Indian J. Genet. Plant Breed., (2025); 85(4 Suppl.) : 696-702
https://doi.org/10.31742/ISGPB.85.suppl.2

ISSN: 0975-6906
www.isgpb.org

RESEARCH ARTICLE

Maximizing genetic gain for rice (*Oryza sativa* L.) grain yield by implementing genomic selection

C. Anilkumar[†*], Rameswar Prasad Sah[†], T.P. Muhammed, Azharudheen[†$], Annamalai Anandan[#], Sasmita Behera, Bishnu Charan Marndi, S.K. Dash, J. Meher and Sanghamitra Samantaray

Abstract

Improving a quantitative trait like grain yield in rice using conventional breeding approaches is time and resource-demanding. Utilizing genomic selection for improving grain yield in rice is assumed to be promising. A founder population genotyped with novel genomic markers was used as a training population. The training population was phenotyped over three years for grain yield. A bi-parental population developed from parents selected from the training population genotyped with the same markers was used as the testing population. Four different predictive models were used on the training population at different marker densities. The results indicated that lower marker densities leads to poor predictive abilities among all models. Increasing marker density improves the prediction ability; however, the increment in predictive ability over the mid-density of markers was relatively low. The candidate genotypes selected based on predicted performance in the testing population showed a 20% higher genetic gain over the testing population mean, a 16% higher gain over the training population mean, and a 150% higher gain over the mid-parent value. The mid-density markers uniformly covering the rice genome uniformly are sufficient to implement genomic selection in rice. Integrating genomic selection into ongoing breeding programs would benefit the breeder in selecting potential candidates for improving grain yield in rice.

Keywords: Genomic selection, Genetic gain, Leave-one-out cross-validation, Marker densities, Predictive abilities.

Introduction

Rice (*Oryza sativa* L.) is the primary food source for most of the population in India and South Asia; hence, the primary rice breeding objectives are to develop high-yielding, nutritious, stress-resilient, and climate-smart cultivars (Chakraborti et al. 2021). As the global population is increasing at an alarming rate, the pressure to produce more food within the available resources is increasing. In this context, improving the genetic potential of rice varieties to produce more grains per unit consumption of resources is a feasible way to meet the global rice demand and achieve sustainability (Saito et al. 2021). However, the conventional breeding approach of hybridizing two parents, developing a population, evaluating progenies, selecting the best progeny based on phenotype, and releasing it as a cultivar consumes time. Despite that, the genetic gain achieved through conventional breeding remains uncertain (Collard and Mackill 2008). However, the utilization of molecular markers in rice crop improvement accelerated the process of cultivar development through marker-assisted selection (MAS) programs. The extent of MAS utilization was limited to the genetic transfer of major genes or large-effect quantitative trait loci (QTL) (Gregorio et al. 2013). The MAS breeding approach was not fully designed to improve quantitative traits like grain yield, which are controlled by several genes, or QTL with minor effects (Jena and Mackill 2008). The genomic selection (GS) approach has been proven promising for accounting for most of the minor

ICAR-Central Rice Research Institute, Cuttack 753 006, Odisha, India

[$] Present address: ICAR-Indian Institute of Spices Research, Calicut 673 102, Kerala, India.

[#] Present address: ICAR- Indian Institute of Seed Science, Regional Station, Bengaluru 560 065, Karnataka, India.

[†] These authors contributed equally to this work

***Corresponding Author:** C. Anilkumar, ICAR-Central Rice Research Institute, Cuttack 753 006, Odisha, India, E-Mail: anilcgpb@gmail.com

How to cite this article: Anilkumar C., Sah R. P., Azharudheen M.T.P., Anandan A., Behera S., Marndi B. C., Dash S. K., Meher J. and Samantaray S. 2025. Maximizing genetic gain for rice (*Oryza sativa* L.) grain yield by implementing genomic selection. Indian J. Genet. Plant Breed.,, **85**(4 Suppl.): 696-702.

Source of support: Nil

Conflict of interest: None.

Received: May 2025 **Revised:** Sept, 2025 **Accepted:** Oct. 2025

genes or QTL controlling a quantitative trait during the selection process (Bernardo, 1994; Meuwissen et al., 2001). The GS approach uses genome-wide marker information and trait phenotype in a population called the training population and estimates the breeding values in the other population called the test population, which is genotyped but not phenotyped.

Since the GS can be implemented at an early breeding stage, it has the potential to accelerate genetic gain by reducing the length of the breeding cycle and accounting for most of the genes or QTL regulating the target trait (Alemu et al. 2024; Crossa et al. 2017). However, high prediction accuracy is the prelude to the successful implementation of GS in breeding programs. The prediction accuracy is known to depend on several factors, like the size of the training population, marker number, relationship between the training and testing populations, precision of phenotyping in the training population, and many other factors (Anilkumar et al. 2022a). Considering the effectiveness of the GS approach, it has been adopted in breeding programs for many small grain cereals like wheat, barley, oats, and rice. Nevertheless, adaptation of the approach in rice breeding is very limited (Anilkumar et al. 2022b). The availability of huge genomic resources and the lower cost of genotyping rice owing to its small genome size make the crop more suitable for implementing GS (Budhlakoti et al. 2022). The marker density required for the crop is much lower than that of any other crop due to its inherited nature of reproduction by self-pollination, as it keeps large linkage disequilibrium blocks. Hence, implementing genomic selection in rice breeding programs provides the advantage of achieving rapid genetic gains with limited resources.

To make genomic selection more resource-efficient, the historical datasets available to the breeder can be utilized as a training set for model development. The major requirement to be met is the relationship between training and testing sets to ensure good prediction accuracy. The training population may be a founder population that was used by the breeder, which was developed from different breeding programs or germplasm sets from which the breeder chose the parents to develop the breeding population, or a portion of the breeding population itself may be used as the training population (Krishnappa et al. 2021). While integrating the genomic selection program with existing breeding activities, breeders do not need to develop populations exclusively for genomic selection. Instead, breeders may explore the possibility of selecting candidates from available breeding populations by using any of the aforementioned populations as a training set that is precisely phenotyped and genotyped (Bassi et al. 2016; Jonas and de Koning, 2015). In this milieu, the objectives of the study were to implement genomic selection for grain yield in an existing breeding program to improve genetic gain in rice. A precisely phenotyped population was used as a training set to develop the genomic selection model and select candidate genotypes from a related bi-parental population.

Materials and methods

Experimental genetic material

The genetic material for the experiment was divided into two sets. Set-1 was composed of a panel of genotypes that comprised advanced breeding lines, released varieties, and some elite germplasm lines. The set-1, with 180 genotypes, was designated as the training population. Set-2 was a biparental population developed from parents selected from Set-1. The bi-parental population was in the $F_{7:8}$ generation and had completely attained homozygous conditions. The bi-parental population was initially developed for a different purpose; however, identifying potential candidates from this population will help carry forward selected candidates and reject the rest of the population, saving resources. Nevertheless, this scenario serves as a good example for the integration of genomic selection into an ongoing breeding program. Set 2, with 150 genotypes, was designated as the testing population. Since both parents were selected from the training population to develop the testing population, there was sufficient genetic relatedness between the training and testing populations to implement genomic selection.

Phenotyping the training population

The training population was evaluated for grain yield over three years at experimental fields of the ICAR-Central Rice Research Institute, Cuttack. The training population was phenotyped during the rainy seasons of 2021, 2022, and 2023. The training population was evaluated in a simple lattice design with two replications, with each replication covering an area of 4 m^2. The row length was 2 m, the spacing between plants was 20 cm, and the rows were 30 cm apart. All agronomic practices to raise a healthy crop were followed. At maturity, the panicles were harvested from plants in a 1 m^2 area in the middle of each plot, excluding border rows. The panicles were threshed separately and dried to 12–13% moisture under natural sunlight conditions before recording the weights. The grain weight per square meter was recorded for each entry in every replication. Further, the weights were converted to per-hectare yield. The same evaluation procedure was followed for three years. Every care was taken during the evaluation process to obtain precise phenotype values.

Genotyping training and testing population

The genomic DNA of genotypes in training and testing populations was extracted from 15-day-old seedlings following the CTAB method (Doyle and Doyle, 1987). The extracted DNA quantity was checked using a NanoDrop spectrophotometer following the manufacturer's guidelines. A set of 307 simple sequence repeat (SSR) markers developed from functionally characterized gene sequences was used in the present study. The gene-based SSR markers were developed from candidate genes reported for grain characters and yield traits in rice. The outline of the development of new SSR markers is detailed in Sah et al. (2023). The 10 µL of PCR reaction mixture consisted of 1 µL of 20ng/ µL template DNA, 3 µL of ready PCR mix (TAKARA®), 1-µL each of forward and reverse primer aliquots and 4 µL of nuclease-free water. The PCR program was set with initial denaturation for 4 minutes at 94°C, followed by 40 cycles of denaturation for 40 seconds at 94°C, 40 seconds of primer annealing at a temperature specific to each primer, 2 minutes of primer elongation at 72°C, and final primer extension for 7 minutes at 72°C. The amplified PCR products were separated on 3.5% agarose gel electrophoresis, and amplicon images were documented in the Zenith (Gel. Pro CCD gel document, Biozen Laboratories, India) gel documentation unit. The amplicons were then recorded as a binary dataset, which was used for genomic prediction and selection analysis.

Statistical analysis

The phenotypic information of grain yield recorded as tons per hectare in two replications over three years of evaluation was subjected to the estimation of best linear unbiased estimates (BLUE) to shrink the possible experimental errors. The BLUE values were estimated using CIMMYT-developed META-R software (Alvarado et al. 2020). The BLUE values were used to estimate the descriptive statistics of the training population using R software.

The genomic prediction model was trained using the marker information in binary form and BLUEs of grain yield. Four different models [Bayes ridge regression (BRR), LASSO, RKHS, and ridge regression BLUP (rrBLUP)] were trained initially to examine the performance of different models. The basic equation considered for developing different models is as follows:

$$Y = X\beta + \sum_{k=1}^{m} Z_k \gamma_k + \varepsilon$$

Where Y is a vector of observed phenotypes, X is a matrix of marker information, β is a vector of fixed effects, m is the number of markers, Z_k is a genotype indicator variable, γ_k is a vector of marker effects and ε is a vector of residuals.

The models were trained following a leave-one-out cross-validation approach implemented through the BGLR and rrBLUP packages in RStudio version 1.4.17. In this approach, one individual of the training population was considered the testing genotype, and the rest of the genotypes were considered the training population to predict the performance of one genotype. For instance, in the present study, one genotype of the training set was predicted using information from 179 genotypes as the training population. The selection of genotypes for testing and training was random, and the process was repeated with 100 iterations for each prediction. After predicting the phenotype of the training set, the predicted grain yield from four models was correlated with observed grain yield to estimate the predictive ability. The cross-validation was performed at different marker densities to evaluate the optimum number of markers for achieving satisfactory predictions in the rice crop. The predictive abilities of models at different marker densities were compared. The marker effects were estimated at high marker density and were used to predict the performance in the testing population. The performance of the testing population was predicted using the RKHS model, which uses a marker-based information-based relationship matrix to predict performance. The top 5 percent of the predicted genotypes were selected to compare the mean with the rest of the testing population and the mean of the training population.

Results and discussion

Phenotypic variation in the training population

The grain yield in the training population ranged from 2.34 to 7.50 t/ha with an average of 4.58 t/ha (Table 1). The skewness, a third-degree statistic, was negligible, and the kurtosis, a fourth-degree statistic, was -0.13. Similarly, the standard error of the population was relatively low (0.07). The standard deviation of the training population was relatively high (0.94). The estimated heritability in a broad sense was high for the grain yield (0.68). The grain yield performance between different types of genetic material in the training population was compared. Advanced breeding lines and released varieties in the training population showed higher grain yields compared to germplasm lines in the population, as expected. The summary statistics, third

Table 1. Descriptive statistics of grain yield (t/ha) in the training population

S. No.	Statistic	value
1	Mean	4.58
2	Standard deviation	0.94
3	Median	4.57
4	Minimum	2.34
5	Maximum	7.50
6	skewness	0.11
7	Kurtosis	-0.13
8	Standard error	0.07
9	Broad sense Heritability (H^2)	0.68

and fourth-degree statistics, indicated the grain yield in the experimental genotypes was controlled by several genes with minor effects.

Estimation of marker effects

The new SSR markers covered the entire rice genome evenly over all the rice chromosomes (Fig. 1). The informativeness of these markers in rice was highlighted in Anilkumar et al. (2023). The marker information and BLUE values of grain yield in the training population were used to estimate marker effects. The marker effects across loci were summed to estimate the genomic estimated breeding values (GEBVs) of the training population. The genomic estimated breeding values of the training population ranged from -1.30 to +0.80. The marker effects were normally distributed, as expected, and are consistent with reports of Budhlakoti et al. (2022). The genotypes used as parents to develop the testing population had genomic estimated breeding values of -0.19 (Thavalakannan, male parent) and +0.12 (Pooja, female parent) (Fig. 2). The GEBVs indicate that the differential distribution of genes for grain yield among parents. Genomic selection helps in the identification of true transgressive segregants derived from this cross for higher grain yield.

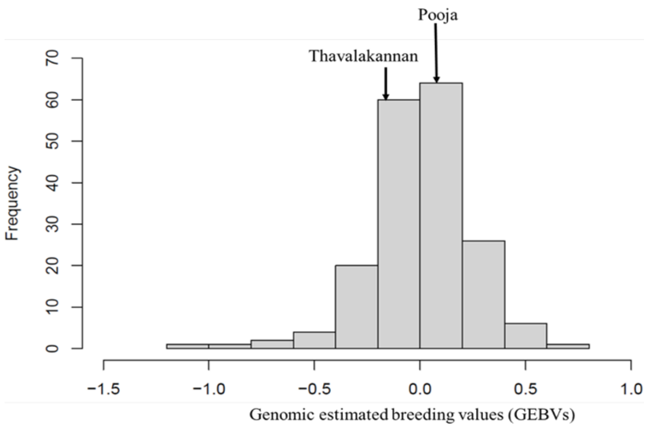

Fig. 2. Genomic estimated breeding values of genotypes for grain yield in the training population. The breeding values of parents used for developing the testing population are indicated with arrows

Cross validation

The cross-validation approach was followed to train the model and examine the predictive ability of each model. The cross-validation was performed at different marker densities. The predictive abilities at lower marker densities (less than 100 markers) were relatively low in all the models. However, there was no significant difference between various models for predictive abilities at low marker densities, suggesting that very low marker density leads to poor predictions. The predictive ability of models increased with an increase in marker density; however, the magnitude of the increment was reduced after 150 markers were used for predictions (Fig. 3). The increment in predictive ability of models was negligible after 225 markers were used for prediction.

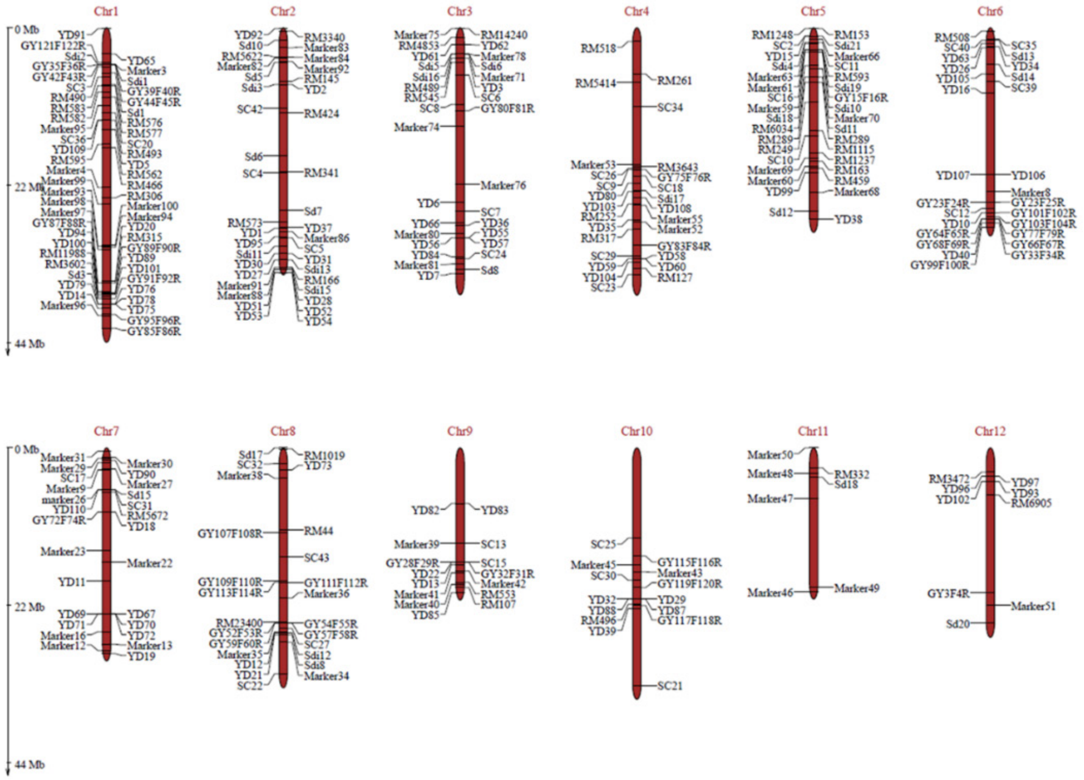

Fig. 1. Distribution of markers on all rice chromosomes in training population

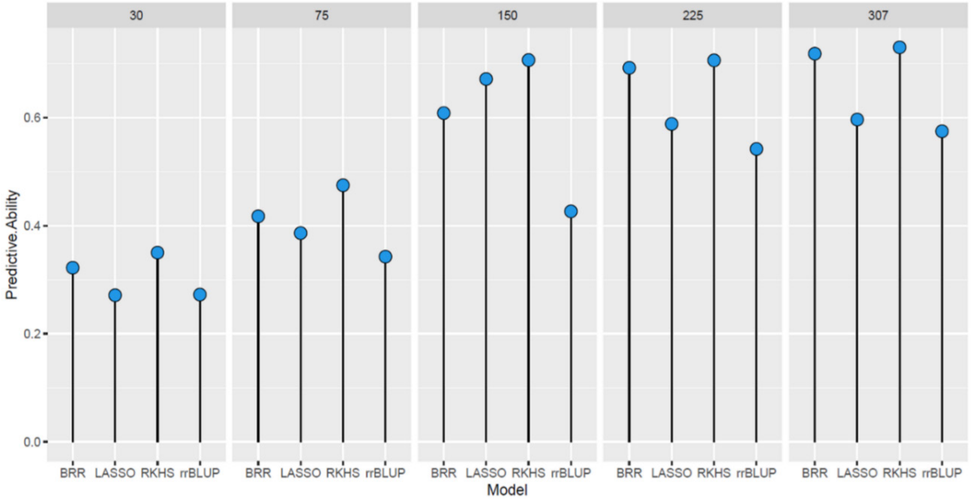

Fig. 3. Comparison of predictive abilities of different models with different marker densities

These results are consistent with earlier reports (de Sousa et al. 2024). The results suggest that near mid-density (~300 or more) markers uniformly spaced in the rice genome are sufficient to implement genomic selection. These results are consistent with those reported by Cerioli et al. (2022). Furthermore, differences in the prediction abilities of different models were negligible at maximum marker numbers; hence, breeders should focus on implementing or integrating genomic selection in breeding programs without wasting resources on optimization procedures. Among the different models used, RKHS showed good prediction ability across all marker densities compared to other models. However, when the maximum number of markers was used, the predictive ability of RKHS was comparable to Bayes ridge regression, and the predictive ability of LASSO was comparable to rrBLUP. The difference in predictive abilities of models when maximum markers were used was negligible, suggesting saturation of predictive abilities of models at higher marker densities. Therefore, it is ideal to use any convenient model to predict and select the candidates based on GEBVs.

Prediction in the testing population

The bi-parental population, which was genotyped with markers and not evaluated for grain yield, was used as the testing population. The predicted performance in the bi-parental population showed the genotypes in the testing set have the genetic potential to produce yields ranging between 3.44 and 5.22 t/ha, with an average of 4.41 t/ha. The median potential grain yield in the predicted population was 4.42 t/ha (Table 2.). The predicted performance of genotypes followed a normal distribution, similar to a typical quantitative trait. The grain yield for the parents of the population was 2.6 t/ha for the male parent and 4.5 t/ha for the female parent. However, the standard deviation of predicted grain yield was low compared to the standard

Table 2. Summary information of predicted grain yield (t/ha) in the testing population

S. No.	Statistic	Testing population	Top 5% predicted
1	Mean	4.41	5.33
2	Standard deviation	0.34	0.21
3	Median	4.42	5.31
4	Minimum	3.44	5.04
5	Maximum	5.22	5.48
6	skewness	0.00	0.00
7	Kurtosis	-0.28	0.13
8	Standard error	0.03	0.01
9	Pooja (female parent yield)	4.50 t/ha	-
10	Thavalakannan (male parent yield)	2.6 t/ha	-

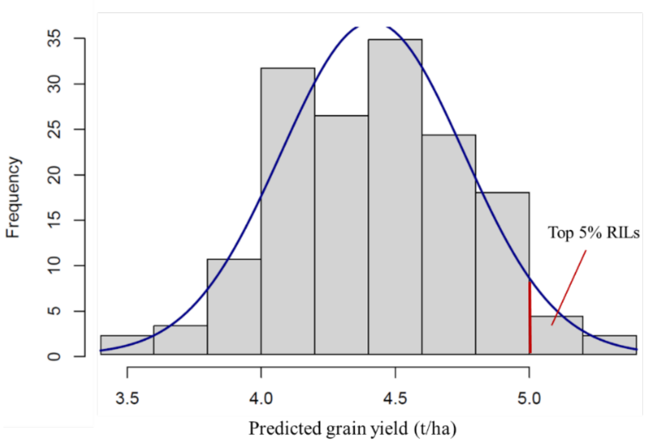

Fig. 4. Predicted performance of RIL population. Red partition in the right corner of the curve indicates top performing 5% RIL genotypes

deviation of grain yield in the training population. From the predicted population, the top 5% of genotypes were selected based on genomic estimated breeding values and predicted performance (Fig. 4). The mean of selected genotypes was higher than the predicted population and the training population. The genetic gain achieved in the selected population was 20% higher than the predicted population, 16.38% higher than the training population, and 150% higher than the mid-parent value. The selected proportion of the genotypes in the testing set may be evaluated at station trials or introduced as founder genotypes for the next cycle of genomic selection for grain yield in rice breeding programs. The genetic gain achieved in selected genotypes was significantly higher than the training population, predicted population, and mid-parent values. Hence, genomic selection helps in identifying the best-performing genotypes and improving genetic gains. The only selected genotypes may be evaluated at the station trial for grain yield instead of the entire population. Further, the selected genotypes can also be used to upgrade the founder population for the next cycle of genomic selection.

Authors' contribution

Conceptualization of research (CA); Designing of the experiments (RPS, SB); Contribution of experimental materials (RPS, AA); Execution of field/lab experiments and data collection (MTA); Analysis of data and interpretation (CA); Preparation of the manuscript (SS, SKD, JM, BCM).

Acknowledgments

The authors acknowledge the support of ICAR-National Rice Research Institute, Cuttack, during the period of this investigation.

References

Alemu A., Åstrand J., Montesinos-López O.A., Sánchez J.I. y, Fernández-Gónzalez J., Tadesse W., Vetukuri R.R., Carlsson A.S., Ceplitis A., Crossa J., Ortiz R. and Chawade, A. 2024. Genomic selection in plant breeding: Key factors shaping two decades of progress. Mol. Plant, **17**: 552–578. https://doi.org/10.1016/j.molp.2024.03.007

Alvarado G., Rodríguez F.M., Pacheco A., Burgueño J., Crossa J., Vargas M., Pérez-Rodríguez P. and Lopez-Cruz M.A., 2020. META-R: A software to analyze data from multi-environment plant breeding trials. The Crop J., Quantitative Genetics in the Omics Era, **8**: 745–756. https://doi.org/10.1016/j.cj.2020.03.010

Anilkumar C., Muhammed Azharudheen T.P., Sah R.P., Sunitha N.C., Devanna B.N., Marndi B.C. and Patra B.C. 2023. Gene based markers improve precision of genome-wide association studies and accuracy of genomic predictions in rice breeding. Heredity, **130**: 335–345. https://doi.org/10.1038/s41437-023-00599-5

Anilkumar C., Sah R., Azharudheen M., Sunitha N.C., Behera S., Marndi B., Sharma T. and Singh A. 2022b. Genomic Selection in Rice: Current Status and Future Prospects. pp. 68–82. https://doi.org/10.1201/9781003214991-4

Anilkumar C., Sunitha N.C., Harikrishna, Devate N.B. and Ramesh S. 2022a. Advances in integrated genomic selection for rapid genetic gain in crop improvement: A review. Planta, **256**: 87. https://doi.org/10.1007/s00425-022-03996-y

Bassi F.M., Bentley A.R., Charmet G., Ortiz R. and Crossa J. 2016. Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). Plant Science, From genomics to breeding, **242**: 23–36. https://doi.org/10.1016/j.plantsci.2015.08.021

Bernardo R. 1994. Prediction of Maize Single-Cross Performance Using RFLPs and Information from Related Hybrids. Crop Sci., **34**: cropsci1994.0011183X003400010003x.

Budhlakoti N., Kushwaha A.K., Rai A., Chaturvedi K.K., Kumar A., Pradhan A.K., Kumar U., Kumar R.R., Juliana P., Mishra D.C. and Kumar S. 2022. Genomic Selection: A Tool for Accelerating the Efficiency of Molecular Breeding for Development of Climate-Resilient Crops. Front. Genet., **13**: 832153. https://doi.org/10.3389/fgene.2022.832153

Cerioli T., Hernandez C.O., Angira B., McCouch S.R., Robbins K.R. and Famoso A.N. 2022. Development and validation of an optimized marker set for genomic selection in southern U.S. rice breeding programs. The Plant Genome, **15**: e20219. https://doi.org/10.1002/tpg2.20219

Chakraborti M., Anilkumar C., Verma R., Fiyaz R, A., Raj R., Patra B., Balakrishnan D., Sarkar S., Mondal N., Kar M., Meher J., Sundaram R. and Venkata SubbaRao L. 2021. Rice breeding in India: eight decades of journey towards enhancing the genetic gain for yield, nutritional quality, and commodity value. ORYZA- An Int. J. Rice, **58**: 69–88. https://doi.org/10.35709/ory.2021.58.spl.2

Collard B.C.Y. and Mackill D.J. 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Phil. Trans. R. Soc. B., **363**: 557–572. https://doi.org/10.1098/rstb.2007.2170

Crossa J., Pérez-Rodríguez P., Cuevas J., Montesinos-López O., Jarquín D., de Los Campos G., Burgueño J., González-Camacho J.M., Pérez-Elizalde S., Beyene Y., Dreisigacker S., Singh R., Zhang X., Gowda M., Roorkiwal M., Rutkoski J. and Varshney R.K. 2017. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends Plant Sci., **22**: 961–975. https://doi.org/10.1016/j.tplants.2017.08.011

de Sousa I.C., Barreto C.A.V., Caixeta E.T., Nascimento A.C.C., Azevedo C.F., Alkimim E.R. and Nascimento M. 2024. The trade-off between density marker panels size and predictive ability of genomic prediction for agronomic traits in Coffea canephora. Euphytica, **220**: 46. https://doi.org/10.1007/s10681-024-03303-8

Doyle J.J. and Doyle J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull., **19**: 11–15.

Gregorio G., Islam R., Vergara G. and Thirumeni S. 2013. Recent advances in rice science to design salinity and other abiotic stress tolerant rice varieties. SABRAO J.Breed. Genet., **45**: 31–41.

Jena K.K. and Mackill D.J. 2008. Molecular Markers and Their Use in Marker-Assisted Selection in Rice. Crop Sci., **48**: 1266–1276. https://doi.org/10.2135/cropsci2008.02.0082

Jonas E. and de Koning D.-J. 2015. Genomic selection needs to be carefully assessed to meet specific requirements in livestock breeding programs. Front. Genet., **6:** 49. https://

doi.org/10.3389/fgene.2015.00049

Krishnappa G., Savadi S., Tyagi B.S., Singh S.K., Mamrutha H.M., Kumar S., Mishra C.N., Khan H., Gangadhara K., Uday G., Singh G. and Singh G.P. 2021. Integrated genomic selection for rapid improvement of crops. Genomics, **113**: 1070–1086. https://doi.org/10.1016/j.ygeno.2021.02.007

Meuwissen T.H., Hayes B.J. and Goddard M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genet., **157**: 1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Sah R.P., Nayak A.K., Anilkumar C., Behera S., Azharudheen T.P.M. and Lavanya, G.R. 2023. cgSSR marker-based genome-wide association study identified genomic regions for panicle characters and yield in rice (*Oryza sativa* L.). J. Sci. Food Agric., **103**: 720–728. https://doi.org/10.1002/jsfa.12183

Saito H., Fukuta Y., Obara M., Tomita A., Ishimaru T., Sasaki K., Fujita D. and Kobayashi N. 2021. Two Novel QTLs for the Harvest Index that Contribute to High-Yield Production in Rice (*Oryza sativa* L.). Rice, **14**: 18. https://doi.org/10.1186/s12284-021-00456-1