THE CROP JOURNAL 8 (2020) 819–829

Available online at www.sciencedirect.com

ScienceDirect

Using genomic data to improve the estimation of general combining ability based on sparse partial diallel cross designs in maize

Xin Wang[a,b], *Zhenliang Zhang*[b], *Yang Xu*[b], *Pengchen Li*[b], *Xuecai Zhang*[c], *Chenwu Xu*[b,*]

[a]*College of Information Engineering, Yangzhou University, Yangzhou 225009, Jiangsu, China*
[b]*Key Laboratory of Plant Functional Genomics of the Ministry of Education/Jiangsu Key Laboratory of Crop Genomics and Molecular Breeding/ Jiangsu Co-Innovation Center for Modern Production Technology of Grain Crops, College of Agriculture, Yangzhou University, Yangzhou 225009, Jiangsu, China*
[c]*International Maize and Wheat Improvement Center (CIMMYT), Mexico D.F. 06600, Mexico*

ARTICLE INFO

Article history:
Received 29 December 2019
Received in revised form 3 May 2020
Accepted 31 May 2020
Available online 3 July 2020

ABSTRACT

Evaluation of general combining ability (GCA) is crucial to hybrid breeding in maize. Although the complete diallel cross design can provide an efficient estimation, sparse partial diallel cross (SPDC) is more flexible in breeding practice. Using real and simulated data sets of partial diallel crosses between 266 maize inbred lines, this study investigated the performance of SPDC designs for estimating the GCA. With different distributions of parental lines involved in crossing (called random, balanced and unbalanced samplings), different numbers of hybrids were sampled as the training sets to estimate the GCA of the 266 inbred lines. In this process, three statistical approaches were applied. One obtained estimations through the ordinary least square (OLS) method, and the other two utilized genomic prediction (GP) to estimate the GCA. It was found that the coefficient of determination of each approach was always higher than the heritability of a target trait, showing that the GCA for maize inbred lines could be accurately predicted with SPDC designs. Both the GP approaches were more accurate than the OLS, particularly in the scenario for a low-heritability trait with a small sample size. Additionally, prediction results demonstrated that a big sample of hybrids could greatly help improve the accuracy. The random sampling of parental lines had little influence on the average accuracy. However, the prediction for lines that never or seldom involved in crossing might suffer from much lower accuracy.

© 2020 Crop Science Society of China and Institute of Crop Science, CAAS. Publishing services 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 author.*
E-mail address: cwxu@yzu.edu.cn (C. Xu).

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

https://doi.org/10.1016/j.cj.2020.04.012

2214-5141 / © 2020 Crop Science Society of China and Institute of Crop Science, CAAS. Publishing services 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/).

1. Introduction

The concept of general combining ability (GCA), originally defined by Sprague and Tatum [1], refers to the average performance of a line in hybrid combinations. It can be estimated using the difference between the average of its hybrids and the general average for all crossings. GCA is mainly a measure of additive effects, which can be directly transmitted from parents to offspring. In recent years, doubled haploid (DH) technology has facilitated the generation of a large number of inbred lines, making GCA evaluation the major bottleneck in hybrid maize breeding [2]. Therefore, the evaluation of GCA is a crucial process for hybrid development, and the maize line with high GCA is an essential component for producing elite hybrids [3].

Estimation of GCA can be obtained easily using the complete diallel cross design or North Carolina Design II (NC II) [4,5]. However, with the development of breeding programs, a great quantity of inbred lines are available. The number of possible crosses grows very rapidly, making these designs time and resource intensive. Partial diallel cross designs in which only a subset of possible crosses is performed are more attractive options [6]. They allow the evaluation of a greater number of inbred lines in crosses [7]. In the classic circulant designs [8], each of n lines is only crossed with s other lines, instead of n–1 lines as in the complete diallel. In this way, each line is guaranteed to be involved in the sampling crosses. Analysis using circulant diallels significantly reduces the number of crosses in which each genitor is involved, and enables the participation of a greater number of genitors [9]. However, in breeding practice, only a small proportion of hybrids are possible to be evaluated in the field, making the partial diallel table very sparse. This means that the average value of s is small, and some complex factors of field trials may bring big fluctuation to s for the parental lines. In this scenario, limited research has been reported to estimate the GCA. So, there is an urgent need to develop procedures to allow the accurate evaluation of GCA based on sparse partial diallel cross (SPDC) designs [6].

Previous studies [10–12] have proposed many statistical methods for diallel analysis. Among them, diallel mating models proposed by Griffing [10] are the most widely used models for investigating the genetic parameters (general and specific combining ability). For a partial diallel cross scheme, traditional studies [13] used the ordinary least squares (OLS) to estimate GCA. However, if the diallel table is very sparse, the great reduction in the ratio of observed hybrids will be striking. Using the traditional OLS to evaluate many inbred lines with a small sample of hybrids can't ensure high accuracy.

Fortunately, with the advances of molecular biology, breeders can accurately understand the genetic structure of breeding populations, and thus greatly improve the estimation of genetic parameters using genomic prediction (GP). Estimated breeding values based on the genotypes of individuals were remarkably accurate [14,15]. Some studies [16,17] used GS models for directly predicting agronomic traits in inbred lines, and some others [18–20] used GS for predicting hybrid performance. Various methods, such as Genomic best linear unbiased prediction (GBLUP), Bayes, the least absolute shrinkage and selection operator (LASSO) and machine learning [21–24] have been developed for GP, and these differ with respect to assumptions about the marker effects. GBLUP and RR-BLUP [21] assign identical variance to all loci and essentially treat all of them as equally important. In BayesA, markers are assumed to have different variances and follows a posterior scaled inverse chi-square distribution [22]. The prior in BayesB assumes that the variance of markers is equal to zero with probability π, and the complement with probability $1-\pi$ follows a scaled inverse chi-square distribution [22]. In Bayes Cπ, the mixture probability π has a prior uniform distribution [25]. LASSO is a popular method for regression that uses a penalty to achieve a sparse solution, and it is somewhat indifferent to closely correlated markers and tends to pick one and ignore the others [26]. Machine learning is also an alternative for GS. It has been employed to enhance the prediction of genetic values for wheat and maize [27,28]. Although various models have been successfully applied to GS, some studies [29–31] showed that not much variation in prediction accuracy among the different models was observed.

In terms of the GCA analyses and hybrid prediction, Bernardo [32–34] was one of the first to advocate the BLUP model [35] in maize. From then on, many studies have been reported to analyze GCA especially when predicting hybrids. For predicting the GCA of a maize testcross population, Riedelsheimer et al. [2] proposed a genomic selection (GS) approach based on RR-BLUP, showing that more efficient predictive procedures could be developed using genomic data. With a linear mixed model using the ASReml-R software [36], Kadam et al. [37] evaluated random inbreds derived from biparental families of maize. Using GBLUP and BayesB, Technow et al. [38] investigated genome properties of the parental line based on the Dent × Flint heterotic pattern. Greenberg et al. [6] developed a hierarchical Bayesian model to estimate quantitative genetic parameter from partial diallel cross designs. Werner et al. [39] considered GCA and specific combining ability (SCA) to apply RR-BLUP and Bayesian models for predicting hybrid performance in oilseed rape using a collection of 220 paternal DH lines and five male-sterile inbred lines. Recently, using GBLUP and a complete diallel cross design with twenty-eight single-crosses formed between eight parental lines, Velez-Torres et al. [40] concluded that GS is a more effective and efficient approach to predict the GCA of maize lines compared with phenotyping method. However, in most of the previous studies, more attention has been paid to the prediction accuracy of hybrids, and the prediction of parental GCA is not the focus. For an SPDC maize population, few studies have been reported to systematically investigate the GP accuracy for GCA.

The purpose of this study is to assess the efficacy of GP for estimating the GCA of maize inbred lines with SPDC systems. As mentioned earlier, the accuracy of various GS models is similar. Considering that GBLUP is more suitable for quantitative traits influenced by polygenes and its high computational efficiency [15], GBLUP was adopted in this study. Using genome-wide SNPs called from a real maize data set of 266 inbred lines, genetic and phenotypic values of all possible hybrids were simulated. And thus different hybrid sample

sizes and different distributions of parental lines involved in crossing were investigated. Such a scheme was implemented to assess the efficacy of GP in estimating the GCA of lines. Then, the utility of statistical approaches was illustrated with an example using an actual trait of maize. The methods we described would be useful in various sets of maize and other crops.

2. Materials and methods

2.1. Materials

2.1.1. Plant materials

The models were fitted to the maize data set from Yangzhou University. Partial diallel crossings between a total of 266 maize inbred lines were performed during the 2017 and 2018 maize growing seasons from field trials on the experimental farms in Yangzhou and Taian, China, and two replicates were made in each environment. Regardless of reciprocals, ear weight (EW) for 945 hybrids were collected to estimate the GCA of the inbred lines. The phenotypes fitting the statistical model were the average performance of each hybrid from two environments. The 266 inbred lines were genotyped, and 319,668 SNPs evenly distributed on chromosomes were called initially. 61,468 genome-wide SNPs were filtered by eliminating the heterozygosity >0.05 and the missing rate > 0.05. Genotypes of the hybrids were inferred from SNPs of their inbred parents.

2.1.2. Simulations

Based on the genotypes mentioned above, a large number of simulations were performed to assess the prediction accuracy of different statistical approaches. Considering additive and dominant genetic effects of markers, several traits of all the 35,245 possible hybrids with 200 QTLs and different heritabilities were randomly simulated. The numbers of QTLs on chromosome 1–10 are 29, 32, 21, 25, 21, 12, 15, 21, 9, and 15, respectively. According to the study of Meuwissen et al. [22], the additive effects of the 200 QTLs were drawn from a gamma distribution with shape parameter $\alpha = 0.4$ and scale parameter $\beta = 1.66$. Half of the additive effects had positive effects and the other half had negative effects. The dominant effects were determined as the product of the absolute additive effects and the degree of dominance, which was drawn from a normal distribution with mean and variance equal to 0.193 and 0.312^2, respectively. In the 200 QTLs, two loci on chromosome 1 and chromosome 3 were found to have weak over-dominant effects. For all the simulated hybrids, the ratio of dominant variance to additive variance was 0.160. Normal independent error deviations with variances calculated were added to meet assumed heritability, and the simulated phenotypes were centered and standardized to unit variance. Finally, three traits with heritability of 0.7, 0.5, and 0.3, named T7, T5, and T3 were obtained.

2.1.3. Sampling designs

Two types of sampling designs were performed to evaluate the difference in predictive accuracy. First, for each trait of the 35,245 hybrids, different numbers of hybrids (m = 500, 1000, and 2000) were sampled as the training sets to estimate the GCA of the 266 parental lines. Second, with a certain sample size of hybrids mentioned above, three distributions of parental lines were designed. One was similar to a circulant diallel table. In this paper it was called balanced sampling, which meant that all parental lines were involved in nearly an equal number of crosses (designated s). Another was called random sampling, which meant that the crossing times of all the lines were random but at least 1. The third was called unbalanced sampling, which meant that only part of the 266 lines (n = 200 or 150) were involved in random crossing. Each sampling method was randomly repeated 20 times to obtain the average results of 20 replicates. The density plot for the crossing times of the 266 lines derived from the 20 samplings with different methods is shown in Fig. 1. Taking the first round of sampling with m = 500 as an example, the detailed sampling scheme is illustrated in Table S1. In this study, for the random samplings, the crossing times (s) of each inbred line was classified into three levels to compare the accuracy for line subsets with different s value. About a quarter of the lines with the lowest s value were defined as low-frequency; a quarter of the lines with the highest s value were defined as high-frequency; the middle half was defined as medium-frequency.

2.2. Methods

The Griffing Model [10] was used for analyzing our SPDC schemes:

$$y_{ij} = \mu + g_i + g_j + s_{ij} + \varepsilon_{ij}$$

where y_{ij} is the phenotypic value of the hybrids between line i and line j ($i, j = 1, 2, \ldots, n$); μ is the overall mean; g_i and g_j are the GCA effects of the ith parent and the jth parent, respectively; s_{ij} is the SCA effect for the cross between the ith and jth parents; and ε_{ij} is the random error effect.

Based on the above model, three statistical approaches were applied. One obtained estimations through the OLS and the other two utilized GP to predict the GCA. In our simulations, knowing the phenotypes of all the possible hybrids, the true GCA of all the 266 inbred lines can be calculated with its definition. The coefficient of determination (the squared Pearson correlation coefficient) between the true GCA and predicted GCA was adopted to evaluate the accuracy of different statistical approaches.

2.2.1. OLS approach

In the matrix form, the vector for m observed hybrids can be represented by:

$$\mathbf{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where $\mathbf{y}$ is an $m \times 1$ vector, $\boldsymbol{X}$ is an $m \times (m + n + 1)$ matrix of coefficients, and $\boldsymbol{\beta}$ is an $(m + n + 1) \times 1$ vector, defined as $(\mu, g_1, g_2, \ldots, g_n, s_1, s_2, \ldots, s_m)^{\mathrm{T}}$ which includes the overall mean, GCA and SCA effects. $\boldsymbol{\varepsilon}$ is an $m \times 1$ random error vector. The restrictions $\sum_{i=1}^{n} g_i = 0$ and $\sum_{k=1}^{m} s_k = 0$ were imposed on the combining ability effects by combining $\begin{bmatrix} 1\ldots1 & 0 \\ 0 & 1\ldots1 \end{bmatrix}$ with the $\boldsymbol{X}$ matrix by rows, and combining

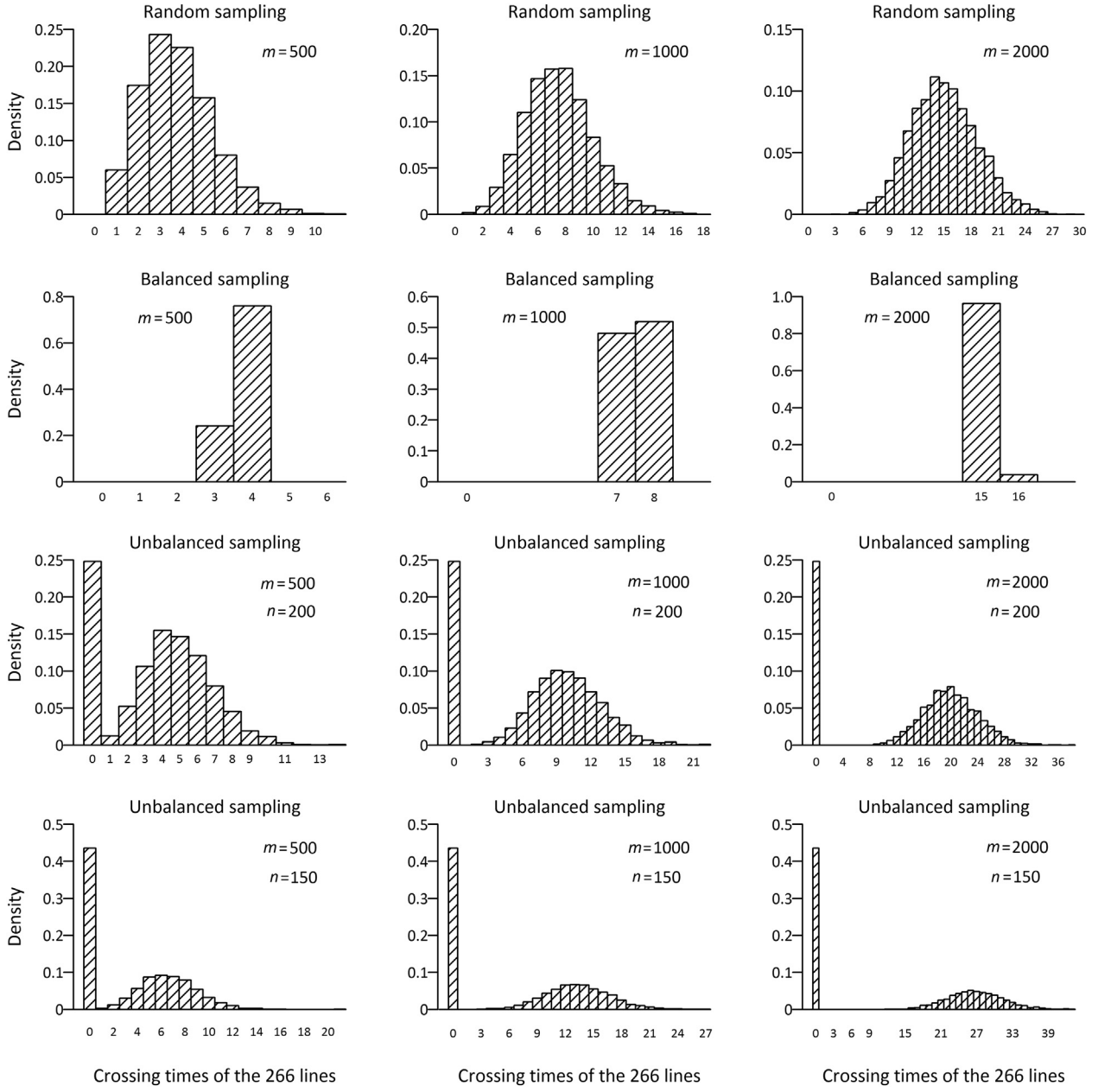

Fig. 1 – Density plot for the crossing times of the 266 inbred lines derived from 20 rounds of random, balanced and unbalanced samplings, respectively.

$(0,0)^{\mathrm{T}}$ with $\mathbf{y}$ by columns. Thus, the OLS solutions could be achieved using the equation: $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\mathbf{y})$. Because the diallel tables in our research were very sparse, $\boldsymbol{X}'\boldsymbol{X}$ was always a singular matrix. In this study, the Moore-Penrose generalized inverse was adopted to calculate $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ using the R package MASS [41].

2.2.2. GP approach

GBLUP is an efficient method using whole-genome markers to predict genetic values and phenotypes of interest. It exploits the genomic relationships between training population and testing population to predict the genomic values for unknown individuals [42]. In this study, two GP approaches using GBLUP were performed for predicting the GCA of the 266 lines. One was designated GP-I, which used the Griffing Model to directly estimate the GCA. The other, designated GP-II, estimated the GCA by predicting the phenotypes of all possible hybrids.

The model of GP-I can be described as:

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{Z}_{g(1)}\,\boldsymbol{g}_{(1)} + \mathbf{Z}_{g(2)}\,\boldsymbol{g}_{(2)} + \mathbf{Z}_{s}\,\mathbf{s} + \boldsymbol{\varepsilon},$$

where $\mathbf{y}$ is an $m \times 1$ vector for hybrid observations; μ is the overall mean; $\mathbf{1}$ is a vector of ones. For each hybrid, $\boldsymbol{g}_{(1)}$ and $\boldsymbol{g}_{(2)}$ are vectors for the GCA effects of two parents, respectively; $\mathbf{s}$ is the vector for the SCA effects, and $\boldsymbol{\varepsilon}$ is the vector for random residuals. $\mathbf{Z}_{g(1)}$, $\mathbf{Z}_{g(2)}$, $\mathbf{Z}_{s}$ are incidence matrices related to $\boldsymbol{g}_{(1)}$, $\boldsymbol{g}_{(2)}$ and $\mathbf{s}$. It is assumed that $\boldsymbol{g}_{(1)} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{G}_{(1)}\sigma^2_{(1)})$, $\boldsymbol{g}_{(2)} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{G}_{(2)}\sigma^2_{(2)})$,

$\boldsymbol{s}\sim N(\boldsymbol{0}, \boldsymbol{G_s}\sigma_s^2)$, $\boldsymbol{\varepsilon}\sim N(\boldsymbol{0}, \boldsymbol{I_m}\sigma_\varepsilon^2)$, where $\sigma_{(1)}^2$ and $\sigma_{(2)}^2$ are the variances of the GCA effects for two parents, respectively; σ_s^2 is the variance of SCA effects; σ_ε^2 is the residual variance; $\boldsymbol{I_m}$ is an $m \times m$ identity matrix; $\boldsymbol{G_{(1)}}$ and $\boldsymbol{G_{(2)}}$ are the additive genetic relationship matrices for two parents, respectively; $\boldsymbol{G_s}$ is the non-additive genetic relationship matrix for the SCA effects. Using information from the genome-wide 61,836 SNPs, $\boldsymbol{G_{(1)}}$, $\boldsymbol{G_{(2)}}$ and $\boldsymbol{G_s}$ were defined as $(\boldsymbol{M_{(1)}M_{(1)}'})/k_{(1)}$, $(\boldsymbol{M_{(2)}M_{(2)}'})/k_{(2)}$ and $(\boldsymbol{M_sM_s'})/k_s$, respectively. $\boldsymbol{M_{(1)}}$ and $\boldsymbol{M_{(2)}}$ are $m \times q$ matrices (q is the number of markers) for two parents which specify genetic values at each locus (–1, 0 and 1 for genotypes A_1A_1, A_1A_2 and A_2A_2, respectively). $\boldsymbol{M_s}$ is an $m \times q$ matrix for hybrids which specifies dummy variables that take value 1 for heterozygous and 0 for homozygous loci. $k_{(1)}$, $k_{(2)}$ and k_s were calculated as trace$(\boldsymbol{M_{(1)}M_{(1)}'})/m$, trace$(\boldsymbol{M_{(2)}M_{(2)}'})/m$ and trace$(\boldsymbol{M_sM_s'})/m$, respectively. Using GBLUP, for each observed hybrid, $\boldsymbol{g_{(1)}}$ and $\boldsymbol{g_{(2)}}$ were predicted. Then arithmetic mean of the predicted GCA of each line was calculated.

The model of GP-II can be described as:

$$\boldsymbol{y} = \mu\boldsymbol{1} + \boldsymbol{Z_a}\,\boldsymbol{a} + \boldsymbol{Z_d}\,\boldsymbol{d} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y}$ is an $m \times 1$ vector for hybrid observations; μ is the overall mean; $\boldsymbol{1}$ is a vector of ones. $\boldsymbol{a}$ is the vector for additive genetic effects of hybrids, $\boldsymbol{d}$ is the vector for dominance effects of hybrids. $\boldsymbol{Z_a}$ and $\boldsymbol{Z_d}$ are incidence matrices related to $\boldsymbol{a}$ and $\boldsymbol{d}$, respectively. It is assumed that $\boldsymbol{a}\sim N(\boldsymbol{0}, \boldsymbol{G_a}\sigma_a^2)$, $\boldsymbol{d}\sim N(\boldsymbol{0}, \boldsymbol{G_d}\sigma_d^2)$ and $\boldsymbol{\varepsilon}\sim N(\boldsymbol{0}, \boldsymbol{I_m}\sigma_\varepsilon^2)$ where σ_a^2 is the additive genetic variance; σ_d^2 is the dominance variance; σ_ε^2 is the residual variance; $\boldsymbol{G_a}$ and $\boldsymbol{G_d}$ are the additive and dominance genetic relationship matrices, respectively. They were defined as $(\boldsymbol{M_aM_a'})/k_a$ and $(\boldsymbol{M_dM_d'})/k_d$, respectively. $\boldsymbol{M_a}$ is an $m \times q$ matrix for hybrids which specifies genetic values (–1, 0 and 1 for genotypes A_1A_1, A_1A_2 and A_2A_2, respectively) at each locus. $\boldsymbol{M_d}$ is an $m \times q$ matrix for hybrids which specifies dummy variables that take value 1 for heterozygous and 0 for homozygous loci. k_a and k_d were calculated as trace$(\boldsymbol{M_aM_a'})/m$ and trace$(\boldsymbol{M_dM_d'})/m$, respectively. The proportions of the variances (σ_a^2, σ_d^2, σ_ε^2) to the total variance ($\sigma^2 = \sigma_a^2 + \sigma_d^2 + \sigma_\varepsilon^2$) were defined as $h_a^2 = \sigma_a^2/\sigma^2$, $h_d^2 = \sigma_d^2/\sigma^2$ and $h_\varepsilon^2 = \sigma_\varepsilon^2/\sigma^2$, respectively. Using GBLUP, for all the possible hybrids, $\boldsymbol{a}$ and $\boldsymbol{d}$ were predicted, and then the predicted phenotypes was derived from the sum of $\boldsymbol{a}$ and $\boldsymbol{d}$. Finally, the predicted GCA of each line could be calculated with the definition. The variance ratios h_a^2, h_d^2 and h_ε^2 were estimated using restricted maximum likelihood (REML). In GP-I, the ratios of $\sigma_{(1)}^2$ and $\sigma_{(2)}^2$ to the total variance were set equal to $0.5h_a^2$, and the ratio of σ_s^2 to the total variance was set equal to h_d^2. The REML algorithm and the GBLUP model were performed using the R program [43].

3. Results

3.1. Comparison of the methods for estimating the GCA

For the traits T7, T5 and T3, based on random and balanced samplings, the prediction results of OLS, GP-I and GP-II using different sample sizes were compared in Table 1. In each case, the trait T7 had the highest accuracy, while the trait T3 had the lowest accuracy, showing that the estimation of GCA was largely dependent on heritability.

It is notable that the accuracy of each approach was always higher than the heritability of a target trait, showing that the GCA for maize lines can be accurately predicted with SPDC designs. This character is very beneficial to genetic improvement in breeding practice. For T7, derived from the random sampling i.e., when the sample size was 500, GP-II gave the highest accuracy (0.8068). It was 4.7% higher than that of GP-I (0.7709) and 9.6% higher than that of OLS (0.7361); when the sample sizes were 1000 and 2000, GP-II gave the highest accuracy as before. At this time, the accuracy of GP-I was only slightly higher than that of OLS and the advantage of GP-II was smaller. For T5, when the sample sizes were 500 and 1000, the statistical approaches showed the similar pattern increasing from OLS to GP-II. When the sample size was 2000, the accuracy of GP-I was once again slightly higher than that of OLS. For T3, no matter which sample size was adopted, GP-II gave much more accurate GCA than GP-I and OLS, and the accuracies obtained by GP-I always substantially exceeded that of OLS.

Obviously, the sample size had a great influence on the accuracy. For all the three traits, statistical approaches with the sample size of 2000 provided the highest accuracies, followed by those with the sample size of 1000, reflecting that a big sample size could substantially contribute to the prediction. Additionally, the sample size affected the significance of GP approaches over OLS. As mentioned above, the smaller the sample size, the higher the level of significance was, showing that GP is particularly beneficial for the estimation of GCA with a sparse partial diallel table. No matter which statistical approach was used, when comparing the random and balanced samplings, their accuracies were

Table 1 – Comparison of accuracy using OLS, GP-I and GP-II with different sampling designs and training set sizes.

Trait	Sample size	OLS		GP-I		GP-II	
		Random sampling	Balanced sampling	Random sampling	Balanced sampling	Random sampling	Balanced sampling
T7	500	0.7361	0.7435	0.7709	0.7685	0.8068	0.8112
	1000	0.8602	0.8711	0.8653	0.8740	0.8891	0.8995
	2000	0.9364	0.9281	0.9340	0.9361	0.9443	0.9444
T5	500	0.5617	0.5758	0.6496	0.6477	0.6856	0.6926
	1000	0.7432	0.7562	0.7790	0.7869	0.8068	0.8149
	2000	0.8704	0.8679	0.8761	0.8776	0.8921	0.8960
T3	500	0.3634	0.3893	0.4659	0.4901	0.5185	0.5369
	1000	0.5668	0.5773	0.6553	0.6551	0.6869	0.6862
	2000	0.7423	0.7434	0.7814	0.7902	0.8021	0.8112

not significantly different, showing that the random sampling commonly used in breeding practice had little impact on the average accuracy of GCA.

In brief, GP-II performed the best in our research and would be a promising approach for estimating the GCA of maize and other crops.

3.2. Accuracy with different numbers of inbred lines involved in crossing

Sometimes, due to insufficient material resources or experimental budget, the number of inbred lines involved in crossing is limited. Therefore, in this study, the effect of line quantity on prediction was explored. To simplify the problem, only the GP-II approach that had the highest accuracy in the previous experiments was investigated. The accuracy for the 266 lines with different numbers of lines (n = 150, 200, and 266) involved in random crossing was plotted against the sampling number of hybrids (Fig. 2). As expected, for each of the three traits, the accuracy was higher for the sampling designs with 266 lines over the sampling designs with 200 and 150 lines. For T7, when the sample size was 500, 1000 and 2000, the highest accuracy obtained with n = 266 was 22.9%, 25.2%, and 30.4% higher than that with n = 150, respectively; For T5, when the sample size was 500, 1000 and 2000, the highest accuracy obtained with n = 266 was 20.8%, 22.3%, and 27.7% higher than that with n = 150, respectively; For T3, when the sample size was 500, 1000 and 2000, the highest accuracy obtained with n = 266 was 14.0%, 17.5%, and 24.4% higher than that with n = 150, respectively. It was clear that the advantage of more inbred lines involved in crossing could be brought into full play when predicting a high-heritability trait with a big sample size of hybrids.

3.3. Accuracy for subsets of inbred lines

The results of above research demonstrated that no significant differences were found between the accuracies of the random and balanced samplings. However, for the random sampling, inbred lines involved in crossing with different frequency must provide different amounts of information. Therefore, taking GP-II for instance, accuracy for lines involved in crossing with different levels of frequency derived from the random sampling were demonstrated in Table 2. It was shown that the lines involved in crossing with high-frequency always got the highest accuracy, and those with low-frequency performed the worst. For predicting T3, with the sample size of 500, the high-frequency lines achieved the biggest increase (55.3%) in accuracy relative to the low-frequency lines. That is to say, although the random sampling brings us similar accuracy to the balanced sampling, the prediction for low-frequency lines may suffer from much lower accuracy, especially for a low-heritability trait with a small sample size of hybrids.

Unbalanced sampling design was also concerned in this research. The accuracies of GP-II for lines involved and not involved in crossing are summarized in Table 3. As expected, the accuracies were always much higher for the involved lines over the non-involved lines. In most cases, the accuracy increases with the sample size. It is interesting to note that the accuracy of non-involved lines is always less sensitive to the sample size than that of the involved lines. In particular

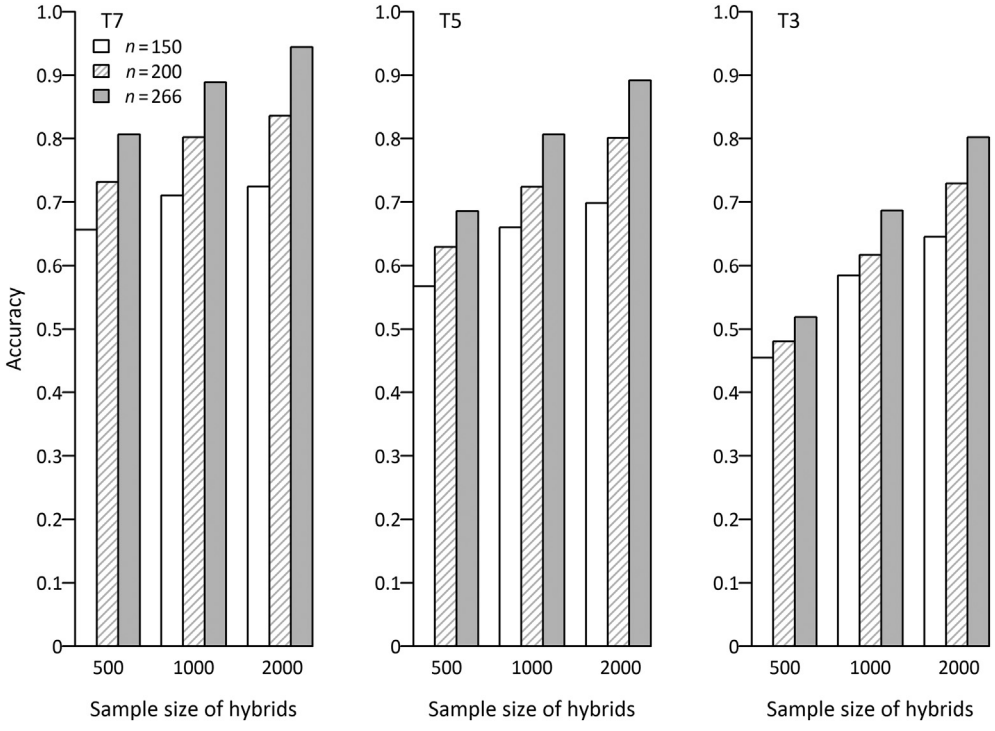

Fig. 2 – Accuracy of GP-II for the 266 inbred lines with different numbers of lines (n = 150, 200, or 266) involved in the random crossing.

Table 2 – Accuracy of GP-II for inbred lines involved in crossing with different levels of frequency.

Trait	Sample size	Lines involved in crossing with low-frequency	Lines involved in crossing with medium-frequency	Lines involved in crossing with high-frequency
T7	500	0.7304	0.8041	0.8675
	1000	0.8372	0.8932	0.9285
	2000	0.9242	0.9463	0.9592
T5	500	0.5882	0.6730	0.7802
	1000	0.7173	0.8196	0.8530
	2000	0.8649	0.8942	0.9129
T3	500	0.4081	0.4954	0.6337
	1000	0.6087	0.6870	0.7581

for T7, the accuracy of the 166 non-involved lines with m = 2000 (0.4105) was 6.4% higher than that with m = 500 (0.3858), while the advantage of the 150 involved lines was 10.5%; in the same way, the accuracy of the 66 non-involved lines with m = 2000 (0.4594) was only 4.3% higher than that with m = 500 (0.4405), while the advantage of the 200 involved lines was 14.5%. For T5 and T3, the pattern of the accuracy was similar to that for T7.

In each case, accuracy of involved lines with n = 200 was lower than that with n = 150. A good interpretation was that the involved lines with n = 150 has higher frequency in crossing, which showed again the contribution of high-frequency to prediction. On the contrary, accuracies of non-involved lines with n = 200 were almost higher than that with n = 150. The reason may be that the advantage of high-frequency cannot work on the non-involved lines, and a sample of hybrids derived from more lines can benefit the prediction.

Table 3 – Accuracy of GP-II for inbred lines involved and not involved in crossing derived from the unbalanced samplings.

Trait	Number of lines involved in crossing	Number of lines not involved in crossing	Sample size	Accuracy for involved lines	Accuracy for non-involved lines
T7	150	116	500	0.8755	0.3858
			1000	0.9359	0.4143
			2000	0.9676	0.4105
	200	66	500	0.8359	0.4405
			1000	0.9182	0.4578
			2000	0.9570	0.4594
T5	150	116	500	0.7511	0.3405
			1000	0.8746	0.3790
			2000	0.9359	0.3925
	200	66	500	0.7166	0.3869
			1000	0.8337	0.3978
			2000	0.9164	0.4454
T3	150	116	500	0.6002	0.2745
			1000	0.7662	0.3486
			2000	0.8699	0.3564
	200	66	500	0.5549	0.2689
			1000	0.7081	0.3416
			2000	0.8349	0.4007

3.4. Prediction results for EW of maize

In addition to the simulated studies, actual EW of 945 hybrids were used to estimate the GCA of the 266 inbred lines. The predicted GCA values were sorted in descending order and the prediction results of the top 20 inbred lines are demonstrated in Table 4. As shown, eight common varieties, including B93, B57, B108, B74, B214, B275, B254, and B167 were screen out by all the three statistical approaches. It is noteworthy that the coefficient of determination between the predicted GCA of GP-II and OLS was the highest (0.8703), and up to sixteen common varieties were selected by the two approaches simultaneously, indicating that the intersection between the results of GP-II and OLS was more reliable in this scenario. Moreover, the absolute GCA values predicted by GP-I were lower than those predicted by OLS and GP-II. The reason may be that the non-additive effects were excluded from the estimates of GCA when using GP-I. Although this exclusion may not affect the predictive accuracy, the loss of non-additive variance will inevitably reduce the absolute values of the predicted GCA.

4. Discussion

With a complete diallel cross scheme, the GCA of inbred line can be easily calculated with its definition [1]. However, because of the large number of possible crosses, Kempthorne and Curnow [7] suggested the partial diallel cross to evaluate inbred lines. Each of the n lines are crossed with s other lines, and there will be $ns/2$ crosses in the whole set. In this scenario, although the OLS was widely used for estimating the GCA [13,44], genomic information was ignored. GS uses all molecular markers for predicting the performance of the candidates, and it has shown tangible genetic gains in maize breeding [14]. Recently, Alves et al. [45] pointed out that GS can be used to estimate genetic parameters accurately in maize hybrids. For maize inbred lines, there is no reason to doubt the advantage of GS over the phenotype-based OLS. Comparisons of the three approaches in the present study have shown that GP-II and GP-I are superior to OLS, showing that prediction with genomic data can help improve the estimation of GCA for inbred lines based on SPDC designs.

Riedelsheimer et al. [2] crossed 285 diverse Dent inbred lines with two testers and predicted the GCA using genomic and metabolic information. Predictive accuracies (the Pearson correlation) ranged from 0.72 to 0.81 for GP, which are similar to ours for T3 and T5. However, a fivefold cross-validation scheme was applied in their prediction and the predicted GCA values of one subset were estimated using the observed GCA values of the other four subsets. In our research, the situation was quite different. The partial diallel table was assumed to be very sparse. Even if the hybrid sample size is equal 2000, the ratio of the sample size (2000) to the possible hybrid size (35,245) is only 5.7%, making it impossible to obtain a GCA observation of any line. In other words, our prediction is based on the SPDC. The sampling hybrids are treated as the training set, and their number is often lager than the number of lines

Table 4 – The top 20 inbred lines of the GCA for EW using OLS, GP-I, and GP-II.

Top 20	OLS		GP-I		GP-II	
	Serial number	Estimated GCA	Serial number	Estimated GCA	Serial number	Estimated GCA
1	B093	193.7	B254	152.3	B057	183.4
2	B079	185.9	B275	148.6	B068	144.4
3	B057	180.2	B057	136.3	B093	139.5
4	B108	171.7	B218	125.5	B131	136.0
5	B074	170.9	B214	122.0	B108	134.9
6	B053	160.4	B241	121.4	B003	133.1
7	B106	147.2	B263	118.4	B016	130.7
8	B214	146.7	B271	118.4	B052	129.8
9	B052	146.0	B167	118.3	B074	129.4
10	B275	140.6	B093	116.5	B079	129.0
11	B063	138.8	B264	115.9	B117	126.7
12	B254	137.7	B211	114.9	B053	125.0
13	B047	136.2	B208	109.4	B167	123.0
14	B167	135.3	B226	106.9	B028	121.8
15	B010	128.4	B117	106.7	B006	121.0
16	B016	125.9	B233	99.4	B047	120.0
17	B003	124.9	B074	98.6	B254	118.6
18	B071	124.7	B270	96.7	B275	116.4
19	B037	124.7	B108	96.4	B106	115.4
20	B068	123.1	B276	95.8	B214	114.3

with identified GCA. Maybe this is the reason why our accuracies are slightly higher than those of Riedelsheimer et al. [2].

Population design plays a vital role in breeding programs, and partial diallel cross is preferable in many cases. For instance, Miranda Filho and Vencovsky [13] estimated the GCA for ear length of maize in a partial diallel cross with $n = 10$ and $s = 3$. Reis et al. [44] estimated the genetic parameters using a partial circulant diallel cross design with $n = 34$ (two groups of parents) and different sizes of s (from 2 to 5). Analysis using cross-validation process showed that the accuracy increased as the value of s increased. Our balanced and unbalanced sampling designs just mimicked the partial diallel cross scheme. With $n = 266$, the hybrid sample size was set to 500, 1000, and 2000, respectively. Correspondingly, the mean of s was 3.8, 7.5, and 15.0. Note that s was much less than n - 1. The inadequate phenotypic information of each parental line in crossing couldn't guarantee the accurate estimation of GCA. Although Vivas et al. [9] declared that it is possible to obtain good agreement (correlation coefficient above 0.8) with $s = 3$, our accuracies with $s = 3.8$ are still lower than those with $s = 7.5$ and $s = 15.0$. When evaluating the efficiency of the circulant diallel, Veiga et al. [46] pointed out that it is advantageous to increase the s value for a low-heritability trait. In our research, with the sample size of 500 for predicting T3, accuracies of OLS were substantially lower than those of GP-I and GP-II, showing that the reduction in the s value decreased the potential accuracy of OLS. However, it is gratifying that the GP approaches using genomic information have been demonstrated to partly compensate the "small s" problem.

The unbalanced sampling design was also worthy of attention in breeding practice. Because of experimental cost and complex factors in field trails, involving all inbred lines in crossing is impossible. Previous studies [19,47] have adopted the strategy of predicting untested single-cross hybrids in maize and rice. However, few GS studies have been undertaken in predicting the GCA of the lines that never participate in crossing. In this case, the traditional phenotype-based OLS is impracticable. Our research has demonstrated that the GCA of the non-involved lines could also be estimated using GP approaches based on SPDC designs. This strategy allows a reliable selection of more inbred lines for their potential to create superior hybrids. But on the other hand, lines never or seldom involved in crossing were found to have lower accuracy than the involved lines. In this respect, our results are in agreement with those reported by Fristche-Neto et al. [48] who have indicated that the number of parents and the crosses per parent in the training sets should be maximized when predicting maize hybrid performance.

Empirical studies [31,49,50] have shown that heritability of target traits and training population size are two important factors affecting the accuracy. In our research, the accuracies always increase with the heritability and the sample size, which was in agreement with previous theoretical results. Reduction of the error effects may also help increase the accuracy. A previous study [43] has shown that the trouble of predicting a low-heritability trait is the significant standard deviation caused by errors. In the present study, for predicting the GCA, the accuracies of GP-II were not only higher than the heritability of target traits, but also higher than the accuracies of hybrid prediction (Table 5), indicating that the prediction of GCA may be more effective than that of hybrids. According to the study of Daetwyler et al. [50], the Pearson correlation coefficient between the true breeding values and estimated breeding values ($r_{g\hat{g}}$) is a function of heritability (h^2). They derived the equation: $r_{g\hat{g}} = \sqrt{\frac{N_p h^2}{N_p h^2 + M_e}}$, where N_p is the number of individuals, and M_e is the number of independent

Table 5 – Accuracy of GP-II for the GCA of 266 inbred lines and for all the potential hybrids with the random samplings.

Trait	Sample size	Accuracy for GCA	Accuracy for hybrids
T7	500	0.8068	0.7754
	1000	0.8891	0.8558
	2000	0.9443	0.9106
T5	500	0.6856	0.6555
	1000	0.8068	0.7742
	2000	0.8921	0.8560
T3	500	0.5185	0.4955
	1000	0.6869	0.6536
	2000	0.8021	0.7594

chromosome segments. Furthermore, Resende et al. [51] pointed out that the Pearson correlation coefficient between the true phenotypic values and estimated phenotypic values of GS can be expressed as $r_{y\hat{y}} = r_{g\hat{g}}h$. Theoretically, $r_{g\hat{g}}$ is less than 1, and $r_{y\hat{y}}$ is less than h. In our manuscript, the accuracy is the squared Pearson correlation coefficient between the true GCA and predicted GCA. By definition, GCA is essentially the average performance of a line in hybrid combinations, and the average of predicted genetic values of the possible hybrids greatly eliminated the influence of errors. This factor may make the accuracy between $r^2_{y\hat{y}}$ and $r^2_{g\hat{g}}$, so the accuracy in our manuscript was higher than the heritability of the trait. This result agrees well with the fact that the accuracy of GS excluding error effects can be much higher than the predictive ability containing error effects [31,50].

Maize breeding involves two critical steps, developing superior inbred lines from breeding populations and identifying elite combinations of two inbred lines [52]. With the development of DH and other technologies, breeders have been able to develop a large number of inbred lines which need to be evaluated by their performance in crosses. However, the number of potential crosses grows rapidly, making the field evaluation of hybrid performance time and resource intensive. Sparse partial diallel tables are becoming more and more common in breeding practice. GCA is mainly a measure of additive effects, and it is in response to selection of inbred lines. Accurate prediction of the GCA will enhance the efficiency of inbred line selection, and then accelerate the hybrid breeding. In actual breeding projects, especially for the scenario with only datasets based on SPDC designs, breeders can evaluate their inbred lines using the methods proposed in the present study. Then, for different heterotic groups, top inbred lines can be selected, and a few corresponding testers can be used to perform the validation by field trials. We believe that in this way, the efficiency and accuracy of maize breeding can be improved.

Previous studies [3,53,54] have used maize introgression lines or recombination inbred lines to perform testcrosses for detecting significant loci of GCA. In all these experiments, phenotypes of hybrids were observed and the true GCA value could be calculated with certainty. However, in most breeding programs, only a small part of possible hybrids can be identified in field trials, and thus the detection work can't be carried out. In such cases, methods proposed in the present study guarantee a reliable estimation for the GCA, providing an opportunity for further detection of significant loci. This strategy may open up a promising research direction for inbred line selection in maize and other crops.

Supplementary data for this article can be found online at https://doi.org/10.1016/j.cj.2020.04.012.

Declaration of competing interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2016YFD0100303), the National Natural Science Foundation of China (31801028, 31902101), the Open Research Fund of State Key Laboratory of Hybrid Rice (Wuhan University) (KF201701), the Science and Technology Innovation Fund Project in Yangzhou University (2019CXJ052) and the Priority Academic Program Development of Jiangsu Higher Education Institutions.

Author contributions

Xin Wang performed the analysis and wrote the paper. Zhenliang Zhang, Yang Xu, and Pengchen Li collected the data. Xuecai Zhang assisted with the analysis. Chenwu Xu conceived and designed the analysis.

REFERENCES

[1] G.F. Sprague, L.A. Tatum, General vs. specific combining ability in single crosses of corn, Agron. J. 34 (1942) 923–932.
[2] C. Riedelsheimer, A. Czedik-Eysenberg, C. Grieder, J. Lisec, F. Technow, R. Sulpice, T. Altmann, M. Stitt, L. Willmitzer, A.E. Melchinger, Genomic and metabolic prediction of complex heterotic traits in hybrid maize, Nat. Genet. 44 (2012) 217–220.
[3] J. Huang, H. Qi, X. Feng, Y. Huang, L. Zhu, B. Yue, General combining ability of most yield-related traits had a genetic basis different from their corresponding traits per se in a set of maize introgression lines, Genetica 141 (2013) 453–461.
[4] H.A. Sadalla, M.O. Barznji, S.A. Kakarash, Full diallel crosses for estimation of genetic parameters in maize, Iraqi J. Agric. Sci. 48 (2017) 30–40.
[5] X.M. Fan, Y.D. Zhang, D.P. Jeffers, Y.Q. Bi, M.S. Kang, X.F. Yin, Combining ability of yellow lines derived from CIMMYT populations for use in subtropical and tropical midaltitude maize production environments, Crop Sci. 58 (2018) 169–179.
[6] A.J. Greenberg, S.R. Hackett, L.G. Harshman, A.G. Clark, A hierarchical Bayesian model for a novel sparse partial diallel crossing design, Genetics 185 (2010) 361–373.
[7] O. Kempthorne, R. Curnow, The partial diallel cross, Biometrics 17 (1961) 229–250.
[8] O. Kempthorne, A class of experimental designs using blocks of two plots, Ann. Math. Statist. (1953) 76–84.
[9] M. Vivas, S.F. Silveira, A.P. Viana, A.T. Amaral, D.L. Cardoso, M.G. Pereira, Efficiency of circulant diallels via mixed models

in the selection of papaya genotypes resistant to foliar fungal diseases, Gen. Mol. Res. 13 (2014) 4797–4804.
[10] B. Griffing, Concept of general and specific combining ability in relation to diallel crossing systems, Aust. J. Biol. Sci. 9 (1956) 463–493.
[11] J. Jinks, B. Hayman, The theory and analysis of diallel crosses, Genetics 43 (1953) 63–85.
[12] C. Gardner, S. Eberhart, Analysis and interpretation of the variety cross diallel and related populations, Biometrics (1966) 439–452.
[13] J.B. Miranda Filho, R. Vencovsky, The partial circulant diallel cross at the interpopulation level, Genet. Mol. Biol. 22 (1999) 249–255.
[14] 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.C. 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.
[15] X. Wang, Y. Xu, Z. Hu, C. Xu, Genomic selection methods for crop improvement: current status and prospects, Crop J. 6 (2018) 330–340.
[16] J. Crossa, G. de los Campos, P. Perez, D. Gianola, J. Burgueno, J. Luis 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.
[17] J. Crossa, Y. Beyene, S. Kassa, P. Pérez, J.M. Hickey, C. Chen, G. de los Campos, J. Burgueño, V.S. Windhausen, E. Buckler, Genomic prediction in maize breeding populations with genotyping-by-sequencing, G3-Genes Genomics Genet. 3 (2013) 1903–1926.
[18] 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.
[19] 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.
[20] X. Wang, Y. Xu, P. Li, M. Liu, C. Xu, Z. Hu, Efficiency of linear selection index in predicting rice hybrid performance, Mol. Breed. 39 (2019) 77.
[21] P.M. VanRaden, Efficient methods to compute genomic predictions, J. Dairy Sci. 91 (2008) 4414–4423.
[22] 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.
[23] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. B (1996) 267–288.
[24] O. González-Recio, G.J. Rosa, D. Gianola, Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits, Livest. Sci. 166 (2014) 217–231.
[25] D. Habier, R.L. Fernando, K. Kizilkaya, D.J. Garrick, Extension of the Bayesian alphabet for genomic selection, BMC Bioinformatics 12 (2011) 186.
[26] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, J. Stat. Software 33 (2010) 1–22.
[27] G. de Los Campos, D. Gianola, G.J. Rosa, K.A. Weigel, J. Crossa, Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods, Genet. Res. 92 (2010) 295–308.
[28] J. González-Camacho, G. de Los Campos, P. Pérez, D. Gianola, J. Cairns, G. Mahuku, R. Babu, J. Crossa, Genome-enabled prediction of genetic values using radial basis function neural networks, Theor. Appl. Genet. 125 (2012) 759–771.
[29] G. Moser, B. Tier, R.E. Crump, M.S. Khatkar, H.W. Raadsma, A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers, Genet. Sel. Evol. 41 (2009) 56.
[30] H.H. Neves, R. Carvalheiro, S.A. Queiroz, A comparison of statistical methods for genomic selection in a mice population, BMC Genet. 13 (2012) 100.
[31] X. Wang, Z. Yang, C. Xu, A comparison of genomic selection methods for breeding value prediction, Sci. Bull. 60 (2015) 925–935.
[32] R. Bernardo, Prediction of maize single-cross performance using RFLPs and information from related hybrids, Crop Sci. 34 (1994) 20–25.
[33] R. Bernardo, Genetic models for predicting maize single-cross performance in unbalanced yield trial data, Crop Sci. 35 (1995) 141–147.
[34] R. Bernardo, Best linear unbiased prediction of maize single-cross performance, Crop Sci. 36 (1996) 50–56.
[35] C.R. Henderson, Best linear unbiased estimation and prediction under a selection model, Biometrics 31 (1975) 423–447.
[36] D.G. Butler, B.R. Cullis, A.R. Gilmour, B.J. Gogel, Mixed Models for S Language Environments: ASReml-R Reference Manual, Queensland Department of Primary Industries and Fisheries, Australia, 2009.
[37] 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 Genomics Genet. 6 (2016) 3443–3453.
[38] F. Technow, T.A. Schrag, W. Schipprack, E. Bauer, H. Simianer, A.E. Melchinger, Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize, Genetics 197 (2014) 1343–1355.
[39] C.R. Werner, L. Qian, K.P. Voss-Fels, A. Abbadi, G. Leckband, M. Frisch, R.J. Snowdon, Genome-wide regression models considering general and specific combining ability predict hybrid performance in oilseed rape with similar accuracy regardless of trait architecture, Theor. Appl. Genet. 131 (2018) 299–317.
[40] M. Velez-Torres, J.J. Garcia-Zavala, M. Hernandez-Rodriguez, R. Lobato-Ortiz, J.J. Lopez-Reynoso, I. Benitez-Riquelme, J.A. Mejia-Contreras, G. Esquivel-Esquivel, J.D. Molina-Galan, P. Perez-Rodriguez, X.C. Zhang, Genomic prediction of the general combining ability of maize lines (*Zea mays* L.) and the performance of their single crosses, Plant Breed. 137 (2018) 379–387.
[41] W.N. Venables, B.D. Ripley, Modern Applied Statistics with S, 4th ed Springer, New York, USA, 2002.
[42] Y. Xu, X. Wang, X. Ding, X. Zheng, Z. Yang, C. Xu, Z. Hu, Genomic selection of agronomic traits in hybrid rice using an NCII population, Rice 11 (2018) 32.
[43] X. Wang, L. Li, Z. Yang, X. Zheng, S. Yu, C. Xu, Z. Hu, Predicting rice hybrid performance using univariate and multivariate GBLUP models based on North Carolina mating design II, Heredity 118 (2017) 302–310.
[44] A.J.S. Reis, L.J. Chaves, J.B. Duarte, E.M. Brasil, Prediction of hybrid means from a partial circulant diallel table using the ordinary least square and the mixed model methods, Genet. Mol. Biol. 28 (2005) 314–320.
[45] F. Alves, A. Granato, G. Galli, D. Lyra, R. Fritsche-Neto, G. de los Campos, Bayesian analysis and prediction of hybrid performance, Plant Methods 15 (2019) 14.
[46] R.D. Veiga, D.F. Ferreira, M.A.P. Ramalho, Efficiency of circulant diallels in parental choice, Pesq. Agropec. Bras. 35 (2000) 1395–1406.
[47] K. Dias, H. Piepho, L. Guimarães, P. Guimarães, S. Parentoni, M. Pinto, R. Noda, J. Magalhães, C. Guimarães, A. Garcia, Novel strategies for genomic prediction of untested single-cross maize hybrids using unbalanced historical data, Theor. Appl. Genet. (2019) 1–13.
[48] R. Fristche-Neto, D. Akdemir, J.L. Jannink, Accuracy of genomic selection to predict maize single-crosses obtained

through different mating designs, Theor. Appl. Genet. 131 (2018) 1153–1162.
[49] 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.
[50] 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.
[51] M.F. Resende, P. Muñoz, M.D. Resende, D.J. Garrick, R.L. Fernando, J.M. Davis, E.J. Jokela, T.A. Martin, G.F. Peter, M. Kirst, Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.), Genetics 190 (2012) 1503–1510.
[52] 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.
[53] H. Qi, J. Huang, Q. Zheng, Y. Huang, R. Shao, L. Zhu, Z. Zhang, F. Qiu, G. Zhou, Y. Zheng, Identification of combining ability loci for five yield-related traits in maize using a set of testcrosses with introgression lines, Theor. Appl. Genet. 126 (2013) 369–377.
[54] Z.Q. Zhou, C.S. Zhang, X.H. Lu, L.W. Wang, Z.F. Hao, M.S. Li, D. G. Zhang, H.J. Yong, H.Y. Zhu, J.F. Weng, X.H. Li, Dissecting the genetic basis underlying combining ability of plant height related traits in maize, Front. Plant Sci. 9 (2018) 1117.