



OPEN ACCESS

EDITED BY

Byoung-Cheorl Kang,
Seoul National University, South Korea

REVIEWED BY

Chenwu Xu,
Yangzhou University, China
Ana M. Casas,
Aula Dei Experimental Station, (CSIC),
Spain
Sung-Chur Sim,
Sejong University, South Korea

*CORRESPONDENCE

Juan Ma
majuanjuan85@126.com

SPECIALTY SECTION

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

RECEIVED 19 June 2022

ACCEPTED 24 October 2022

PUBLISHED 11 November 2022

CITATION

Ma J, Cao Y, Wang Y and Ding Y
(2022) Development of the maize
5.5K loci panel for genomic
prediction through genotyping
by target sequencing.
Front. Plant Sci. 13:972791.
doi: 10.3389/fpls.2022.972791

COPYRIGHT

© 2022 Ma, Cao, Wang and Ding. This
is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction
in other forums is permitted, provided
the original author(s) and the
copyright owner(s) are credited and
that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is
permitted which does not comply with
these terms.

Development of the maize 5.5K loci panel for genomic prediction through genotyping by target sequencing

Juan Ma*, Yanyong Cao, Yanzhao Wang and Yong Ding

Institute of Cereal Crops, Henan Academy of Agricultural Sciences, Zhengzhou, China

Genotyping platforms are important for genetic research and molecular breeding. In this study, a low-density genotyping platform containing 5.5K SNP markers was successfully developed in maize using genotyping by target sequencing (GBTS) technology with capture-in-solution. Two maize populations (Pop1 and Pop2) were used to validate the GBTS panel for genetic and molecular breeding studies. Pop1 comprised 942 hybrids derived from 250 inbred lines and four testers, and Pop2 contained 540 hybrids which were generated from 123 new-developed inbred lines and eight testers. The genetic analyses showed that the average polymorphic information content and genetic diversity values ranged from 0.27 to 0.38 in both populations using all filtered genotyping data. The mean missing rate was 1.23% across populations. The Structure and UPGMA tree analyses revealed similar genetic divergences (76–89%) in both populations. Genomic prediction analyses showed that the prediction accuracy of reproducing kernel Hilbert space (RKHS) was slightly lower than that of genomic best linear unbiased prediction (GBLUP) and three Bayesian methods for general combining ability of grain yield per plant and three yield-related traits in both populations, whereas RKHS with additive effects showed superior advantages over the other four methods in Pop1. In Pop1, the GBLUP and three Bayesian methods with additive-dominance model improved the prediction accuracies by 4.89–134.52% for the four traits in comparison to the additive model. In Pop2, the inclusion of dominance did not improve the accuracy in most cases. In general, low accuracies (0.33–0.43) were achieved for general combining ability of the four traits in Pop1, whereas moderate-to-high accuracies (0.52–0.65) were observed in Pop2. For hybrid performance prediction, the accuracies were moderate to high (0.51–0.75) for the four traits in both populations using the additive-dominance model. This study suggests a reliable genotyping platform that can be implemented in genomic selection-assisted breeding to accelerate maize new cultivar development and improvement.

KEYWORDS

maize, genotyping by target sequencing, genomic prediction, hybrid prediction, general combining ability

Introduction

Genotyping platforms are prerequisite for genomic research, genetic analysis, and marker-assisted breeding in animals and plants. Compared with other marker types, single nucleotide polymorphisms (SNPs), as the most extensive and stable genomic variations of multiple species, are ideal markers for genotyping because of their advantages in ultra-high-throughput detection and easy integration of genotypic data (Zhang et al., 2020a).

Array-based and sequencing-based technologies (next-generation sequencing) are the major genotyping platforms which are available for the screening of SNP markers. In the former technology, the fixed nature of SNPs on an array is helpful for cross-project comparisons because the same markers are used (Rasheed et al., 2017). However, when new SNPs are required, the array-based genotyping platform can be expensive because the array must be redesigned (Rasheed et al., 2017). Sequencing-based technologies contain three strategies to obtain SNP markers. Whole genome resequencing, identifying all sequence variability, is still high-cost for genotyping large populations with the aim to perform genetic and breeding studies. Reduced-representation genome sequencing (restriction-site associated DNA and genotyping-by-sequencing), a partial or selective sequencing, is simple, quick, and low-cost (Davey et al., 2011; Andrews et al., 2016). The two strategies may not allow comparisons across projects because different sequencing technologies and analysis pipelines affect the selection of SNPs detected (Torkamaneh et al., 2016; BurrIDGE et al., 2018).

Genotyping by target sequencing (GBTS), a newly developed sequencing-based genotyping platform, involves the capture of target genomic loci by probes (Guo et al., 2021). GBTS integrates the advantages of array-based and partial sequencing, and possesses the characteristics of customized flexibility, high throughput, and low cost (Guo et al., 2019). The technology also allows cross-project comparisons due to the target genomic loci. GBTS mainly contains multiplex PCR-based (GenoPlex) (Zhang et al., 2020a) and probe-in-solution-based target sequencing (GenoBaits) (Guo et al., 2019). Recently, Guo et al. (2021) improved the latter system and developed a multiple SNP (mSNP) approach where mSNPs can be captured from a single amplicon. GBTS has been successfully utilized for genotyping, genetic diversity analysis, quantitative trait locus mapping, genome-wide association study, and traditional marker-assisted selection in wheat (BurrIDGE et al., 2018), maize (Guo et al., 2019; Guo et al., 2021), pepper (Du et al., 2019), cucumber (Zhang et al., 2020a), faba bean (Wang et al., 2021), and broccoli (Shen et al., 2021). However, the application of GBTS in genomic prediction for parent and hybrid performance was rarely reported.

Hybrid breeding plays a great role in improving maize and many other crops. It mainly involves the development of inbred

lines with high general combining ability (GCA) and specific combining ability (SCA) and the identification of hybrids with high yield potentials (Zhang et al., 2022). The estimation of GCA and SCA needs to conduct multi-environment trials using specific mating designs, such as the diallel cross and North Carolina II design. Therefore, the process of hybrid breeding not only requires a vast of field resources to evaluate the performances of all possible combinations among many inbred lines, but laborious work for the identification of hybrid performance. In fact, only a small proportion of crosses can be tested in the field and abundant crosses with potentials may not have the chance to be evaluated.

Genomic selection, first proposed by Meuwissen et al. (2001), aims to estimate breeding values of untested populations only having genotyping data and select inbred lines or hybrids with high yield potentials based on the information of training population which is genotyped and phenotyped. The application of GS in hybrid breeding projects can help predict the performance of untested crosses and conduct selections with the aid of genotyping platforms according to the genotypic and phenotypic information of tested populations, which can accelerate the breeding process of developing high GCA parental lines and high-yielding hybrids. The genomic prediction for combining ability and hybrid performance has been reported in maize (de Oliveira et al., 2020; Zhang et al., 2022), rice (Cui et al., 2020), wheat (Zhao et al., 2015), sorghum (Ishimori et al., 2020), and canola (Knoch et al., 2021) using genotypic data derived from the array-based and partial sequencing-based genotyping platforms. However, these genotyping platforms are still high-cost for GS-assisted breeding programs although moderate-to-high prediction accuracies were revealed in those studies, which may guarantee a reliable prediction for the performance of unevaluated lines.

Although several SNP genotyping platforms were developed through GBTS in maize, no GBTS system was evaluated in genomic selection for GCA and hybrid performance. In addition, the current GBTS platform still needed to be specifically customized and optimized according to different applications in genetic and molecular breeding. In the present study, we designed a low-density GBTS panel from diverse resources and evaluated its applications in genotyping, population structure classification, and genomic prediction for GCA and hybrid performance.

Materials and methods

Design of the maize GBTS-based 5.5K loci panel

To build a reliable and genome-wide genotyping array, we selected 5,521 target SNPs from diverse resources (Table 1). In a

TABLE 1 The number of target SNPs selected from different resources.

Category [#]	Number
Heterosis-related SNPs (Dataset I and II)	1,973
Published references	487
113 known genes (Dataset II)	184
Dataset I	898
Dataset II	776
Dataset III	836
maize GBTS-based 48K loci panel	367

[#] Dataset I: the whole genome resequencing data of four inbred lines Zheng58, Chang7-2, Zheng588, and ZhengH71; Dataset II: the RNA-seq data of Zheng58, Chang7-2, Zheng588, and ZhengH71 and their F₁ hybrids (Zhengdan958, Zhengdan1002, Zheng588/Chang7-2, and Zheng58/ZhengH71) at seven seed developmental stages; Dataset III: the RNA-seq data of Qi319, Ye478, B104, AJ525, A350, A314, and LH209 in terms of multiple tissues containing roots, leaves, seeds, and young ears.

previous study (Wang et al., 2022), four inbred lines Zheng58, Chang7-2, Zheng588, and ZhengH71 were sequenced through the whole genome resequencing technology (Dataset I), and the four inbred lines and their F₁ hybrids (Zhengdan958, Zhengdan1002, Zheng588/Chang7-2, and Zheng58/ZhengH71) were sequenced through RNA-seq technology at seven seed developmental stages (Dataset II). A total of 1,973 heterosis-related SNPs were selected from the above data using two following criteria: (1) SNPs were retained when the corresponding genes were significantly differentially expressed between F₁ hybrids and one of their corresponding parents in at least one F₁ hybrid and one developmental stage using edgeR ($\log_2[\text{fold change}] > 1$, $\text{FDR} < 0.05$), and these differentially expressed genes were significantly correlated with the mid-parent heterosis of hundred-kernel weight (HKW) and fresh HKW in at least one developmental stage using weighted gene co-expression network analysis (Supplementary Figure 1), and (2) SNPs showed allele-specific expression in at least one F₁ hybrid were selected according to the method of a previous study (Shao et al., 2019). Allele-specific expression, the imbalance between expression levels of two parental alleles in a hybrid, has been considered as a mechanism of heterosis (Shao et al., 2019). We found 653 SNPs were related with mid-parent heterosis, 1,772 SNPs were considered allele-specific expression, and 452 SNPs were passed the both selected criteria.

A great number of genes have been reported to regulate maize development and related agronomic traits. To cover these functional genetic loci, we included 184 synonymous SNPs or SNPs located at exonic or UTR regions from 113 known genes based on the Dataset II (Table 1, Supplementary Table 1). In addition, 487 SNPs were selected from important loci associated with grain yield and yield-related traits from published references (Yang et al., 2014; Liu et al., 2017a; Zhang et al., 2017; Pang et al., 2019; Liu et al., 2020; Ma et al., 2021; Ma and Cao, 2021). To make markers as evenly distribute across the genome as possible, 898 SNPs were selected from Dataset I, 776

SNPs were selected from Dataset II, 836 SNPs were selected from the RNA-seq data of Qi319, Ye478, B104, AJ525, A350, A314, and LH209 in terms of multiple tissues containing roots, leaves, seeds, and young ears (Dataset III), and 367 SNPs were selected from maize GBTS-based 48K loci panel (China Golden Marker, Beijing) after the quality control of GC content (40–60%) and the filtration of multi-copy SNPs. The distribution of 5.5K target markers on ten chromosomes was demonstrated in Figure 1A.

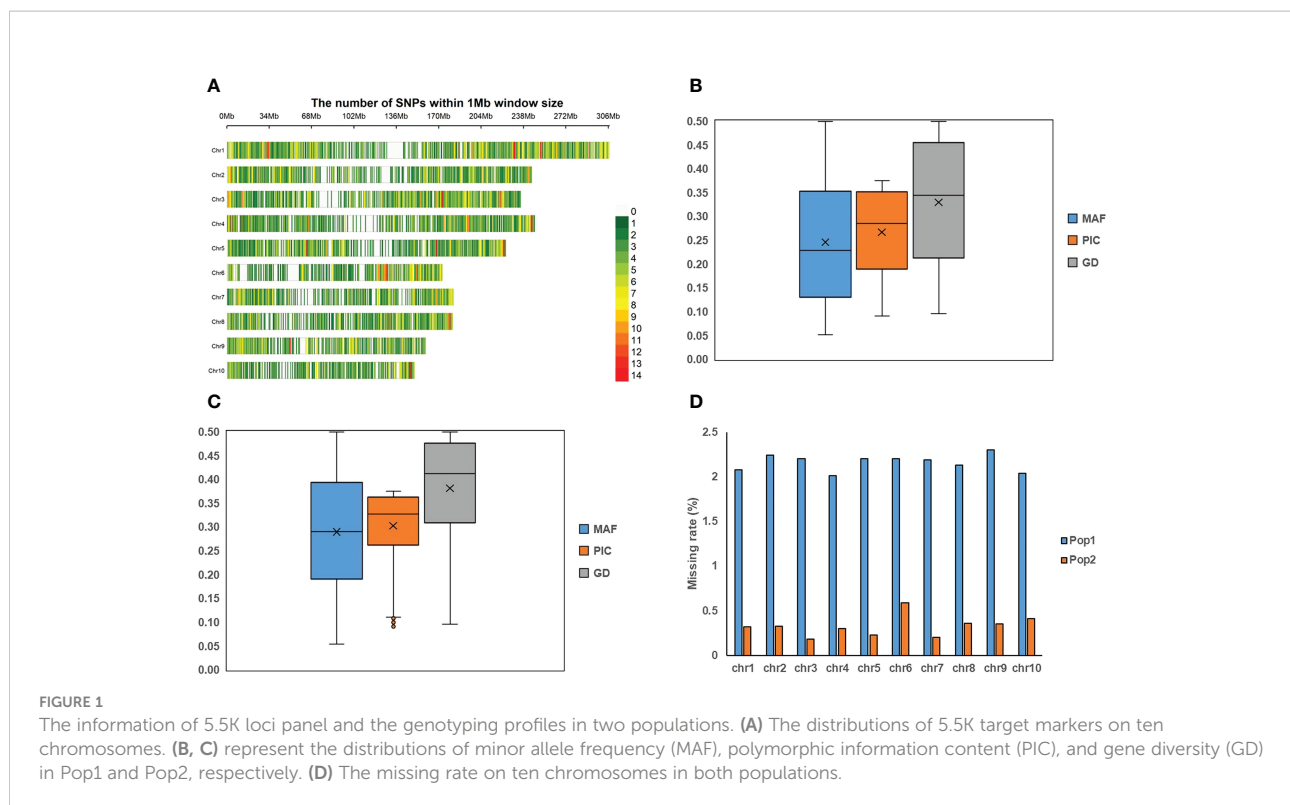
Plant materials, field trials, and evaluation of agronomic traits

To verify the effectiveness of the low-density GBTS genotyping platform, two populations (Pop1 and Pop2) were used in the present study. Pop1, a genetic population, contained 254 inbred lines from China (150) and USA (104), of which 250 inbred lines and four testers (Zheng58, Chang7-2, PH6WC, and PH4CV) produced 942 F₁ hybrids using North Carolina II mating design. Pop2, a breeding population, consisted of 123 new-developed inbred lines and eight testers (Chang7-2, PH4CV, Nongxi531, M119, M189, 20H1419, L119A, and S110T), which generated 540 F₁ hybrids. Two hybrid populations were evaluated in field experiments at Xinxiang and Zhoukou, Henan, but Pop1 and Pop2 were grown in 2020 and 2021, respectively. Entries were evaluated in one-row plot using randomized complete block design with two replicates. The plot size was 4 m and 3.3 m in length in Xinxiang and Zhoukou, respectively, all with 0.60 m between rows and 0.22 m between plants. Traits determined were grain yield per plant (GYP), ear weight (EW), HKW, and kernel number per row (KNR) in both populations. The analysis of variance was calculated following a linear mixed model.

$y = \mu + E + R + GCA_L + GCA_T + SCA + GCA_L \times E + GCA_T \times E + SCA \times E + \epsilon$, where y indicates the phenotypic value of hybrids, μ denotes the overall mean, E represents environment effect, R represents replicates. GCA_L and GCA_T are effects of inbred lines and testers, respectively; SCA is the effect of the combinations of inbred lines and testers; $GCA_L \times E$, $GCA_T \times E$, and $SCA \times E$ indicate GCA_L , GCA_T , and SCA interaction effects with environment, respectively. The variance components and GCA effects were calculated using R package lme4. The heritability of GCA effects was calculated using the following formula which was modified from Liu et al. (2021).

$$H_{GCA}^2 = \frac{\sigma_{GCA_L}^2 + \sigma_{GCA_T}^2}{\sigma_{GCA_L}^2 + \sigma_{GCA_T}^2 + \frac{\sigma_{SCA}^2}{e} + \frac{\sigma_{GCA_L \times E}^2}{e} + \frac{\sigma_{GCA_T \times E}^2}{e} + \frac{\sigma_{SCA \times E}^2}{le} + \frac{\sigma_e^2}{ler}}$$

where $\sigma_{GCA_L}^2$, $\sigma_{GCA_T}^2$, σ_{SCA}^2 , $\sigma_{GCA_L \times E}^2$, $\sigma_{GCA_T \times E}^2$, $\sigma_{SCA \times E}^2$, and σ_e^2 represent the GCA variance of inbred lines, the GCA variance of testers, the variance of SCA, the interaction variance between the GCA of inbred lines and environment, the interaction variance between the GCA of testers and environment, the interaction variance between SCA and environment, and residual variance,



respectively, and t , r , and e are the numbers of tester lines, replicates, and environments, respectively. Heritability at per mean level and best linear unbiased estimate (BLUE) values of hybrid traits were calculated using QTL IciMapping v4.2 software (Meng et al., 2015).

Genotyping and analyses based on the maize 5.5K loci panel

All parental lines of the two populations were used for genotyping. CTAB method was adopted to extract genomic DNA from fresh leaves. The length of each probe for the 5.5K loci panel was 100 bp to cover the SNP regions, which can capture approximately 250-400 bp sequence. The major processes of GBTS based on liquid-phase probe hybridization were as follows according to Wang et al. (2021): (1) Genomic DNA was fragmented and added a sequencing adapter, (2) The biotin-labelled RNA probe was combined with the DNA fragments that had already been attached to the adapter sequence, (3) Streptavidin-coated magnetic beads were combined with the double stranded complex of biotin-labelled RNA probe and DNA (probe excess), (4) Washing to obtain the DNA of the target region to remove nonspecific hybridization and improve the capture efficiency, and (5) The eluted DNA products were amplified by PCR and sequenced using Illumina NovaSeq 6000 platform (China Golden Marker, Beijing). BWA

software was used to align the filtered reads to B73 RefGen v4 (<http://www.gramene.org/>). GATK v4.1.2.0 (McKenna et al., 2010) was used to detect variants. Vcftools and PLINK software were used to filter $\text{minDP} < 11$, $\text{minGQ} < 20$, minor allele frequency (MAF) ≤ 0.05 , missing rate $> 10\%$, and heterozygous rate $> 1\%$. The polymorphic information content (PIC) was calculated according to the following equation that was proposed by Botstein et al. (1980).

$$PIC = 1 - \left(\sum_{i=1}^n P_i^2 \right) - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2P_i^2 P_j^2$$
, where P_i and P_j are the population frequencies of the i th and the j th allele. Gene diversity (GD) was estimated as:

$$GD = 1 - \left(\sum_{i=1}^n P_i^2 \right)$$

Population structure analysis

All filtered genotyping data were used for population structure analysis. Population structure was inferred using the Bayesian Markov Chain Monte Carlo (MCMC) program in Structure v2.3.4 (Pritchard et al., 2000). The number of subgroups (K) was set from 1 to 8 in Pop1, whereas that was 1 to 10 in Pop2. The length of burnin period and the number of MCMC replicates after burnin were 5,000 and 50,000, respectively. The Structure output was visualized by Structure Harvester (Earl and vonHoldt, 2012), and delta K was used to

determine the optimal number of subgroups. The FullSearch algorithm in CLUMPP v1.1.2 (Jakobsson and Rosenberg, 2007) was used to estimate cluster membership coefficient matrices from the optimal subgroup. To verify the optimal number of clusters, unweighted pair-group method with arithmetic means (UPGMA) tree was performed using the software TASSEL v5.2.60 (Bradbury et al., 2007). The circular tree was demonstrated using R package ggtree.

Genomic prediction for general combining ability and hybrid performance

Five models including Bayes A, Bayes C, Bayesian least absolute shrinkage and selection operator (Bayesian LASSO), genomic best linear unbiased prediction (GBLUP), and reproducing kernel Hilbert space (RKHS) were adopted for genomic prediction using all filtered genotyping data. For RKHS, three kernels were used and their bandwidth parameter h was set at 0.1, 0.5, and 2.5. For the GCA prediction, the genotypes were coded by -1 for one homozygote, 0 for the heterozygote, and 1 for the other homozygote. Randomized imputation was used for missing markers, according to the known genotype frequency (Ma and Cao, 2021). The above five GS methods were used to perform hybrid phenotypic prediction using additive (A) and additive plus dominance (AD) model. In the A model, the homozygous genotypes were coded as -1 and 1, and the heterozygous genotypes were coded as 0. For the mating type A1A1 × A1A2 and A2A2 × A1A2, these hybrids were coded as -0.5 and 0.5, respectively. When the mating type was A1A2 × A1A2, their hybrids were coded as 0. For the dominance model, the homozygous genotypes were coded as 0, and the heterozygous genotypes were coded as 1. For the mating type A1A1 × A1A2 and A2A2 × A1A2, the hybrids were all coded as 0.5.

All GS models and prediction strategies were performed using the R package, BGLR (Pérez and de los Campos, 2014). For all models, the number of Gibbs iterations was 12,000, and the burn-in was 3,000. A 10-fold cross-validation scheme was used and repeated 100 times for all prediction methods and models. In the 10-fold cross validation, 90% inbred lines or hybrids were selected as the training set to predict the remaining 10% inbred lines or hybrids as the testing set. The average correlation coefficient between genomic estimated breeding values and phenotypic values in the testing set was used to estimate the accuracies of different GS models.

Genomic prediction across projects

The fixed nature of the GBTS technology allowed the comparisons between different projects. The genotyping of the

two populations was conducted in two batches, therefore we found common SNPs between the filtered genotyping data of both populations. Based on these common SNPs, the prediction accuracy of GCA and hybrid performance was calculated using the RKHS method. As all filtered genotyping data, the same cross-validation and parameters were used.

Genomic prediction for potentially functional markers

Among the 5,521 target SNPs, 2,644 SNPs were identified from the weighted gene co-expression network analysis, allele-specific expression analysis, known genes, and published references (Table 1), which were defined as potentially functional markers. These functional SNPs existed in the filtered genotyping data were used as marker subset to predict GCA and hybrid performance using the RKHS method. To validate the performance of these markers, the same number of other target SNPs was also used to conduct the genomic prediction.

Results

The performance of hybrids and general combining ability

Hybrid phenotypes and parental GCA were analyzed in this study. Descriptive statistics were shown in Supplementary Figure 2. Genetic correlations showed GYP and EW showed positive and high correlations in both hybrid populations, with r value ranging from 0.98 to 0.99 (Supplementary Figure 3). HKW and KNR had low or no correlations, but they were positively and significantly correlated with GYP and EW, with r value ranging from 0.36 to 0.64 in both hybrid populations. As the hybrid trait *per se*, high correlations ($r = 0.98$) were also observed between the GCA effects of GYP and EW in the two populations. For GCA effects of other traits, similar correlation values were found as those of the hybrid traits.

Analysis of variance showed that significant GCA, SCA, GCA-by-environment interaction, and SCA-by-environment interaction variances were revealed for all traits except for one source in KNR of Pop1 (Supplementary Table 2). In Pop1, the variances of SCA were much higher than those of GCA for GYP and EW. The contrast trend was observed in Pop2. The heritabilities of traits in hybrids ranged from 0.55 for KNR in Pop2 to 0.71 for HKW in Pop1. In Pop1, the heritabilities of GCA effects for GYP and EW were low ($H^2 = 0.16-0.17$), whereas those were high ($H^2 = 0.82-0.84$) in Pop2. For other traits, the heritabilities of GCA effects ranged from 0.50 to 0.84. Big variations observed in the heritabilities of GCA for GYP and EW across populations could be attributed to genetic backgrounds, environmental effects, or the interactions between them.

Profiles of genotyping using the maize GBTS-based 5.5K loci panel

Based on the 5,521 target SNPs using the GBTS technology with capture-in-solution, 75,876 and 33,971 raw SNPs were detected in Pop1 and Pop2, respectively. After filtering, 20,210 and 11,734 high-quality SNPs were generated for Pop1 and Pop2, respectively, which were used for the further genetic analyses. The number of SNPs per chromosome ranged from 1,383 (chromosome 10) to 3,446 (chromosome 1) in Pop1, whereas that ranged from 376 (chromosome 6) to 2,391 (chromosome 1) in Pop2 (Supplementary Figure 4A). The mean MAF after filtering across all SNPs was 0.25 and 0.29 in Pop1 and Pop2, respectively (Figures 1B, C). The mean missing rate after filtering across all SNPs was 2.16% and 0.30% in Pop1 and Pop2, respectively (Figure 1D). The average PIC and GD values were 0.27 and 0.33 in Pop1, respectively, whereas those of Pop2 were 0.30 and 0.38, respectively. Among these filtered markers, 4,865 and 2,237 target SNPs were found in Pop1 and Pop2, respectively (Supplementary Figure 5A). The mean MAF, PIC, and GD values of target SNPs were slightly higher than those of all filtered markers in both populations (Figures 1B, C, Supplementary Figures 5B, C). The mean missing rate of the

target SNPs was 0.12-0.87%, which was much lower than that of the full set of filtered markers (Figure 1D and Supplementary Figure 5D). These results highlighted the high quality of those target SNPs.

Genetic structures of two maize populations

The Structure and CLUMMP analyses revealed that Pop1 and Pop2 were divided into five and six sub-populations, respectively, based on the optimal number of K (Supplementary Figure 6). In Pop1, Cluster 1 and Cluster 2 mainly belonged to non-Stiff Stalk, Cluster 3 mainly represented Stiff Stalk, Cluster 4 indicated Tang Si Ping Tou, and Cluster 5 inferred as modified Reid group (Figure 2A). For the breeding population Pop2, the number of inbred lines within sub-populations ranged from 10 (Cluster 3) to 42 (Cluster 1) (Figure 2B). Similar genetic divergences (76-89%) were also observed using the circular UPGMA tree in both populations (Figure 2), which indicated that the GBTS-based 5.5K loci panel can be used for genetic analyses and assisted for the inference of germplasm origins.

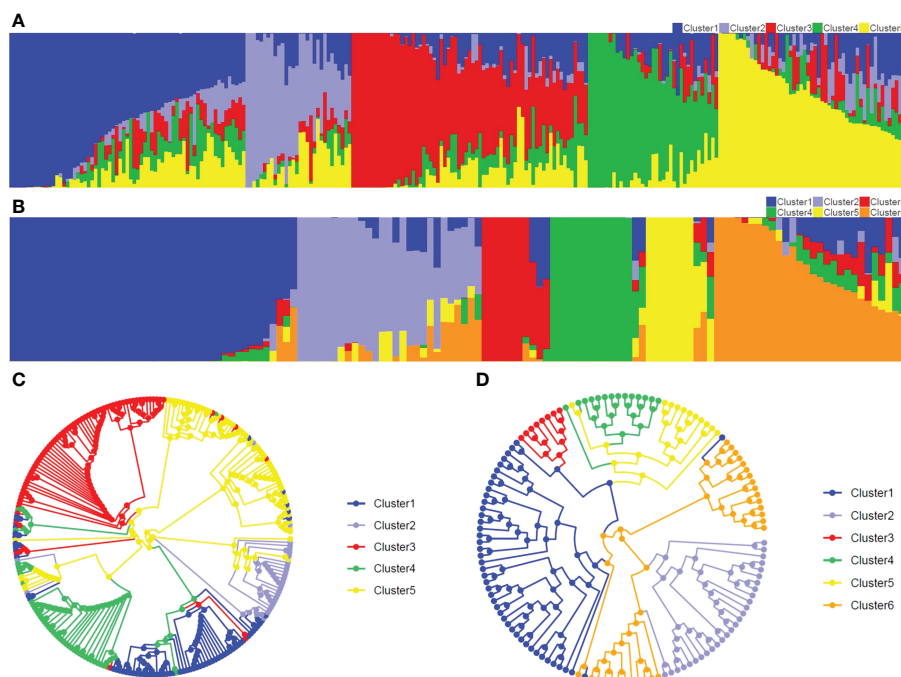


FIGURE 2

Population structure and UPGMA tree in two populations. (A, B) represent population structure in Pop1 ($K=5$) and Pop2 ($K=6$), respectively. Abscissa and ordinate represent the inbred lines and the membership percentage of inbred line, respectively. (C, D) represent circular UPGMA tree in Pop1 and Pop2, respectively.

Prediction accuracies for general combining ability and hybrid performance

To detect the prediction power of the GBTS panel, we used the classical parameter model GBLUP, three Bayesian models, and a semi-parameter model RKHS to predict GCA effects of the observed traits within populations. The prediction accuracies

ranged from 0.33 (GYP) to 0.43 (HKW) in Pop1, whereas those varied from 0.52 (HKW) to 0.65 (EW) in Pop2 (Figure 3). Regardless of populations and traits, GBLUP and the three Bayesian methods resulted in similar predictive performance for GCA effects, with the difference values ranging from 0 to 0.01. In general, the prediction accuracy of RKHS was slightly lower than that of the other four methods, with the percentage decrease ranging from 1.18 to 6.96%. Moderate-to-high

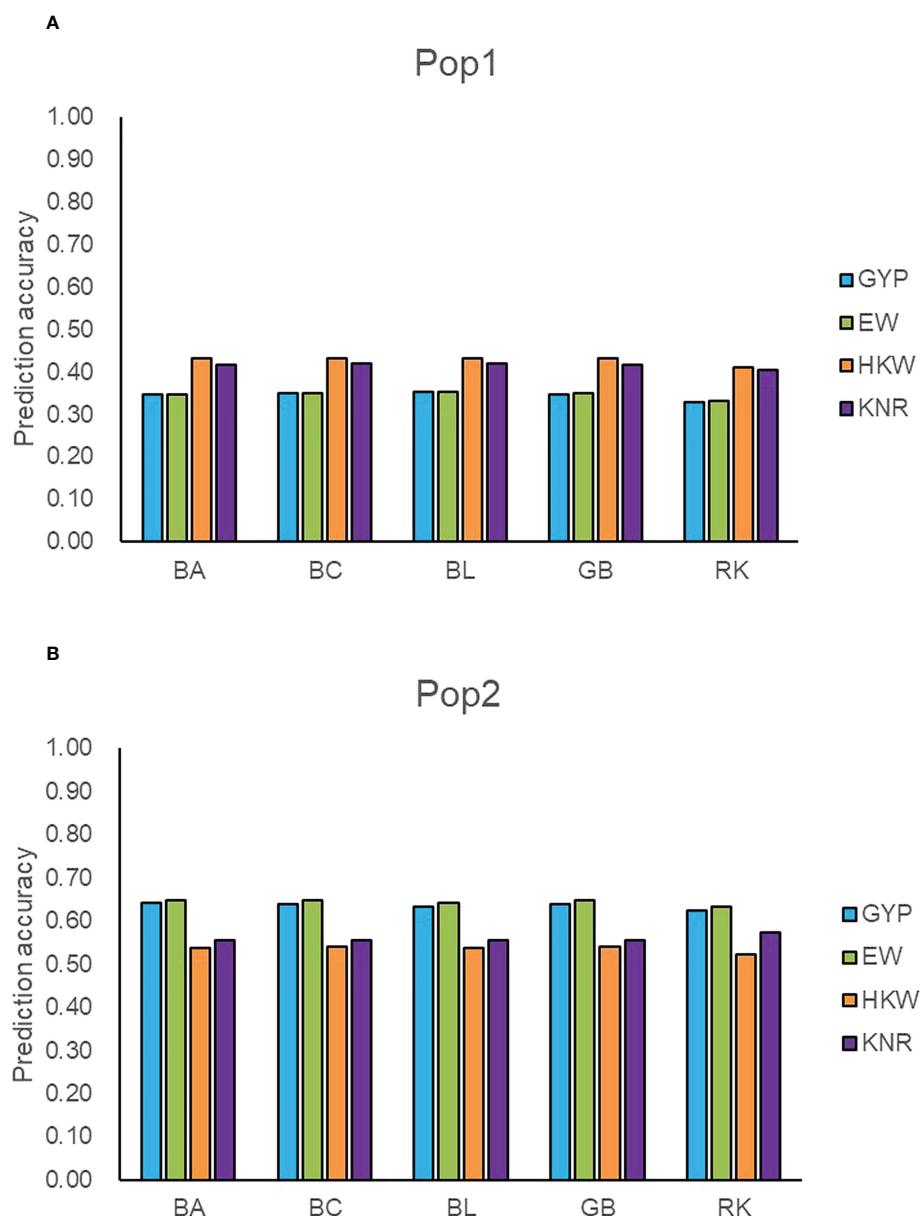


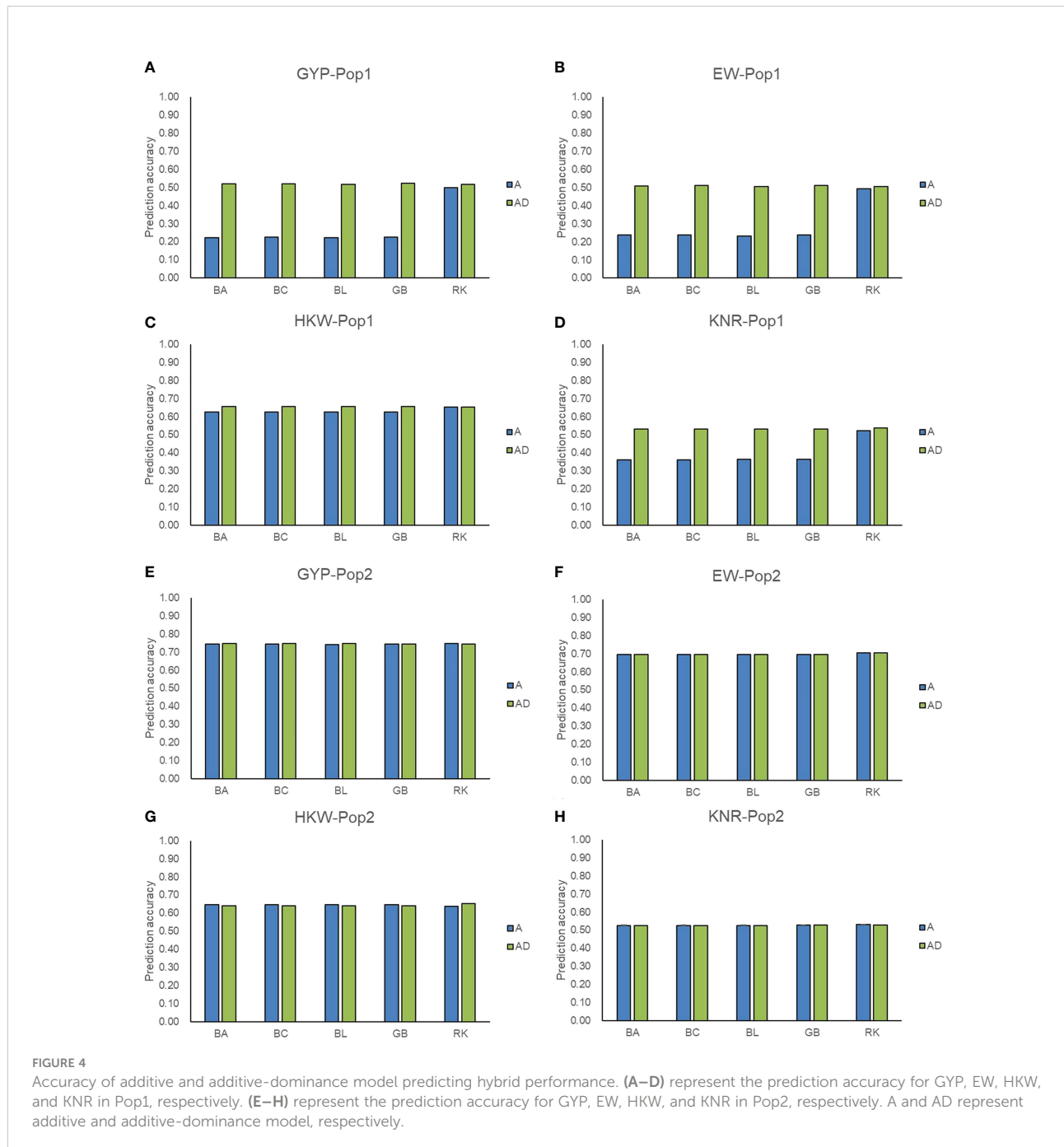
FIGURE 3

Accuracy of five models predicting general combining ability for four traits in two populations. (A) The prediction accuracy of GYP, EW, HKW, and KNR in Pop1. (B) The prediction accuracy of GYP, EW, HKW, and KNR in Pop2. GYP, EW, HKW, and KNR are abbreviations of grain yield per plant, ear weight, thousand-kernel weight, and kernel row number, respectively. BA, BC, BL, GB, and RK denote Bayes A, Bayes C, Bayesian LASSO, GBLUP and RKHS, respectively.

accuracies were obtained for GCA effects in the breeding population (Pop2), which suggested that the maize GBTS-based 5.5K loci panel can be used for GS-assisted selection for high GCA lines.

Except for GCA effects, the predict power of hybrid phenotypes was also evaluated for the genotyping panel. The above five GS methods incorporating additive effect only and additive plus dominance effect were adopted. In Pop1, the prediction accuracies for the performance of hybrids ranged

from 0.22 (GYP) to 0.65 (HKW) in the five methods including additive effects, whereas those were improved and ranged from 0.51 (EW) to 0.66 (HKW) in the AD model (Figure 4). Compared with the A model, the AD model of GBLUP and the three Bayesian methods improved the accuracy by 114.35–134.52% for GYP and EW in Pop1. Compared with the A model, the percentage increase ranged from 45.66 to 46.82% for KNR when the dominance was incorporated into GBLUP and the three Bayesian methods, whereas a small percentage increase



(4.89–5.15%) was observed for HKW. For RKHS, the addition of dominance slightly improved the accuracies of hybrid performance for GYP, EW, and KNR in Pop1. Due to high accuracies (0.53–0.75) achieved in the A model for all four traits in Pop2, the incorporation of dominance did not improve the accuracies of hybrid performance in most scenarios. For the A model only, RKHS showed the superior performance over the other four methods, improving the accuracy by 42.68–126.44% for GYP, EW, and KNR in Pop1. Regardless of GS methods and models, moderate-to-high accuracy values were achieved for the hybrid prediction in both populations, which supported the reliability of GS-assisted selection of excellent hybrids using the maize 5.5K loci genotyping panel.

Cross-project comparisons and prediction accuracy for potentially functional markers

Based on all filtered SNPs, the RKHS method showed superior performance over the other four methods particularly when the additive effect was considered in Pop1 and showed similar or slightly lower advantages in other circumstances, therefore the method was used for the cross-project comparisons and the prediction for potentially functional markers. Due to the fixed nature of target genomic loci, 7,743 SNPs were simultaneously detected between 20,210 and 11,734 SNPs, accounting for 38–66% of those filtered genotyping data (Supplementary Figure 4B). The prediction accuracies of 7,743 SNPs ranged from 0.35 to 0.64 for GCA effects of GYP and EW in the two populations, which was slightly higher than those of all filtered markers (Figure 3 and Supplementary Figure 7A). Compared with all filtered SNPs, slightly lower accuracy was found in GCA effects of HKW and KNR in Pop1. The 7,743 SNPs enabled high accuracy (0.71) for GYP with the A and AD model in Pop2, although the value was smaller than that of all markers (Figure 4E and Supplementary Figure 7C). For other circumstances, similar prediction abilities were observed between common SNPs and the full set of markers. These findings agreed with Spindel et al. (2015) who pointed out that ~ 7,000 (approximately 1 SNP for every 0.2 cM) SNPs were sufficient for GS. These suggested that the GBTS-based 5.5K loci genotyping platform can be used for cross-project comparisons in terms of genomic prediction.

Among the 5,521 target SNPs, 2,644 SNPs were associated with mid-parent heterosis, allele-specific expression, known genes, and yield-related traits and were considered as potentially functional markers. A total of 2,021 target SNPs were overlapped between the two populations among the filtered genotyping data (Supplementary Figure 5A), of which 906 target SNPs were potentially functional markers. The prediction ability was compared when these potentially functional markers and the same number of other target SNPs were used for genomic prediction. For GYP and EW, the prediction accuracy of

potentially functional markers was consistently higher than that of randomly selected markers in terms of GCA and hybrid performance in Pop1 (Figures 5A, C, E). The good performance was also observed for GCA prediction in both populations and for hybrid prediction with the A model in Pop1 for KNR (Figures 5A–C). These might prove that the selection strategies for functional SNPs were valid. Zhang et al. (2020b) found that the employment of functional genes information such as the number of favorable alleles and genotypes enabled accurately predicting maize yield. However, these potentially functional markers did not show any advantage in the remaining scenarios (Figure 5). Most of these target SNPs achieved similar accuracies as the full set of filtered SNPs (Figures 3–5), which again highlighted the reliability of target SNPs.

Discussion

Evaluation of genotyping, genetic diversity, and population structure

High-throughput genotyping technology is very important for effective crop breeding programs. GBTS technology integrates the advantages of array-based and partial sequencing, showing advantages in customized flexibility, high throughput, and low cost. In maize, a series of high-quality GBTS panels, including 1–20K SNP (GenoBaits) and 1–40K mSNP, were developed, which made the technology an effective and efficient tool for genotyping and population structure classification (Guo et al., 2019; Guo et al., 2021). In the present study, the GBTS-based 5.5K loci panel was developed mainly from the whole genome resequencing and transcriptome sequencing of Huanghuaihai maize germplasm. Two populations containing 383 accessions were genotyped using the platform. Due to the target region sequencing, the number of detected raw SNPs was approximately 14-fold and six-fold as that of the target SNPs for Pop1 and Pop2, respectively. Wang et al. (2021) found that 1,579,411 SNPs were identified and further filtered according to the Faba_bean_130K targeted next-generation sequencing genotyping platform. These all showed that the GBTS technology can detect a large number of SNPs in comparison to the capacity of target SNPs. Like the genotyping-by-sequencing or restriction-site associated DNA, the GBTS was based on next-generation sequencing, therefore the genotyping results were affected by the size of restriction fragment length, population background, and population size when samples were genotyped in different batches. Based on the filtered genotyping data, the mean missing rate across populations was 1.23%, which was lower than that of populations genotyped using GBTS-based 1–20K panels (Guo et al., 2019). The average PIC and GD values ranged from 0.27 to 0.38, which could be considered high. The biallelic nature of the SNP markers limited the range of PIC and GD values from 0 to 0.5 (Eltaher et al., 2018; Shen et al., 2021).

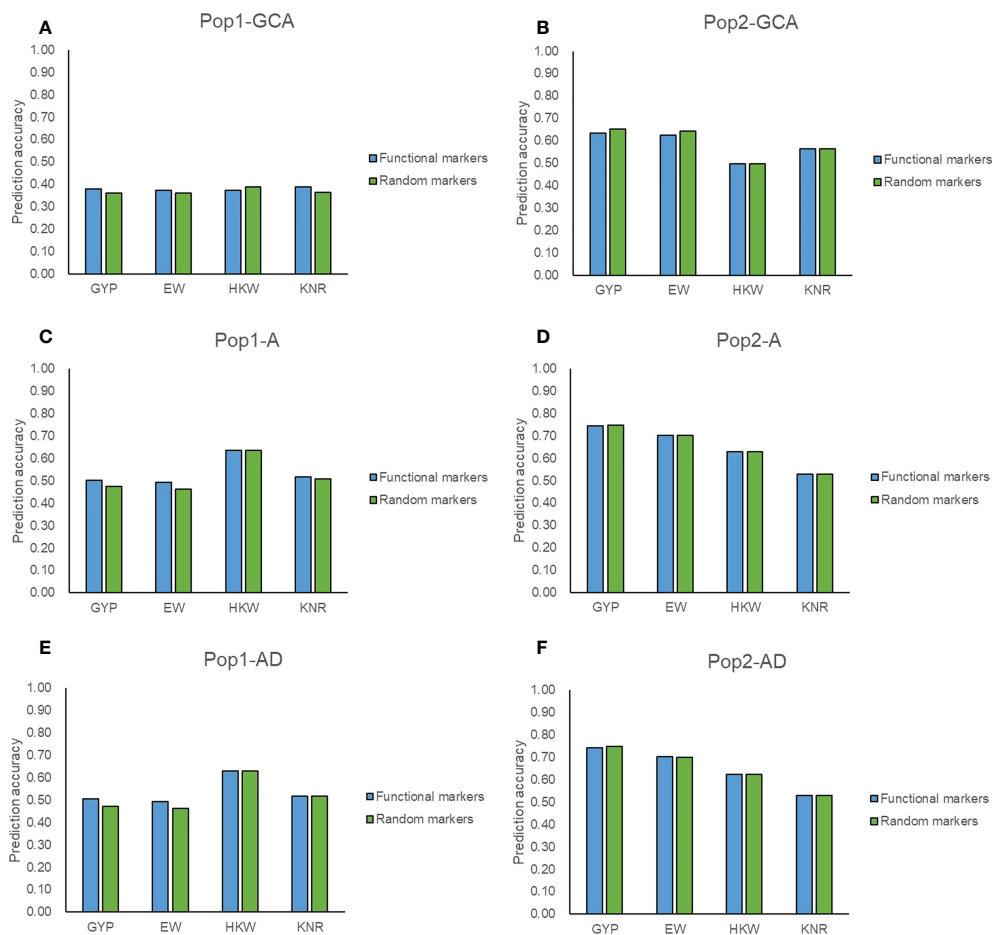


FIGURE 5

Accuracy of target SNPs in two populations. (A, B) represent the general combining ability prediction using RKHS method in Pop1 and Pop2, respectively. (C, D) represent hybrid performance prediction using RKHS with additive model (A) in Pop1 and Pop2, respectively. (E, F) represent hybrid performance prediction using RKHS with additive-dominance model (AD) in Pop1 and Pop2, respectively. Functional markers represent 906 target SNPs associated with mid-parent heterosis, allele-specific expression, known genes, and yield-related traits. Random markers represent 906 target SNPs which are randomly selected from 1,115 target SNPs.

All these proved that the GBTS-based 5.5K loci panel is available for identifying the genetic diversity of maize germplasms.

For Pop1, 254 maize accessions were divided into five subpopulations by the Structure analysis, 76% of which was in agreement with the phylogenetic tree (Figures 2A, C). For Pop2, Structure and phylogenetic tree results agreed with each other with 11% exceptions (Figures 2B, D). For Structure analysis, a sub-population membership percentage was produced and the highest percentage was used to assign one individual to one group, whereas a fixed branch position was assigned to each accession for UPGMA analysis (Wang et al., 2009; Shen et al., 2021). This discrepancy between the two methods of grouping might result in some biases. The similar results showed that the GBTS-based 5.5K loci panel can be assisted for the population structure classification.

Potential applications in genomic selection

GS has obvious advantages for improving genetic gains in animal and plant breeding, but the price of genotyping can be prohibitive for many species (Kriaridou et al., 2020). Therefore, the development of cost-effective and user-friendly genotyping platform that is suitable for genomic selection is valuable for breeding programs with limited funds and resources. We demonstrated the potential of the GBTS-based 5.5K loci panel in genomic prediction in terms of GCA and hybrid performance using one genetic population and one breeding population. Regardless of GCA or hybrid performance prediction, higher accuracies were observed in Pop2 than in Pop1, especially for GYP and EW. The phenomena also occurred even if only

common SNPs, such as 7,743 overlapped SNPs, 906 functional target SNPs, and randomly selected target SNPs, were used for prediction. Obvious variations in prediction accuracies for hybrid performance were also revealed in different maize breeding populations (Windhausen et al., 2012; Li et al., 2021). The variances of GCA and SCA varied across populations indicated that the genetic structure of GYP and EW was very different in two populations. Therefore, the differences in genetic basis of traits among different populations might be an important factor influencing the prediction accuracies (Li et al., 2021). The significant environment effect and genotype-by-environment effect were revealed across traits and populations through analysis of variance, indicating that the environment factor might affect the prediction results because the two populations were grown in different years.

Decomposing of variances of hybrid performance into GCA and SCA variances could reflect the role of additive and non-additive effects. In Pop1, the inclusion of the dominance effects could effectively improve the prediction accuracy of hybrid phenotypes in GBLUP and the three Bayesian methods. In particular, the AD model boosted the prediction accuracy for GYP and EW by more than two-fold compared with the A model, which agreed with a previous study where GBLUP with AD effects doubled the predictive capacity for maize grain yield under water-stressed trial in comparison to the A model (Dias et al., 2018). In general, these results were consistent with the size of SCA variances (Supplementary Table 2). Incorporating dominance effects improved the prediction accuracy considerably for convergent parent populations, where dominance generated major contributions of SCA effects to the genetic variance among inter-population hybrids (Technow et al., 2012). In Pop2, the AD model didn't not improve predictive performance in comparison to the A model in most circumstances because the GCA variances (additive variances) was large. High level of additive variances also explained the reason that the A model achieved high accuracies in Pop2. In line with our findings, Ferrão et al. (2020) demonstrated that the inclusion of dominance effects increased the predictive ability of grain yield because dominance explained a large portion of the phenotypic variance for grain yield; when the additive variance was large, the A model yielded better results for grain moisture. The superiority of GBLUP-AD and Gaussian kernel regression depended on the level of dominance variance in sorghum (Ishimori et al., 2020). In addition, the loss in accuracy that was induced by the inclusion of dominance or epistatic effects was most likely caused by more pronounced interactions of environments with dominance and epistatic effects than with additive effects (Liu et al., 2017b).

In most instances, RKHS did not improve the prediction accuracies when the dominance effect was included especially in Pop2. RKHS *per se* can capture non-additive effects in hybrid

populations (Gianola et al., 2006), even if only additive genomic matrix was fitted into the method. For trait *per se*, RKHS gave consistently high predictive performance than some parameter models (Ma and Cao, 2021). However, our findings showed that this method was not good for GCA prediction. Alves et al. (2019) found that although RKHS was in all cases the one that had the highest proportion of variance explained, the predictive performance of this model was not the highest one.

The prediction accuracy (0.62-0.64) of GCA effects for GYP in Pop2 based on 11,734 SNPs was higher than that (0.49-0.55) achieved using 39,659 SNPs from the DArT-seq platform (Zhang et al., 2022). For hybrid prediction, high prediction accuracy (0.74-0.75) was achieved for GYP using 11,734 SNPs in Pop2. In some previous studies, lower accuracies (0.03-0.67) were achieved for GYP or grain yield per hectare in hybrid populations with higher marker densities (21,475-52,811) which were obtained from the genotyping-by-sequencing, 50K Illumina chip, maize 500k Affymetrix chip, and Affymetrix genotyping array of 616 K SNPs platforms (Supplementary Table 3, Dias et al., 2018; Alves et al., 2019; Schrag et al., 2019; de Oliveira et al., 2020; Dias et al., 2020; Ferrão et al., 2020; Costa-Neto et al., 2021). Using fewer markers, a moderate accuracy was achieved with the AD model for hybrid GYP in Pop1, which was comparable to that achieved in some previous studies (Supplementary Table 3). Heffner et al. (2010) concluded that GS can significantly accelerate genetic gains through shortening the breeding cycle if moderate selection accuracies are obtained. Several studies showed that GBTS can significantly reduce the cost of genotyping by at least half compared with the array-based and genotyping-by-sequencing platforms (Guo et al., 2019; Bernardo et al., 2020; Guo et al., 2021). All these indicated that the GBTS-based 5.5K loci panel is sufficient for predicting GCA effects and hybrid performance and will be a reliable, efficient, and low-cost genotyping platform for GS-assisted breeding in selecting high GCA lines and high-yielding hybrids in maize.

Data availability statement

The data presented in the study are deposited in the Sequence Read Archive repository, accession number PRJNA728476, PRJNA649667, PRJNA681326, PRJNA843481, and PRJNA843784, and in the China National GeneBank DataBase Sequence Archive repository, accession number CNP0003317.

Author contributions

JM designed the study, collected the phenotypic data, analyzed the phenotypic and genomic data including

population structure and genomic prediction analyses, wrote and revised the manuscript. JM, YC, YW, and YD provided the DNA resequencing and RNA sequencing data for GBTS panel development. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Science and Technology Project of Henan Province (222102110043) and by the Science-Technology Foundation for Outstanding Young Scientists of Henan Academy of Agricultural Sciences (2020YQ04).

Acknowledgments

We would like to thank Dr. Jiankang Wang, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences (CAAS) for providing valuable suggestions for this manuscript.

References

- Alves, F. C., Granato, Í., Galli, G., Lyra, D. H., Fritsche-Neto, R., and de Los Campos, G. (2019). Bayesian Analysis and prediction of hybrid performance. *Plant Methods* 15, 14. doi: 10.1186/s13007-019-0388-x
- Andrews, K. R., Good, J. M., Miller, M. R., and Luikart, G. (2016). And hohlenlohe, p Harnessing the power of RADseq for ecological and evolutionary genomics. *A. Nat. Rev. Genet.* 17, 81–92. doi: 10.1038/nrg.2015.28
- Bernardo, A., Amand, P. S., Le, H. Q., Su, Z., and Bai, G. (2020). Multiplex restriction amplicon sequencing: a novel next-generation sequencing based marker platform for high-throughput genotyping. *Plant Biotechnol. J.* 18, 254–265. doi: 10.1111/pbi.13192
- Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphism. *Am. J. Hum. Genet.* 32, 314–331. doi: 10.1016/0165-1161(81)90274-0
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
- Burridge, A. J., Wilkinson, P. A., Winfield, M. O., Barker, G., and Edwards, K. J. (2018). Conversion of array-based single nucleotide polymorphic markers for use in targeted genotyping by sequencing in hexaploid wheat (*Triticum aestivum*). *Plant Biotechnol. J.* 16, 867–876. doi: 10.1111/pbi.12834
- Costa-Neto, G., Fritsche-Neto, R., and Crossa, J. (2021). Nonlinear kernels, dominance, and envirotyping data increase the accuracy of genome-based prediction in multi-environment trials. *Heredity* 126, 92–106. doi: 10.1038/s41437-020-00353-1
- Cui, Y., Li, R., Li, G., Zhang, F., Zhu, T., Ali, J., et al. (2020). Hybrid breeding of rice via genomic selection. *Plant Biotechnol. J.* 18, 57–67. doi: 10.1111/pbi.13170
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12, 499–510. doi: 10.1038/nrg3012
- de Oliveira, A. A., Resende, M. F. R., Ferrão, L. F. V., Amadeu, R. R., Guimares, L., Guimares, C. T., et al. (2020). Genomic prediction applied to multiple traits and environments in second season maize hybrids. *Heredity* 125, 60–72. doi: 10.1038/s41437-020-0321-0
- Dias, K. O. G., Gezan, S. A., Guimarães, C. T., Nazarian, A., Silva, L. C., Parentoni, S. N., et al. (2018). Improving accuracies of genomic predictions for

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.972791/full#supplementary-material>

- drought tolerance in maize by joint modeling of additive and dominance effects in multi-environment trials. *Heredity* 121, 24–37. doi: 10.1038/s41437-018-0053-6
- Dias, K. O. G., Piepho, H. P., Guimares, L. J. M., Guimares, P. E. O., Parentoni, S. N., Pinto, M. O., et al. (2020). Novel strategies for genomic prediction of untested single-cross maize hybrids using unbalanced historical data. *Theor. Appl. Genet.* 133, 443–455. doi: 10.1007/s00122-019-03475-1
- Du, H., Yang, J., Chen, B., Zhang, X., and Wen, C. (2019). Genotyping pepper varieties using target SNP-seq reveals that population structure clusters according to fruit shape. *BMC Plant Biol.* 19, 578. doi: 10.21203/rs.2.10821/v1
- Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7
- Eltaher, S., Sallam, A., Belamkar, V., Emara, H. A., Nower, A. A., Salem, K. F. M., et al. (2018). Genetic diversity and population structure of F_{3:6} Nebraska winter wheat genotypes using genotyping-by-sequencing. *Front. Genet.* 9. doi: 10.3389/fgenet.2018.00076
- Ferrão, L. F. V., Marinho, C. D., Munoz, P. R., and Resende, M. F. R. (2020). Improvement of predictive ability in maize hybrids by including dominance effects and marker × environment models. *Crop Sci.* 60, 666–677. doi: 10.1002/csc.2.20096
- Gianola, D., Fernando, R. L., and Stella, A. (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* 173, 1761–1776. doi: 10.1534/genetics.105.049510
- Guo, Z., Wang, H., Tao, J., Ren, Y., Xu, C., Wu, K., et al. (2019). Development of multiple snp marker panels affordable to breeders through genotyping by target sequencing (GBTS) in maize. *Mol. Breed.* 39, 37–48. doi: 10.1007/s11032-019-0940-4
- Guo, Z., Yang, Q., Huang, F., Zheng, H., Sang, Z., Xu, Y., et al. (2021). Development of high-resolution multiple-SNP arrays for genetic analyses and molecular breeding through genotyping by target sequencing and liquid chip. *Plant Commun.* 2, 100230. doi: 10.1016/j.xplc.2021.100230
- Heffner, E. L., Lorenz, A. J., Jannink, J. L., and Sorrells, M. E. (2010). Plant breeding with genomic selection: gain per unit time and cost. *Crop Sci.* 50, 1681–1690. doi: 10.2135/cropsci2009.11.0662
- Ishimori, M., Hattori, T., Yamazaki, K., Takanashi, H., Fujimoto, M., Kajiya-Kanegae, H., et al. (2020). Impacts of dominance effects on genomic prediction of sorghum hybrid performance. *Breed. Sci.* 70, 605–616. doi: 10.1270/jsbbs.20042

- Jakobsson, M., and Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23, 1801–1806. doi: 10.1093/bioinformatics/btm233
- Knoch, D., Werner, C. R., Meyer, R. C., Riewe, D., Abbadi, A., Lücke, S., et al. (2021). Multi-omics-based prediction of hybrid performance in canola. *Theor. Appl. Genet.* 134, 1147–1165. doi: 10.1007/s00122-020-03759-x
- Kriaridou, C., Tsairidou, S., Houston, R. J. D., and Robledo, D. (2020). Genomic prediction using low density marker panels in aquaculture: performance across species, traits, and genotyping platforms. *Front. Genet.* 11. doi: 10.3389/fgene.2020.00124
- Liu, J., Huang, J., Guo, H., Lan, L., Wang, H., Xu, Y., et al. (2017a). The conserved and unique genetic architecture of kernel size and weight in maize and rice. *Plant Physiol.* 175, 774–785. doi: 10.1104/pp.17.00708
- Liu, X., Hu, X., Li, K., Liu, Z., Wu, Y., Feng, G., et al. (2021). Identifying quantitative trait loci for the general combining ability of yield-relevant traits in maize. *Breed. Sci.* 71, 217–228. doi: 10.1270/jsbs.20008
- Liu, Y., Yi, Q., Hou, X., Hu, Y., and Huang, Y. (2020). Identification of quantitative trait loci for kernel-related traits and the heterosis for these traits in maize (*Zea mays* L.). *Mol. Genet. Genomics* 295, 1–13. doi: 10.1007/s00438-019-01608-1
- Liu, P., Zhao, Y., Liu, G., Wang, M., Hu, D., Hu, J., et al. (2017b). Hybrid performance of an immortalized F₂ rapeseed population is driven by additive, dominance, and epistatic effects. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00815
- Li, D., Xu, Z., Gu, R., Wang, P., Xu, J., Du, D., et al. (2021). Genomic prediction across structured hybrid populations and environments in maize. *Plants (Basel)* 10, 1174–1185. doi: 10.3390/plants10061174
- Ma, J., and Cao, Y. (2021). Genetic dissection of grain yield of maize and yield-related traits through association mapping and genomic prediction. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.690059
- Ma, J., Wang, L., Cao, Y., Wang, H., and Li, H. (2021). Association mapping and transcriptome analysis reveal the genetic architecture of maize kernel size. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.632788
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL icimapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* 3, 269–283. doi: 10.1016/j.cj.2015.01.001
- Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829. doi: 10.1017/S0016672301004931
- Pang, J., Fu, J., Zong, N., Wang, J., Song, D., Zhang, X., et al. (2019). Kernel size-related genes revealed by an integrated eQTL analysis during early maize kernel development. *Plant J.* 98, 19–32. doi: 10.1111/tpj.14193
- Pérez, P., and de los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198, 483–495. doi: 10.1534/genetics.114.164442
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1093/genetics/155.2.945
- Rasheed, A., Hao, Y., Xia, X., Khan, A., Xu, Y., Varshney, R. K., et al. (2017). Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. *Mol. Plant* 10, 1047–1064. doi: 10.1016/j.molp.2017.06.008
- Schrag, T. A., Schipprack, W., and Melchinger, A. E. (2019). Across-years prediction of hybrid performance in maize using genomics. *Theor. Appl. Genet.* 132, 933–946. doi: 10.1007/s00122-018-3249-5
- Shao, L., Xing, F., Xu, C., Zhang, Q., Che, J., Wang, X., et al. (2019). Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the genetic basis of heterosis. *Proc. Natl. Acad. Sci. U.S.A.* 116, 5653–5658. doi: 10.1073/pnas.1820513116
- Shen, Y., Wang, J., Shaw, R. K., Yu, H., Sheng, X., Zhao, Z., et al. (2021). Development of GBTS and KASP panels for genetic diversity, population structure, and fingerprinting of a large collection of broccoli (*Brassica oleracea* L. var. *italica*) in China. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.655254
- Technow, F., Riedelsheimer, C., Schrag, T. A., and Melchinger, A. E. (2012). Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. TAG. theoretical and applied genetics. *Theor. Appl. Genet.* 125, 1181–1194. doi: 10.1007/s00122-012-1905-8
- Torkamaneh, D., Laroche, J., and Belzile, F. (2016). Genome-wide SNP calling from genotyping by sequencing (GBS) data: a comparison of seven pipelines and two sequencing technologies. *PLoS One* 11, e0161333. doi: 10.1371/journal.pone.0161333
- Wang, C., Liu, R., Liu, Y., Hou, W., Wang, X., Miao, Y., et al. (2021). Development and application of the Faba_bean_130K targeted next-generation sequencing SNP genotyping platform based on transcriptome sequencing. *Theor. Appl. Genet.* 134, 3195–3207. doi: 10.1007/s00122-021-03885-0
- Wang, Y., Nie, L., Ma, J., Zhou, B., Han, X., Cheng, J., et al. (2022). Transcriptomic variations and network hubs controlling seed size and weight during maize seed development. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.828923
- Wang, M. L., Zhu, C., Barkley, N. A., Chen, Z., Erpelding, J. E., Murray, S. C., et al. (2009). Genetic diversity and population structure analysis of accessions in the US historic sweet sorghum collection. *Theor. Appl. Genet.* 120, 13–23. doi: 10.1007/s00122-009-1155-6
- Windhausen, V. S., Atlin, G. N., Hickey, J. M., Crossa, J., Jannink, J. L., Sorrells, M. E., et al. (2012). Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. *G3 (Bethesda)* 2, 1427–1436. doi: 10.1534/g3.112.003699
- Yang, N., Lu, Y., Yang, X., Huang, J., Zhou, Y., Ali, F., et al. (2014). Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. *PLoS Genet.* 10, e1004573. doi: 10.1371/journal.pgen.1004573
- Zhang, M., Cui, Y., Liu, Y. H., Xu, W., Sze, S. H., Murray, S. C., et al. (2020b). Accurate prediction of maize grain yield using its contributing genes for gene-based breeding. *Genomics* 112, 225–236. doi: 10.1016/j.ygeno.2019.02.001
- Zhang, A., Pérez-Rodríguez, P., Vicente, F. S., Palacios-Rojas, N., Dhliwayo, T., Liu, Y., et al. (2022). Genomic prediction of the performance of hybrids and the combining abilities for line by tester trials in maize. *Crop J.* 10, 109–116. doi: 10.1016/j.cj.2021.04.007
- Zhang, J., Yang, J., Zhang, L., Luo, J., Zhao, H., Zhang, J., et al. (2020a). A new SNP genotyping technology target SNP-seq and its application in genetic analysis of cucumber varieties. *Sci. Rep.* 10, 5623–5634. doi: 10.1038/s41598-020-62518-6
- Zhang, C., Zhou, Z., Yong, H., Zhang, X., Hao, Z., Zhang, F., et al. (2017). Analysis of the genetic architecture of maize ear and grain morphological traits by combined linkage and association mapping. *Theor. Appl. Genet.* 130, 1011–1029. doi: 10.1007/s00122-017-2867-7
- Zhao, Y., Li, Z., Liu, G., Jiang, Y., Maurer, H. P., Würschum, T., et al. (2015). Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15624–15629. doi: 10.1073/pnas.1514547112