Genome-wide Association Study and Genomic Prediction for Yield and Grain Quality Traits of Hybrid Rice

Peiyi Yu
Huazhi Biotechnology Co. Ltd., Changsha
Changrong Ye
Huazhi Biotechnology Co. Ltd., Changsha https://orcid.org/0000-0002-4095-1068
Le Li
Huazhi Biotechnology Co. Ltd., Changsha
Hexing Yin
Huazhi Biotechnology Co. Ltd., Changsha
Jian Zhao
Huazhi Biotechnology Co. Ltd., Changsha
Yongka Wang
Huazhi Biotechnology Co. Ltd., Changsha
Zhe Zhang
Huazhi Biotechnology Co. Ltd., Changsha
Weiguo Li
Huazhi Biotechnology Co. Ltd., Changsha
Yu Long
Huazhi Biotechnology Co. Ltd., Changsha
Xueyi Hu
Huazhi Biotechnology Co. Ltd., Changsha
Jinhua Xiao
Huazhi Biotechnology Co. Ltd., Changsha
Gaofeng Jia
Huazhi Biotechnology Co. Ltd., Changsha
Bingchuan Tian (tianbc@higentec.com)
Huazhi Biotechnology Co. Ltd., Changsha

Research Article

Keywords: genomic selection, molecular breeding, yield, grain quality, hybrid rice

Posted Date: February 16th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1355596/v1

Version of Record: A version of this preprint was published at Molecular Breeding on March 18th, 2022. See the published version at https://doi.org/10.1007/s11032-022-01289-6.

Abstract

Genomic selection is an efficient tool for breeding selection, especially for quantitative traits controlled by multiples genes with low heritability. To validate the application of genomic selection in hybrid rice breeding, the yield, grain quality and agronomic traits of 404 hybrid rice breeding lines were investigated, and the same accessions were genotyped by using a 56K SNP chip. There were wide variances among the tested accessions for all the measured traits, and most of the traits were correlated. A total of 67 significant loci were identified for the yield and agronomic traits, and 123 significant loci were identified for the grain quality traits by GWAS. Two of these loci associated with increasing grain yield but decreasing grain quality. The GEBVs of all the yield, quality and agronomic traits were calculated by using 15 different prediction algorithms. The plant height, panicle length, thousand grain weight, grain length and width ratio, amylose content and alkali value have higher predictability than other traits. However, the predictability of different GS models is different for different traits. This study provided useful information for genomic selection of specific trait using proper markers and prediction models.

Introduction

The use of heterosis in hybrid rice has become increasingly important since the beginning of hybrid rice extension in China (Ma GH and Yuan LP 2015). Hybrid rice has contributed greatly to food security in China and the world. In recent years, the average yield of rice in China has increased from 3.5 ton/ha in1975 to 6 ton/ha in 1995, and to 7 ton/ha in 2018 (FAOSTAT), and the grain quality of rice has been improved. About 50% of newly registered rice varieties in China (national approval) have grain quality of grade one (Lu F et al. 2019). However, it is time and labor consuming for developing a hybrid variety by conventional breeding even though marker assisted selection (MAS) has been used. Future breeding of hybrid rice will benefit from the use of new breeding technology integrated with genetics, genomics, computational science and artificial intelligence.

Rice is a model species for genomic study of monocotyledonous plant. The genome of rice was fully sequenced in 2005 (International Rice Genome Sequencing Project and Sasaki T 2005), and more than 3000 genes have been cloned and analyzed (Yao W et al. 2018). Large number of molecular markers have been developed for MAS of important traits such as plant height, blast resistance, leaf blight resistance, submergence tolerance and fragrance (Jena KK and Mackill DJ 2008). However, the success of MAS heavily depended on level of heritability and genetic architectures of the selected traits. MAS is not effective for traits controlled by large number of genes/QTLs with small contribution. With the development of high-throughput sequencing and chip technology, genome-wide association study (GWAS) has been used for identification of useful genes/QTLs, and genomic selection (GS) or genome-wide selection (GWS) has been proposed as a promising tool and applied for animal and plant genetic improvement (Meuwissen THE et al. 2001). GS has higher genetic gain than marker assisted selection for complex traits controlled by large number of QTLs (Crossa J et al. 2017b). However, GS has not been successfully used in hybrid rice breeding yet.

Genomic selection uses genotypes and phenotypes of target traits from individuals in a training population to establish prediction models, and uses the models to predict genomic estimated breeding values (GEBVs) of individuals based on their genotypes in a test population (Crossa J et al. 2017a). The hypothesis is based on the assumption that with high density SNP markers distributed throughout the whole genome, at least one SNP can be found in a linkage disequilibrium state with the quantitative genetic loci affecting the target trait, so that the effect of each QTL can be reflected by SNP markers (Meuwissen T 2007). The statistical models of genome selection can be roughly divided into two categories. The first is the direct method, which takes the individual as the random effect, the genetic relationship matrix constructed by the genetic information of the reference population and the predicted population as the variance covariance matrix, to estimate the variance components through the iterative method, and obtain the predicted breeding value of the individual. The second is the indirect method, which first estimates the marker effect in the reference group, and then accumulates the marker effect combined with the genotype information from the prediction group to obtain the individual estimated breeding value of the prediction group (Zhang Z et al. 2011; Misztal I and Legarra A 2017). Different prediction modes use different statistical methods, thus, the efficiency of the models need to be compared and validated before using for breeding selection.

Genomic selection has been successfully used in animal breeding programs to increase the rate of genetic gain of dairy cattle, pig, dairy goat, layer chicken, and fish (García-Ruiz A et al. 2016; Samorè AB and L. 2015; Mucha S et al. 2015; Wolc A et al. 2015; López M et al. 2015). In recent year, simulations and experimental studies have been conducted to validate the efficiency of this method in breeding of plants. Be specific to rice, the predictive ability of heading date, culm length, panicle length, panicle number, grain length and grain width varied from 0.4 to 0.8 in a population of 110 rice cultivars using nine prediction methods (Onogi A et al. 2015). The highest predictive abilities for spikelets per panicle, heading date, plant height and protein content was 0.44–0.7 in a diverse population of 413 rice inbred lines from 82 countries genotyped with a 44 K SNP chip (Isidro J et al. 2015). The GEBVs of other traits such as grain shape, grain yield, nitrogen balance index, panicle weight, grain weight, and blast resistance have been predicted using inbred lines or cultivars (Spindel J et al. 2015; Yabe S et al. 2018; Iwata H et al. 2015; Grenier C et al. 2015; Hassen M et al. 2018; Huang M et al. 2019). Genomic prediction has also been conducted for grain yield, thousand grain weight, and index of different traits of hybrid rice (Wang X et al. 2017; Xu S et al. 2014; Xu Y et al. 2018; Wang W et al. 2018; Cui Y et al. 2020). The predicted GEBVs from different populations were similar, thus, genomic selection is a reliable method for rice breeding.

In this study, we investigated the agronomic traits, yield related traits, and grain quality related traits of 404 hybrid rice lines that genotyped by using a 56K SNP chip, and conducted genome wide association study and genomic prediction for 20 traits using 15 statistical methods. The objectives of this study were to validate the predictability of different models and to find best-fit statistical methods for prediction of different traits.

Materials And Methods

Plant materials

A total of 404 hybrid rice accessions were planted in the filed in Changsha (N28.31, E113.31, A80m) from 2014 to 2019. Hybrid rice varieties Fengliangyou 4 (FLY4) and Fyou498 (FY498) were used as common check variety and panted with the tested hybrids every year (Table 1).

Table 1
Number of hybrids investigated.

Year	Number of hybrids	Check variety
2014	37	FLY4, FY498
2015	46	FLY4, FY498
2016	50	FLY4, FY498
2017	102	FLY4, FY498
2018	72	FLY4, FY498
2019	97	FLY4, FY498
Total	404	-

Field experiments and phenotyping data collection

A randomized complete block design (RCBD) with three replications was used for field experiments. For each hybrid, 250 seedlings were transplanted into a 13.3 m^2 plot (5 m X 2.66 m) with a density of 0.2 m X 0.266 M. At maturity, growth period (GP, days from seeding to harvest), number of tillers (TN), plant height (PH, cm), panicle length (PL, cm), number of grains per panicle (GN), spikelet fertility (SF, %), and thousand grain weight (TGW, g) were measured. Panicles from 1 m^2 area were harvested and dried for calculating the grain yield (YLD, Kg/ha).

The grain quality was evaluated following the industry standard for rice variety (NY/ T 593–2013, Ministry of Agriculture, China). The evaluated traits include BRR (brown rice rate, %), WRR (white rice rate, %), WWRR (whole white rice rate or head rice rate, %), GL (grain length, mm), GLWR (grain length/width ratio), CP (chalk percentage, %), CD (chalk degree), AC (amylose content, %), GC (gel consistency, mm), ALK (alkali value), TRANS (transparency), OG (overall grade of grain quality).

Genotyping data collection

Rice seeds were germinated in petri dishes at 28 ^{o}C in an incubator. Leaf samples were collected from two-week-old seedlings and grounded in a motor with liquid nitrogen, and genomic DNA was extracted by using standard CTAB extraction protocol (Doyle JJ and Doyle JL 1987). The quality of DNA sample was checked by using electrophoresis on 1% agarose gel, and the concentration of DNA was measured by using a UV-Vis spectrophotometer (Nanodrop 8000, Thermo Fisher Scientific, USA). These high quality DNA samples were then used for fragmentation, hybridization with 56K SNP chip and imaging in a GeneTitan Multi-Channel (MC) Instrument (Thermo Fisher Scientific, USA) following the user manual. The rice 56K SNP chip was design by Huazhi Biotechnology Co. Ltd., which includes 56897 SNPs from the dataset of the 3000 rice genome project (3K RGP)(Li J et al. 2014).

Data analysis

Pearson correlations among traits were calculated by using Minitab 17 (Minitab LLC). GWAS was conducted by using TASSEL 5.0 (Bradbury P et al. 2007). Genotypic data containing 34832 high quality SNPs from the 56K chip (56897 SNPs) was used for the analysis. The kinship matrix with centered IBS (default) was generated using genotyping data. A united data file with genotyping and phenotyping data of the hybrids was created by using union join. The united file along with kinship matrix were analyzed for marker-trait association using mixed linear model (MLM). The compression level was set to optimum level, and variance component estimation was set to P3D. A criteria for claiming a QTL was $p < 1x10-4$ (-log10 p-value > 4.0). The identified QTLs were named using the CGSNL nomenclature (McCouch S and CGSNL (Committee on Gene Symbolization 2008).

Genomic selection models were built by using big scale ridge regression (bigRR), best line unbiased prediction (GBLUP), least absolute shrinkage and selection operation (LASSO), ridge regression BLUP (rrBLUP), sparse partial least square regression GBLUP(SPLS), reproducing Kernel Hilbert Space (RKHS), BayesA, BayesB, BayesC, bayesian ridge regression (BRR), random forest classifier (RFC), random forest regression (RFR), support vector regression (SVR), support vector linear classifier (SVC), and bayesian regularized neural network (BRNN) in R/Python with default settings (Table S1). For example, the grain yield from multi-year and multi-site was calculated by using mixed linear model of IME4 program in R (Bates D et al. 2015):

$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ij}$

y_{ij} is the yield of i_{th} variety in j_{th} environment;

μ is the overall average yield;

α_i is the varietal effect i_{th} variety;

β_j is the environmental effect of j_{th} environment;

$(\alpha\beta)_{ij}$ is the interaction effect of i_{th} variety and j_{th} environment;

ε_{ij} is the residual error;

Variety is fixed effect, while environment and variety and environment interaction are random effects.

GEBVs of different traits were calculated by using SOMMER for GBLUP model and G2P for other models in R program (Covarrubias-Pazaran G 2016).

The predictive ability of the models was validated by using 5x cross validation method; all data were randomly divided into 5 groups, with one group being used as the validation set, and the other 4 groups being used as the training set, until the complete prediction of all data.

The predictive ability of a models was compared by Pearson correlation coefficient and Spearman correlation coefficient between predicted GEBVs and actual values of the hybrids. Mean square error (MSE) and the maximum 10% yield through five-fold cross-validation were also calculated for each trait.

Results

Statistics of the phenotypic data

Eight yield related traits and 12 grain quality related traits were investigated. There was significant variation for each trait. Most of the traits showed normal distribution, except TN, GLWR and CD with higher kurtosis values than other traits (Table 2).

Table 2
Statistics of the agronomic traits and grain quality traits.

Traits	Count	Mean	StDev	Minimum	Maximum	Skewness	Kurtosis
GP (days)	298	134.0	13.5	112.8	157.7	0.1	-1.3
TN	298	16.4	2.1	11.5	31.8	1.5	9.5
PH (cm)	393	114.4	9.5	92.7	135.7	0.1	-0.9
PL (cm)	393	24.3	1.6	19.4	29.3	0.1	-0.1
GN	393	187.2	23.5	129.7	258.9	-0.2	-0.1
SF (%)	393	82.9	3.7	65.4	90.2	-0.7	0.6
TGW (g)	393	25.2	2.4	17.5	32.8	0.5	0.0
YLD (Kg/ha)	393	9185.8	1163.6	5910.0	11256.0	-1.0	0.3
BRR (%)	392	79.3	1.7	72.3	83.1	-0.9	1.3
WRR (%)	392	70.1	1.8	64.5	75.6	-0.4	-0.2
WWRR (%)	392	58.6	5.9	37.1	70.3	-0.8	0.7
GL (%)	392	6.6	0.3	5.7	7.7	0.3	0.5
GLWR	392	3.2	0.2	2.3	4.2	0.9	3.1
CP (%)	392	23.4	10.7	5.0	70.0	0.9	1.5
CD	392	6.2	3.5	0.7	23.4	1.4	3.4
AC (%)	392	15.6	2.5	12.0	24.9	1.3	1.0
GC	392	67.9	14.6	30.0	90.0	-0.9	0.1
ALK	392	5.3	1.2	3.0	7.0	-0.2	-1.1
TRANS	392	1.3	0.5	1.0	3.0	1.1	0.2
OG	392	4.3	1.1	2.0	5.0	-1.0	-0.8

Traits: GP (growth period, days), TN (number of tillers), PH (plant height, cm), PL (panicle length, cm), GN (number of grains per panicle), SF (spikelet fertility, %), TGW (thousand grain weight, g), YLD (yield, Kg/ha), BRR (brown rice rate,%), WRR (white rice rate,%), WWRR (whole white rice rate, %), GL (grain length, mm), GLWR (grain length/width ratio), CP (chalk percentage, %), CD (chalk degree), AC (amylose content, %), GC (gel consistency, mm), ALK (alkali value), TRANS (transparency), OG (overall grade of grain quality).

Based on Pearson correlation, spikelet fertility was not correlated with number of tillers (TN) and thousand grain weight (TGW). Other yield related traits were correlated (Table 3). For the grain quality related traits, whole white rice rate (WWRR), grain length (GL), chalk degree (CD) and overall grade (OG) were not correlated with amylose content (AC) and gel consistency (GC). And whole white rice rate (WWRR), gel consistency (GC), alkali value (ALK), and transparency (TRANS) were not correlated with grain length/width ratio (GLWR). While, most of other grain quality traits were correlated (Table 4).

Table 3

Pearson correlation among agronomic traits. Upper number is Pearson correlation, lower number is p value.

	GP	**TN**	**PH**	**PL**	**GN**	**SF**	**TGW**
TN	-0.456						
	0.000						
PH	0.339	-0.513					
	0.000	0.000					
PL	0.417	-0.547	0.662				
	0.000	0.000	0.000				
GN	0.703	-0.532	0.545	0.392			
	0.000	0.000	0.000	0.000			
SF	0.317	-0.090	0.289	0.119	0.269		
	0.000	0.119	0.000	0.040	0.000		
TGW	0.337	-0.550	0.324	0.458	0.135	0.007	
	0.000	0.000	0.000	0.000	0.020	0.899	
YLD	0.676	-0.218	0.407	0.291	0.660	0.602	0.291
	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 4
Pearson correlation among grain quality traits. Upper number is Pearson correlation, lower number is p value.

	BR	WR	WWRR	GL	GLWR	CP	CD	AC	GC	ALK	TRANS
WR	0.702										
	0.000										
WWRR	0.251	0.502									
	0.000	0.000									
GL	0.066	-0.052	-0.233								
	0.190	0.301	0.000								
GLWR	-0.103	-0.171	-0.061	0.547							
	0.043	0.001	0.232	0.000							
CP	-0.199	-0.263	-0.422	-0.126	-0.309						
	0.000	0.000	0.000	0.013	0.000						
CD	-0.242	-0.265	-0.382	-0.161	-0.246	0.941					
	0.000	0.000	0.000	0.001	0.000	0.000					
AC	0.247	0.154	-0.058	0.066	-0.218	0.146	0.083				
	0.000	0.002	0.254	0.192	0.000	0.004	0.100				
GC	-0.162	-0.200	-0.002	-0.068	0.089	-0.091	-0.048	-0.859			
	0.001	0.000	0.974	0.178	0.078	0.072	0.343	0.000			
ALK	0.230	0.187	0.161	0.045	0.085	-0.378	-0.408	0.212	-0.237		
	0.000	0.000	0.001	0.377	0.091	0.000	0.000	0.000	0.000		
TRANS	-0.248	-0.266	-0.174	-0.256	-0.032	0.435	0.453	-0.195	0.132	-0.346	
	0.000	0.000	0.001	0.000	0.529	0.000	0.000	0.000	0.009	0.000	
OG	-0.162	-0.104	-0.260	-0.127	-0.222	0.617	0.604	0.016	0.000	-0.496	0.261
	0.001	0.039	0.000	0.012	0.000	0.000	0.000	0.758	0.993	0.000	0.000

Diversity of the hybrids tested

All 404 hybrids were genotyped by using a 56K SNP chip which includes 56897 SNPs. After filtering, 34832 high quality SNPs remained. There are 1992–4736 SNP markers on each chromosome (Fig. 1). Phylogenetic tree from these high-quality SNPs showed that all the hybrids were *indica* rice, with the exception of three hybrids which have larger genetic distance from others, possibly due to introgression from *japonica* rice (Fig. 2).

GWAS of yield related traits and grain quality traits

A total of 67 significant loci were identified for the yield related traits, and clusters of loci for different traits were identified on chromosome 1, 3, 4, 5, 6, 9, 11 and 12 (Fig. 3, Table S2). QTLs for grain yield were identified on chromosomes 3, 4, 5, 6, 9, 10 and 12, and most of them were collocated with QTLs for plant height (PH), panicle length (PL), number of grains per panicle (GN), spikelet fertility (SF) and thousand grain weight (TGW).

A total of 123 significant loci were identified for the grain quality traits, and clusters of loci for different traits were identified on chromosome 1, 2, 5, 6, 7, 9, 11 and 12 (Fig. 4, Table S3). QTLs for overall grade of grain quality were identified on chromosomes 2, 5, 6 and 12, and most of them were collocated with QTLs for alkali value (ALK), chalk percentage (CP), chalk degree (CD), amylose content (AC), gel consistency (GC) and transparency (TRANS).

We found that two SNP markers on chromosome 5 (AX-155748928) and chromosome 12 (AX-154698806) were significantly associated with different yield and grain quality traits (Table 5). There are few accessions with GG genotype for both markers AX-155748928 and AX-154698806, and the means of phenotypic traits were not different from AG genotype. When compared the homozygote AA genotype, the heterozygotes (AG) of both markers AX-155748928 and AX-154698806 were higher in plant height, number of grains per panicle, yield, chalkiness, transparency and overall grade (low quality), but lower in white rice rate and alkali value (Figure S1).

Table 5
Summary of significant association between two SNP markers and yield and grain quality related traits.

Trait	Locus	Marker	Chr	Position	F	p	R2
GN	qGN5.1	AX-155748928	5	5417532	29.22	1.13E-07	0.0749
YLD	qYLD5.1	AX-155748928	5	5417532	37.23	2.56E-09	0.0954
ALK	qALK5.1	AX-155748928	5	5417532	15.54	9.61E-05	0.0400
CD	qCD5.1	AX-155748928	5	5417532	23.29	2.01E-06	0.0601
CP	qCP5.1	AX-155748928	5	5417532	24.02	1.41E-06	0.0617
OG	qOG5.1	AX-155748928	5	5417532	18.44	2.23E-05	0.0474
WR	qWR5.1	AX-155748928	5	5417532	16.76	5.17E-05	0.0438
GN	qGN12.1	AX-154698806	12	13961623	11.87	1.00E-05	0.0626
PH	qPH12.1	AX-154698806	12	13961623	24.40	1.09E-10	0.1265
TGW	qTGW12.1	AX-154698806	12	13961623	13.07	3.25E-06	0.0742
YLD	qYLD12.1	AX-154698806	12	13961623	19.72	7.24E-09	0.1068
ALK	qALK12.1	AX-154698806	12	13961623	12.15	7.72E-06	0.0641
CD	qCD12.1	AX-154698806	12	13961623	20.64	3.14E-09	0.1091
CP	qCP12.1	AX-154698806	12	13961623	22.83	4.44E-10	0.1209
OG	qOG12.1	AX-154698806	12	13961623	16.20	1.80E-07	0.0846
TRANS	qTRANS12.1	AX-154698806	12	13961623	13.70	1.81E-06	0.0726
WR	qWR12.1	AX-154698806	12	13961623	14.67	7.36E-07	0.0785

Genomic selection models for yield and grain quality traits

The GEBVs of all yield and grain quality traits were calculated by using 15 different prediction algorithms, and 5x cross validation was used to evaluate the prediction accuracy. The prediction ability of different models can be seen from the correlation heat map of different GS models (Fig. 5). BayesA, BayesB, BayesC, RKHS, rrBLUP and BRR were highly correlated.

For the same trait, the prediction abilities of different GS models were different. Also, for the same GS prediction model, the prediction abilities varied for different traits (Fig. 6). The plant height, panicle length, thousand grain weight, grain length and width ratio, amylose content and alkali value had higher predictability than other traits. Thousand grain weight could be well predicted by all the models, while the transparency of the grain had very low predictability. The predictabilities for grain length and width ratio, chalk percentage, amylose content, alkali value and gel consistency significantly varied among models. The predictability for grain yield ranged from 0.22 to 0.35 (average 0.31) (Table S4).

By comparing the GS models without or with the significant SNP markers from the GWAS analysis, most of the models had higher predictability when the significant SNP markers from GWAS were considered (Fig. 7, Table S5).

Discussions

QTL and genes affecting rice yield and grain quality

It is generally accepted that rice grain yield and quality are two negatively related traits. The high yield varieties usually have low grain quality. Rice breeders have been trying to balance these traits in the breeding process (Xiao N et al. 2021). However, the genetic linkage between grain yield and grain quality have not been dissected. In this study, a total of 67 QTLs for grain yield and 123 QTLs for grain quality were identified by comprehensive evaluation of related traits. Among these loci, we found two SNP markers that were significantly associated with various yield and grain quality traits. The heterozygotes (AG) of markers AX-155748928 on chromosome 5 (chr5:5417532) and AX-154698806 on chromosome 12 (chr12:13961623) had high yield but low grain quality. The SNP marker AX-155748928 was located in the exon region of gene LOC_Os05g09590 with unknown function (Putative uncharacterized protein). Based on the online database analysis (Proost S and Mutwil M 2017), this gene was highly expressed at seed development stage S4-S5 (11–29 days after pollination) (Figure S2). Gene LOC_Os05g09590 was significantly associated with grain chalkiness (Misra G et al. 2019). The SNP marker AX-154698806 was located between genes LOC_Os12g24450 and LOC_Os12g24460 (5849 bp upstream of LOC_Os12g24460). As a putative unclassified retrotransposon protein, LOC_Os12g24450 was highly expressed at seed development stage S5 (21–29 days after pollination). And LOC_Os12g24460, a putative uncharacterized protein, was highly expressed at seed development stage S4 (11–20 days after pollination) (Figure S2). Further validation of the effects of these genes on rice yield and grain quality should find ways to improve both grain yield and quality.

Predictive ability of GS models

High prediction accuracy is a prerequisite for successful application of genomic selection. The prediction accuracy is often measured by the correlation between observed phenotypes and the predicted GEBVs or predicted phenotypes of cross-validation (Xu S 2017). The predictive ability is influenced by several factors such as population size, variation within the training population and between the training and the test populations, heritability of a trait, marker density, and statistical method (Crossa J et al. 2017a; Robertsen CD et al. 2019). In this study, the predictive abilities of a number of models were measured by Pearson and Spearman correlations between predicted GEBVs and actual values of the hybrid rice accessions. The predictive ability of the same trait varied among models, and the predictive ability of the same model also showed varying performance on different traits. No single model can be used for a good estimation of all the traits. The genetic structures of traits were complex and diverse. In practical breeding, multiple GS models should be used for prediction of these traits.

Previous studies showed that the significant markers GWAS have obvious effect on genomic prediction and can be used to assist in deciding what model strategies should be considered (Wilson S et al. 2021). In this study, when comparing the predictabilities of models with or without considering the SNP markers from GWAS, almost all the models have higher predictability with the consideration of markers from GWAS. Thus, markers associated with the trait should be considered in the genomic selection models for better predict accuracy.

Predictive accuracy and potential application of GS models

At present, there is no model that can be widely applied to all traits. Though the stability and accuracy of GS models are continuously improved over time, there are still two main challenges, namely, computational accuracy and computational efficiency. The direct method (represented by GBLUP) had the higher calculation efficiency, but lower calculation accuracy when compared with the indirect method (represented by Bayes B). The other factors perplex the direct method were the setup of parameters which highly depend on researchers' experiences, as various parameter setups have profound effects on the final results. Similarly, though the indirect method has high accuracy. but it is difficult to effectively guide breeding practice because of the large amount of calculation in the process of parameter solution and the inability to realize parallel operation. In this study, Bayes B and RKHS had high predictability for plantheight, panicle length, thousand grain weight, grain length, grain length and width ratio, amylose content and alkali value; while BRNN, GBLUP, rrBLUP, RFC, SPLS, SVC and SVR had low predictability for other traits (Table S4). The predictive accuracy was low (0.22–0.35) for grain yield, but high for yield component traits such as number of grains per panicle, spikelet fertility and thousand grain weight. Thus, it will be useful to predict the yield related traits such as grain weight rather than the yield itself. The predictive accuracies for grain quality traits were higher, with the only exception of transparency.

The goal of GS for hybrid breeding is using the genotypes of the parents to predict the performance of the hybrids, which will significantly reduce the number of crosses for field test (Xu S et al. 2014; Xu Y et al. 2018; Labroo MR et al. 2021). Although only F1 hybrids were investigated in this study, in practical breeding, only the parents (sterile lines and restore lines) need to be genotyped, then genotypes of the F1 hybrids can be simulated and used for prediction of yield and grain quality traits by the GS models. This will significantly reduce the number of crosses to be made and the hybrids to be tested, thus, genomic selection is more efficient for hybrid rice breeding.

Conclusions

In this study, a total of 404 hybrid rice accessions were genotyped by 56K SNP chip, and 20 traits of related to yield and grain quality were investigated. Sixty seven significant loci were identified for the yield and agronomic traits, and 123 significant loci were identified for the grain quality traits by genome-wide association study. Two of these loci associated with increasing grain yield but decreasing grain quality. Genomic selection models of 15 different prediction algorithms were established using the genotypic and phenotypic data. The GS models are useful for plant height, panicle length, thousand grain weight, grain length and width ratio, amylose content and alkali value, but the predictability for other traits are low. Use of proper model for specific trait is important for successful genomic selection. The GS models could be used for prediction of some important traits of hybrid rice through the genotypes of the parental varieties.

Declarations

Acknowledgements: -

Author contributions: P Yu and C Ye wrote the original draft; L Li, H Yin, J Zhao and Y Wang performed the experiments; Z Zhang, W Li and Y Long analyzed the data; X Hu, J Xiao, G Jia and B Tian designed and supervised this study. All the authors reviewed the manuscript.

Funding: This study was financially supported by the national key research and development project of Ministry of Science and Technology (2017YFD0102002-4).

Data Availability Statement: The data supporting the findings of this study are available within the article and its supplementary materials.

Code availability: Not applicable

Ethics approval: This article does not contain any studies with animals performed by any of the authors.

Consent to participate: Not applicable.

Consent for publication: All authors are consent to publication.

Competing interests: The authors declare no competing interests.

References

1. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67 (1):1-48
2. Bradbury P, Zhang Z, Kroon D, Casstevens T, Ramdoss Y, Buckler E (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23 (19):2633-2635
3. Covarrubias-Pazaran G (2016) Genome-assisted prediction of quantitative traits using the R package sommer. PLOS ONE 11 (6):e0156744
4. Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, Campos G, Burgueño J, González-Camacho J, Pérez-Elizalde S, Beyene Y, Dreisigacker S, Singh R, Zhang X, Gowda M, Roorkiwal M, Rutkoski J, Varshney R (2017a) Genomic selection in plant breeding: methods, models, and perspectives. Trends in Plant Science 22 (11):961-975
5. Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, Campos G, Burgueño J, González-Camacho M, Pérez-Elizalde S, Beyene Y, Dreisigacker S, Singh R, Zhang X, Gowda M, Roorkiwal M, Rutkoski J, Varshney K (2017b) Genomic selection in plant breeding: methods, models, and perspectives. Trends in Plant Science 22:961-975
6. Cui Y, Li R, Li G, Zhang F, Zhu T, Zhang Q, Ali J, Xu S (2020) Hybrid breeding of rice via genomic selection. Plant Biotechnology Journal 18 (1):57-67
7. Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry Bulletin 19:11-15
8. García-Ruiz A, Cole J, anRaden P, Wiggans G, Ruiz-López F, Tassell C (2016) Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. PANS 113 (28):3995-4004
9. Grenier C, Cao T, Ospina Y, Quintero C, Châtel M, Tohme J, Courtois B, Ahmadi N (2015) Accuracy of genomic selection in a rice synthetic population developed for recurrent selection breeding. PLOS ONE 10 (8):e0136594

10. Hassen M, Cao T, Bartholomé J, Orasen G, Colombi C, Rakotomalala J, Razafinimpiasa L, Bertone C, Biselli C, Volante A, Desiderio F, Jacquin L, Valè G, Ahmadi N (2018) Rice diversity panel provides accurate genomic predictions for complex traits in the progenies of biparental crosses involving members of the panel. Theoretical and Applied Genetics 131:417-435
11. Huang M, Balimponya E, Mgonja E, McHale L, Kihupi A, Wang G, Sneller C (2019) Use of genomic selection in breeding rice (Oryza sativa L.) for resistance to rice blast (Magnaporthe oryzae). Molecular Breeding 39:114
12. International Rice Genome Sequencing Project, Sasaki T (2005) The map-based sequence of the rice genome. Nature 436:793-800
13. Isidro J, Jannink J, Akdemir D, Poland J, Heslot N, Sorrells M (2015) Training set optimization under population structure in genomic selection. Theoretical and Applied Genetics 128:145-158
14. Iwata H, Ebana K, Uga Y, Hayashi T (2015) Genomic prediction of biological shape: elliptic fourier analysis and kernel partial least squares (PLS) regression applied to grain shape prediction in rice (Oryza sativa L.). PLOS ONE 10 (3):e0120610
15. Jena KK, Mackill DJ (2008) Molecular Markers and Their Use in Marker-Assisted Selection in Rice. Crop Science 48 (4):1266-1276
16. López M, Neira R, Yáñez J (2015) Applications in the search for genomic selection signatures in fish. Frontiers in Genetics 5:458
17. Labroo MR, Ali J, Aslam MU, de Asis EJ, dela Paz MA, Sevilla MA, Lipka AE, Studer AJ, Rutkoski JE (2021) Genomic prediction of yield traits in single-cross hybrid rice (Oryza sativa L.). Frontiers in Genetics 12:692870
18. Li J, Wang J, Zeigler R (2014) The 3,000 rice genomes project: new opportunities and challenges for future rice research. GigaScience 3:8
19. Lu F, Yang F, Fan T, Liu J, Li Q, Wang L, Long X (2019) Analysis of rice variety approval data from 1977 to 2018. China Seed Industry 2:29-40
20. Ma GH, Yuan LP (2015) Hybrid rice achievements, development and prospect in China. Journal of Integrative Agriculture 14 (2):197-205
21. McCouch S, CGSNL (Committee on Gene Symbolization NaL, Rice Genetics Cooperative), (2008) Gene nomenclature system for rice. Rice 1:72-84
22. Meuwissen T (2007) Genomic selection : marker assisted selection on a genome wide scale. Animal Breeding and Genetics 124 (6):321-322
23. Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157 (4):1819-1829
24. Misra G, Anacleto R, Badoni S, Butardo V, Molina JL, Graner A, Demont M, Morell M, Sreenivasulu N (2019) Dissecting the genome-wide genetic variants of milling and appearance quality traits in rice. Journal of Experimental Botany 70 (19):5115-5130
25. Misztal I, Legarra A (2017) Invited review: efficient computation strategies in genomic selection. Animal 11 (5):731-736
26. Mucha S, Mrode R, MacLaren-Lee I, Coffey M, Conington J (2015) Estimation of genomic breeding values for milk yield in UK dairy goats. Dairy Science 98 (11):8201-8208
27. Onogi A, Ideta O, Inoshita Y, Ebana K, Yoshioka T, Yamasaki M, Iwata H (2015) Exploring the areas of applicability of whole-genome prediction methods for Asian rice (Oryza sativa L.). Theoretical and Applied Genetics 128:41-53
28. Proost S, Mutwil M (2017) PlaNet: Comparative Co-Expression Network Analyses for Plants. Methods in Molecular Biology 1533:213-227

29. Robertsen CD, Hjotrtshøj RL, Janss LL (2019) Genomic selection in cereal breeding. Agronomy 9:1-16
30. Samorè AB, L. F (2015) Genomic selection in pigs: state of the art and perspectives. Italian Journal of Animal Science 15 (2):211-232
31. Spindel J, Begum H, Akdemir D, Virk P, Collard B, Redoña E, Atlin G, Jannink J, S. M (2015) Genomic selection and association mapping in rice (Oryza sativa): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite tropical rice breeding lines. PLOS Genetics 11 (2):e1004982
32. Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F, Mansueto L, Copetti D, Sanciangco M, Palis KC, Xu J, Sun C, Fu B, Zhang H, Gao Y, Zhao X, Shen F, Cui X, Yu H, Li Z, Chen M, Detras J, Zhou Y, Zhang X, Zhao Y, Kudrna D, Wang C, Li R, Jia B, Lu J, He X, Dong Z, Xu J, Li Y, Wang M, Shi J, Li J, Zhang D, Lee S, Hu W, Poliakov A, Dubchak I, Ulat VJ, Borja FN, Mendoza JR, Ali J, Li J, Gao Q, Niu Y, Yue Z, Naredo MEB, Talag J, Wang X, Li J, Fang X, Yin Y, Glaszmann JC, Zhang J, Li J, Hamilton RS, Wing RA, Ruan J, Zhang G, Wei C, Alexandrov N, McNally KL, Li Z, H. L (2018) Genomic variation in 3010 diverse accessions of Asian cultivated rice. Nature 557:43-49
33. Wang X, Li L, Yang Z, Zheng X, Yu S, Xu C, Z. H (2017) Predicting rice hybrid performance using univariate and multivariate GBLUP models based on North Carolina mating design II. Heredity 118:302-310
34. Wilson S, Zheng C, Maliepaard C, Mulder HA, Visser RGF, Burgt A, Eeuwijk F (2021) Understanding the effectiveness of genomic prediction in tetraploid potato. Frontiers in Plant Science 12:1-13
35. Wolc A, Zhao H, Arango J, Settar P, Fulton J, O'Sullivan N, Preisinger R, Stricker C, Habier D, Fernando R, Garrick D, Lamont S, J. D (2015) Response and inbreeding from a genomic selection experiment in layer chickens. Genetics Selection Evolution 47:59
36. Xiao N, Pan C, Li Y, Wu Y, Cai Y, Lu Y, Wang R, Yu L, Shi W, Kang H, Zhu Z, Huang N, Zhang X, Chen Z, Liu J, Yang Z, Ning Y, A. L (2021) Genomic insight into balancing high yield, good quality, and blast resistance of japonica rice. Genome Biology 22:283
37. Xu S (2017) Predicted residual error sum of squares of mixed models: an application for genomic pPrediction. G3: Genes, Genomes, Genetics 7 (3):895-909
38. Xu S, Zhu D, Zhang Q (2014) Predicting hybrid performance in rice using genomic best linear unbiased prediction. PNAS 111 (34):12456-12461
39. Xu Y, Wang X, Ding X, Zheng X, Yang Z, Hu Z (2018) Genomic selection of agronomic traits in hybrid rice using an NCII population. Rice 11 (1):32
40. Yabe S, Yoshida H, Kajiya-Kanegae H, Yamasaki M, Iwata H, Ebana K, Hayashi T, Nakagawa H (2018) Description of grain weight distribution leading to genomic selection for grain-filling characteristics in rice. PLOS ONE 13 (11):e0207627
41. Yao W, Li G, Yu Y, Ouyang Y (2018) FunRiceGenes dataset for comprehensive understanding and application of rice functional genes. GigaScience 7 (1):gix119
42. Zhang Z, Zhang Q, Ding X (2011) Advances in genomic selection in domestic animals. Chinese Science Bulletin 56 (25):2655-2663

Figures

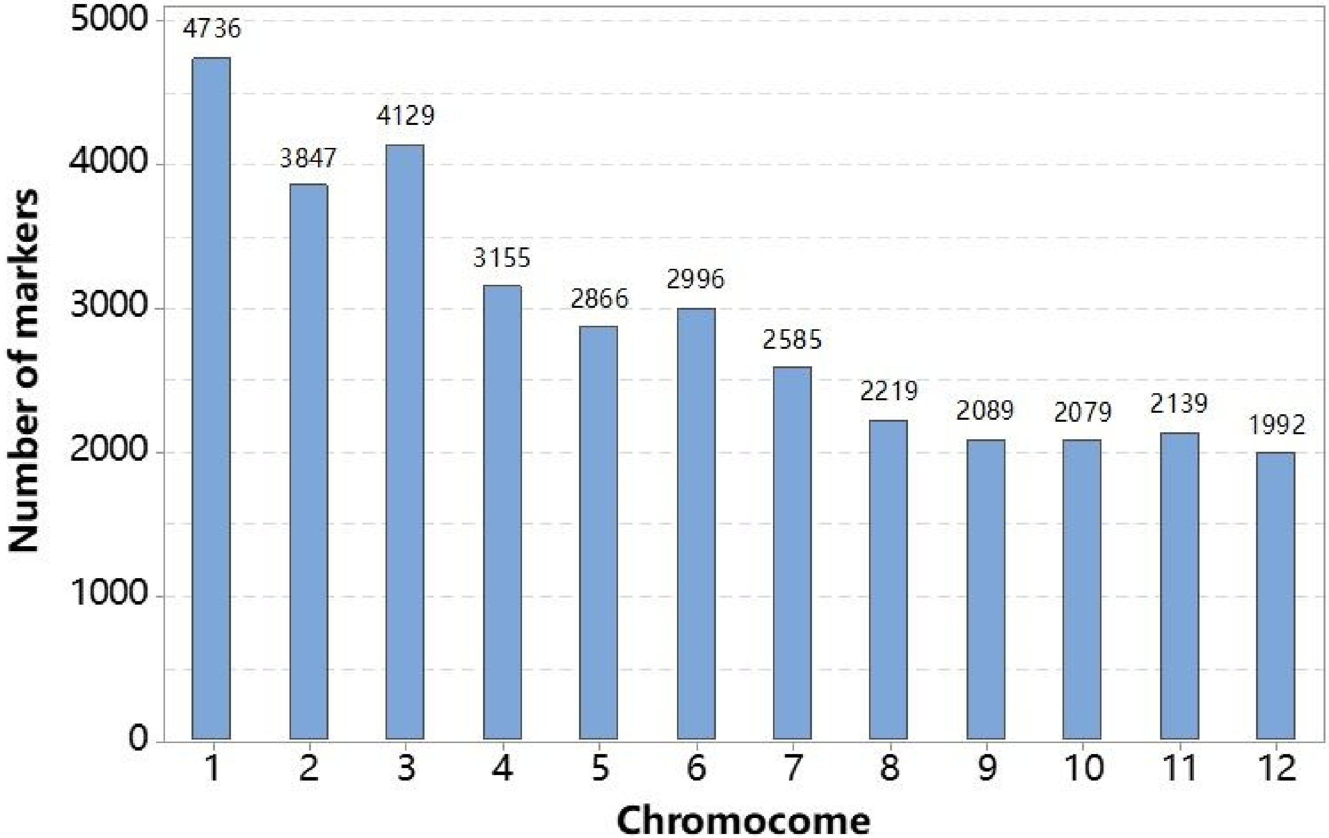

Figure 1

Number of markers on each chromosome. Total number of high-quality SNP markers is 34832.

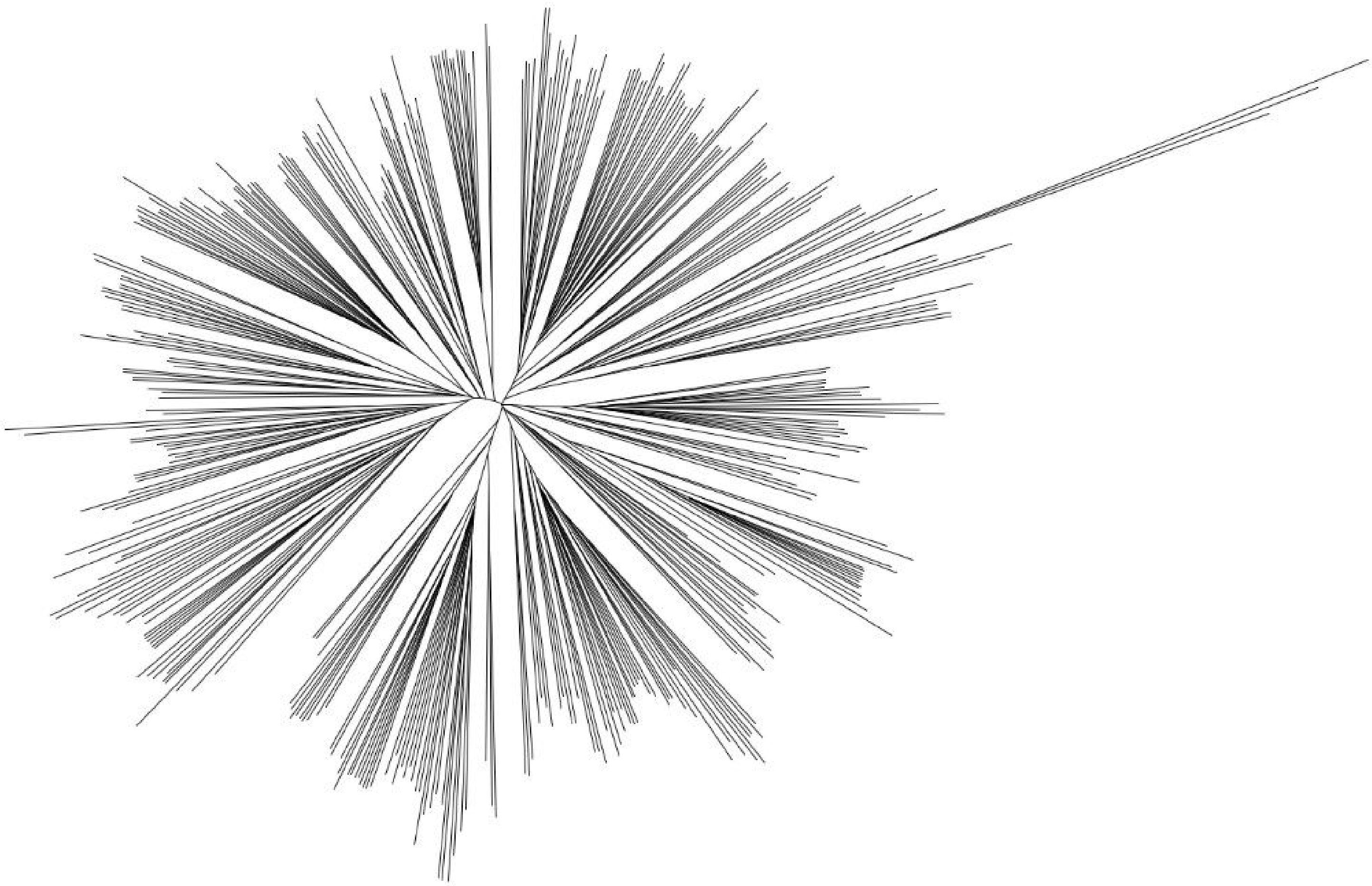

Figure 2

Phylogenetic tree of the tested hybrids.

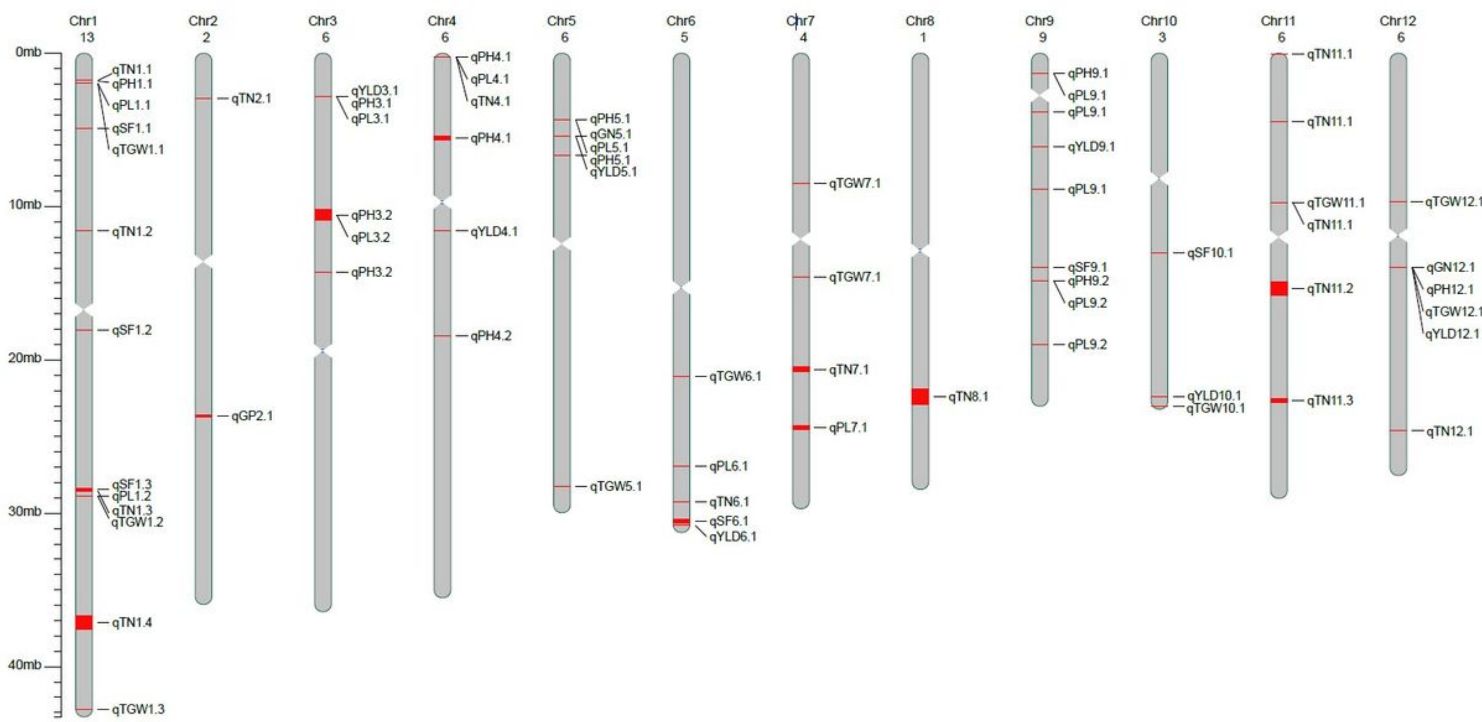

Figure 3

QTLs for agronomic traits identified by genome-wide association study (GWAS).

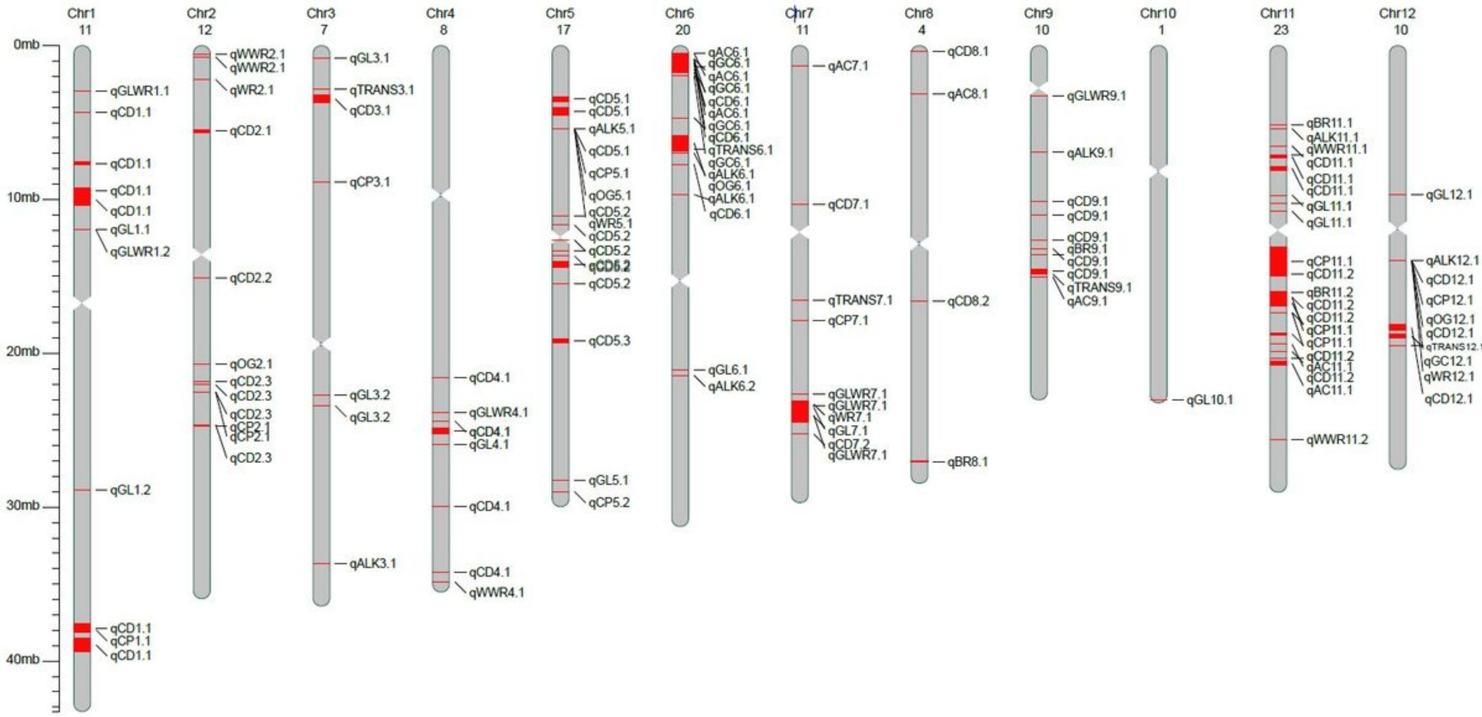

Figure 4

QTLs for grain quality traits identified by GWAS.

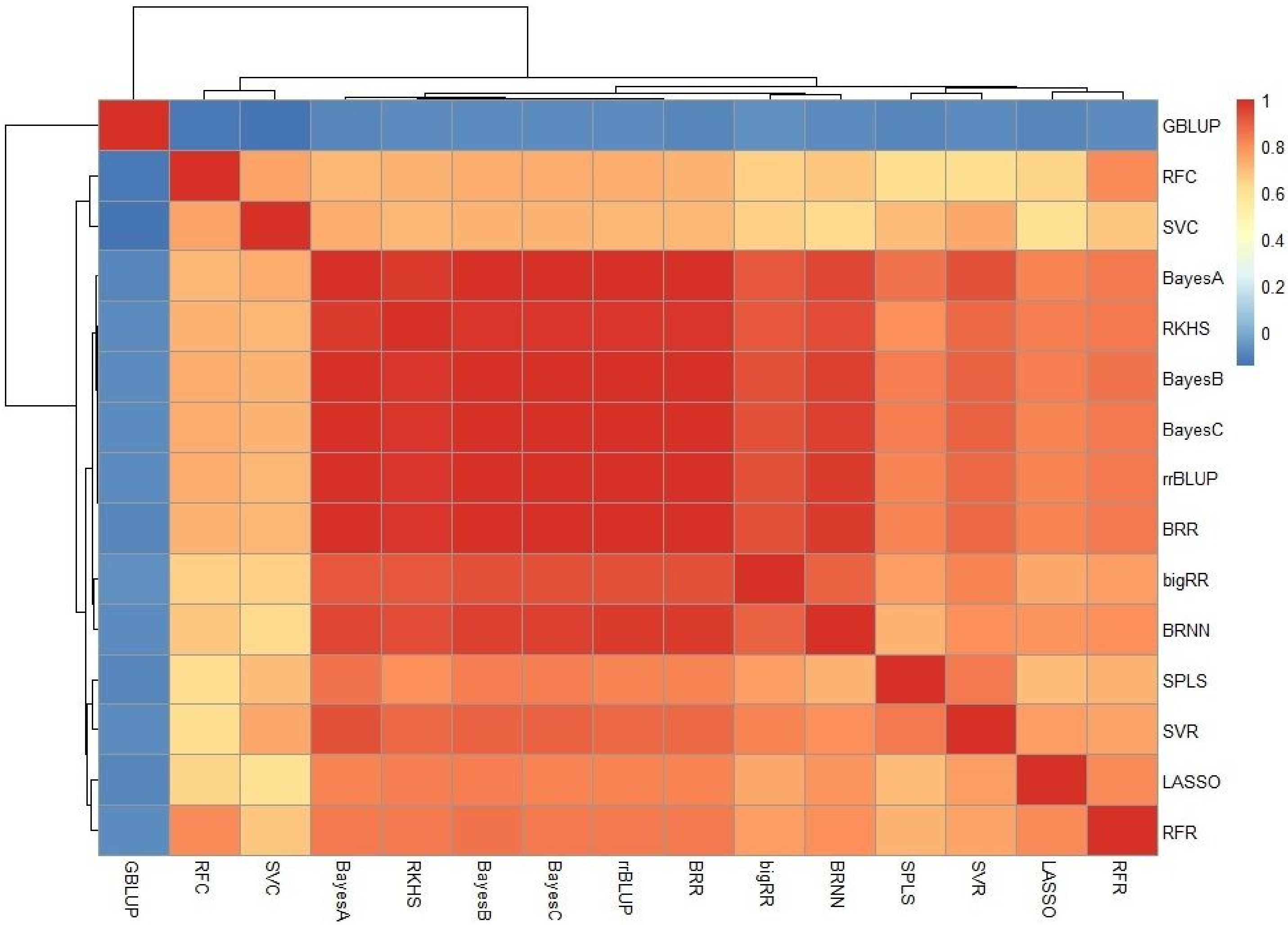

Figure 5

The correlation heat map of different GS models.

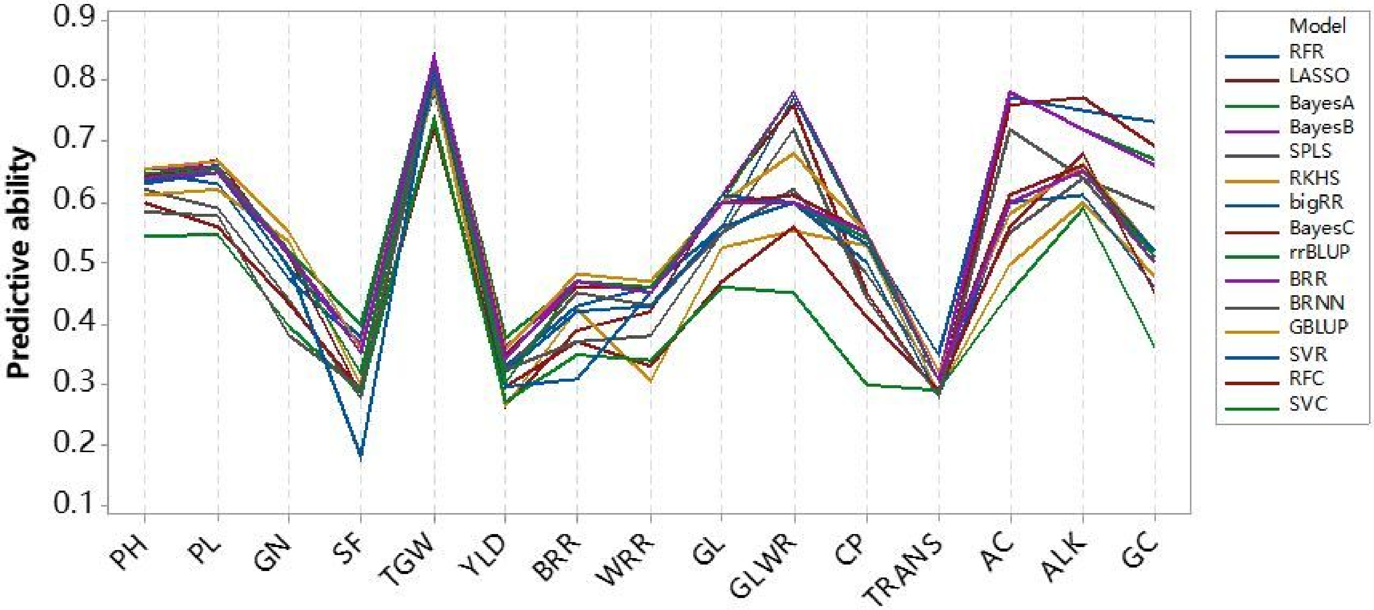

Figure 6

Comparison of predictive ability of different GS models for different traits.

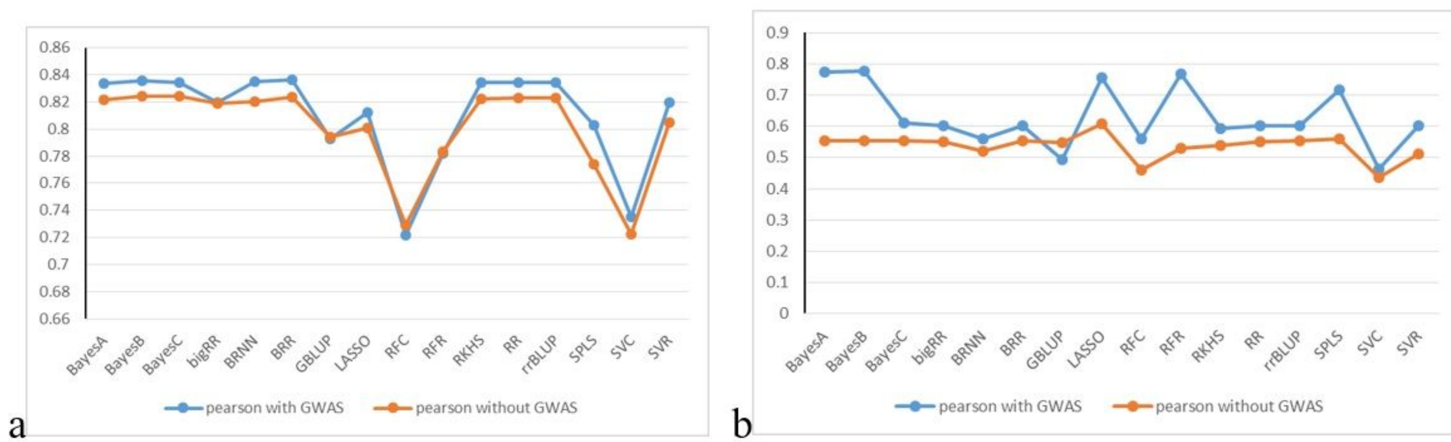

Figure 7

Comparison of predictability of different GS models without or with the significant SNP markers from the GWAS analysis. a. thousand grain weight, b. amylose content.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalFigures.pdf
- SupplementalTable14.pdf
- SupplementalTable5.pdf