growing environments and the number of field replicates per environment, respectively; δ_g^2 represents genetic variance; $\delta_g e^2$ represents genotype-by- environment interaction variance; and δ_e^2 is residual variance [31]. Correlations and analyses of variance (ANOVA) for phenotypic data were calculated using the *cor* and *aov* R functions, respectively.

2.3. Genotyping and genotypic data analysis

The Illumina single-nucleotide polymorphism (SNP) chip (GenoBaits Maize 45K Panel) from Beijing Boruidi Biotechnology was used to genotype the maize inbred lines [32]. SNPs that did not meet the following requirements were removed from further analysis: (i) fewer than 10% missing values; (ii) minor-allele frequency (MAF) larger than 5%; and (iii) no more than three heterozygous genotypes. After this filter, 23,692 SNPs remained. Chromosomal locations of SNPs were established using the B73 reference genome (B73 RefGen_v4). Using TASSEL [33] version 5, the genetic distances among the 34 parental lines were calculated as 1 minus the identity-by-state similarity. Using the *prcomp* function in R, principal component analysis (PCA) was performed on the parental lines.

2.4. Genomic prediction of the maize hybrid performance

2.4.1. Genomic best linear unbiased prediction (GBLUP)

In a genomic prediction where $\boldsymbol{b}$ includes the genomic relationship matrix, the GBLUP model is the GP model most frequently employed in plant breeding [34]. A linear mixed model was fitted:

$$Y = 1_n \mu + Z_L b + \epsilon$$

The fixed-effects design matrix $\boldsymbol{X} = 1_n$, the vector of length n corresponds to the general mean $\boldsymbol{\beta} = \beta_0$, $\boldsymbol{b} = (b_1, b_2, \ldots, b_J)^{\mathrm{T}}$ contains the genotypic effects of $\boldsymbol{J}$ lines, and $\mathbf{Z}$ is the incidence matrix design for the random line effects ($\boldsymbol{Z_L}$), where $\boldsymbol{b} \sim N_J(\mathbf{0}, \sigma_g^2 \boldsymbol{G})$, and $\boldsymbol{R} = \sigma^2 \boldsymbol{I}_n$.

2.4.2. Additive and the non-additive model

Hybrid phenotypes were predicted by estimating the BLUPs for general combining abilities (GCAs) in males and females ($\mathrm{GCA_{female}}$ and $\mathrm{GCA_{male}}$) and specific combining abilities (SCAs) of crosses along with their variance components ($\sigma^2 u_1$, $\sigma^2 u_2$, and $\sigma^2 u_s$). This can be expressed with the same model used in the diallel experiment:

$$y = X\beta + Z_1 u_1 + Z_2 u_2 + Z_3 u_s + \varepsilon$$

The mixed-model equations for this model are:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z_1 & X'R^{-1}Z_2 & X'R^{-1}Z_3 \\ Z_1'R^{-1}X & Z_1R^{-1}Z_1+G_1^{-1} & Z_1'R^{-1}Z_2 & Z_1'R^{-1}Z_3 \\ Z_2'R^{-1}X & Z_2'R^{-1}Z_1 & Z_2'R^{-1}Z_2+G_2^{-1} & Z_2'R^{-1}Z_3 \\ Z_3'R^{-1}X & Z_3'R^{-1}Z_1 & Z_3'R^{-1}Z_2 & Z_3'R^{-1}Z_3+G_3^{-1} \end{bmatrix} \begin{bmatrix} X'R^{-1}y \\ Z_1'R^{-1}y \\ Z_2'R^{-1}y \\ Z_3'R^{-1}y \end{bmatrix} = \begin{bmatrix} \beta \\ u_1 \\ u_2 \\ u_S \end{bmatrix}$$

where $\boldsymbol{\beta}$ is the vector of fixed effects; $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, and $\boldsymbol{u}_s$ are the respective BLUPs for $\text{GCA}_{\text{female}}$, GCA_{male}, and SCA effects; $\boldsymbol{X}$ and $\boldsymbol{Z}$ s are incidence matrices for fixed and random effects, respectively; $\boldsymbol{R}$ is a matrix of residuals (here $\mathbf{I}\boldsymbol{\sigma}_{\mathbf{e}}^{\mathbf{2}}$); and $\mathbf{G_1^{-1}}$, $\mathbf{G_2^{-1}}$, and $\mathbf{G_3^{-1}}$ are the inverses of the variance–covariance matrices of random effects. The models were tested considering additive (A) ($\boldsymbol{u}_1$, $\boldsymbol{u}_2$, genomic relationship matrix as additive effects) and non-additive ($\boldsymbol{u}_s$, this kinship matrix simulates dominance (D) + epistatic (E) effects), which were implemented using the R package sommer [35].

2.4.3. Genotype × environment interaction (G × E) model

The G × E model is an expansion of the GBLUP model that incorporates environmental effects, genotypic effects, and genotype × environment interaction effects:

$$Y = 1_n\mu + X_E\beta_E + Z_Lb_1 + Z_{EL}b_2 + \epsilon$$

It is split into the general mean ($1_n\mu$) and the environment effects term ($\boldsymbol{X}_E\boldsymbol{\beta}_E$), $\boldsymbol{X} = [1_n\boldsymbol{X}_E]$ and $\boldsymbol{\beta} = \left(\mu, \boldsymbol{\beta}_E^{\mathrm{T}}\right)^{\mathrm{T}}$. Similarly, for the random effects, $\boldsymbol{Z} = [\boldsymbol{Z}_L\boldsymbol{Z}_{EL}]$ and $\boldsymbol{b} = \left[\boldsymbol{b}_1^{\mathrm{T}}, \boldsymbol{b}_2^{\mathrm{T}}\right]^{\mathrm{T}}$, where $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are vectors of random genotypic effects and of random G × E interaction effects, with the incidence matrixes $\boldsymbol{Z}_L$ and $\boldsymbol{Z}_{EL}$, respectively. For $\boldsymbol{b}_1$, the same distribution as the GBLUP model was assumed, $\boldsymbol{b}_1 \sim N_J\left(0, \sigma_g^2\boldsymbol{G}\right)$, and for the second random effect, $\boldsymbol{b}_2 \sim N_J(0, \Sigma_E \otimes G)$, where, $\Sigma_E \otimes G$ is the relationship matrix of the G × E interaction term, with Σ_E the genetic variance–covariance matrix between I environments; and the ith element of the diagonal of Σ_E, σ_{Ei}^2, is the genetic variance in environment i, i = 1, …, I, and $\sigma_{Eik}\boldsymbol{G}$ is the genetic variance–covariance matrix for lines in environments i and k, where σ_{Eik} is the element (i, k) of Σ_E.

When genomic prediction was performed using the G × E model method, two parameters of TRN:TST ratio and population size, were set at nine and seven levels, respectively. In scenario 1, TRN: TST ratio was set to 1:9, 1:7, 1:5, 1:3, 1:1, 3:1, 5:1, 7:1, or 9:1. In scenario 2, the population size was set to 20, 60, 100, 140, 180, 220, or 285.

2.4.4. Bayesian genomic multi-trait linear regression model (multi-trait model)

For the multi-trait (MT) model, a multivariate Bayesian Gaussian model with an unstructured variance–covariance matrix was employed [36]. The MT model was

$$Y = 1_J\mu^{\mathrm{T}} + XB + Z_1b_1 + E$$

where $\boldsymbol{Y} = [\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_J]^{\mathrm{T}}$, $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_J]^{\mathrm{T}}$, $\boldsymbol{b}_1 = [\mathbf{g}_1, \ldots, \mathbf{g}_J]^{\mathrm{T}}$, and $\boldsymbol{E} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_J]^{\mathrm{T}}$.

Note that under this notation, $\boldsymbol{E}^{\mathrm{T}} \sim \boldsymbol{MN}_{nT} \times_J (\mathbf{0}, \boldsymbol{R}, \boldsymbol{I}_J)$ or equivalently $\boldsymbol{E}^{\mathrm{T}} \sim \boldsymbol{MN}_{nT} \times_J (\mathbf{0}, \boldsymbol{I}_J, \boldsymbol{R})$, and $\boldsymbol{b}_1^{\mathrm{T}} \sim \boldsymbol{MN}_{nT} \times_J (\mathbf{0}, \Sigma_{\mathrm{T}}, \boldsymbol{G})$ or $\boldsymbol{b}_1 \sim \boldsymbol{MN}_{nT} \times_J (\mathbf{0}, \boldsymbol{G}, \sum_{\mathrm{T}})$.

2.5. Cross-validation scheme and evaluation metrics

2.5.1. Cross-validation scheme

A 10-fold cross-validation scheme was implemented and repeated 100 times to compare prediction performances. The dataset was divided into ten disjoint genotype subsets, one of which was left out for validation and the other nine of which served as the training population for estimating model parameters and predicting the genotypes in the validation population.

2.5.2. Evaluation metrics

To evaluate the prediction accuracy of the G × E model, four evaluation indicators were added to the Pearson's correlation coefficient: the Kendall rank correlation coefficient (KCC), Spearman rank correlation coefficient (SCC), squared R coefficient of determination (R^2), and mean squared error (MSE).

Pearson's correlation coefficient (PCC, r, R):

$$\mathrm{PCC}(X,Y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

The Kendall rank correlation coefficient (KCC, tau):

$$\mathrm{KCC}(X,Y) = \frac{N1-N2}{n(n-1)/2}$$

The Spearman rank correlation coefficient (SCC, rho):

$$\mathrm{SCC}(X,Y) = \frac{\mathrm{cov}(R(X),R(Y))}{\sigma R(X)\sigma R(Y)}$$

Coefficient of determination, R squared (R^2, r^2):

$$R^2(X,Y) = 1 - \frac{\sum_{i=1}^{n}(x_i-y_i)^2}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$$

Mean squared error (MSE):

$$\mathrm{MSE}(X,Y) = \frac{\sum_{i=1}^{n}(x_i-y_i)^2}{N}$$

3. Results

3.1. Phenotypic data

All nine traits showed broad phenotypic variation among environments (Fig. 1). GY was significantly higher in SPS than in SUS, especially in LN, the location with the highest GY of 6.27 kg per plot, and in HN with the lowest GY of 2.95 kg per plot. The ear traits in SPS were significantly different from those in SUS, but TipL was not significant in the five locations.

The heritabilities across the nine traits ranged from 0.46 to 0.86 in SUS and from 0.53 to 0.82 in SPS (Table 1). The heritabilities of yield in SUS and SPS were 0.64 and 0.53, respectively. Of the nine traits, EarL, EarD, TipL, RowN, and 100KW showed moderate to high heritabilities. The heritabilities of SP (0.46 and 0.56) were relatively low.

For all traits, genotype variance was greater than genotype × environment interaction variance. By combining phenotypic data from all environments in SUS and SPS, the Pearson's correlations between traits were calculated using the BLUP values (Fig. 2). GY and EarW were significantly correlated in SUS and SPS (r = 0.96 and 0.94, respectively). GY and RowN were not correlated (r = −0.01 and −0.09, respectively). RowN and 100KW were significantly negatively correlated (r = −0.52) in both SUS and SPS. Correlations between the same traits in differing environments were all positive, with a range of coefficients from 0.09 to 0.67 (Fig. S1). These results confirm the high stability of these phenotypic data, reflecting the accuracy of phenotypes collected.

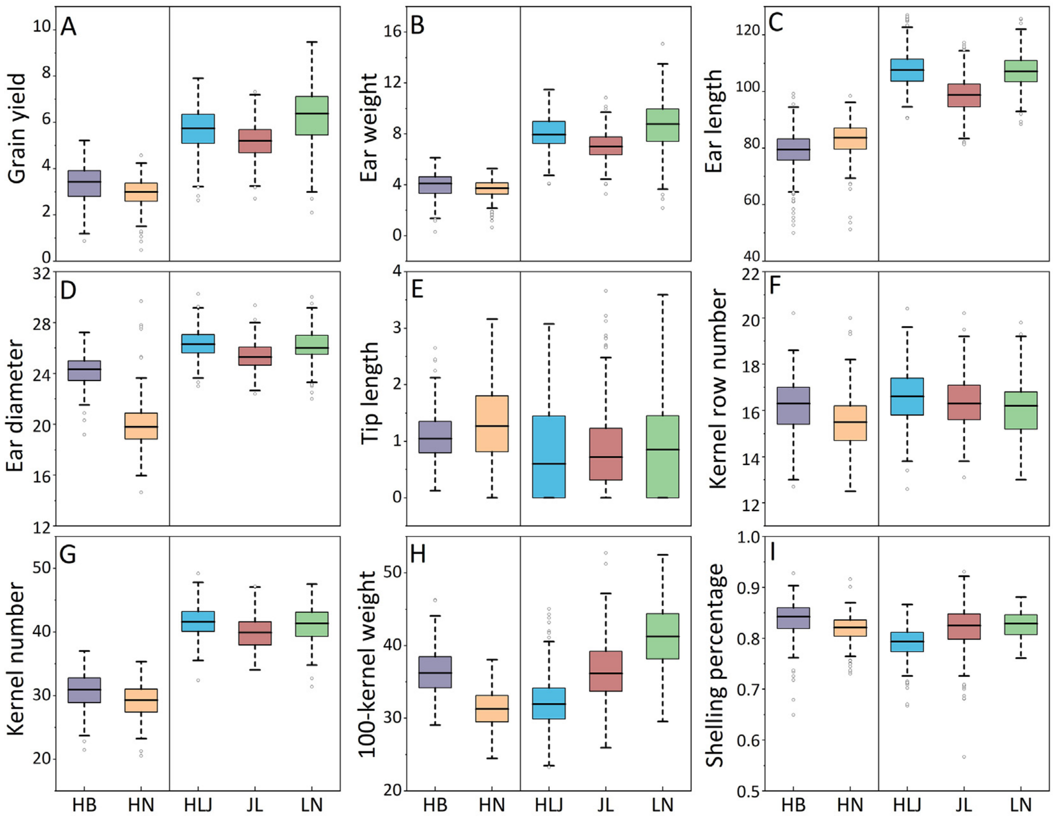

Fig. 1. Phenotypic distributions of the nine traits (A-I) evaluated in five locations. For each box, the upper and lower boundaries represent the 25th and 75th percentiles, respectively. The middle horizontal lines represent the median. Whiskers represent 1.5 × the interquartile range. The dots beyond the whiskers represent outliers. HB, Hebei; HN, Henan; HLJ, Heilongjiang; JL, Jilin; LN, Liaoning.

Table 1
Analysis of variance (ANOVA) and estimated heritabilities of nine traits in SUS and SPS.

Traits	SUS					SPS				
	H^2	V_G	V_E	V_P	$V_{G\times E}$	H^2	V_G	V_E	V_P	$V_{G\times E}$
GY	0.64	0.41	0.28	0.69	0.27	0.53	0.10	0.06	0.16	0.15
EarW	0.69	0.18	0.03	0.21	0.35	0.57	0.92	0.69	1.61	0.40
EarL	0.75	20.28	8.77	29.05	17.15	0.78	28.65	25.85	54.5	3.20
EarD	0.80	0.97	9.07	10.04	0.04	0.75	1.03	0.27	1.30	0.17
TipL	0.77	0.13	0.02	0.15	0.11	0.79	0.35	7E-04	4E-01	0.09
RowN	0.83	1.05	0.26	1.31	0.08	0.82	1.07	0.05	1.12	0.04
KerN	0.53	1.78	1.22	3.00	1.93	0.73	4.31	0.90	5.21	0.71
100KW	0.86	5.43	12.81	18.24	1.77	0.74	8.68	20.11	28.79	2.27
SP	0.46	3E-04	2E-04	5E-04	1E-04	0.56	5E-04	1E-04	6E-04	4E-04

SUS, summer maize area; SPS, spring maize area; H^2, heritability; V_G, genotype variance; V_E, environment variance; V_P, phenotypic variance; and $V_{G\times E}$, genotype × environment variance; GY, grain yield; EarW, ear weight; EarL, ear length; EarD, ear diameter; TipL, tip length; RowN, ear row number; KerN, kernel rows; 100KW, 100-kernel weight; SP, shelling percentage.

3.2. SNP marker data and genetic relationships

The distribution map of SNP density in each chromosome is shown in Fig. 3A. The number of SNPs per chromosome ranged from 1744 on chromosome 10 to 3551 on chromosome 1. The mean heterozygosity rate after filtering across all SNPs was 2.24% and the mean missing rate was 1.32%. The mean heterozygosity rate across all the inbred lines in the SS group was 2.22% and that in the NSS group was 2.26%. The mean heterozygosity rate across all the hybrids formed between the SS and NSS inbred lines was 55.48%. The mean MAF after filtering across all SNPs was 0.19. The mean genetic distance across all the inbred lines calculated using these SNPs was 0.38.

Genetic relationships between the genotyped materials are illustrated in Fig. 3. The values of the first two principal components were 16.2% and 12.2%. Two groups appeared in the PCA plot, and lines from the same heterotic group were clustered, suggesting that the lines adequately represented the genetic diversity of their heterotic groups.

3.3. Prediction accuracy of nine traits using the GBLUP model

Prediction accuracies of the traits are shown in Fig. 4. The prediction accuracies of GY using the GBLUP model were 0.51 and 0.46 in SUS and SPS, respectively. For the remaining yield-related traits, the GBLUP prediction accuracies ranged from 0.49 to 0.86 and from

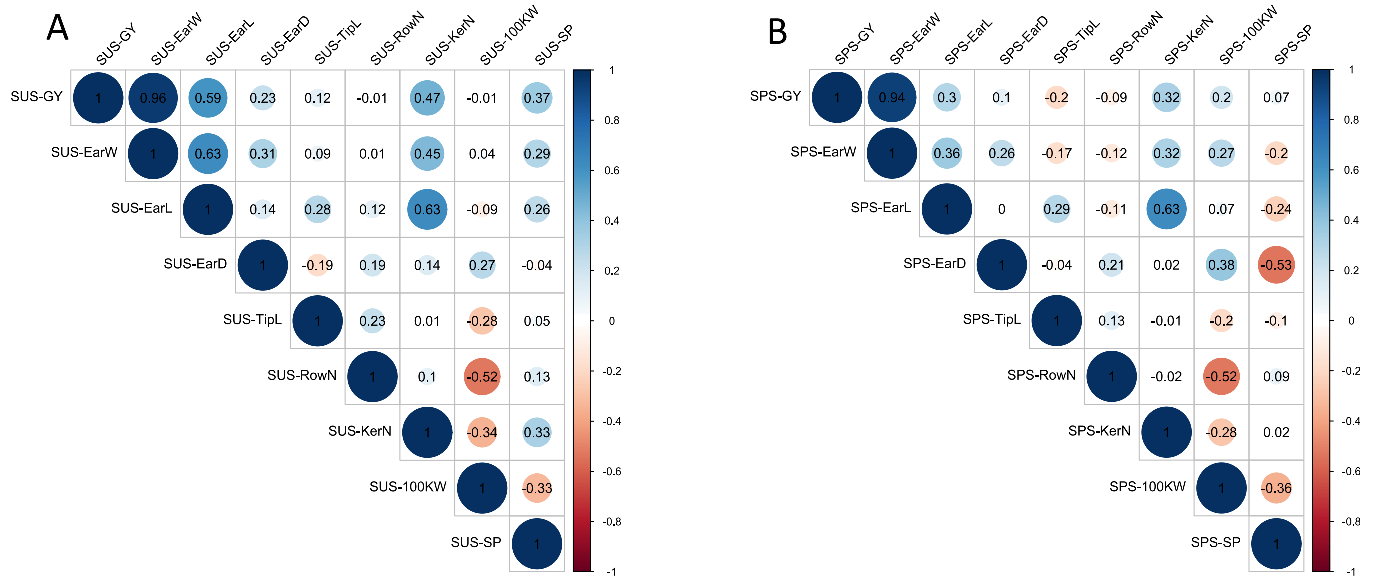

Fig. 2. Scatter plot matrix with Pearson correlations traits using best linear unbiased predictions (BLUPs) and combining the summer maize area, SUS (A), and the spring maize area, SPS (B). GY, grain yield; EarW, ear weight; EarL, ear length; EarD, ear diameter; TipL, tip length; RowN, ear row number; KerN, kernel rows; 100KW, 100-kernel weight; SP, shelling percentage.

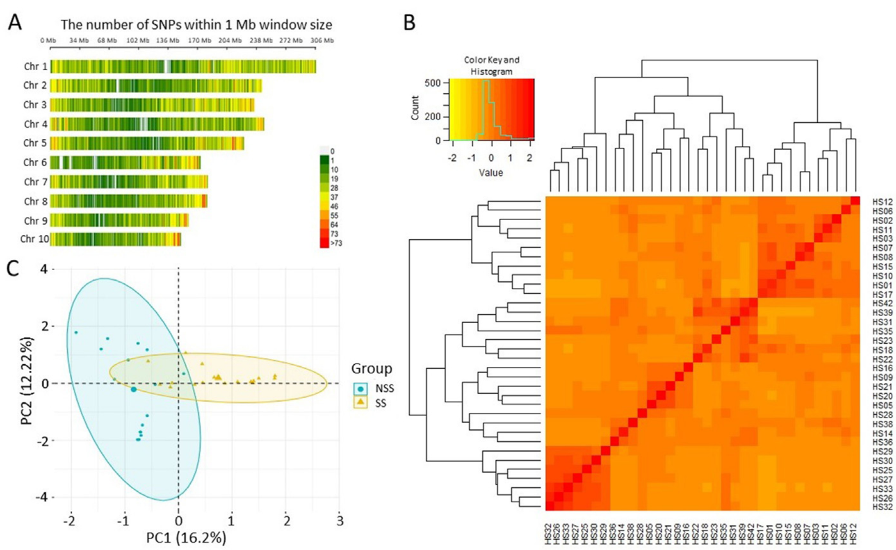

Fig. 3. SNP density distribution map using 23,692 SNPs (A) of 34 parental inbred lines, and heat map of the kinship matrix (B), and genetic relationships illustrated with a principal components analysis (PCA) plot (C).

0.53 to 0.89 in SUS and SPS. The prediction accuracies of all target traits were consistent with their heritabilities. The prediction accuracies of seven traits were higher in SPS than in SUS, but prediction accuracies of GY and 100KW in SUS were slightly higher than those in SPS. The prediction accuracies by environment are presented in Table S3. The prediction accuracy obtained using the phenotypes of a single environment was significantly lower than that from fitting multiple environments.

3.4. Prediction accuracies estimated from the prediction model considering multiple genetic effects

The accuracies of nine traits from the GP model considering additive effects (A) in the 10-fold cross-validation scheme was showed in Fig. 4C. The mean prediction accuracies across all nine traits ranged from 0.49 to 0.87. Compared to GBLUP, the prediction accuracy based on the A model was improved for each trait. The accuracy of the GBLUP model predicting SPS-GY was 0.46, whereas the prediction accuracy of SPS-GY based on the A model was 0.50. However, the prediction accuracies of traits with high heritability did not change significantly.

The model was extended based on the A model to include non-additive effects, ADE (additive + dominant + epistatic effects), the accuracies of nine traits from the ADE model in the 10-fold cross-validation scheme were showed in Fig. 4D. The prediction accuracy of SUS-GY from the ADE model was 0.63, which was 0.17 and 0.12 higher than that from the GBLUP and A model, respectively. The prediction accuracy of SPS-GY from the ADE model was 0.42. The accuracies for the other traits based on the ADE model (0.49–0.91) were not significantly improved. Genetic-effect models differed only marginally from GBLUP with respect to the predictive accuracy for the traits with high heritabilities.

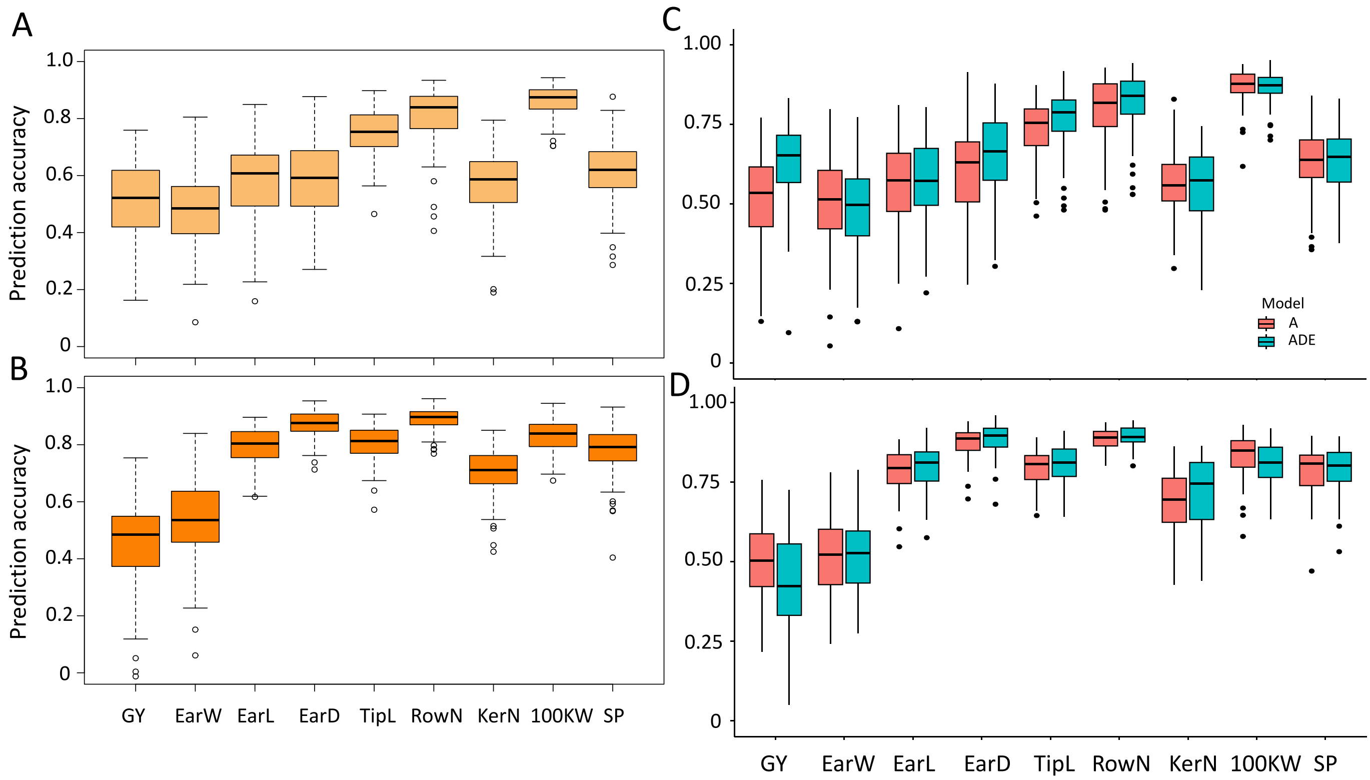

Fig. 4. Prediction accuracies of nine traits of the hybrids, estimated with the GBLUP model and prediction models considering multiple genetic effects (A, ADE) in SUS (A and C) and SPS (B and D).

The prediction accuracy of the genetic-effect model in a single environment was generally comparable to that of the GBLUP model (Table S3).

3.5. G × E and multi-trait prediction models in the hybrid population

To exploit correlation information available from other environments and traits, the prediction accuracy of hybrids using the interaction effect model of genotype and environments was considered, and the multi-trait prediction model was used to predict the hybrid performance of nine traits (Fig. 5A, B). In the G × E model, the prediction accuracies of across all nine traits in SUS and SPS ranged from 0.52 to 0.83 and 0.48 to 0.83, respectively. The prediction accuracies of GY in SUS and SPS were 0.57 and 0.55, respectively, which were greater than those estimated from the GBLUP and genetic effect models (A, ADE). The prediction accuracies of two traits with high heritabilities, KerN and TipL, decreased in the G × E model, and the prediction accuracies of these two traits in the G × E model were respectively only 0.55–0.57 and 0.63–0.73.

Above results also appeared in the multi-trait model, which did improve the prediction accuracy of other traits (Fig. 5A, B). The prediction accuracy of GY in the multi-trait model was higher than that from the GBLUP (0.53) and A models (0.54). The G × E model outperformed the other four models for the prediction accuracy of GY (Figs. 4, 5). The multi-trait model was also used to evaluate the prediction accuracy of nine traits in a single environment. The prediction accuracy was similar to those of the GBLUP, A, and ADE models (Table S3).

3.6. Effect of TRN:TST ratio and population size on prediction accuracy

Based on the superiority of the G × E model in predicting yield, the accuracy of yield prediction under varying TRN:TST ratios and population sizes were evaluated. In scenario 1 (Fig. 5C, D), population size was constant and the TRN population size increased from 28 to 257 over nine ratios of TRN and TST samples. The ratio 1:1 appeared to be a turning point at which the precision and stability of both methods began to increase, indicating that the optimum size of TRN was 50% of the total number of populations used. In scenario 2 (Fig. 5E, F), TRN:TST of 1:1 was constant and population size decreased from 285 to 20. With the decrease in population size, the prediction accuracy dropped from 0.37 to 0.08 (SUS) and from 0.38 to −0.11 (SPS) on average, suggesting that a large TRN population size is required for GP of maize hybrid yield.

3.7. Analysis of evaluation metrics

In scenario 1 (Table 2), TRN:TST of 1:1 seemed to be a turning point for the evaluation of the four indicators PCC, KCC, SCC, and R^2, after which a large change occurred. In scenario 2, when the population size was 220, these four indicators were stable. Given that the MSE index is an evaluation of the applicability of the model, this index does not change much under changing conditions. However, the MSE showed that the G × E model was more suitable for the prediction in SUS (Fig. 5A, B).

4. Discussion

4.1. Implications of GP for hybrid maize breeding based on two divergent heterotic groups, SS and NSS

Genomic selection (GS) has been shown to be a promising genomic tool in various fields, since Meuwissen [3] proposed the concept in 2001. In recent years, GP has been intensively evaluated in maize breeding programs [37–41]. The adoption of GS breeding, however, faces challenges in increasing the predictive power of complex traits. Theorical and empirical findings [42,43] imply that

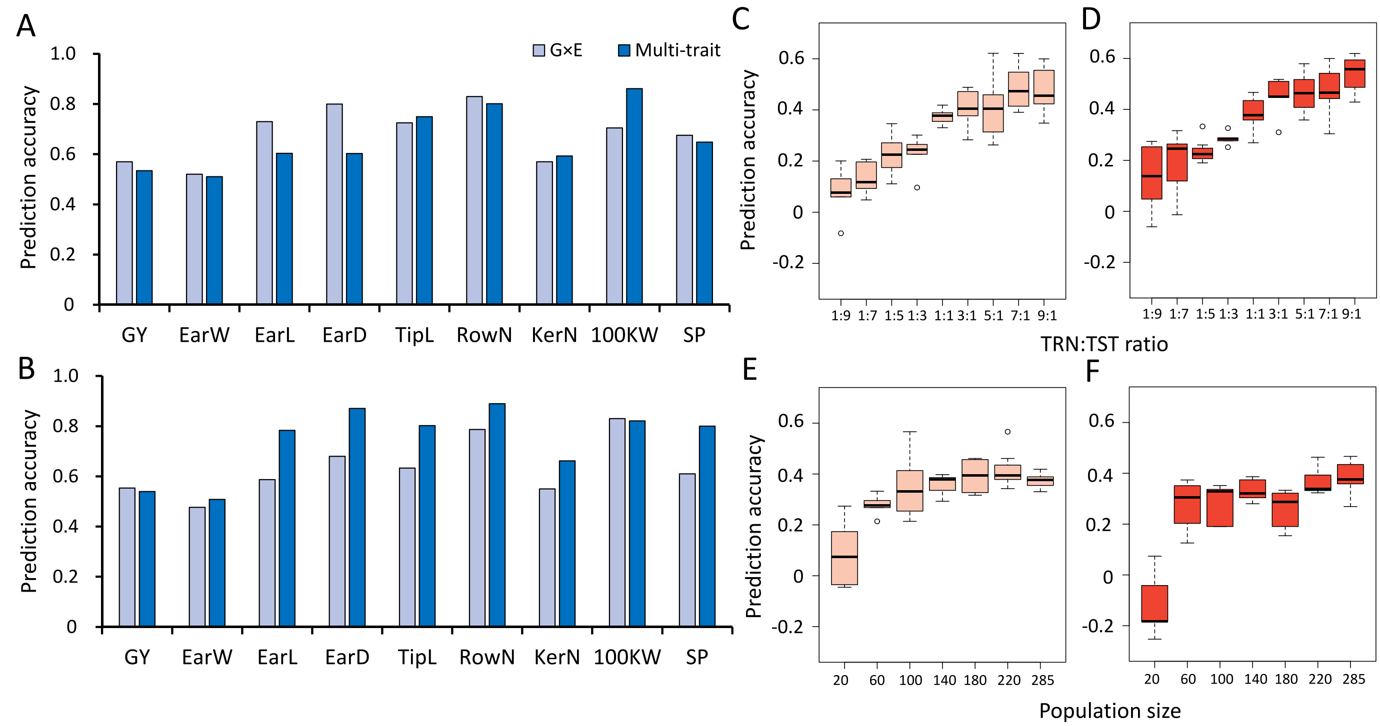

Fig. 5. Prediction accuracies of nine traits of the hybrids estimated with the model of G × E and multiple traits and prediction accuracy of yield using G × E model in multiple scenarios in SUS and SPS. Prediction accuracies of nine traits of the hybrids estimated with the model of G × E and multiple traits in SUS (A) and SPS (B). The population size was 285, the prediction accuracy of the G × E model was compared for nine settings of TRN:TST ratio in SUS (C) and SPS (D). The TRN:TST ratio was 1:1, the prediction accuracy of the G × E model was compared for six settings of population size in SUS (E) and SPS (F).

Table 2
Comparison of evaluation metrics for prediction accuracy of nine TRN:TST ratio designs and seven population sizes.

Item	TRN:TST ratio									Population size						
	1:9	1:7	1:5	1:3	1:1	3:1	5:1	7:1	9:1	20	60	100	140	180	220	285
Summer maize sowing area																
PCC	0.08	0.13	0.22	0.23	0.37	0.41	0.41	0.49	0.57	0.09	0.28	0.34	0.39	0.39	0.36	0.37
KCC	0.02	0.06	0.13	0.15	0.25	0.27	0.26	0.29	0.29	0.03	0.27	0.24	0.22	0.26	0.27	0.25
SCC	0.02	0.08	0.2	0.22	0.36	0.4	0.38	0.42	0.42	0.06	0.36	0.36	0.31	0.34	0.4	0.36
R^2	0.02	0.02	0.06	0.06	0.14	0.16	0.18	0.25	0.24	0.04	0.07	0.08	0.13	0.16	0.14	0.14
MSE	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.03	0.03	0.03
Spring maize sowing area																
PCC	0.03	0.2	0.24	0.29	0.39	0.45	0.46	0.48	0.55	−0.12	0.28	0.29	0.33	0.26	0.36	0.39
KCC	0.05	0.12	0.14	0.25	0.25	0.32	0.32	0.29	0.35	−0.03	0.29	0.27	0.17	0.21	0.21	0.25
SCC	0.07	0.18	0.21	0.37	0.37	0.47	0.47	0.43	0.51	−0.04	0.35	0.34	0.25	0.31	0.31	0.4
R^2	0.03	0.05	0.06	0.17	0.17	0.23	0.22	0.23	0.32	0.01	0.11	0.06	0.12	0.1	0.14	0.17
MSE	0.28	0.27	0.26	0.23	0.23	0.21	0.21	0.22	0.2	0.27	0.22	0.21	0.23	0.22	0.21	0.23

PCC, Pearson correlation coefficient; KCC, Kendall rank correlation coefficient; SCC, Spearman rank correlation coefficient; R^2, Coefficient of determination, R squared; and MSE, Mean squared error.

the accuracy of GP depends mostly on two factors: the accuracy of estimates of SNP effects and the amount of genetic variance at causal variants explained by SNPs in linkage disequilibrium (LD) with these causal variants. Hybrid maize breeding benefits from genetically distinct and complementary heterotic groups [44,45]. This has been shown based on quantitative genetic theory [46] but also empirically [47]. Our study was based on two genetically divergent heterotic groups, SS and NSS (Fig. 3C), which were used for hybrid maize breeding in SUS and SPS. We evaluated the potential and limits of GP of the hybrid performance for this heterotic pattern in a partial diallel cross design based on genotypic and phenotypic data.

The potential of GP for hybrid maize breeding was previously demonstrated in predicting F_1 hybrids between recombinant inbred lines [48] or doubled haploid lines [49] produced from biparental populations. A recent study [50] demonstrated the ability of GP to identify superior single-cross combinations between two heterotic populations. Our results were similar to these and substantiated the ability of GP to identify excellent single-cross combinations. The prediction accuracy of hybrid yield in the partial-diallel hybrid design was higher than in a previous study [51]. Our findings will serve as a basis for continuing to increase training population size and can accumulate years and multiple partial design breeding data to predict the performance of untested hybrid combinations, saving costs, shortening the cycle, and increasing the efficiency of breeding.

4.2. Variance of phenotypic data influences prediction accuracy

GBLUP constructs the kinship G matrix using DNA marker information and calculates the breeding value of GBLUP, one of the commonly used prediction models, for individuals [52]. The prediction accuracy of GBLUP for traits with high heritability was as expected, but the prediction accuracy of traits with low heritability, such as GY and SP, was not satisfactory. When GBLUP was used to predict the same trait in SUS and SPS, the prediction accuracy varied with conditions. Combined with our phenotypic ANOVA results, the prediction accuracy varied with the variance of the trait. The phenotypic variance of SUS was generally smaller than that of SPS, and under the GBLUP model, the prediction accuracy of SPS was higher than that of SUS. The findings that the prediction accuracy of SUS-GY was higher than that of SPS-GY and that the phenotypic variance of SUS-GY was higher than that of SPS-GY show the extent to which the variance affects prediction accuracy. The prediction accuracy of SPS-TipL using the GBLUP model was 0.81, while the prediction accuracy of SPS-TipL based on the G $\times$ E model was 0.63. The interaction variance of 0.09 between genotype and environment in SPS-TipL indicates that G $\times$ E interaction variance also affects the prediction accuracy of the G $\times$ E model.

4.3. Modeling of non-additive effects improves predictive accuracy of traits

Previous studies [53,54] have shown that simulating genetic effects in maize prediction, such as dominance and epistasis, may improve prediction accuracy, reducing breeding costs and achieving genetic gains. We found that non-additive effect models increased prediction accuracy, except for SUS-EW, SPS-GY, and SPS-100KW. Studies in hybrid wheat [17], rapeseed (*Brassica napus* L.) [20], and maize [16,48] also showed that the non-additive effect model improved prediction accuracy, in contrast to the findings [5] for GP in rice hybrid performance. The prediction accuracies of the non-additive model for SUS-EW, SPS-GY, and SPS-100KW were lower than those of the additive model. This disparity may be due to the limited information provided by the non-additive variance, and increased noise of the prediction models, leading to decreased prediction accuracy. Thus, prediction accuracy can be improved by inclusion of non-additive effects in the model.

4.4. Value of extending the prediction model

Multi-trait models could be useful in establishing better genomic selection strategies when the objective is to anticipate challenging or expensive qualities that are connected with cheaply evaluated secondary traits. A multi-trait model could improve the prediction performance for low-heritability characters when there is some correlation between them and high-heritability traits [55–57]. Adding the interaction effect of genotype and environment can also improve prediction accuracy. Our prediction accuracy was improved, especially for yield by expanding the model and adding multiple traits and G $\times$ E effects. Although the multi-trait approach has evident advantages, larger data sets and greater computational power are needed because there are more factors that need to be estimated (such as genetic and error covariances), which could reduce the precision of genomic prediction [53,58]. When these techniques are used to increase the accuracy of genomic prediction, additional elements such as genetic correlation, heritability, training population composition, and quantity of data sets should be taken into account. Some of our results have similar problems. The prediction accuracy of SPS-EarL and SPS-TipL using the G $\times$ E model was lower than that of GBLUP. However, the prediction accuracy of low-heritability traits was improved by expanding the model.

CRediT authorship contribution statement

Ping Luo: Methodology, Software, Investigation, Writing-original draft preparation, Writing-review and editing. **Xinhai Li:** Conceptualization, Methodology, Writing-review and editing, Project administration. **Zhuanfang Hao:** Conceptualization, Methodology, Writing-review and editing, Project administration. **Xuecai Zhang:** Conceptualization, Methodology, Writing-original draft preparation, Writing-review and editing. **Junjie Fu:** Conceptualization, Methodology, Writing-original draft preparation, Writing-review and editing, Software. **Houwen Wang:** Investigation, Methodology, Writing-original draft preparation. **Zhiyong Ni:** Writing-original draft preparation, Methodology. **Ruisi Yang, Fei Wang, Hongjun Yong, Lin Zhang, Wei Song, Jie Yang, Yong Chen, Liwei Wang, Runze Zhang, Xiaoyu Chen, Meng Zhao, Wenwei Gao and Wenjie Li:** Investigation. **Mingshun Li, Jianfeng Weng, Zhiqiang Zhou, Degui Zhang, Jienan Han, Zhaodong Meng:** Methodology, Writing-review and editing. All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (32272049, 32261143757), Sustainable Development International Cooperation Program from Bill & Melinda Gates Foundation (2022YFAG1002), the National Key Research and Development Program of China (2020YFE0202300), the Agricultural Science & Technology Innovation Program (CAAS-ZDRW202109), and the China Scholarship Council. The funding agencies were not involved in the design of the study; the collection, analysis, and interpretation of the data; or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data for this article can be found online at https://doi.org/10.1016/j.cj.2023.09.009.

References

[1] B.M. Prasanna, J.E. Cairns, P.H. Zaidi, Y. Beyene, D. Makumbi, M. Gowda, C. Magorokosho, M. Zaman-Allah, M. Olsen, A. Das, M. Worku, J. Gethi, B.S. Vivek, S.K. Nair, Z. Rashid, M.T. Vinayan, A.B. Issa, F. San Vicente, T. Dhliwayo, X. Zhang, Beat the stress: breeding for climate resilience in maize for the tropical rainfed environments, Theor. Appl. Genet. 134 (2021) 1729–1752.

[2] S.A. Prado, C.G. López, M.L. Senior, L. Borrás, The genetic architecture of maize (*Zea mays* L.) kernel weight determination, G3-Genes Genomes Genet. 4 (2014) 1611–1621.

[3] T.H. Meuwissen, B.J. Hayes, M.E. Goddard, Prediction of total genetic value using genome-wide dense marker maps, Genetics 157 (2001) 1819–1829.

[4] J.M. Hickey, T. Chiurugwi, I. Mackay, W. Powell, Implementing Genomic Selection in Implementing Genomic Selection in CGIAR Breeding Programs Workshop Participants, Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery, Nat. Genet. 49 (2017) 1297–1303.

[5] 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.

[6] J. Crossa, P. Perez-Rodriguez, J. Cuevas, O. Montesinos-Lopez, D. Jarquin, G. de Los Campos, J. Burgueno, J.M. Gonzalez-Camacho, S. Perez-Elizalde, Y. Beyene, S. Dreisigacker, R. Singh, X. Zhang, M. Gowda, M. Roorkiwal, J. Rutkoski, R.K. Varshney, Genomic selection in plant breeding: methods, models, and perspectives, Trends Plant Sci. 22 (2017) 961–975.

[7] X. Wang, Y. Xu, Z. Hu, C. Xu, Genomic selection methods for crop improvement: current status and prospects, Crop J. 6 (2018) 330–340.

[8] J. Crossa, L. Campos Gde, P. Perez, D. Gianola, J. Burgueno, J.L. Araus, D. Makumbi, R.P. Singh, S. Dreisigacker, J. Yan, V. Arief, M. Banziger, H.J. Braun, Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers, Genetics 186 (2010) 713–724.
[9] J. Crossa, Y. Beyene, S. Kassa, P. Perez, J.M. Hickey, C. Chen, G. de los Campos, J. Burgueno, V.S. Windhausen, E. Buckler, J.L. Jannink, M.A. Lopez Cruz, R. Babu, Genomic prediction in maize breeding populations with genotyping-by-sequencing, G3-Genes Genomes Genet. 3 (2013) 1903–1926.
[10] R. Bernardo, Prediction of maize single-cross performance using RFLPs and information from related hybrids, Crop Sci. 34 (1994) 20–25.
[11] D.N. Duvick, Biotechnology in the 1930s: the development of hybrid maize, Nat. Rev. Genet. 2 (2001) 69–74.
[12] N.M. Springer, R.M. Stupar, Allelic variation and heterosis in maize: how do two halves make more than a whole?, Genome Res 17 (2007) 264–275.
[13] Y. Zhao, J. Zeng, R. Fernando, J.C. Reif, Genomic prediction of hybrid wheat performance, Crop Sci. 53 (2013) 802–810.
[14] Y. Zhao, Z. Li, G. Liu, Y. Jiang, H.P. Maurer, T. Würschum, H.-P. Mock, A. Matros, E. Ebmeyer, R. Schachschneider, Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) 15624–15629.
[15] Y. Jiang, R.H. Schmidt, Y. Zhao, J.C. Reif, A quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat, Nat. Genet. 49 (2017) 1741–1746.
[16] F. Technow, C. Riedelsheimer, T.A. Schrag, A.E. Melchinger, Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects, Theor. Appl. Genet. 125 (2012) 1181–1194.
[17] Y. Zhao, Z. Li, G. Liu, Y. Jiang, H.P. Maurer, T. Wurschum, H.P. Mock, A. Matros, E. Ebmeyer, R. Schachschneider, E. Kazman, J. Schacht, M. Gowda, C.F. Longin, J. C. Reif, Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) 15624–15629.
[18] J. Dudley, G. Johnson, Epistatic models improve prediction of performance in corn, Crop Sci. 49 (2009) 763–770.
[19] C.F. Azevedo, M.D.V. de Resende, F.F. e Silva, J.M.S. Viana, M.S.F. Valente, M.F.R. Resende, P. Muñoz, Ridge, Lasso and Bayesian additive-dominance genomic models, BMC Genet. 16 (2015) 1–13.
[20] P. Liu, Y. Zhao, G. Liu, M. Wang, D. Hu, J. Hu, J. Meng, J.C. Reif, J. Zou, Hybrid performance of an immortalized F_2 rapeseed population is driven by additive, dominance, and epistatic effects, Front. Plant Sci. 8 (2017) 815.
[21] X. Liu, H. Wang, X. Hu, K. Li, Z. Liu, Y. Wu, C. Huang, Improving genomic selection with quantitative trait loci and nonadditive effects revealed by empirical evidence in maize, Front. Plant Sci. 10 (2019) 1129.
[22] R.E. Lorenzana, R. Bernardo, Accuracy of genotypic value predictions for marker-based selection in biparental plant populations, Theor. Appl. Genet. 120 (2009) 151–161.
[23] G. Su, O.F. Christensen, T. Ostersen, M. Henryon, M.S. Lund, Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers, PLoS ONE 7 (2012) e0045293.
[24] Y. Jiang, J.C. Reif, Modeling epistasis in genomic selection, Genetics 201 (2015) 759–768.
[25] H. Zhang, L. Yin, M. Wang, X. Yuan, X. Liu, Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations, Front. Genet. 10 (2019) 189.
[26] O.A. Montesinos-López, A. Montesinos-López, J.C. Montesinos-López, J. Crossa, F.J. Luna-Vázquez, J. Salinas-Ruiz, A Bayesian multiple-trait and multiple-environment model using the matrix normal distribution, in: M.A. El-Esawi (Ed.), Physical Methods for Stimulation of Plant and Mushroom Development, IntechOpen, London, UK, 2018, pp. 19–33.
[27] E. Pollak, J. van der Werf, R. Quaas, Selection bias and multiple trait evaluation, J. Dairy Sci. 67 (1984) 1590–1595.
[28] L. Schaeffer, Sire and cow evaluation under multiple trait models, J. Dairy Sci. 67 (1984) 1567–1580.
[29] S. Yan, A. Loladze, N. Wang, S. Sun, M.I. Chilvers, M. Olsen, J. Burgueño, C.D. Petroli, T. Molnar, F. San Vicente, X. Zhang, B.M. Prasanna, Association mapping of resistance to tar spot complex in maize, Plant Breed. 141 (2022) 745–755.
[30] D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4, J. Stat. Softw. 67 (2015) 1–48.
[31] A.R. Hallauer, M.J. Carena, J.d. Miranda Filho, Quantitative Genetics in Maize Breeding, Springer, New York, USA, 2010.
[32] Z. Guo, H. Wang, J. Tao, Y. Ren, C. Xu, K. Wu, C. Zou, J. Zhang, Y. Xu, Development of multiple SNP marker panels affordable to breeders through genotyping by target sequencing (GBTS) in maize, Mol. Breed. 39 (2019) 37.
[33] P.J. Bradbury, Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, E.S. Buckler, TASSEL: software for association mapping of complex traits in diverse samples, Bioinformatics 23 (2007) 2633–2635.
[34] P.M. Vanraden, Genomic measures of relationship and inbreeding, Interbull Bull. 37 (2007) 33.
[35] G. Covarrubias-Pazaran, Genome-assisted prediction of quantitative traits using the R package sommer, PLoS ONE 11 (2016) e0156744.
[36] B. Lado, D. Vazquez, M. Quincke, P. Silva, I. Aguilar, L. Gutierrez, Resource allocation optimization with multi-trait genomic prediction for bread wheat (*Triticum aestivum* L.) baking quality, Theor. Appl. Genet. 131 (2018) 2719–2731.
[37] N. Wang, H. Wang, A. Zhang, Y. Liu, D. Yu, Z. Hao, D. Ilut, J.C. Glaubitz, Y. Gao, E. Jones, M. Olsen, X. Li, F. San Vicente, B.M. Prasanna, J. Crossa, P. Perez-Rodriguez, X. Zhang, Genomic prediction across years in a maize doubled haploid breeding program to accelerate early-stage testcross testing, Theor. Appl. Genet. 133 (2020) 2869–2879.
[38] Y. Xu, Y. Ma, X. Wang, C. Li, X. Zhang, P. Li, Z. Yang, C. Xu, Kernel metabolites depict the diversity of relationship between maize hybrids and their parental lines, Crop J. 9 (2021) 181–191.
[39] A. Zhang, P. Pérez-Rodríguez, F. San Vicente, N. Palacios-Rojas, T. Dhliwayo, Y. Liu, Z. Cui, Y. Guan, H. Wang, H. Zheng, M. Olsen, B.M. Prasanna, Y. Ruan, J. Crossa, X. Zhang, Genomic prediction of the performance of hybrids and the combining abilities for line by tester trials in maize, Crop J. 10 (2022) 109–116.
[40] A. Zhang, S. Chen, Z. Cui, Y. Liu, Y. Guan, S. Yang, J. Qu, J. Nie, D. Dang, C. Li, X. Dong, J. Fan, Y. Zhu, X. Zhang, J. Crossa, H. Cao, Y. Ruan, H. Zheng, Genomic prediction of drought tolerance during seedling stage in maize using low-cost molecular markers, Euphytica 218 (2022) 154.
[41] X. Zhang, P. Perez-Rodriguez, J. Burgueno, M. Olsen, E. Buckler, G. Atlin, B.M. Prasanna, M. Vargas, F. San Vicente, J. Crossa, Rapid cycling genomic selection in a multiparental tropical maize population, G3-Genes Genomes Genet. 7 (2017) 2315–2326.
[42] G. de Los Campos, D. Gianola, D.B. Allison, Predicting genetic predisposition in humans: the promise of whole-genome markers, Nat. Rev. Genet. 11 (2010) 880–886.
[43] M. Goddard, Genomic selection: prediction of accuracy and maximisation of long term response, Genetica 136 (2009) 245–257.
[44] Y. Zhao, M. Gowda, W. Liu, T. Wurschum, H.P. Maurer, F.H. Longin, N. Ranc, J.C. Reif, Accuracy of genomic selection in European maize elite breeding populations, Theor. Appl. Genet. 124 (2012) 769–776.
[45] D.C. Kadam, S.M. Potts, M.O. Bohn, A.E. Lipka, A.J. Lorenz, Genomic prediction of single crosses in the early stages of a maize hybrid breeding pipeline, G3-Genes Genomes Genet. 6 (2016) 3443–3453.
[46] J.C. Reif, F.M. Gumpert, S. Fischer, A.E. Melchinger, Impact of interpopulation divergence on additive and dominance variance in hybrid populations, Genetics 176 (2007) 1931–1934.
[47] R. Desper, O. Gascuel, Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle, J. Comput. Biol. 9 (2002) 687–705.
[48] T. Guo, H. Li, J. Yan, J. Tang, J. Li, Z. Zhang, L. Zhang, J. Wang, Performance prediction of F_1 hybrids between recombinant inbred lines derived from two elite maize inbred lines, Theor. Appl. Genet. 126 (2013) 189–201.
[49] Y. Beyene, M. Gowda, M. Olsen, K.R. Robbins, P. Perez-Rodriguez, G. Alvarado, K. Dreher, S.Y. Gao, S. Mugo, B.M. Prasanna, J. Crossa, Empirical comparison of tropical maize hybrids selected through genomic and phenotypic selections, Front. Plant. Sci. 10 (2019) 1502.
[50] G. Li, Y. Dong, Y. Zhao, X. Tian, T. Würschum, J. Xue, S. Chen, J.C. Reif, S. Xu, W. Liu, Genome-wide prediction in a hybrid maize population adapted to Northwest China, Crop J. 8 (2020) 830–842.
[51] G. Yu, Y. Cui, Y. Jiao, K. Zhou, X. Wang, W. Yang, Y. Xu, K. Yang, X. Zhang, P. Li, Z. Yang, Y. Xu, C. Xu, Comparison of sequencing-based and array-based genotyping platforms for genomic prediction of maize hybrid performance, Crop J. 11 (2023) 490–498.
[52] G. Covarrubias-Pazaran, B. Schlautman, L. Diaz-Garcia, E. Grygleski, J. Polashock, J. Johnson-Cicalese, N. Vorsa, M. Iorizzo, J. Zalapa, Multivariate GBLUP improves accuracy of genomic selection for yield and fruit weight in biparental populations of *Vaccinium macrocarpon* ait, Front. Plant. Sci. 9 (2018) 1310.
[53] X. Zhang, P. Perez-Rodriguez, K. Semagn, Y. Beyene, R. Babu, M.A. Lopez-Cruz, F. San Vicente, M. Olsen, E. Buckler, J.L. Jannink, B.M. Prasanna, J. Crossa, Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs, Heredity 114 (2015) 291–299.
[54] M. Ali, L. Zhang, I. DeLacy, V. Arief, M. Dieters, W.H. Pfeiffer, J. Wang, H. Li, Modeling and simulation of recurrent phenotypic and genomic selections in plant breeding under the presence of epistasis, Crop J. 8 (2020) 866–877.
[55] N. Budhlakoti, D.C. Mishra, A. Rai, S.B. Lal, K.K. Chaturvedi, R.R. Kumar, A comparative study of single-trait and multi-trait genomic selection, J. Comput. Biol. 26 (2019) 1100–1112.
[56] O.A. Montesinos-Lopez, A. Montesinos-Lopez, J. Crossa, F.H. Toledo, O. Perez-Hernandez, K.M. Eskridge, J. Rutkoski, A genomic bayesian multi-trait and multi-environment model, G3-Genes Genomes Genet. 6 (2016) 2725–2744.
[57] Y. Jia, J.L. Jannink, Multiple-trait genomic selection methods increase genetic value prediction accuracy, Genetics 192 (2012) 1513–1522.
[58] A.J. Lorenz, K.P. Smith, Adding genetically distant individuals to training populations reduces genomic prediction accuracy in barley, Crop Sci. 55 (2015) 2657–2667.