



# Genomic Prediction of Yield Traits in Single-Cross Hybrid Rice (*Oryza sativa* L.)

Marlee R. Labroo<sup>1</sup>, Jauhar Ali<sup>2\*</sup>, M. Umair Aslam<sup>2</sup>, Erik Jon de Asis<sup>2</sup>,  
Madonna A. dela Paz<sup>2</sup>, M. Anna Sevilla<sup>2</sup>, Alexander E. Lipka<sup>1</sup>, Anthony J. Studer<sup>1</sup> and  
Jessica E. Rutkoski<sup>1</sup>

<sup>1</sup> Department of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, United States, <sup>2</sup> Rice Breeding Platform, International Rice Research Institute, Los Baños, Philippines

## OPEN ACCESS

### Edited by:

Shiori Yabe,  
Institute of Crop Science (NARO),  
Japan

### Reviewed by:

Motoyuki Ishimori,  
The University of Tokyo, Japan  
Hiromi Kajiya-Kanegae,  
National Agriculture and Food  
Research Organization (NARO), Japan

### \*Correspondence:

Jauhar Ali  
j.ali@irri.org

### Specialty section:

This article was submitted to  
Statistical Genetics and Methodology,  
a section of the journal  
Frontiers in Genetics

**Received:** 09 April 2021

**Accepted:** 09 June 2021

**Published:** 30 June 2021

### Citation:

Labroo MR, Ali J, Aslam MU,  
de Asis EJ, dela Paz MA, Sevilla MA,  
Lipka AE, Studer AJ and Rutkoski JE  
(2021) Genomic Prediction of Yield  
Traits in Single-Cross Hybrid Rice  
(*Oryza sativa* L.).  
*Front. Genet.* 12:692870.  
doi: 10.3389/fgene.2021.692870

Hybrid rice varieties can outyield the best inbred varieties by 15 – 30% with appropriate management. However, hybrid rice requires more inputs and management than inbred rice to realize a yield advantage in high-yielding environments. The development of stress-tolerant hybrid rice with lowered input requirements could increase hybrid rice yield relative to production costs. We used genomic prediction to evaluate the combining abilities of 564 stress-tolerant lines used to develop Green Super Rice with 13 male sterile lines of the International Rice Research Institute for yield-related traits. We also evaluated the performance of their F<sub>1</sub> hybrids. We identified male sterile lines with good combining ability as well as F<sub>1</sub> hybrids with potential further use in product development. For yield per plant, accuracies of genomic predictions of hybrid genetic values ranged from 0.490 to 0.822 in cross-validation if neither parent or up to both parents were included in the training set, and both general and specific combining abilities were modeled. The accuracy of phenotypic selection for hybrid yield per plant was 0.682. The accuracy of genomic predictions of male GCA for yield per plant was 0.241, while the accuracy of phenotypic selection was 0.562. At the observed accuracies, genomic prediction of hybrid genetic value could allow improved identification of high-performing single crosses. In a reciprocal recurrent genomic selection program with an accelerated breeding cycle, observed male GCA genomic prediction accuracies would lead to similar rates of genetic gain as phenotypic selection. It is likely that prediction accuracies of male GCA could be improved further by targeted expansion of the training set. Additionally, we tested the correlation of parental genetic distance with mid-parent heterosis in the phenotyped hybrids. We found the average mid-parent heterosis for yield per plant to be consistent with existing literature values at 32.0%. In the overall population of study, parental genetic distance was significantly negatively correlated with mid-parent heterosis for yield per plant ( $r = -0.131$ ) and potential yield ( $r = -0.092$ ), but within female families the correlations were non-significant and near zero. As such, positive parental genetic distance was not reliably associated with positive mid-parent heterosis.

**Keywords:** hybrid rice, genomic selection, general combining ability, breeding, best linear unbiased prediction

## INTRODUCTION

Hybrid crop varieties are economically valued for increased vigor, yield, yield stability, and uniformity in species including maize, sugar beet, and cotton (Hochholdinger and Baldauf, 2018). Rice (*Oryza sativa* L.) is a self-pollinated crop that has traditionally been cultivated as an inbred, but the introduction of male sterility into cultivated germplasm in the 1970s enabled hybrid breeding (Virmani and Wan, 1988; Yuan et al., 1989; Nalley et al., 2017). Public hybrid rice varietal development to date has resulted primarily from identification of superior single crosses rather than the systematic breeding of heterotic pools (Lu and Xu, 2010; Spielman et al., 2013). Developing heterotic pools for rice by reciprocal recurrent selection methods may increase the rate of genetic gain for hybrid rice breeding compared to evaluating random crosses, because reciprocal recurrent selection can concurrently improve the additive value of the populations while exploiting heterosis due to dominance (Comstock et al., 1949).

Existing hybrid rice varieties may outyield the best inbred varieties by 10 to 30% with appropriate management (Spielman et al., 2013). However, adoption of hybrid rice varieties is low outside of China, in part because the hybrid yield advantage of temperate *japonica* varieties used in China is much greater than that observed in tropical *indica* varieties to date (Janaiah and Xie, 2010; Longin et al., 2012; Spielman et al., 2013). In some countries, socioeconomic factors such as lack of irrigation systems, paved roads, certified seed suppliers, seed marketing, farmer education, and available credit to purchase seed and fertilizer have limited hybrid rice adoption (Mottaleb et al., 2015; Abebrese et al., 2019). Of the agronomic factors that influence hybrid rice adoption, poor grain quality has been a longstanding challenge, but breeding progress since the early 2000s has produced some acceptable varieties (Spielman et al., 2013). Surveys of farmers suggest that poor quality is not the primary determinant of hybrid rice rejection (Spielman et al., 2013; Feng et al., 2017). Farmers choose not to grow hybrid rice for many reasons, including the high cost of seed, poor seed quality, and lack of hybrid seed availability (Spielman et al., 2013). However, the key agronomic reason for limited adoption is that hybrid rice varieties require more intensive management of irrigation, fertilizer, weeds, and other biotic stressors to provide a yield advantage over inbred varieties in otherwise high-yielding environments (Spielman et al., 2013; Mottaleb et al., 2015; Nalley et al., 2016). Therefore, the development of stress-tolerant hybrids with lowered input requirements could spur hybrid adoption and unlock hybrid yield advantages.

In this study, we evaluated the general combining abilities (GCAs) of the existing male sterile lines of the International Rice Research Institute (IRRI) with stress-tolerant germplasm used in the development of Green Super Rice varieties, as well as the performance, or genetic value, of their F<sub>1</sub> hybrids (Sprague and Tatum, 1942; Ali et al., 2018; Yu et al., 2020). The founders of the Green Super Rice program were selected for multiple stress tolerances, including salinity, submergence, tungro disease, anaerobic germination conditions, and low water and nitrogen inputs (Ali et al., 2018). We sought to identify any outstanding

F<sub>1</sub> hybrids—which may be as stress-tolerant and yet higher-yielding than existing Green Super Rice lines—to advance for further testing for varietal release. We also sought to identify male and female lines with superior GCA which could be used to initiate the development of heterotic pools from IRRI germplasm, presumably stacked with alleles conferring stress tolerance. In addition to phenotypic evaluation, we used genomic prediction to evaluate non-phenotyped parental lines and hybrid crosses.

We also tested whether parental genetic distance was correlated with mid-parent heterosis using a large sample of hybrids and genome-wide molecular markers. Mid-parent heterosis due to dominance is expected to be positively correlated with parental squared difference in allele frequency (SDAF) by quantitative genetic theory (Falconer and Mackay, 1996; Amuzu-Aweh et al., 2013). It has also been posited that genetic divergence in founders of heterotic pools may lead to improved gain in reciprocal recurrent selection programs, even though in practice heterotic pools have been developed from closely related germplasm in species such as maize (Melchinger, 1999; Tracy and Chandler, 2006; Rembe et al., 2019). A previous study of rice which used > 100k genome wide markers and six parental lines found a curvilinear relationship of genetic distance and mid-parent heterosis, with mid-parent heterosis increasing with genetic distance to a point and then declining (Waters et al., 2015). Due to past lack of availability of molecular markers, other studies used fewer than 500 markers and found positive correlations of genetic distance and heterosis using 10 or 22 parents (Xiao et al., 1996; Kwon et al., 2002). However, another study using 319 markers found no correlation of genetic distance and heterosis in progeny of 13 parents (Boeven et al., 2020). In other species, such as wheat, whether parental genetic distance is correlated with heterosis varies, with different findings among studies and populations (Melchinger, 1999; Boeven et al., 2020). We wished to test whether parents of hybrids with high SDAF tended to produce hybrids with high mid-parent heterosis in our rice population of study.

## MATERIALS AND METHODS

### Plant Materials and Population Design

The plant materials for prediction comprised 13 female lines, 564 male lines, and their 10,716 possible F<sub>1</sub> hybrids (**Supplementary File 1**). Twelve of the female lines were wild-abortive cytoplasmic male sterile (CMS), and one female line, A07, was thermosensitive genic male sterile (TGMS). The 564 male tester lines were backcross introgression lines (BILs) from 11 families. Each family of BILs was generated by crossing one of the 11 diverse males to a common female, Weed Tolerant Rice 1 (WTR-1), and advancing the backcrosses to the BC<sub>1</sub>F<sub>5</sub> generation under stringent selection for multiple stress tolerances as described in Ali et al. (2018). The recurrent parent, WTR-1, was a restorer line, but the male lines likely segregated for fertility restoration.

Of the 10,716 possible F<sub>1</sub> hybrids, a random subset of 938 were made to comprise the genomic prediction model training set by crossing six female lines to 137 males. To avoid unintentional

selection for synchronous flowering, the female parents had two planting dates. None of the 137 male lines were completely crossed to all six females. However, in pairwise comparisons of females, all females had some overlap with each other female in males crossed (**Supplementary Table 1**). In total, 85 of the males were crossed to a single female, 108 of the males were crossed to 2 females, 124 of the males were crossed to 3 females, 60 of the males were crossed to 4 females, and 5 of the males were crossed to 5 females. All female lines were manually emasculated to prevent contamination by selfing and to expose the stigma.

Two groups of lines were used to estimate mid-parent heterosis and commercial relative performance but were not included for prediction. The five maintainer (B) lines of the five CMS female parents were used to estimate mid-parent heterosis. Six inbred lines were used as commercial checks to estimate commercial relative performance: five of the donor parents, Y 134 (DP 6), Khazar (DP 8), OM 997 (DP 10), M 401 (DP17), and X 21 (DP19), and the recurrent parent, WTR-1.

## Field Experimental Design

The F<sub>1</sub> hybrids, their inbred parents, the female maintainer lines, and the commercial checks were phenotyped in an unbalanced randomized complete block design (RCBD) in two environments, irrigated lowland and irrigated upland, at the IRRI farm (approximately 14°09'50.7" N, 121°15'50.5" E) in the dry season of 2018. After establishment in seedbeds on January 8, 2018, seedlings were transplanted at the three-leaf stage. Transplanting occurred on January 31, 2018, at the irrigated lowland site and on February 8, 2018, at the irrigated upland site. The plants were harvested the week of May 14, 2018. Basal NPK fertilizer was applied at a rate of 30 kg/ha, and zinc was applied at a rate of 5 kg/ha. N fertilizer was also applied at 28–30 days after transplanting and at the panicle initiation stage (42 days after transplanting) at a rate of 35 kg/ha. Rat fences were installed at both locations; bird pressure was controlled by farmworkers in the lowland environment, and by a bird net at the upland environment. Both environments were hand-weeded. Both environments were irrigated, but the lowland environment was continuously flooded to a depth of ~10 cm, whereas water depth was allowed to vary in the upland field. Insect pressure was controlled by application of Regent® pesticide (fipronil). Temperatures were sufficient throughout the growing season to ensure seed set in the TGMS female line A07.

The field layout was designed in PBTtools 1.4, which depends on the R package agricolae (IRRI, 2014; de Mendiburu, 2020). There were two blocks per environment with one replicate per genotype per block, but replicates were missing for some genotypes. Genotypes were replicated in single-row plots due to limited availability of hybrid seed, with five plants per row, and plants were spaced to 25 × 20 cm within rows. Measurements were only recorded for plants at 20 cm spacing within rows; i.e., edge plants were not measured, nor were plants with missing neighbor plants within the row.

The following traits or trait derivatives were phenotyped: plant height, number of tillers, panicle dry weight, panicle length, proportion of spikelets filled, yield per plant, and yield potential per plant (**Supplementary File 1**). For all genotype replicates,

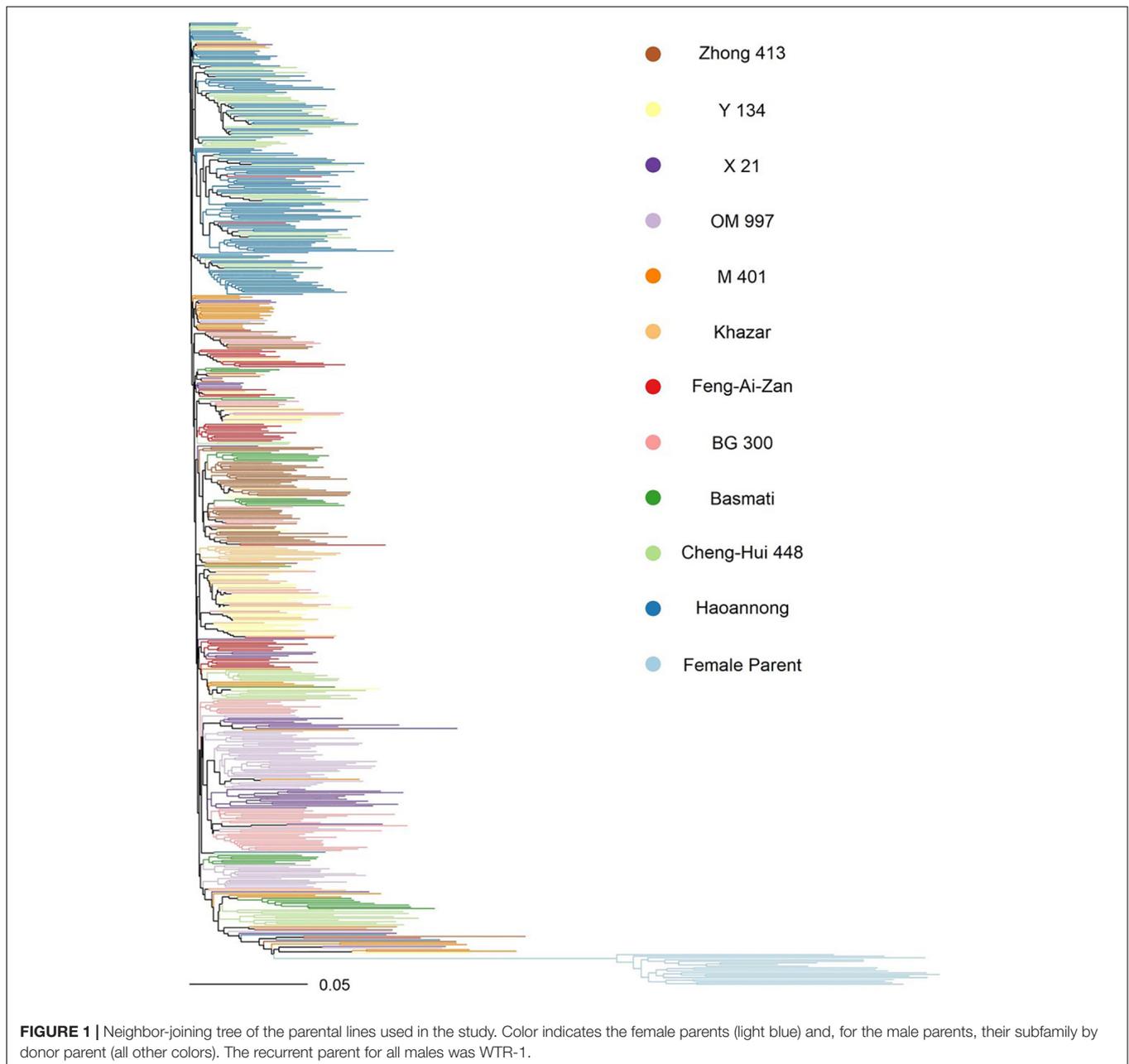
phenotypes were averaged across measured plants in a single-row plot; plants were subsamples, but were not treated as subsamples in downstream modeling because subsampling variance was not of interest. Plant height was measured from the base of the plant to the panicle tip after flowering. All tillers were assumed to be productive based on observations in a subset of samples. For panicle measurements and yield estimates, three random panicles were harvested from each sampled plant, totaling nine panicles per replicate. Panicle length was measured from the pedicel to the panicle tip and averaged across all panicles in a replicate. For a given replicate, yield per plant was calculated as panicle dry weight times tiller number. For a given replicate, yield potential per plant was calculated as average panicle dry weight divided by proportion of spikelets filled times tiller number. Yield potential per plant and proportion of spikelets filled were only measured in the irrigated lowland environment due to cost.

## Molecular Marker Generation and Analysis

Genome-wide molecular markers were generated for the parents with tunable genotyping-by-sequencing® (tGBS) by data2Bio and its subsidiary, Freedom Markers, in Ames, Iowa (Ali et al., 2017; Ott et al., 2017). In general, each individual DNA sample was double-digested with restriction enzymes, then the resulting fragments were ligated to a uniquely barcoded adapter at the 5' end. At the 3' end, the fragments were ligated to a universal sequencing adapter. However, in the first subsequent PCR amplification of the library, the complementary primer for the universal sequencing adapter was extended by 1–3 nucleotides; only fragments in which the genomic sequence complemented the extension were amplified. Then, the libraries were amplified with Ion Proton sequencing primers.

The male and female parents were sequenced in separate Ion Proton runs. For the male parents, a total of ten Ion Proton runs were done; the female parents were sequenced in two Ion proton runs as part of a larger set. The Ion Proton sequencing reads were trimmed by the manufacturer to remove adapter sequence and bases with PHRED quality scores less than 15 in the software Lucy (Chou and Holmes, 2001). For the female parent A07, additional RAD-sequencing was done in-house. In brief, DNA was extracted from mature leaf tissue of the A07 parent and digested separately with one of three enzyme combinations: *ApeKI-PstI*, *HinPII-PstI*, or *ApeKI* only. Then, each digestion was ligated separately to unique barcoded adapters and subsequently pooled. Fragments were selected for sizes ranging from 200–500 bp, and the libraries were amplified by PCR. Then, the libraries were sequenced for single-read 100 bp reads with an Illumina NovaSeq6000 SP. All reads were aligned to the Nipponbare IRGSP-1.0 v7 reference genome in GSNAP 2017-11-15 (Kawahara et al., 2013; Wu et al., 2016). Then, variants were called in BCftools 1.7 (SAMtools, 2018).

Variant sites were filtered separately within the male and female parent sets in TASSEL 5.0 (Glaubitz et al., 2014). Within sets, only the 2 most major alleles of the variant were considered, and at most 50% of the individuals were permitted to be heterozygous. In the females, the minimum site count was



2 (corresponding to a minimum site presence of 10% in all females), and the minimum minor allele count was 2, yielding 77,709 polymorphic sites among the females. In the males, the minimum site count was 188 (corresponding to a minimum site presence of 33% in the males), and the minimum minor allele count was 3, yielding 148,922 total polymorphic sites among the males. The genotypes were imputed separately for males and females in Beagle 5.0 (Browning et al., 2018). Principal components analysis of the male and female parent genotypes was done using 20,000 sites common to both using the *glPca* function in the R package *adegenet* (Jombart and Ahmed, 2011). The  $F_1$  hybrid genotypes were inferred from the same 20,000 sites common to males and females using the *build.HMM* function in

the R package *sommer* version 3.8 (Covarrubias-Pazaran, 2016, 2018). A phylogenetic neighbor-joining tree of the male and female parent genotypes was constructed in TASSEL 5.0 using the same common 20,000 sites, and the tree was visualized in the R package *ape* version 5.5 (Figure 1; Paradis and Schliep, 2019).

The additive relationship matrices of each the females, males, and  $F_1$  hybrids, denoted respectively as  $G_F$ ,  $G_M$ , and  $G_H$ , were calculated with the *A.mat* function in *sommer* by the method of Endelman and Jannink (2012). The 77,709 imputed female sites were used to estimate  $G_F$ , and the 148,922 imputed male sites were used to estimate  $G_M$ . The hybrid genotypes inferred from the 20,000 sites common to males and females via the *build.HMM* function were used to estimate  $G_H$ . The specific combining ability

(SCA) relationship matrix  $\mathbf{G}_{FM}$  was the Kronecker product of the male and female additive relationship matrices (Bernardo, 2002). Pairwise SDAF, a hypothetical predictor of mid-parent heterosis, was calculated as:

$$SDAF_{ij} = \frac{\sum_{n=1}^N (p_{in} - p_{jn})^2}{N} \quad (1)$$

Where  $SDAF_{ij}$  was the squared difference in allele frequency between the  $i^{\text{th}}$  female parent and the  $j^{\text{th}}$  male parent,  $p_{in} - p_{jn}$  was the difference allele frequency between the  $i^{\text{th}}$  female parent and  $j^{\text{th}}$  male parent at the  $n^{\text{th}}$  variant site, and  $N$  was the total number of variant sites.

## Modeling and Statistical Analysis

All linear models were fit in a single step with the *mmer* function in sommer (Covarrubias-Pazarán, 2016, 2018). For a given trait, i.e., plant height, tiller number, panicle length, average yield per plant, proportion of spikelets filled, or potential yield per plant, the genotype replicates (i.e., single-row plots) were considered the experimental unit. Genetic variances for plant height, tiller number, panicle length, and average yield per plant traits were estimated with models of the following form:

$$Y_{ijkl} = \mu + H_i + E_j + HE_{ij} + B_{(k)j} + \varepsilon_{ijkl} \quad (2)$$

where  $Y_{ijkl}$  was the random phenotypic response of the  $i^{\text{th}}$  single-cross hybrid genotype in the  $k^{\text{th}}$  block nested in the  $j^{\text{th}}$  environment from the  $l^{\text{th}}$  replicate,  $\mu$  was the grand mean,  $H_i$  was the random effect of the  $i^{\text{th}}$  hybrid genotype with  $N(0, \mathbf{I}\sigma_H^2)$ ,  $E_j$  was the random effect of the  $j^{\text{th}}$  environment with  $N(0, \mathbf{I}\sigma_E^2)$ ,  $HE_{ij}$  was the random interaction of the  $i^{\text{th}}$  hybrid genotype and the  $j^{\text{th}}$  environment with  $N(0, \mathbf{I}\sigma_{HE}^2)$ ,  $B_{(k)j}$  was the effect of the  $k^{\text{th}}$  block nested within the  $j^{\text{th}}$  environment with  $N(0, \mathbf{I}\sigma_B^2)$ , and  $\varepsilon_{ijkl}$  was the random error associated with each replicate with  $N(0, \mathbf{I}\sigma_e^2)$ .

The genetic variance for proportion of spikelets filled and potential yield per plant was estimated using the following model in (3), without the environment term and its associated interactions, because the traits were only phenotyped in the irrigated lowland environment:

$$Y_{ijk} = \mu + H_i + B_j + \varepsilon_{ijk} \quad (3)$$

$Y_{ijk}$  was the random phenotypic response of the  $i^{\text{th}}$  hybrid genotype in the  $j^{\text{th}}$  block from the  $k^{\text{th}}$  replicate,  $\mu$  was the grand mean,  $H_i$  was the random effect of the  $i^{\text{th}}$  hybrid genotype with  $N(0, \mathbf{I}\sigma_H^2)$ ,  $B_j$  was the random effect of the  $j^{\text{th}}$  block with  $N(0, \mathbf{I}\sigma_B^2)$ , and  $\varepsilon_{ijk}$  was the random error associated with each replicate with  $N(0, \mathbf{I}\sigma_e^2)$ .

The entry-mean heritability was estimated for each of height, tiller number, panicle length, and yield per plant by (4) following the method of Holland et al. (2003) for unbalanced RCBDs:

$$H^2 = \frac{\sigma_H^2}{\sigma_H^2 + \frac{\sigma_{HE}^2}{h_j} + \frac{\sigma_e^2}{h_t}} \quad (4)$$

where  $\sigma_H^2$  was the variance among hybrid genotypes from models of the form in (2),  $\sigma_{HE}^2$  was the variance of the interaction of the hybrid genotype and environment,  $\sigma_e^2$  was the error variance,  $h_j$  was the harmonic mean of the number of total observations of each hybrid genotype within an environment, and  $h_t$  was the harmonic mean of the total number of observations per hybrid genotype.

For proportion of spikelets filled and potential yield, which were phenotyped in a single environment, entry-mean heritability was estimated by (5) also following Holland et al. (2003), using the following equation with the terms as described in (3):

$$H^2 = \frac{\sigma_H^2}{\sigma_H^2 + \frac{\sigma_e^2}{h_t}} \quad (5)$$

Additive genetic variances were estimated from models of the form in (6) for plant height, tiller number, panicle length, and yield per plant. The terms of (6) are the same as in (2), but in (6) the random effect  $H$  was assumed to have a multivariate normal (MVN) distribution with  $H \sim MVN(0, \mathbf{G}_H\sigma_H^2)$ , where  $\mathbf{G}_H$  was the additive genomic relationship matrix of the  $F_1$  hybrids:

$$Y_{ijkl} = \mu + H_i + E_j + HE_{ij} + B_{(k)j} + \varepsilon_{ijkl} \quad (6)$$

Additive genetic variances for proportion of spikelets filled and potential yield were estimated from model of the form in (7), with the same terms as (3), but the random effect  $H$  was assumed to have a multivariate normal (MVN) distribution with  $H \sim MVN(0, \mathbf{G}_H\sigma_H^2)$ , where  $\mathbf{G}_H$  was the additive genomic relationship matrix of the  $F_1$  hybrids:

$$Y_{ijk} = \mu + H_i + B_j + \varepsilon_{ijk} \quad (7)$$

Narrow-sense heritability, or the proportion of additive genetic variance out of total phenotypic variance, was estimated on a single-plant basis for all traits. Variance components were estimated from the models of the form in (6) for plant height, tiller number, panicle length, and yield per plant, and narrow-sense heritability was estimated with (8):

$$h^2 = \frac{\sigma_H^2}{\sigma_H^2 + \sigma_{HE}^2 + \sigma_e^2} \quad (8)$$

For proportion of spikelets filled and potential yield, narrow-sense heritability was estimated using (9) with variances estimated from the models of the form in (7):

$$h^2 = \frac{\sigma_H^2}{\sigma_H^2 + \sigma_e^2} \quad (9)$$

For each trait, genomic best linear unbiased predictions (GBLUPs) of hybrid genetic value, male GCA, and female GCA were each estimated using two separate models: the genomic GCA model and the genomic GCA + SCA model. Model fits were compared with the Akaike information criterion (AIC) and Bayesian information criterion (BIC). The genomic GCA model was:

$$Y_{ijklm} = \mu + F_i + M_j + E_k + B_{(l)k} + FE_{ik} + ME_{jk} + \varepsilon_{ijklm} \quad (10)$$

where  $Y_{ijklm}$  was the random phenotypic response of a single-cross hybrid of the  $i^{\text{th}}$  female line and the  $j^{\text{th}}$  male line observed in the  $k^{\text{th}}$  environment,  $l^{\text{th}}$  block, and  $m^{\text{th}}$  replicate,  $\mu$  was the grand mean,  $F_i$  was the random GCA effect of the  $i^{\text{th}}$  female parent with  $F \sim MVN(0, \mathbf{G}_F\sigma_F^2)$  where  $\mathbf{G}_F$  was the additive genomic relationship matrix among females,  $M_j$  was the random GCA effect of the  $j^{\text{th}}$  male parent with  $M \sim MVN(0, \mathbf{G}_M\sigma_M^2)$  where  $\mathbf{G}_M$  was the additive genomic relationship matrix among males,  $E_k$  was the effect of the  $k^{\text{th}}$  environment with  $N(0, \mathbf{I}\sigma_E^2)$ ,  $B_{(l)k}$  was the effect of the  $l^{\text{th}}$  block nested in the  $k^{\text{th}}$  environment with  $N(0, \mathbf{I}\sigma_B^2)$ ,  $FE_{ik}$  was the random interaction of the  $i^{\text{th}}$  female and the  $k^{\text{th}}$  environment with  $N(0, \mathbf{I}\sigma_{FE}^2)$ ,  $ME_{jk}$  was the random interaction of the  $j^{\text{th}}$  male and the  $k^{\text{th}}$  environment with  $N(0, \mathbf{I}\sigma_{ME}^2)$ , and  $\varepsilon_{ijklm}$  was the random error of each observation with  $N(0, \mathbf{I}\sigma_\varepsilon^2)$ .

The genomic GCA + SCA model was:

$$Y_{ijklm} = \mu + F_i + M_j + E_k + B_{(l)k} + FE_{ik} + ME_{jk} + FM_{ij} + \varepsilon_{ijklm} \tag{11}$$

where terms were as described above, and  $FM_{ij}$  was the additional random SCA interaction effect of the  $i^{\text{th}}$  female and the  $j^{\text{th}}$  male, with  $FM \sim MVN(0, \mathbf{G}_{FM}\sigma_{FM}^2)$ .  $\mathbf{G}_{FM}$  was the Kronecker product of  $\mathbf{G}_F$  and  $\mathbf{G}_M$  (Bernardo, 2002).

Best linear unbiased predictions (BLUPs) of hybrid genetic value and male and female GCAs were also estimated without genomic information to 1) estimate the predictive ability and prediction accuracy of the genomic prediction models, and 2) estimate the accuracy of phenotypic selection. For the GCA model, all terms remained the same as in (10), but the distribution of the random effects of  $F$  and  $M$  were simply assumed to be  $N(0, \mathbf{I}\sigma_F^2)$  and  $N(0, \mathbf{I}\sigma_M^2)$  respectively.

$$Y_{ijklm} = \mu + F_i + M_j + E_k + B_{(l)k} + FE_{ik} + ME_{jk} + \varepsilon_{ijklm} \tag{12}$$

Similarly, for the GCA + SCA model, all terms remained the same as in (11), but the distribution of the random effects  $F$ ,  $M$  and  $FM$  were assumed to be  $N(0, \mathbf{I}\sigma_F^2)$ ,  $N(0, \mathbf{I}\sigma_M^2)$ , and  $N(0, \mathbf{I}\sigma_{FM}^2)$  respectively:

$$Y_{ijklm} = \mu + F_i + M_j + E_k + B_{(l)k} + FE_{ik} + ME_{jk} + FM_{ij} + \varepsilon_{ijklm} \tag{13}$$

There are multiple methods to estimate genomic prediction accuracy (Estaghevrou et al., 2013). Here, predictive ability was Pearson's correlation of an estimated value and a true value. Prediction accuracy was considered to be predictive ability divided by the square root of the reliability of the estimated value (Mrode, 2014). This method of estimating prediction accuracy, which is well-established in animal breeding programs, was chosen because it is relatively unbiased, precise, and stable compared to other methods (Estaghevrou et al., 2013). Predictive abilities of the genomic GCA and genomic GCA + SCA models described in (10) and (11) were each estimated for male GCA by ten-fold cross-validation (Resende et al., 2012). Sample size was

inadequate to estimate genomic predictive ability and accuracy for female GCA. In each fold, the hybrid progeny phenotypes of 38 of the males were masked from the training set, and the masked training set was used to train the prediction model. Predictive abilities for male GCA, or Pearson's correlation of the GBLUP of male GCA estimated in the training set fold and the BLUP of male GCA estimated from all available observations in the full dataset, were then calculated for the 38 masked males and averaged across folds for each model. Prediction accuracy for male GCA for each model was the predictive ability divided by the square root of the reliability of the genomic prediction. For each model, the square root of the reliability of the genomic prediction was the correlation of the GBLUP of male GCA and BLUP of male GCA when all available observations were used for estimation of both, as in (10), (11), (12), and (13).

Predictive abilities and prediction accuracies were also estimated for hybrid genetic values for each phenotypic response following Technow et al. (2014). For each of 500 replications, four females and 127 males which had been crossed to at least one of the four females were randomly sampled. Then, a training set was formed by randomly sampling 150 hybrids which had phenotypic records available, with the constraint that the randomly sampled male and female lines had at least one hybrid progeny in the training set. Predictive ability, here Pearson's correlation of the hybrid genetic values estimated from the training fold and the observed hybrid genetic values, was recorded for hybrids which were not included in the training set. Predictive ability was recorded separately for hybrids which had both parents included in the training set (T2), one parent included in the training set (T1), only the female parent included in the training set (T1F), only the male parent included in the training set (T1M), and neither parent included in the training set (T0). The prediction accuracy was the predictive ability divided by the square root of the reliability of the genomic prediction, here the correlation of the genomic BLUP of hybrid genetic value and BLUP of hybrid genetic value when all available observations were used for estimation of both.

The accuracies of phenotypic selection for each trait were estimated for male GCA, female GCA, and hybrid genetic value as the square root of the reliabilities of their respective BLUPs (Falconer and Mackay, 1996; Mrode, 2014). Reliabilities of male GCA and female GCA were estimated from each the GCA model in (12) and the GCA + SCA model in (13). Reliabilities of hybrid genetic values were estimated from model (2) and (3). To estimate each reliability, the prediction error variances (PEV) of the appropriate BLUPs (i.e., of hybrid genetic value, female GCA, or male GCA) were obtained in sommer by inverting the coefficient matrix of the relevant model. For the BLUPs of male GCA for a given model and trait, the reliability was the average of:

$$1 - \frac{PEV_j}{\sigma_M^2} \tag{14}$$

where  $PEV_j$  was the prediction error variance of the  $j^{\text{th}}$  BLUP of male GCA and  $\sigma_M^2$  was the estimated male GCA variance. For the

BLUPs of female GCA for a given model and trait, the reliability was the average of:

$$1 - \frac{PEV_i}{\sigma_F^2} \tag{15}$$

where  $PEV_i$  was the prediction error variance of the  $i^{\text{th}}$  BLUP of female GCA, and  $\sigma_F^2$  was the estimated female GCA variance. For the BLUPs of hybrid genetic value for a given trait from model (2), the reliability was the average of:

$$1 - \frac{PEV_i}{\sigma_H^2} \tag{16}$$

where  $PEV_i$  was the prediction error variance of the  $i^{\text{th}}$  BLUP of hybrid genetic value, and  $\sigma_H^2$  was the estimated variance among hybrids.

To assess correlation of SDAF with mid-parent heterosis, mid-parent heterosis of each  $F_1$  hybrid was estimated using BLUPs of genetic values from the following model:

$$Y_{ijklm} = \mu + F_i + G_j + E_k + GE_{jk} + B_{(l)k} + \varepsilon_{ijklm} \tag{17}$$

$Y_{ijklm}$  was the random phenotypic response and  $\mu$  was the grand mean. In the vein of Xiang et al. (2016) and Liang et al. (2018),  $F_i$  was the fixed effect of an indicator of whether the genotype was an inbred or  $F_1$  hybrid to account for the possibility of differing inbred and hybrid group means in the presence of heterosis.  $G_j$  was the random effect of the  $j^{\text{th}}$  inbred parent, inbred commercial check, or  $F_1$  hybrid genotype with  $N(0, \mathbf{I}\sigma_G^2)$ .  $E_k$  was the random effect of the  $k^{\text{th}}$  environment with  $N(0, \mathbf{I}\sigma_E^2)$ ,  $GE_{jk}$  was the random interaction effect of the  $j^{\text{th}}$  genotype and the  $k^{\text{th}}$  environment with  $N(0, \mathbf{I}\sigma_{GE}^2)$ ,  $B_{(l)k}$  was the random effect of the  $l^{\text{th}}$  block nested within the  $k^{\text{th}}$  environment with  $N(0, \mathbf{I}\sigma_B^2)$ , and  $\varepsilon_{ijklm}$  was the random error associated with each observation with  $N(0, \mathbf{I}\sigma_\varepsilon^2)$ . The environment term and its associated interactions were dropped in estimation of yield potential and proportion of spikelets filled, because they were observed only in the irrigated lowland environment.

Mid-parent heterosis of each  $F_1$  hybrid was obtained as:

$$MPH = \frac{\hat{H} - \widehat{MP}}{\widehat{MP}} \tag{18}$$

where  $MPH$  was mid-heterosis,  $\hat{H}$  was the BLUP of the  $F_1$  genotype value from (17), and  $\widehat{MP}$  was the mid-parent value, i.e., the mean of the BLUPs from (17) of its parental genotype values (Supplementary File 1). The BLUP of the  $i^{\text{th}}$  genotype value was the sum of  $\mu$ ,  $\hat{F}_i$ , and  $\hat{G}_j$  from (17). Because the CMS parents do not set seed, the corresponding maintainer (B) line phenotype was used to estimate heterosis. Pearson’s correlation of SDAF with mid-parent heterosis was estimated among all hybrids in the study and also separately within families of hybrids with the same female parent. Student’s  $t$  test of significance was conducted at  $\alpha = 0.05$  for each correlation, with the null hypothesis that a given correlation did not significantly differ from zero and the alternate hypothesis that the given correlation significantly differed from zero.

Commercial relative performance (commercial heterosis) was estimated for each  $F_1$  hybrid with phenotypic observations against each check as:

$$CRP = \frac{\hat{H} - \hat{C}}{\hat{C}} \tag{19}$$

where  $CRP$  was commercial relative performance,  $\hat{H}$  was the BLUP of the  $F_1$  hybrid genotype value from (17), and  $\hat{C}$  was the BLUP of the commercial check genotype value from (17). The BLUP of the  $i^{\text{th}}$  genotype value was the sum of  $\mu$ ,  $\hat{F}_i$ , and  $\hat{G}_j$  from (17).

## RESULTS

### Summary Statistics and Heritabilities

Mean phenotypic values for height, tiller number, panicle length, proportion of spikelets filled, yield per plant, and potential yield were respectively 85 cm, 14 tillers, 226 mm, 0.757, 36 g per plant, and 54 g per plant (Supplementary Figure 1; Table 1). Highest entry-mean heritability observed was for height at 0.906, and lowest entry-mean heritability observed was for potential yield at 0.311 (Table 2). Narrow-sense heritabilities on a single-plant basis were greatest for the proportion of spikelets filled at 0.864 and least for potential yield at 0.271 (Table 2). Because the narrow-sense heritability of proportion of spikelets filled was high, perhaps due to differing genetic architectures between TGMS and CMS lines for the trait, we also estimated the narrow-sense heritability in the CMS lines only as 0.922 by removing observations of the TGMS line A07 from the model. Principal components analysis of the parental genotypes showed clustering of the males and females, but divergence of the male and female parents was not due to historical reciprocal recurrent selection (Supplementary Figure 2).

### Model Fit, Predictive Ability, and Prediction Accuracy

For all traits, model fit was superior for the genomic GCA + SCA model compared to the genomic GCA model as assessed by

**TABLE 1 |** Trait mean phenotypic values and standard deviations overall and within environments for the  $F_1$  hybrids.

Trait	Mean ± Standard Deviation, Overall	Mean ± Standard Deviation, Irrigated Lowland	Mean ± Standard Deviation, Irrigated Upland
Height (cm)	85 ± 15	89 ± 13	70 ± 11
Tiller Number	14 ± 5	16 ± 4	10 ± 3
Panicle Length (mm)	226 ± 24	231 ± 22	213 ± 25
Proportion of Spikelets Filled*		0.757 ± 0.149	
Yield per Plant (g)	36 ± 19	44 ± 17	18 ± 7
Potential Yield* (g)		54 ± 20	

\*Proportion of spikelets filled and potential yield were only phenotyped in the irrigated lowland environment.

**TABLE 2** | Estimates and their standard errors for each trait of entry-mean heritability of the F<sub>1</sub> hybrids and narrow-sense heritability on a single-plant basis.

Trait	Entry-Mean Heritability	Narrow-Sense Heritability
Height	0.906 ± 0.007	0.719 ± 0.029
Tiller Number	0.505 ± 0.034	0.376 ± 0.054
Panicle Length	0.824 ± 0.014	0.542 ± 0.046
Proportion of Spikelets Filled	0.780 ± 0.014	0.864 ± 0.014*
Yield per Plant	0.433 ± 0.038	0.328 ± 0.055
Potential Yield	0.311 ± 0.046	0.271 ± 0.063

\*In the CMS lines only, the narrow-sense heritability for proportion of spikelets filled was 0.922 ± 0.008.

either AIC or BIC (**Supplementary Tables 2, 3**). However, the predictive abilities and accuracies of the genomic GCA + SCA models were not substantially different from the genomic GCA models (**Supplementary Tables 4, 5; Tables 3, 4**). Mean prediction accuracies for male GCA ranged from 0.215 to 0.318 for the genomic GCA models and 0.233 to 0.332 in the genomic GCA + SCA models (**Table 4**). Prediction accuracies for untested females were not estimated. For hybrids in the T0 set, mean genomic GCA model accuracies ranged 0.039 to 0.394, while the genomic GCA + SCA model accuracies ranged from 0.043 to 0.490. For hybrids in the T1 set, mean genomic GCA model accuracies ranged from 0.476 to 0.806, while the genomic GCA + SCA model accuracies ranged from 0.509 to 0.827. For hybrids in the T1F set, mean genomic GCA model accuracies ranged from 0.310 to 0.908, while the genomic GCA + SCA model accuracies ranged from 0.364 to 0.943. For hybrids in the T1M set, mean genomic GCA model accuracies ranged from 0.537 to 0.742, while the genomic GCA + SCA model accuracies ranged from 0.423 to 0.785. For hybrids in the T2 set, mean genomic GCA model

accuracies ranged from 0.769 to 0.948, while the genomic GCA + SCA model accuracies ranged from 0.772 to 0.956 (**Table 3**). However, in the hybrid T0 case, accuracy for yield per plant was increased from 0.215 to 0.490 by inclusion of the SCA effect. No other trait had more than a 10% increase in accuracy by inclusion of the SCA effect in the T0 case, and substantial increases in accuracy with inclusion of the SCA effect were not observed for yield per plant in scenarios besides T0. Except in the case of proportion of spikelets filled, accuracy in the T1F scenarios was always substantially higher than the T1M scenarios.

The accuracy of phenotypic selection for hybrid genetic value ranged from 0.566 to 0.952 among traits (**Table 5**). For the GCA model, the accuracy of phenotypic selection for male GCA ranged from 0.484 to 0.861, and the accuracy of phenotypic selection for female GCA ranged from 0.853 to 0.910 (**Tables 6, 7; Supplementary Tables 6, 7**). For the GCA + SCA model, the accuracy of phenotypic selection for male GCA ranged from 0.253 to 0.798, and the accuracy of phenotypic selection for female GCA ranged from 0.850 to 0.910 (**Tables 6, 7; Supplementary Tables 6, 7**).

## Hybrid Genetic Value and Parental GCA

The genomic GCA + SCA model was used to rank hybrid genetic values. The maximum predicted F<sub>1</sub> hybrid yield was 43.352 grams per plant, which scaled to 8.670 tons per hectare (**Supplementary Table 8**). The maximum predicted F<sub>1</sub> hybrid potential yield was 77.401 grams per plant, scaling to 15.479 tons per hectare (**Supplementary Table 9**). Over half of the top 20 F<sub>1</sub> hybrids in terms of yield per plant had phenotypic observations available, though the top-ranked hybrid did not. The relative performance of the F<sub>1</sub> hybrids compared to the commercial inbred checks (commercial heterosis) ranged from -43.9% to 70.0% for yield per plant (**Supplementary Figure 3; Table 8**). The maximum

**TABLE 3** | Mean prediction accuracies ± standard error thereof in cross-validation of the genomic prediction models for hybrids.

Trait	T0	T1	T1F	T1M	T2
<b>Genomic GCA model</b>					
Height	0.345 ± 0.013	0.793 ± 0.009	0.908 ± 0.007	0.562 ± 0.006	0.948 ± 0.003
Tiller Number	0.162 ± 0.005	0.628 ± 0.014	0.644 ± 0.015	0.537 ± 0.008	0.812 ± 0.007
Panicle Length	0.394 ± 0.012	0.806 ± 0.009	0.906 ± 0.008	0.574 ± 0.005	0.942 ± 0.005
Proportion of Spikelets Filled	0.039 ± 0.003	0.476 ± 0.003	0.310 ± 0.006	0.742 ± 0.003	0.769 ± 0.003
Yield per Plant	0.215 ± 0.012	0.688 ± 0.008	0.741 ± 0.008	0.646 ± 0.007	0.820 ± 0.005
Potential Yield	0.380 ± 0.015	0.719 ± 0.011	0.842 ± 0.009	0.566 ± 0.011	0.855 ± 0.006
<b>Genomic GCA + SCA model</b>					
Height	0.414 ± 0.019	0.820 ± 0.010	0.943 ± 0.008	0.498 ± 0.014	0.956 ± 0.005
Tiller Number	0.190 ± 0.007	0.640 ± 0.016	0.676 ± 0.016	0.442 ± 0.012	0.772 ± 0.013
Panicle Length	0.446 ± 0.020	0.827 ± 0.011	0.940 ± 0.011	0.423 ± 0.018	0.928 ± 0.011
Proportion of Spikelets Filled	0.043 ± 0.003	0.509 ± 0.004	0.364 ± 0.007	0.785 ± 0.005	0.809 ± 0.004
Yield per Plant	0.490 ± 0.022	0.715 ± 0.009	0.795 ± 0.009	0.532 ± 0.014	0.822 ± 0.006
Potential Yield	0.415 ± 0.016	0.723 ± 0.012	0.847 ± 0.010	0.525 ± 0.014	0.851 ± 0.007

For each of the 500 cross-validation folds, 4 female parents and 127 male parents were chosen to have hybrid progeny included in the training set. The total number of hybrid genotypes sampled for training was 150. Accuracies are reported for hybrids which were not included in the training set which had neither parent in the training set (T0), one parent in the training set (T1), the female parent only in the training set (T1F), the male parent only in the training set (T1M), and both parents in the training set (T2).

**TABLE 4** | Model prediction accuracy and standard error for male GCA for each trait as estimated by ten-fold cross-validation.

Trait	Genomic GCA model	Genomic GCA + SCA model
Height	0.232 ± 0.056	0.233 ± 0.067
Tiller Number	0.215 ± 0.060	0.261 ± 0.065
Panicle Length	0.224 ± 0.043	0.269 ± 0.046
Proportion of Spikelets Filled	0.318 ± 0.083	0.332 ± 0.079
Yield per Plant	0.219 ± 0.072	0.241 ± 0.079
Potential Yield	0.292 ± 0.078	0.233 ± 0.096

**TABLE 5** | Reliabilities and accuracies of phenotypic selection for hybrid performance.

Trait	Reliability	Accuracy
Height	0.906	0.952
Tiller Number	0.533	0.730
Panicle Length	0.828	0.910
Proportion of Spikelets Filled	0.791	0.889
Yield per Plant	0.466	0.682
Potential Yield	0.321	0.566

**TABLE 6** | Accuracies of phenotypic selection for male GCA.

Trait	GCA model	GCA + SCA model
Height	0.809	0.627
Tiller Number	0.644	0.570
Panicle Length	0.680	0.253
Proportion of Spikelets Filled	0.861	0.798
Yield per Plant	0.640	0.562
Potential Yield	0.484	0.454

**TABLE 7** | Accuracies of phenotypic selection for female GCA.

Trait	GCA model	GCA + SCA model
Height	0.906	0.906
Tiller Number	0.860	0.858
Panicle Length	0.910	0.910
Proportion of Spikelets Filled	0.898	0.876
Yield per Plant	0.853	0.850
Potential Yield	0.900	0.899

genomic predicted GCA for yield per plant in females and males respectively were 36.341 and 34.047; both of the female and male lines top-ranked for GCA had phenotypic observations available (**Supplementary Tables 10, 11**).

## Mid-Parent Heterosis and Parental SDAF

Average mid-parent heterosis was positive, though not extremely so, for all traits except height and proportion of spikelets filled (**Figure 2** and **Table 9**). Yield per plant and its component trait, tiller number, showed the highest average heterosis; average heterosis for yield was 32.0%. Parental SDAF ranged from 0.200 to 0.285 in the phenotyped F<sub>1</sub> hybrids (**Supplementary Figure 4**).

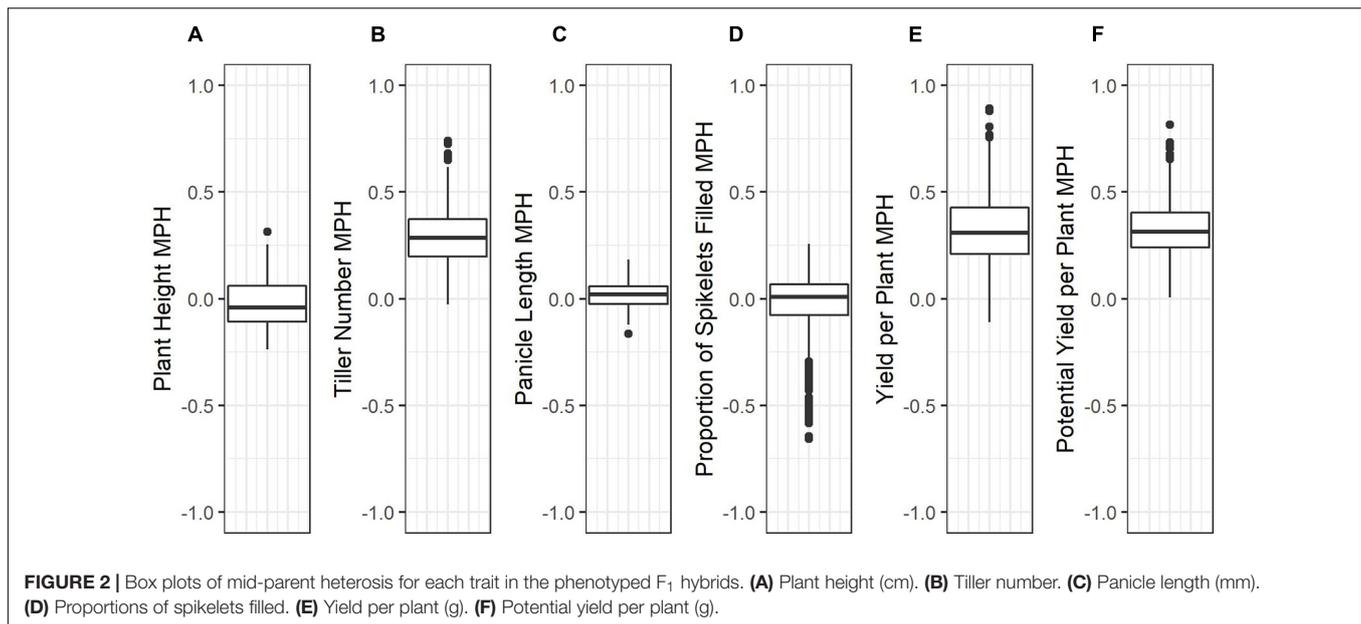
**TABLE 8** | Mean, standard deviation, and range of relative yield per plant of the F<sub>1</sub> hybrids compared to each inbred check.

Check	Mean Relative Performance ± SD	Range
WTR-1	0.206 ± 0.143	−0.238–0.679
Y 134 (DP 6)	0.191 ± 0.141	−0.248–0.659
Khazar (DP 8)	0.283 ± 0.147	−0.218–0.724
OM 997 (DP 10)	0.283 ± 0.152	−0.190–0.786
M 401 (DP 17)	0.179 ± 0.140	−0.256–0.642
X 21 (DP 19)	0.289 ± 0.153	−0.186–0.796

Overall, in all hybrids, parental SDAF was significantly correlated with mid-parent heterosis for all traits except proportion of spikelets filled (**Figure 3** and **Table 10**). Interestingly, the direction of the correlation was negative for all traits except tiller number. The strongest correlation of mid-parent heterosis and SDAF for hybrids overall was for panicle length. However, when hybrids were grouped into families by female parent, there were no significant correlations between parental SDAF and mid-parent heterosis.

## DISCUSSION

The objectives of this study were (1) to identify high-yielding F<sub>1</sub> hybrids from crosses of IRRI male sterile lines with stress-tolerant male lines, (2) to identify parental lines with high GCA for use in future reciprocal recurrent genomic selection programs, and (3) to evaluate genomic prediction and phenotypic selection accuracies in our hybrid breeding population, with the end goal of developing stress-tolerant hybrid rice varieties. Compared to inbred commercial checks (which were also progenitors of the male lines), phenotyped F<sub>1</sub> hybrids showed genetic yield advantages of up to 80% and warrant further testing (**Supplementary Figure 3**; **Table 8**). Although the genetic yield of the top-performing F<sub>1</sub> hybrid observed in the study environment was 8.670 tons per hectare, this measure pertains to the study environment only—which included both standard irrigated conditions and stressful upland conditions—and only the plant population densities used, which were lower than those observed in farmers' fields (**Supplementary Table 8**). Relative to the mid-parent, the F<sub>1</sub> hybrids showed on average 32.0% mid-parent heterosis for yield, which is consistent with literature averages of 10 to 30% in rice (**Figure 2** and **Table 9**; Janaiah and Xie, 2010; Longin et al., 2012; Spielman et al., 2013). However, the maximum mid-parent heterosis observed for yield was 89.2%, and heterosis up to 48.3% was observed within a single standard deviation of the mean (**Figure 2** and **Table 9**). Because heterosis is present in the F<sub>1</sub> hybrids and SCA as well as GCA variance was detected, recurrent selection for GCA in the male and female lines tested should allow development of heterotic pools. However, we did not evaluate the relative efficiency of hybrid vs. line breeding for our population given the estimated GCA and SCA variance detected in our population or other relevant factors, and further investigation of this topic is warranted.



## Fertility Restoration and Genetic Architecture of Spikelet Filling

Interestingly, narrow-sense heritability for the proportion of spikelets filled was relatively high at 0.600 (Table 2). Although variation in grain fill is generally driven by environmental factors in rice, with starch synthesis and deposition depending on available assimilate, in rice hybrids absence of grain fill can be due to spikelet sterility from lack of fertility-restoring (*Rf*) alleles (Peng et al., 1998; Tang et al., 2017). Because fertility restoration is only relevant in hybrids of CMS parents, not TGMS parents, we estimated narrow-sense heritability for proportion of spikelets filled in the CMS-derived hybrids only as 0.922 (Table 2). The relatively high proportion of phenotypic variance explained by additive genetic variance for this trait in CMS lines suggests that segregation for fertility restoration played a role in the proportion of spikelets filled in the CMS-derived hybrids and as such in observed yield per plant. Concordantly, selection accuracy for proportion of spikelets filled was substantially higher in the T1M hybrids than T1F hybrids in cross-validation, though T1F hybrids had higher accuracies than T1M for all other traits. Screening the population for known major fertility restoration alleles at *Rf3* and *Rf4* may allow the use of marker-assisted selection to improve selection accuracy (Tang et al., 2017). It may also be possible to select for fertility restoration by genomic prediction rather than mapping fertility-restoring alleles of more minor or modifying effect. Selection for fertility restoration may effectively unlock observed yield potential in future hybrids and fix *Rf* alleles in the male heterotic pool.

## Accuracies of Genomic Prediction for Hybrid Genetic Value and Parent GCA

Genomic prediction model accuracies were high in unobserved T2 F<sub>1</sub> hybrids, for which both parents were included in the training set (Table 3). All F<sub>1</sub> hybrids surveyed were

closely related; closely related individuals have smaller effective population size, which reduces the effective number of loci controlling traits and is expected to increase prediction accuracy (Supplementary Figure 5; Daetwyler et al., 2010). Accuracy appeared to be driven primarily by estimation of the female line effects, and accuracy in T2 hybrids was not substantially different from T1F hybrids (Table 3). The exception was proportion of spikelets filled, for which the male (restorer) line effects were more relevant. As expected, accuracy in the T0 hybrids was low, though positive and improved substantially for yield per plant by the inclusion of SCA effects in the model (Table 3).

Accuracy was low for genomic estimated male GCA (< 0.300) despite that the males all shared a recurrent parent and as such a large proportion of their genomes ( $75\% \pm$  Mendelian sampling and selection; Supplementary Figure 6; Table 4). Although the male donor progenitors were diverse, multiple male lines per donor were sampled. Low accuracies of genomic predictions of male GCA may have been due to highly unbalanced crossing of males to females, with no single male crossed to all females. It was not possible to estimate accuracy for genomic estimated GCA in the females, which were also closely related but more extensively phenotyped (Supplementary Figure 7).

**TABLE 9** | Mean, standard deviation, and range of mid-parent heterosis for each trait.

Trait	Mean Mid-Parent Heterosis $\pm$ SD	Range
Height	$-0.019 \pm 0.106$	$-0.237$ – $-0.313$
Tiller Number	$0.292 \pm 0.132$	$-0.026$ – $0.741$
Panicle Length	$0.017 \pm 0.054$	$-0.163$ – $0.183$
Proportion of Spikelets Filled	$-0.034 \pm 0.160$	$-0.657$ – $-0.258$
Yield per Plant	$0.320 \pm 0.163$	$-0.110$ – $0.892$
Potential Yield	$0.327 \pm 0.131$	$0.005$ – $0.817$

Phenotypic accuracies were similar to or lower than genomic prediction accuracies for hybrid performance in the T2 case (Table 5; Rutkoski et al., 2015). Notably, accuracy of genomic prediction of hybrid yield (0.820) was greater than accuracy of the phenotypes (0.682). However, for male GCA, the phenotypic accuracies greatly exceeded those of genomic prediction for both the GCA and the GCA + SCA models (Table 6). It was not possible to compare genomic and phenotypic selection accuracies for female GCA.

## Inconsistent Correlations of Mid-Parent Heterosis and Parental SDAF

Considering all hybrids in the study, we observed parental SDAF to be negatively correlated with mid-parent heterosis and hybrid genetic value for all traits surveyed except tiller number, for which SDAF was positively correlated with mid-parent heterosis (Figure 3 and Table 10). In many species, parental SDAF (or other measures of genetic distance) is positively correlated with heterosis due to release from inbreeding depression to a point, as dominant alleles mask deleterious recessive alleles in hybrids (Falconer and Mackay, 1996). However, as genetic distance increases, outbreeding depression eventually prevails as favorable epistatic combinations of genes are separated (Lynch and Walsh, 1998). A common manifestation of outbreeding depression is fertility barriers (Edmands, 2002). In the case of yield per plant, we cannot eliminate the possibility that genetic distance is correlated with absence of wide-cross compatibility alleles known to affect seed set, given the inter-subspecific diversity present in the male lines (Ji et al., 2005). However, given the intense selection on the male lines, it seems possible that wide-cross compatibility in the males may have also been positively and indirectly selected with yield. Genetic distance could also be correlated with absence of fertility restoring alleles by chance. However, yield potential corrects for fertility restoration by estimating yield as if all spikelets were filled to the average weight observed in the study, and overall mid-parent heterosis for yield potential was also negatively correlated with parental SDAF. Importantly, though unsurprisingly given the relationships of the BC<sub>1</sub>F<sub>5</sub> male lines, the correlation of parental SDAF and mid-parent heterosis was not observed within female families of hybrids (Table 10). This suggests that the negative correlations observed in hybrids overall were due to differences in female genetic distance from the average male. More crucially for practical purposes, whether genetic distance is indicative of mid-parent heterosis depends on the population defined, even in closely related hybrids.

## Future Directions for IRRI Hybrid Rice Breeding

Based on the study findings, we caution against the conventional wisdom that increased genetic distance between parents alone will always confer improved hybrid performance or positive heterosis. Increased genetic distance in the potential founders of heterotic pools of rice screened was not reliably associated with desired positive heterosis for yield, even though the pools could be genetically distinguished. In this population, and probably in rice more generally, empirical selection for GCA

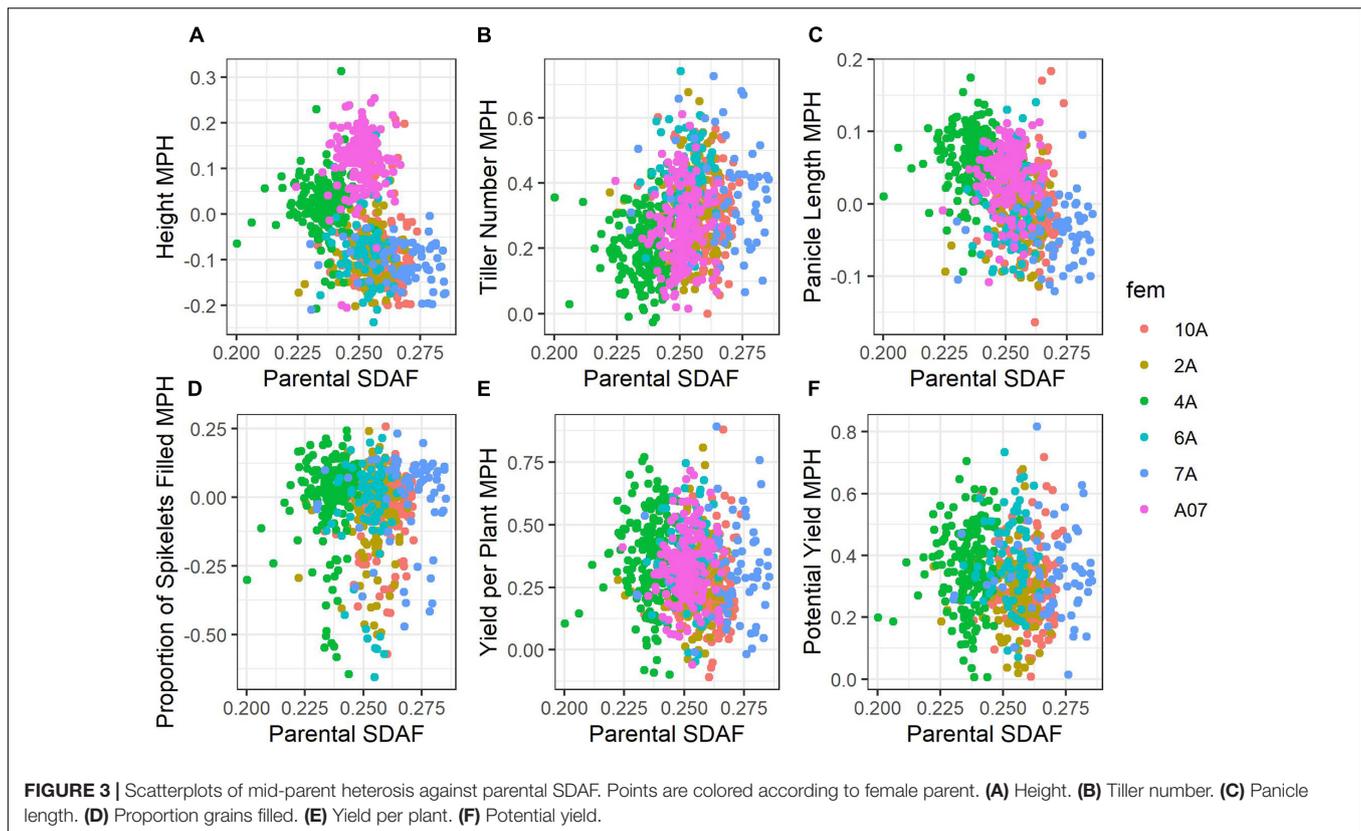
**TABLE 10** | Pearson's correlation coefficient of mid-parent heterosis and SDAF with 95% confidence intervals of the coefficient and *t*-tests of significance conducted at  $\alpha = 0.05$ .

Trait	<i>r</i> ± 95% CI	<i>t</i>	<i>df</i>	<i>P</i>
<b>Height</b>				
<i>Overall</i>	-0.330 ± 0.062	-9.791	783	< 0.001
10A	-0.041 ± 0.157	-0.512	154	0.610
2A	0.007 ± 0.176	0.077	122	0.939
4A	0.031 ± 0.145	0.419	181	0.676
6A	-0.044 ± 0.227	-0.381	73	0.705
7A	-0.073 ± 0.237	-0.592	66	0.556
A07	0.084 ± 0.146	1.118	177	0.265
<b>Tiller Number</b>				
<i>Overall</i>	0.299 ± 0.064	8.725	778	< 0.001
10A	-0.143 ± 0.155	-1.786	152	0.076
2A	0.058 ± 0.176	0.644	121	0.521
4A	0.027 ± 0.145	0.357	180	0.722
6A	0.032 ± 0.227	0.269	73	0.788
7A	-0.084 ± 0.239	-0.677	65	0.501
A07	0.045 ± 0.146	0.601	177	0.549
<b>Panicle Length</b>				
<i>Overall</i>	-0.430 ± 0.057	-13.309	783	< 0.001
10A	-0.042 ± 0.157	-0.516	154	0.607
2A	-0.084 ± 0.175	-0.935	122	0.352
4A	0.081 ± 0.144	1.094	181	0.275
6A	0.066 ± 0.226	0.568	73	0.572
7A	0.013 ± 0.238	0.103	66	0.918
A07	0.058 ± 0.146	0.772	177	0.441
<b>Proportion of Spikelets Filled</b>				
<i>Overall</i>	-0.015 ± 0.081	-0.363	588	0.717
10A	0.127 ± 0.156	1.578	151	0.117
2A	0.063 ± 0.177	0.687	120	0.493
4A	0.103 ± 0.146	1.364	174	0.174
6A	0.027 ± 0.228	0.23	72	0.818
7A	0.083 ± 0.242	0.663	63	0.510
<b>Yield per Plant</b>				
<i>Overall</i>	-0.131 ± 0.069	-3.698	781	< 0.001
10A	0.014 ± 0.157	0.169	154	0.866
2A	0.041 ± 0.176	0.450	122	0.653
4A	0.116 ± 0.144	1.562	180	0.120
6A	0.040 ± 0.227	0.346	73	0.730
7A	-0.022 ± 0.238	-0.176	66	0.861
A07	0.104 ± 0.146	1.389	176	0.167
<b>Potential Yield</b>				
<i>Overall</i>	-0.092 ± 0.08	-2.225	585	0.026
10A	-0.031 ± 0.159	-0.386	151	0.700
2A	-0.005 ± 0.178	-0.059	120	0.953
4A	0.087 ± 0.148	1.141	171	0.255
6A	0.042 ± 0.228	0.360	72	0.720
7A	-0.008 ± 0.244	-0.060	63	0.952

*Overall correlations as well as correlations within female families are reported. Correlations are not available within the A07 female family for potential yield or proportion of spikelets filled because phenotypic observations of the A07 line were not available at the location in which the traits were phenotyped, irrigated lowland.*

is preferable to selection based on genetic distance to breed high-performing hybrid rice.

For hybrid performance, genomic prediction accuracies were similar to or higher than phenotypic accuracies. Therefore,



genomic prediction could be useful for product development in the population of study. Most notably, inclusion of genomic information increased prediction accuracy for hybrid yield per plant by approximately 13.8% compared to phenotype alone (Tables 3, 5). Observed accuracies of prediction of unobserved hybrids with at least one parent in the training population (T1) were also positive and substantial, suggesting that on average genomic prediction could allow identification of further crosses with high value in the population of study. Genomic prediction accuracies for hybrids with neither parent observed (T0) were not as high as in the T1 case, but were nonetheless positive.

In contrast, genomic prediction accuracies for male GCA were substantially lower than phenotypic accuracies. For yield, genomic prediction accuracies were approximately three times less than phenotypic selection accuracies. However, reciprocal recurrent genomic selection for GCA can reduce cycle length by two-thirds compared to reciprocal recurrent phenotypic selection, because parents can be immediately recycled using genomic predictions of their GCA, leading to a cycle length of one (Powell et al., 2020). In conventional reciprocal recurrent selection, it is necessary to cross new parents to the opposing pool and phenotype the inter-pool crosses to estimate GCA before intra-pool recycling is possible, which increases the cycle length to three (Rembe et al., 2019). The genomic prediction accuracies observed in the study would provide comparable genetic gain to phenotypic selection if used to reduce the breeding cycle length to one, assuming that reduction in cycle length has no effect on genomic prediction accuracy and that accuracy of genomic

prediction of female GCA (which could not be estimated, but for which more observations per female were available) is the same or higher than male GCA (Rembe et al., 2019; Powell et al., 2020). Increases in accuracies of genomic prediction of GCA relative to phenotypic selection are likely possible, as the training set of related hybrids would build over time in a closed population, and more complete and informative crossing designs could provide phenotypes. The potential of hybrid breeding strategies in IRRI germplasm would benefit from further assessment by simulation (Faux et al., 2016; Gaynor et al., 2020).

## DATA AVAILABILITY STATEMENT

The phenotypic datasets generated and analyzed for this study can be found in **Supplementary File 1**. The raw genotype data used in this study can be found at NCBI under accession PRJNA479931, excluding the female genotype data. The female genotype data for this article are not publicly available as they are the property of the International Rice Research Institute. Requests to access the female genotype data should be directed to JA at [j.ali@irri.org](mailto:j.ali@irri.org).

## AUTHOR CONTRIBUTIONS

ML planned the study, executed the field trial and analysis, and wrote the manuscript. JA, MA, EA, MP, and MS developed

plant materials used. EA, MP, and MS directed field operations. AL supervised the statistical analysis. AS supervised the marker analysis. JR supervised the quantitative genetic analysis. JA conceived of the study, directed the study, and provided breeding insights. All authors edited and approved the final manuscript.

## FUNDING

Funding for this study was provided by the United States Agency for International Development and Purdue University (Grant No. 208452). Green Super Rice (GSR) materials used in this study were developed under the funding of the GSR Project (IDOPP1130530) to JA as a research sub-grant by the Bill & Melinda Gates Foundation (BMGF). BMGF supports the open access publication fees of this manuscript.

## REFERENCES

- Abebrese, S. O., Martey, E., Dartey, P. K. A., Akromah, R., Gracen, V. E., Offei, S. K., et al. (2019). Farmer preferred traits and potential for adoption of hybrid rice in Ghana. *Sustain. Agric. Res.* 8, 38–48. doi: 10.5539/sar.v8n3.p38
- Ali, J., Aslam, U. M., Tariq, R., Murugaiyan, V., Schnable, P. S., Li, D., et al. (2018). Exploiting the genomic diversity of rice (*Oryza sativa* L.): SNP-typing in 11 early-backcross introgression-breeding populations. *Front. Plant Sci.* 9:849. doi: 10.3389/fpls.2018.00849
- Ali, J., Xu, J. L., Gao, Y. M., Ma, X. F., Meng, L. J., Wang, Y. et al. (2017). Harnessing the hidden genetic diversity for improving multiple abiotic stress tolerance in rice (*Oryza sativa* L.). *PLoS One* 12:e0172515. doi: 10.1371/journal.pone.0172515
- Amuzu-Aweh, E. N., Bijma, P., Kinghorn, B. P., Vereijken, A., Visscher, J., van Arendonk, J. A., et al. (2013). Prediction of heterosis using genome-wide SNP-marker data: application to egg production traits in white Leghorn crosses. *Heredity* 111, 530–538. doi: 10.1038/hdy.2013.77
- Bernardo, R. (2002). *Breeding for Quantitative Traits in Plants*. Woodbury, MN: Stemma Press.
- Boeven, P. H., Zhao, Y., Thorwarth, P., Liu, F., Maurer, H. P., Gils, M., et al. (2020). Negative dominance and dominance-by-dominance epistatic effects reduce grain-yield heterosis in wide crosses in wheat. *Sci. Adv.* 6:eay4897. doi: 10.1126/sciadv.aay4897
- Browning, B. L., Zhou, Y., and Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* 103, 338–348. doi: 10.1016/j.ajhg.2018.07.015
- Chou, H. H., and Holmes, M. H. (2001). DNA sequence quality trimming and vector removal. *Bioinformatics* 17, 1093–1104. doi: 10.1093/bioinformatics/17.12.1093
- Comstock, R. E., Robinson, H. F., and Harvey, P. H. (1949). A breeding procedure designed to make maximum use of both general and specific combining ability 1. *Agron. J.* 41, 360–367. doi: 10.2134/agronj1949.00021962004100080006x
- Covarrubias-Pazarán, G. (2016). Genome assisted prediction of quantitative traits using the R package sommer. *PLoS One* 11:e0156744. doi: 10.1371/journal.pone.0156744
- Covarrubias-Pazarán, G. (2018). Software update: moving the R package sommer to multivariate mixed models for genome-assisted prediction. *bioRxiv [Preprint]* doi: 10.1101/354639
- Daetwyler, H. D., Pong-Wong, R., Villanueva, B., and Woolliams, J. A. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185, 1021–1031. doi: 10.1534/genetics.110.116855
- de Mendiburu, F. (2020). *agricolae: Statistical Procedures for Agricultural Research. R Package Version 1.3-3*.
- Edmands, S. (2002). Does parental divergence predict reproductive compatibility? *Trends Ecol. Evol.* 17, 520–527. doi: 10.1016/s0169-5347(02)02585-5

## ACKNOWLEDGMENTS

We thank the IRRI hybrid rice breeding team, past and present, for their contributions to the study. We thank Rey de la Cueva and Benvenido Bacani for technical expertise and assistance. We thank Erik J. Sacks for co-authoring Grant No. 208452 which partially funded the study. We thank Giovanni Eduardo Covarrubias-Pazarán for developing the open-source R package *sommer* and associated vignettes. We thank our reviewers for their time and helpful insights.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.692870/full#supplementary-material>

- Endelman, J. B., and Jannink, J. L. (2012). Shrinkage estimation of the realized relationship matrix. *G3* 2, 1405–1413. doi: 10.1534/g3.112.004259
- Estaghirou, S. B. O., Ogotu, J. O., Schulz-Streeck, T., Knaak, C., Ouzunova, M., Gordillo, A., et al. (2013). Evaluation of approaches for estimating the accuracy of genomic prediction in plant breeding. *BMC Genomics* 14:860. doi: 10.1186/1471-2164-14-860
- Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction into Quantitative Genetics*. Essex: Prentice Hall.
- Faux, A. M., Gorjanc, G., Gaynor, R. C., Battagin, M., Edwards, S. M., Wilson, D. L., et al. (2016). AlphaSim: software for breeding program simulation. *Plant Genome* 9:Plantgenome2016.02.0013.
- Feng, F., Li, Y., Qin, X., Liao, Y., and Siddique, K. H. (2017). Changes in rice grain quality of Indica and Japonica type varieties released in China from 2000 to 2014. *Front. Plant Sci.* 8:1863. doi: 10.3389/fpls.2017.01863
- Gaynor, R. C., Gorjanc, G., and Hickey, J. M. (2020). AlphaSimR: an R-package for breeding program simulations. *bioRxiv [Preprint]* doi: 10.1101/2020.08.10.245167
- Glaubit, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., et al. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS One* 9:e90346. doi: 10.1371/journal.pone.0090346
- Hochholdinger, F., and Baldauf, J. A. (2018). Heterosis in plants. *Curr. Biol.* 28, R1089–R1092.
- Holland, J. B., Nyquist, W. E., and Cervantes-Martínez, C. T. (2003). Estimating and interpreting heritability for plant breeding: an update. *Plant Breed. Rev.* 22, 9–112. doi: 10.1002/9780470650202.ch2
- IRRI (2014). *Biometrics and Breeding Informatics, PBGB Division*. Los Baños: International Rice Research Institute.
- Janaiah, A., and Xie, F. (2010). *Hybrid rice adoption in India: farm-level impacts and challenges. IRRI Technical Bulletin*. Los Baños, Philippines: International Rice Research Institute, 20.
- Ji, Q., Lu, J., Chao, Q., Gu, M., and Xu, M. (2005). Delimiting a rice wide-compatibility gene S5n to a 50 kb region. *Theor. Appl. Genet.* 111, 1495–1503. doi: 10.1007/s00122-005-0078-0
- Jombart, T., and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27, 3070–3071. doi: 10.1093/bioinformatics/btr521
- Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., et al. (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6:4. doi: 10.1186/1939-8433-6-4
- Kwon, S. J., Ha, W. G., Hwang, H. G., Yang, S. J., Choi, H. C., Moon, H. P., et al. (2002). Relationship between heterosis and genetic divergence in ‘Tongil’-type rice. *Plant Breed.* 121, 487–492. doi: 10.1046/j.1439-0523.2002.00760.x
- Liang, Z., Gupta, S. K., Yeh, C. T., Zhang, Y., Ngu, D. W., Kumar, R., et al. (2018). Phenotypic data from inbred parents can improve genomic prediction in pearl millet hybrids. *G3* 8, 2513–2522. doi: 10.1534/g3.118.200242

- Longin, C. F. H., Mühleisen, J., Maurer, H. P., Zhang, H., Gowda, M., and Reif, J. C. (2012). Hybrid breeding in autogamous cereals. *Theor. Appl. Genet.* 125, 1087–1096. doi: 10.1007/s00122-012-1967-7
- Lu, Z. M., and Xu, B. Q. (2010). On significance of heterotic group theory in hybrid rice breeding. *Rice Sci.* 17, 94–98. doi: 10.1016/s1672-6308(08)60110-9
- Lynch, M., and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*, Vol. 1. Sunderland, MA: Sinauer, 535–557.
- Melchinger, A. E. (1999). “Genetic diversity and heterosis,” in *Genetics and Exploitation of Heterosis in Crops*, eds J. G. Coors and S. Pandey (Madison: American Society of Agronomy, Inc.), 99–118. doi: 10.2134/1999.geneticsandexploitation.c10
- Mottaleb, K. A., Mohanty, S., and Nelson, A. (2015). Factors influencing hybrid rice adoption: a Bangladesh case. *Aust. J. Agric. Resour. Econ.* 59, 258–274. doi: 10.1111/1467-8489.12060
- Mrode, R. A. (2014). *Linear Models for the Prediction of Animal Breeding Values*. Wallingford: CABI.
- Nalley, L., Tack, J., Barkley, A., Jagadish, K., and Brye, K. (2016). Quantifying the agronomic and economic performance of hybrid and conventional rice varieties. *Agron. J.* 108, 1514–1523. doi: 10.2134/agronj2015.0526
- Nalley, L., Tack, J., Durand, A., Thoma, G., Tsiboe, F., Shew, A., et al. (2017). The production, consumption, and environmental impacts of rice hybridization in the United States. *Agron. J.* 109, 193–203. doi: 10.2134/agronj2016.05.0281
- Ott, A., Liu, S., Schnable, J. C., Yeh, C. T., Wang, C., and Schnable, P. S. (2017). Tunable genotyping-by-sequencing (tGBS) enables reliable genotyping of heterozygous loci. *bioRxiv [Preprint]* doi: 10.1101/100461
- Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. doi: 10.1093/bioinformatics/bty633
- Peng, S., Yang, J., Garcia, F. V., Laza, M. R. C., Visperas, R. M., Sanico, A. L., et al. (1998). “Physiology-based crop management for yield maximization of hybrid rice,” in *Advances in Hybrid Rice Technology*, eds S. S. Virmani, E. A. Siddiq, and K. Muralidaran (Los Baños: IRRI), 157–176.
- Powell, O. M., Gaynor, R. C., Gorjanc, G. M., Werner, C. R., and Hickey, J. M. (2020). A two-part strategy using genomic selection in hybrid crop breeding programs. *bioRxiv [Preprint]* doi: 10.1101/2020.05.24.113258
- Rembe, M., Zhao, Y., Jiang, Y., and Reif, J. C. (2019). Reciprocal recurrent genomic selection: an attractive tool to leverage hybrid wheat breeding. *Theor. Appl. Genet.* 132, 687–698. doi: 10.1007/s00122-018-3244-x
- Resende, M. F. R., Munoz, P., Resende, M. D. V., Garrick, D. J., Fernando, R. L., Davis, J. M., et al. (2012). Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics* 190, 1503–1510. doi: 10.1534/genetics.111.137026
- Rutkoski, J., Singh, R. P., Huerta-Espino, J., Bhavani, S., Poland, J., Jannink, J. L., et al. (2015). Genetic gain from phenotypic and genomic selection for quantitative resistance to stem rust of wheat. *Plant Genome* 8, 1–10.
- SAMtools. (2018). *BCFtools*. <https://github.com/samtools/bcftools/releases/tag/1.7>.
- Spielman, D. J., Kolady, D. E., and Ward, P. S. (2013). The prospects for hybrid rice in India. *Food Secur.* 5, 651–665. doi: 10.1007/s12571-013-0291-7
- Sprague, G. F., and Tatum, L. A. (1942). General vs. specific combining ability in single crosses of corn I. *Agron. J.* 34, 923–932. doi: 10.2134/agronj1942.00021962003400100008x
- Tang, H., Xie, Y., Liu, Y. G., and Chen, L. (2017). Advances in understanding the molecular mechanisms of cytoplasmic male sterility and restoration in rice. *Plant Reprod.* 30, 179–184. doi: 10.1007/s00497-017-0308-z
- Technow, F., Schrag, T. A., Schipprack, W., Bauer, E., Simianer, H., and Melchinger, A. E. (2014). Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. *Genetics* 197, 1343–1355. doi: 10.1534/genetics.114.165860
- Tracy, W. F., and Chandler, M. A. (2006). “The historical and biological basis of the concept of heterotic patterns in corn belt dent maize,” in *Plant Breeding: The Arnel R Hallauer International Symposium*, eds K. R. Lamkey and M. Lee (Ames, IA: Blackwell Publishing), 219–233. doi: 10.1002/9780470752708.ch16
- Virmani, S. S., and Wan, B. H. (1988). “Development of CMS lines in hybrid rice breeding,” in *Hybrid Rice*, ed. IRRI (Manila: International Rice Research Institute), 103–114.
- Waters, D. L., Subbaiyan, G. K., Mani, E., Singh, S., Vaddadi, S., Baten, A., et al. (2015). Genome wide polymorphisms and yield heterosis in rice (*Oryza sativa* subsp. indica). *Trop. Plant Biol.* 8, 117–125. doi: 10.1007/s12042-015-9156-x
- Wu, T. D., Reeder, J., Lawrence, M., Becker, G., and Brauer, M. J. (2016). “GMAP and GSNAP for genomic sequence alignment: enhancements to speed, accuracy, and functionality,” in *Statistical Genomics* E. Mathé, and S. Davis (New York, NY: Humana Press), 283–334. doi: 10.1007/978-1-4939-3578-9\_15
- Xiang, T., Christensen, O. F., Vitezica, Z. G., and Legarra, A. (2016). Genomic evaluation by including dominance effects and inbreeding depression for purebred and crossbred performance with an application in pigs. *Genet. Sel. Evol.* 48, 1–14.
- Xiao, J., Li, J., Yuan, L., McCouch, S. R., and Tanksley, S. D. (1996). Genetic diversity and its relationship to hybrid performance and heterosis in rice as revealed by PCR-based markers. *Theor. Appl. Genet.* 92, 637–643. doi: 10.1007/s001220050173
- Yu, S., Ali, J., Zhang, C., Li, Z., and Zhang, Q. (2020). Genomic breeding of green super rice varieties and their deployment in Asia and Africa. *Theor. Appl. Genet.* 133, 1427–1442. doi: 10.1007/s00122-019-03516-9
- Yuan, L., Virmani, S. S., and Changxiong, M. (1989). “Hybrid rice: achievements and outlook,” in *Proceedings of the International Rice Research Conference*, (Hangzhou: IRRI),

**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.

Copyright © 2021 Labroo, Ali, Aslam, de Asis, dela Paz, Sevilla, Lipka, Studer and Rutkoski. 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.