Available online at www.sciencedirect.com

ScienceDirect

Rice Science, 2022, 29(5): 462−472

Research Paper

Genetic Dissection of Grain Size Traits Through Genome-Wide Association Study Based on Genic Markers in Rice

Amrit Kumar NAYAK[3], Anilkumar C[1], Sasmita BEHERA[1], Rameswar Prasad SAH[1], Gera Roopa LAVANYA[3], Awadhesh KUMAR[2], Lambodar BEHERA[1], Muhammed Azharudheen TP[1]

([1]*Division of Crop Improvement, Indian Council of Agricultural Research-National Rice Research Institute, Cuttack 753006, India;* [2]*Division of Crop Physiology and Biochemistry, ICAR-National Rice Research Institute, Cuttack 753006, India;* [3]*Department of Genetics and Plant Breeding, Sam Higginbottom University of Agriculture, Technology and Sciences, Prayagraj 211007, India*)

Abstract: Grain size plays a significant role in rice, starting from affecting yield to consumer preference, which is the driving force for deep investigation and improvement of grain size characters. Quantitative inheritance makes these traits complex to breed on account of several alleles contributing to the complete trait expression. We employed genome-wide association study in an association panel of 88 rice genotypes using 142 new candidate gene based SSR (cgSSR) markers, derived from yield-related candidate genes, with the efficient mixed-model association coupled mixed linear model for dissecting complete genetic control of grain size traits. A total of 10 significant associations were identified for four grain size-related characters (grain weight, grain length, grain width, and length-width ratio). Among the identified associations, seven marker trait associations explain more than 10% of the phenotypic variation, indicating major putative QTLs for respective traits. The allelic variations at genes *OsBC1L4*, *SHO1* and *OsD2* showed association between 1000-grain weight and grain width, 1000-grain weight and grain length, and grain width and length-width ratio, respectively. The cgSSR markers, associated with corresponding traits, can be utilized for direct allelic selection, while other significantly associated cgSSRs may be utilized for allelic accumulation in the breeding programs or grain size improvement. The new cgSSR markers associated with grain size related characters have a significant impact on practical plant breeding to increase the number of causative alleles for these traits through marker aided rice breeding programs.

Key words: best linear unbiased predictor estimate; candidate gene based SSR; efficient mixed-model association approach; genome-wide association study; VanRaden kinship

Rice is a major food crop that provides nourishment to billions of people across the world, accounting for almost half of the total daily calorie intake (Collard et al, 2008). This is also shown by the worldwide spread of rice production and cultivation in various ecologies throughout the year, regardless of the season (Patra et al, 2020; Chakraborti et al, 2021). As the world's population continues to grow at an alarming rate, the daily need for rice as a food source grows as well (Mohanty, 2013). Development of new cultivars and management packages which maximize the yield under differential rice growing environments with sustainable use of inputs is the primary target to meet the demand (Norton et al, 2018). Utilizing advanced breeding tools to counter the change in climate and increase production levels is the need of the hour (Gobu et al, 2020). Three-dimensional grain shape, as assessed by grain length (GL), grain width (GW) and length by

Received: 26 August 2021; **Accepted:** 26 January 2022
Corresponding author: Muhammed Azharudheen TP (md.azharudheen@icar.gov.in)

Peer review under responsibility of China National Rice Research Institute
http://dx.doi.org/10.1016/j.rsci.2022.07.006

width ratio, contributes to grain weight, which is the most significant yield determinant (Bai et al, 2010; Sanghamitra et al, 2018; Gao et al, 2020). Grain type has a direct relationship with grain yield and quality in rice. After grain number per panicle and panicle number per plant, grain size is a significant component of yield in rice. It also has a direct positive correlation with grain shape characters such as GL and GW (Evans, 1972; Xu, 2002). Determining the genetic basis of these traits is a prelude to their improvement. Hence, more than a decade of research has been done on these traits to understand their pattern of inheritance (Fu et al, 1994; Zhou et al, 2000; Hussain et al, 2020).

Rice grain characteristics have been extensively researched and recognized as a complicated inheritance system governed by multiple genes with modest and cumulative effects. To better understand the genetics of these characteristics, researchers applied both forward and reverse genetics methods (Hu et al, 2016; Hu et al, 2018). Several mapping experiments utilizing populations derived from bi-parental crosses have been performed to uncover genomic regions responsible for grain size. More than 500 QTLs for grain size related traits, including GL, GW and 1000-grain weight (TGW), have been mapped across all the rice chromosomes (Huang et al, 2013); however, only a few of these are fine-mapped. Advances in rice functional genomics have made it possible to characterize some of the genes that either positively or negatively regulate grain size characteristics in rice (Zhao et al, 2018). Furthermore, only a limited number of QTLs are directly used in practical plant breeding. The discovered QTLs are frequently not transferrable to other genetic backgrounds since the estimated effects are restricted to the two parents under investigation, as most genetic mapping research relies on conventional linkage mapping utilizing populations generated from bi-parental crosses. Since bi-parental populations account for only a small portion of genetic variation of a quantitative trait, the identified QTL effect compounds with epistatic effect, environmental interaction and pleiotropic effect on the trait. The genetic variation for quantitative traits like grain size should be captured by following an approach that exploits historic recombination events through linkage disequilibrium (Mather et al, 2004).

Advances in molecular marker technology, ease of genotyping at cheaper cost, and improved biometrical analysis platforms have assisted plant breeders to adopt new strategies for identification of QTLs for complex traits (Katara et al, 2021). The constraints of conventional bi-parental linkage mapping may be addressed by utilizing genome-wide association mapping to identify QTL by considering historic and ancestral recombination frequency (Yu et al, 2017). Genome-wide association study (GWAS) is the most successful method for identifying causative alleles for complicated traits like yield using well-distributed DNA markers across the genome (Korte and Farlow, 2013). The efficacy of GWAS is determined by ancestral linkage disequilibria between markers and phenotype-causing alleles. For identification of QTL for grain size, GWAS has been shown to be a strong supplementary approach to bi-parental linkage mapping in rice (Duan et al, 2017; Ma et al, 2019). Considering the availability of huge allelic diversity for grain related traits, GWAS can be the most promising approach for simultaneous mapping of QTLs for several traits with high precision (Huang and Han, 2014). GWAS utilizes dissimilarities among natural populations and identifies the new gene complexes for quantitative traits by whole genome scan with DNA markers. Allele diversity existing in natural populations along with historic recombination frequency considered for mapping enhances map resolution (Rafalski, 2010). Recently, a few research outcomes have proved the importance and efficiency of GWAS in identification of genomic regions for grain size characters in rice (Ponce et al, 2020). Hence, the approach can be considered for effective identification of causative alleles for grain size characters with a sufficient number of well distributed DNA markers.

A high number of DNA markers distributed across all chromosomes is required for a successful GWAS programme (Alqudah et al, 2020). Despite the fact that single nucleotide polymorphism (SNP) markers are the most prevalent in the rice genome, the high cost of genotyping prevents researchers from using them in their studies. Rice researchers most often employ simple sequence repeats (SSRs) for a variety of purposes, ranging from diversity studies (Garris et al, 2005; Anandan et al, 2021) to gene characterization investigations (Lu et al, 2005). The multi-allelic distribution, co-dominant inheritance pattern and high polymorphic informativeness, even with sparser coverage of the genome, made SSRs the most suitable for GWAS studies (Cho et al, 2000; Ching et al, 2002; Varshney et al, 2005). The candidate gene based SSR (cgSSR) markers can solve the uncertainty of linkage of random SSRs with a complex trait and increase the

resolution and precision of mapping through GWAS (Molla et al, 2015). By comprehending the benefits of utilizing SSRs, Vieira et al (2016) believe that the user-friendly and cost-effective nature of SSR markers has encouraged researchers to utilize them in practical plant breeding programmes.

In this study, genome-wide association mapping was conducted with a statistically strong and diverse association panel evaluated over three cropping seasons to identify significant marker-trait associations for grain size related characters. To guarantee the accuracy of findings, a set of cgSSR markers that had been newly developed based on seed dimension-related genes and grain yield-related genes were applied in the study. The findings of this study may be useful in further elucidating the genetic basis of rice grain size as well as in marker-assisted breeding programmes for improving grain yield in rice.

RESULTS

Phenotype variation

A wide range of observations for grain-size traits was recorded over three cropping seasons, which was reflected in across season genotype best linear unbiased predictor (BLUP) estimates. BLUP value based phenotypic variance, mean and other descriptive statistics estimated are presented in Table 1. BLUP estimates for TGW ranged from 10.6 to 31.9 g with an average of 21.50 g; while GL ranged from 5.21 to 10.59 mm with a mean of 8.39 mm. Similarly, GW ranged from 1.65 to 3.26 mm, with an average of 2.62 mm. Length-width ratio (LWR) ranged from 2.01 to 5.59, finding an average of 3.31. Third degree statistics-skewness and fourth degree statistics-kurtosis were employed to measure the distribution of phenotypes in the population. The skewness of the population for all the traits was negligible except for LWR, which showed positive significant skewness. However, kurtosis for all the traits was less than three, indicating platykurtic distribution of phenotypes in the population. The distribution pattern of phenotypes was depicted

Table 1. Phenotype variation and distribution pattern of four grain size-related traits.

Trait	Phenotype				Skewness	Kurtosis	Shapiro-Wilks '*p*'
	Min	Max	Mean ± SE	PV			
TGW	10.6	31.9	21.50 ± 0.04	0.18	0.07	0.41	0.197
GL	5.2	10.6	8.39 ± 0.14	1.82	-0.09	-0.51	0.754
GW	1.7	3.3	2.62 ± 0.04	0.14	-0.22	-0.74	0.094
LWR	2.01	5.59	3.31 ± 2.01	0.52	0.67	0.24	0.016

TGW, 1000-grain weight (g); GL, Grain length (mm); GW, Grain width (mm); LWR, Length-width ratio; PV, Phenotypic variance.

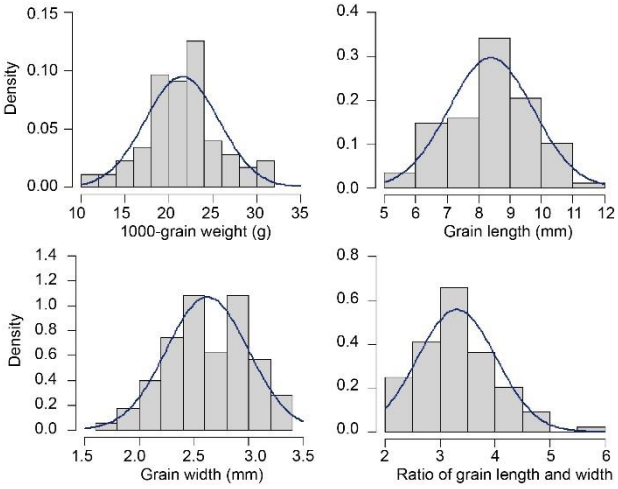

Fig. 1. Variation and distribution pattern of grain size and related traits in association panel.

using frequency distribution plots with a normal curve (Fig. 1), and Shapiro-Wilk's test (Shapiro and Wilk, 1965) for normality was also performed. *P* values suggested non-significance except for LWR, which was significant at the 0.05 level but non-significant at the 0.01 level (Table 1), thus population for above traits was normally distributed.

The correlation analysis performed to understand the linear relationship between grain traits is presented in Fig. 2. However, it was significant and strong

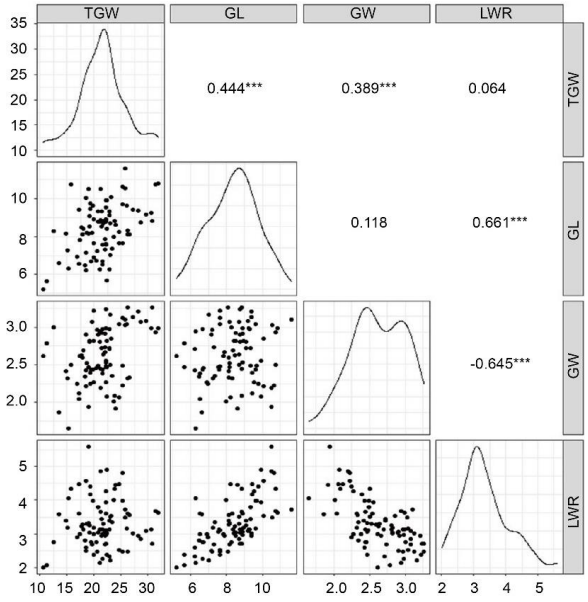

Fig. 2. Correlation coefficients and trend of distribution among grain size characters estimated based on across season best linear unbiased predictor values of phenotypes.

TGW, 1000-grain weight; GL, Grain length; GW, Grain width; LWR, Length-width ratio.

***, $P < 0.001$ by Pearson's correlation approach.

between TGW and GL as well as GW, and non-significant with LWR. Similarly, a negligible positive relationship was found between GL and GW as well as between TGW and LWR. The correlation coefficient between LWR and GW was negative significance, indicating that an inverse relationship exists between these two variables, while LWR has a positive and significant relationship with GL.

Genotype analysis

The 142 cgSSR markers were genotyped on individuals of the association panel. These markers amplified a total of 715 alleles in the population. The number of alleles ranged from 2 to 15, with an average of 6.3 alleles per locus. The robustness of cgSSRs was tested by estimating allele frequency and gene diversity. The major allele frequency ranged from 0.15 (M34) to 0.72 (M4) with allelic diversity ranging from 0 (M78) to 0.89 (M111). To test the informativeness of cgSSRs, the polymorphic information content (PIC) of each marker was estimated as a function of alleles in relation to their frequency in the population (Guo and Elston, 1999), and a PIC value of > 0.5 was considered as significantly high. Only 15 cgSSR markers showed a PIC value of < 0.5. The remaining 127 cgSSR markers expressed a PIC value of > 0.5 with the highest PIC value of 0.89 (M111).

Population structure and kinship analysis

Before performing GWAS analysis, we used genotype data of 142 markers to ascertain the population structure. Structure analysis was performed at three different levels, first by STRUCTURE analysis and prediction of the number of subpopulations through estimation of Δ*K*. The value of Δ*K* was found three upon 10 000 burn-in and 100 000 Markov Chain Monte Carlo (MCMC) with five iterations. Thus, it indicated the presence of three sub-populations within the association panel (Fig. 3-A). The largest sub-population consisted of 37 individuals; the second sub-population had 34, and the lowest had 17 individuals. Second, principal component analysis (PCA) detected the presence of three sub-populations, indicated by three significant components explaining the maximum variation of the population (Fig. 3-B). Third, the relatedness among individuals estimated through the VanRaden kinship algorithm using the genome association and prediction integrated tool

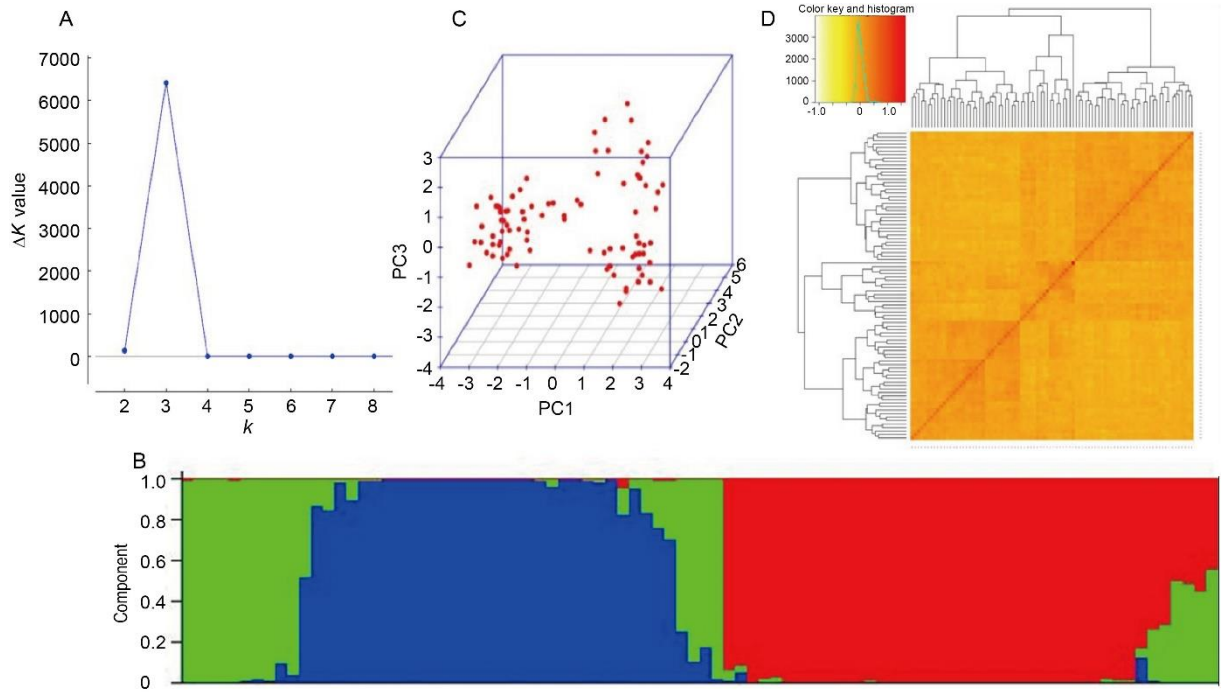

Fig. 3. Population structure analysis.

A, Magnitude of Δ*K* values with *k* ranging from 2 to 8 (*x*-axis) in association mapping panel.

B, Population structure of association panel based on 142 new candidate gene based SSR markers at *K* = 3. Different color columns represent different sub-populations.

C, 3D representation of principle component (PC) analysis showing three sub-populations.

D, Heat map of kinship matrix. The heat map shows the level of relatedness among the population. The darker areas show the level of relatedness between varieties and the dendrogram depicts clustering of sub-populations.

(GAPIT) was also explained by the presence of three sub-groups within the association panel (Fig. 3-C). The bar diagram representing the distribution of genotypes within and between sub-populations is presented in Fig. 3-D.

Association analysis

The genotypic information from 142 cgSSR markers assayed on individuals with four grain size related characters (TGW, GL, GW and LWR) was subjected to association analysis using the mixed linear model (MLM), following the efficient mixed-model association (EMMA) approach. A total of 10 significant marker trait associations (MTAs) at $P < 0.05$ were identified, distributed on five chromosomes (Table 2 and Fig. 4). Four significant MTAs, two on chromosome 5 (M69 and Sdi21) and one each on chromosomes 4 (M55) and 6 (Sd14), were identified for TGW. These MTAs were independent of each other and explained the phenotypic variances of 11.01%, 9.54%, 10.00% and 10.23%, respectively. A significant and solitary QTL was identified for GL on chromosome 4 through association of marker M55, explained 6.34% of the phenotypic variation. A total of four MTAs for GW were identified, two on chromosome 1 (Sdi1 and M99) and one each on chromosome 5 (M69) and chromosome 8 (M35), explaining 13.07%, 11.00%, 10.56% and 13.25% of the phenotypic variances, respectively. Similarly, only one MTA (Sdi1) was identified for LWR on chromosome 1, explaining 8.00% of the phenotypic variance. Among these ten putative QTLs, seven explain more than 10% of phenotypic variation and can be considered as major QTLs. We also identified a few markers associated with more than one trait. Marker M69 on chromosome 5 was associated with TGW and GW explaining 11.01% and 10.56% of phenotypic variations. Marker M55 on chromosome 4 was associated with TGW and GL with explained phenotypic variances of 10.00% and 6.34%, respectively. Similarly, marker Sdi1 on chromosome 1 was found associated with GW and LWR, explaining 13.07% and 8.00% of the phenotypic variances, respectively (Table 2). The graphical representation of results was done by developing Manhattan plots and quantile-quantile (Q-Q) plots for each trait using the GAPIT package (Fig. 4).

Table 2. Significant marker-trait associations identified for four grain size-related traits based on mixed line model.

Trait	Marker	Chr	Position (bp)	*P*-value	PVE (%)	Known gene
TGW	M69	5	18 724 905	0.01	11.01	*OSBC1L4*
	Sd14	6	5 315 178	0.02	10.23	*OsC1*
	M55	4	25 489 003	0.02	10.00	*SHO1*
	Sdi21	5	1 160 267	0.04	9.54	*RSR1*
GL	M55	4	25 489 003	0.04	6.34	*SHO1*
GW	M35	8	26 439 584	0.01	13.25	*NPP1*
	Sdi1	1	5 236 623	0.01	13.07	*OsD2*
	M99	1	25 382 698	0.02	11.00	*Rd*
	M69	5	18 724 905	0.02	10.56	*OSBC1L4*
LWR	Sdi1	1	5 236 623	0.02	8.00	*OsD2*

TGW, 1000-grain weight; GL, Grain length; GW, Grain width; LWR, Length-width ratio; Chr, Chromosome; PVE, Phenotypic variation explained.

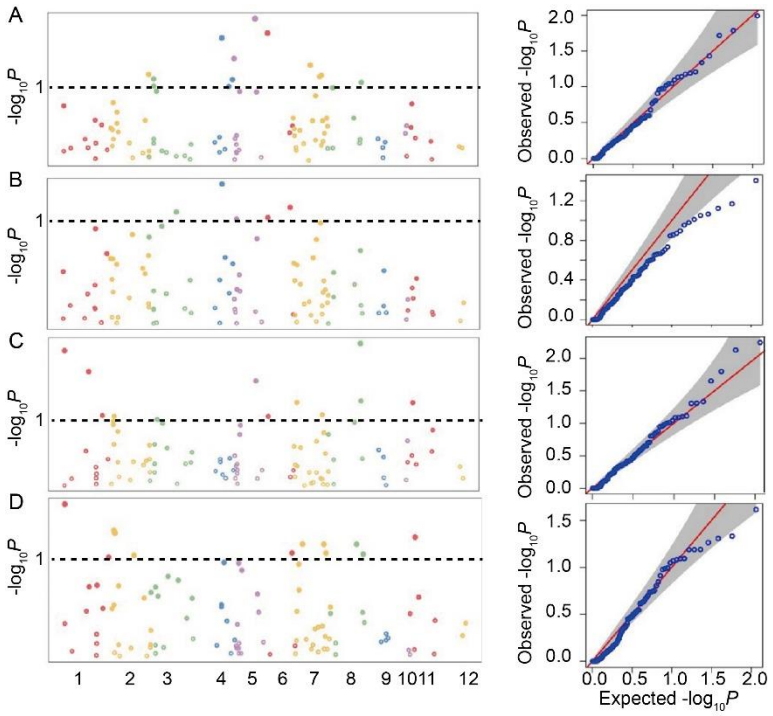

Fig. 4. Manhattan plots and Quantile-quantile plots for markers associated with grain traits across the genome.

A, 1000-grain weight; **B**, Grain length; **C**, Grain width; **D**, Length-width ratio.

In Manhattan plots, *x*-axis represents chromosomes and explains chromosome-wise marker distribution, and -log$_{10}P$ values on *y*-axis indicates significant associations. Quantile-quantile plots show deviation of observed -log$_{10}P$ values and expected -log$_{10}P$ values indicating the significant marker trait associations.

DISCUSSION

Identification of genomic regions associated with quantitative traits is a pre-requisite for deploying them in practical breeding to enhance the trait specific breeding. Improving grain size characters in rice has drawn the attention of researchers since it has a significant impact on grain yield. Several researchers attempted to map the

genomic regions controlling these traits and identify the underlying genes (Meng et al, 2016; Ponce et al, 2020). However, associating candidate gene-based markers to genomic regions controlling grain size traits has a significant impact as it will assist to address more than one trait simultaneously (Molla et al, 2019). Genome-wide association analysis in a set of germplasm accessions offers several advantages over bi-parental mapping in QTL identification (Wu et al, 2015). However, only a few such studies have been reported for grain size traits (Hussain et al, 2020; Ponce et al, 2020). There is abundant scope to explore natural variation that exists in germplasm accession for grain size related characters and improvement by incorporating identified QTL into breeding lines. In this study, a set of 88 highly potential genotypes were considered to constitute the association panel and evaluated over three years for grain size characters, whereas a set of 142 new cgSSR markers developed from different seed dimension-related genes and grain yield-related genes in rice were used to identify significant association of these new cgSSR markers with grain size traits.

Phenotype variation

Significantly wider phenotypic variation was recorded over three years for grain size traits. High phenotypic variation recorded for these traits from the population suggests an abundance of allelic variation for grain size traits. The BLUP values of four traits estimated over years showed normal distribution patterns, indicating the complex inheritance pattern of these traits (Fig. 1). Negligible skewness or zero skewness in a symmetric distribution shows the presence of additive gene interaction, while platykurtic distribution indicates the involvement of multiple genes in the development of certain grain size characters (Table 1) (Azharudheen et al, 2022). Variation analysis, skewness and kurtosis results supported the composition of the association panel for identification of putative QTLs through marker trait association for grain size traits using GWAS. The phenotypic correlation coefficients between TGW and GL, TGW and LWR were found to be positive. These results are consistent with reports by Tan et al (2000) and Ponce et al (2020). The strong correlation between TGW and GL indicates these traits have a significantly higher effect on grain weight than on other grain size traits (Xing and Zhang, 2010). Whereas, a negligible or weak correlation was observed between GW and LWR, and GL and GW, while GW and LWR recorded a strong negative relationship. These results were similar to the results obtained by Qiu et al (2015).

Genotype analysis

The marker assay with 142 new cgSSR markers showed greater diversity existing within the association panel. Upon genotyping, 88 individuals from the association panel with 142 markers amplified 715 alleles. The number of alleles ranged from 2 to 15, with an average of 6.3 alleles amplified per locus. Thus, the abundance of alleles per locus indicates genetic diversity within the association panel coupled with low gene flow, and this is consistent with previous reports (Rahman et al, 2007; Raju et al, 2016). The higher PIC values recorded by new cgSSR markers suggested the efficiency of these markers utilized in marker trait association studies. Since these cgSSR markers were generated from genic regions, they are more helpful for assaying genetic target traits even at smaller numbers.

It's crucial to have control over population structure in GWAS to prevent spurious marker trait associations. The origin, selection pressure and reproductive behavior of genotypes all have an impact on familial relatedness among individuals in an association panel (Atwell et al, 2010). The cgSSR markers applied were greatly efficient in controlling the population structure of the association panel, since they produced abundant allele for each trait. The relatedness among individuals in the association panel resulted in the identification of three sub-populations at $\Delta K = 3$ and the results are similar to earlier reports (Zhang et al, 2013; Wang et al, 2014). These sub-populations arise due to allelic sharing between sub-populations attributed to allelic accumulation due to spontaneous mutation over time (Agrama et al, 2007). PCA confirmed the presence of three sub-populations in the association panel. However, the kinship matrix generated by the VanRaden algorithm plotted as a heat map showing relatedness values between -0.5 to +0.5 indicates poor relationships existing between individuals in the association panel. These results assisted to understand the population structure of the panel before proceeding to GWAS for identification of putative genomic regions for grain size traits. Based on the information about population structure, the MLM with the EMMA approach (Mather et al, 2004) has been selected for association analysis, which detects marker trait associations while simultaneously addresses

population structure to reduce the chances of false positives (Zhang et al, 2014; Wang et al, 2016).

Association analysis

The structured association analysis following the MLM approach was performed with four phenotypes evaluated over three years and genotyped with 142 new cgSSR markers. We identified 10 significant marker trait associations distributed on five chromosomes. The markers M69, Sd14, M55 and Sdi21 are derived from *OsBC1L4*, *OsC1*, *SHO1* and *RSR1* genes, respectively, associated with TGW (Table 2). Except for Sdi21 (9.54%), the other associations explained more than 10% of the phenotypic variances, and hence they can be considered as major putative QTLs for grain weight. *OsBC1L4* is reported as responsible for leaf development in rice (Gao et al, 2020), and *SHO1* is involved in reproductive organ development (Song et al, 2012). *RSR1* is reported to be involved in starch biosynthesis in rice grain, and a mutation in *RSR1* produces larger seed size, increased seed mass and yield (Fu and Xue, 2010). However, markers derived from gene sequences of these markers showed association with TGW, indicating the contribution of genes related to developmental stages to grain weight. The marker M55 is derived from the *SHO1* gene, associated with GL, indicating the importance of *SHO1* in grain elongation. Markers M35, Sd1, M99 and M69, derived from genes *NPP1*, *OsD2*, *Rd* and *OsBC1L4*, showed association with GW. The gene *NPP1* is responsible for starch biosynthesis (Kaneko et al, 2014), and *OsD2* is responsible for grain shape (Seo et al, 2020) and the *Rd* gene is responsible for seed coat development and pigmentation (Jan et al, 2020). All these genes are related to grain characters and showed association with GW, indicating accuracy of association and the importance of new cgSSR markers. The marker Sd1 is derived from *OsD2* gene, associated with LWR and involved in grain shape development. The marker is directly associated with the grain shape gene. Hence, it can be effectively utilized in marker-aided plant breeding. The markers M69, M55 and Sd1 showed multiple trait associations, suggesting the involvement of respective genes in the development of more than one character through interaction of gene products in trait expression. These markers can be used effectively as surrogates for improvement of more than one respective associated character. The association results have been depicted using Manhattan plots and Q-Q plots. Manhattan plots indicate the distribution of markers on chromosomes and the significance of association based on $-\log_{10}P$ values on the *y*-axis (Fig. 4). Similarly, the Q-Q plot is a graphical depiction of the observed *P* values departure from the null hypothesis: each marker's observed *P* values are ordered from greatest to smallest and display against predicted values from a theoretical χ^2 distribution. If the observed and expected *P* values co-inside and fall on the middle line, it indicates acceptance of the null hypothesis and no significant association. In this study, *P* values differed from those predicted which indicated that those markers had a strong association with the trait (Fig. 4). Early separation of observed *P* values from expected indicates a large number of moderately significant marker trait associations, which is very rare.

Some grain size related genes have been identified and cloned over decades. However, the functional role and interaction with other genes in trait expression are still in the dark room. Association studies to identify markers linked to these traits are limited to random markers. Therefore, development and utilization of genic markers for association studies assists to understand the importance of interaction of genes in trait expression and utilize respective genic markers to hasten the process of trait improvement. In this study, we identified genic markers associated with grain size traits, which can be directly utilized for marker-assisted plant breeding programs. The markers associated with more than one trait and markers derived from genes responsible for other developmental processes can be used for simultaneous improvement of more than one grain related traits via marker-assisted breeding programs. Further, these genic markers associated with several grain size traits can be used to accumulate several causative alleles for enhancing grain size related traits through trait introgression breeding programs. At the outset, these results of association analysis have greater significance in practical plant breeding programs focusing on improving grain size-related traits.

METHODS

Association panel

The rice varieties, developed and released over the last three decades in India for varied ecologies, along with a modest number of diverse germplasm accessions, constituted the association panel. A total of 88 genotypes were considered to perform GWAS for identification of significant marker-trait association with GL, GW, LWR and TGW. The details of genotypes are presented in Table S1.

Field experimentation and phenotyping

The field trial for evaluation of the association panel was conducted over three seasons, the wet season of 2018, 2019 and 2020 at experimental plots of ICAR-National Rice Research Institute, Cuttack, India. Prior to starting the experiment, genotypes were tested for purity by planting a single row of true to type panicle in the previous season, and the procedure has been repeated every season to preserve genotype purity (Sahu et al, 2020). The genotypes were initially planted in nursery beds to ensure uniform germination, and healthy seedlings of 21-day-old were transplanted to the main field. The main field experiment was laid out in a randomized complete block design with two replications. Each genotype was planted in three-meter rows having a 20 cm gap between rows accommodating thirty plants in each row. The recommended inputs were provided to grow a healthy crop under irrigated conditions. At maturity, paddy from all the genotypes was harvested separately and dried under natural sunlight for 2 d, and then oven dried to reduce moisture content to 12%–14%. After equilibrating moisture content, a sample of 20 g paddy from each genotype was considered for measuring grain phenotypes. GL, GW and LWR were measured using Annadarpan (CDAC and RRS, West Bengal). Two sets of 50 grains from each genotype were considered for measuring these traits. A thousand random grains of each genotype taken from each replication were weighed on a high precision analytical balance (Sartorius Secura analytical balance, with readability to 0.1 mg to 320 g) to record TGW.

Molecular assay and genotyping

Genomic DNA of individuals in the association panel was estimated using the CTAB method (Doyle and Doyle, 1987). The absorbance ratio at 260 : 280 nm under spectrometer was employed for testing the quality of genomic DNA. Further, isolated DNA was quantified using Nanodrop (Thermo Scientific, USA) and the final concentration was adjusted to 20 ng/μL with 1× TE. The cgSSR markers derived from seed dimensions, grain yield and yield related characters (unpublished) were used. A total of 142 cgSSR markers distributed over 12 chromosomes (Fig. S1) were assayed on the association panel to generate genotype data. These markers were developed from genic sequences of grain weight and other yield related traits in rice. Simple sequence identification tool was applied to identify SSRs within the gene sequences, appropriate motif length and number of repeats was customized to minimum four bases with five repeats. The 10 μL final volume of the PCR reaction mixture was constituted of 1 μL of genomic DNA, 4 μL of premix, 1 μmol/L each of forward and reverse primers and 3 μL of nuclease free water. Amplification was done using a 384 well Thermocycler (Agilent technologies® Surecycler8800) by adopting the following PCR program. Initial template denaturation at 94 ℃ for 4 min followed by 40 cycles of amplification each with 40 s of denaturation at 94 ℃, 40 s of primer annealing (at appropriate Tm) and 1 min of elongation at 72 ℃, and 7 min of final extension. PCR amplified products of all genotypes were separated on a 3.5% agarose gel following a standard electrophoresis procedure. Gel documentation system (Zenith Gel.Pro9 CCD gel doc, Biozen Laboratories, India) was employed for gel image scanning and amplicons were phenotyped using CLIQS Gel image analysis software, version 1.0 from Totallab® by comparing each amplicon with a 50 bp DNA ladder.

Statistical analysis

Phenotype analysis

Observation of grain size traits recorded from each replication as replication mean over three years were considered for estimation of BLUP values. The BLUP estimates help to reduce mean squared error under multi-season evaluation trials by shrinking phenotypes over seasons (Hill and Rosenberger, 1985; Piepho et al, 2008). The BLUP values for genotypes across seasons were estimated using META-R software developed by CIMMYT (Alvarado et al, 2020). Only BLUP values estimated for each trait were considered for further analysis. To ensure the best suitability of the panel for association analysis, phenotypic distribution patterns and descriptive statistics were analyzed using RStudio version 1.4.17 (R Core Team, 2021). The correlation coefficients among traits were calculated following Pearson's correlation approach and plotted using the 'corrplot' package in R software (Wei and Simko, 2021).

Allele diversity and population structure analysis

Higher level of allele diversity, PIC of markers and appropriate population structure are most important for perfect association analysis to avoid false positives. The allelic diversity, allele frequency and PIC of markers on the GWAS panel were assessed using PowerMarker V3.25 (Liu and Muse, 2005). Population structure was estimated from genotypic data of 142 cgSSR markers using Bayesian model based software STRUCTURE 2.2 developed by Pritchard et al (2000). The length of the burn-in period and MCMC were set at 10 000 and 100 000, respectively. To identify the optimum sub-populations in the panel, an admixture ancestry model of an ad hoc statistic ΔK (Evanno et al, 2005) starting from K = 1 to K = 10 was applied with five replications in each K. By harvesting results from structure harvester (Earl and vonHoldt, 2012), the optimum value for K = 3 was determined, thus indicating the association panel could be divided into three sub-populations. Further, PCA was performed using the R package 'factoextra' (Kassambara and Mundt, 2017) to confirm the number of sub-populations. The familial relationship among individuals of the association panel was assessed using the VanRaden kinship algorithm (VanRaden, 2008) and the heat map of the kinship matrix was plotted using the GAPIT package of R software (Lipka et al, 2012).

Association analysis

Association analysis between BLUP estimates of four phenotypes and cgSSR marker genotype data on the association

panel was performed using the GAPIT package (Lipka et al, 2012), implemented in R software. GAPIT analyses the association between markers and traits while addressing population structure and kinship (Yu et al, 2006). To guarantee appropriate association, MLM method applying the EMMA algorithm (Kang et al, 2008) coupled with population structure adjustment was used. The marker $P < 0.05$ was considered to declare a significant association between marker and trait.

ACKNOWLEDGEMENT

Authors are greatly thankful to ICAR-National Rice Research Institute for financial support.

SUPPLEMENTAL DATA

The following materials are available in the online version of this article at http://www.sciencedirect.com/journal/rice-science; http://www.ricescience.org.

Fig. S1. Distribution of 142 candidate gene based SSRs on 12 rice chromosomes.

Table S1. List of genotypes used for association mapping study.

REFERENCES

Agrama H A, Eizenga G C, Yan W. 2007. Association mapping of yield and its components in rice cultivars. *Mol Breed*, **19**(4): 341–356.

Alqudah A M, Sallam A, Stephen Baenziger P, Börner A. 2020. GWAS: Fast-forwarding gene identification and characterization in temperate cereals: Lessons from barley: A review. *J Adv Res*, **22**: 119–135.

Alvarado G, Rodríguez F M, Pacheco A, Burgueño J, Crossa J, Vargas M, Pérez-Rodríguez P, Lopez-Cruz M A. 2020. META-R: A software to analyze data from multi-environment plant breeding trials. *Crop J*, **8**(5): 745–756.

Anandan A, Mahender A, Sah R P, Haque S, Pradhan S K, Roy P S, Singh O N, Ali J. 2021. Genetic diversity and population structure among an assorted group of genotypes pertinent to reproductive stage drought stress in rice (*Oryza sativa* L.). *Acta Sci Agric*, **5**(3): 77–89.

Atwell S, Huang Y S, Vilhjálmsson B J, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone A M, Hu T T, Jiang R, Muliyati N W, Zhang X, Amer M A, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker J R, Faure N, Kniskern J M, Jones J D, Michael T, Nemri A, Roux F, Salt D E, Tang C, Todesco M, Traw M B, Weigel D, Marjoram P, Borevitz J O, Bergelson J, Nordborg M. 2010. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature*, **465**: 627–631.

Azharudheen T P M, Nayak A K, Behera S, Anilkumar C, Marndi B C, Moharana D, Singh L K, Upadhyay S, Sah R P. 2022. Genome-wide association analysis for plant type characters and yield using cgSSR markers in rice (*Oryza sativa* L.). *Euphytica*, **218**(6): 1–13.

Bai X F, Luo L J, Yan W H, Kovi M R, Zhan W, Xing Y Z. 2010. Genetic dissection of rice grain shape using a recombinant inbred line population derived from two contrasting parents and fine mapping a pleiotropic quantitative trait locus *qGL7*. *BMC Genet*, **11**: 16.

Chakraborti M, Anilkumar C, Verma R L, Abdul Fiyaz R, Reshmi Raj K R, Patra B C, Balakrishnan D, Sarkar S, Mondal N P, Kar M K, Meher J, Sundaram R M, Rao L S. 2021. Rice breeding in India: Eight decades of journey towards enhancing the genetic gain for yield, nutritional quality, and commodity value. *Oryza*, **58**: 69–88.

Ching A, Caldwell K S, Jung M, Dolan M, Smith O S, Tingey S, Morgante M, Rafalski A J. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genet*, **3**: 19.

Cho Y G, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch S R, Park W D, Ayres N, Cartinhour S. 2000. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor Appl Genet*, **100**(5): 713–722.

Collard B C Y, Vera Cruz C M, McNally K L, Virk P S, MacKill D J. 2008. Rice molecular breeding laboratories in the genomics era: Current status and future considerations. *Int J Plant Genomics*, **2008**: 524847.

Dai X, You C, Chen G, Li X, Zhang Q, Wu C. 2011. *OsBC1L4* encodes a COBRA-like protein that affects cellulose synthesis in rice. *Plant Mol Biol*, **75**(4): 333–345.

Duan P G, Xu J S, Zeng D L, Zhang B L, Geng M F, Zhang G Z, Huang K, Huang L J, Xu R, Ge S, Qian Q, Li Y H. 2017. Natural variation in the promoter of *GSE5* contributes to grain size diversity in rice. *Mol Plant*, **10**(5): 685–694.

Earl D A, vonHoldt B M. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour*, **4**(2): 359–361.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol Ecol*, **14**(8): 2611–2620.

Evans L T. 1972. Storage Capacity as a Limitation on Grain Yield in Rice Breeding. Manila, the Philippines: International Rice Research Institute.

Fu F F, Xue H W. 2010. Coexpression analysis identifies Rice Starch Regulator1, a rice AP2/EREBP family transcription factor, as a novel rice starch biosynthesis regulator. *Plant Physiol*, **154**(2): 927–938.

Fu F H, Wang F, Huang W J, Peng H P, Wu Y Y, Huang D J. 1994. Genetic analysis on grain characters in hybrid rice. *Acta Agron Sin*, **20**(1): 39–45.

Furukawa T, Maekawa M, Oki T, Suda I, Iida S, Shimada H, Takamure I, Kadowaki K I. 2007. The *Rc* and *Rd* genes are involved in proanthocyanidin synthesis in rice pericarp. *Plant J*, **49**(1): 91–102.

Gao D W, Sun W Q, Wang D W, Dong H L, Zhang R, Yu S B. 2020. A xylan glucuronosyltransferase gene exhibits pleiotropic effects on cellular composition and leaf development in rice. *Sci Rep*, **10**(1): 3726.

Garris A J, Tai T H, Coburn J, Kresovich S, McCouch S. 2005. Genetic structure and diversity in *Oryza sativa* L. *Genetics*, **169**(3): 1631–1638.

Gobu R, Shiv A, Anilkumar C, Basavaraj P S, Harish D, Adhikari S, Vinita R, Umesh H, Sujatha M. 2020. Accelerated crop

breeding towards development of climate resilient varieties. Climate change and Indian Agriculture: Challenges and Adaptation Strategies. *In*: Rao C S, Srinivas T, Rao R V S, Rao N S, Vinayagam S S, Krishnan P. Climate Change and Indian Agriculture: Challenges and Adaptation Strategies, ICAR-National Academy of Agricultural Research Management, Hyderabad, Telangana, India: 49–69.

Guo X, Elston R C. 1999. Linkage information content of polymorphic genetic markers. *Hum Hered*, **49**(2): 112–118.

Hill Jr R R, Rosenberger J L. 1985. Methods for combining data from gemrplasm evaluation trials 1. *Crop Sci*, **25**(3): 467–470.

Hu X X, Wang C, Fu Y P, Liu Q, Jiao X Z, Wang K J. 2016. Expanding the range of CRISPR/Cas9 genome editing in rice. *Mol Plant*, **9**(6): 943–945.

Hu Z J, Lu S J, Wang M J, He H H, Sun L, Wang H R, Liu X H, Jiang L, Sun J L, Xin X Y, Kong W, Chu C C, Xue H W, Yang J S, Luo X J, Liu J X. 2018. A novel QTL *qTGW3* encodes the GSK3/SHAGGY-like kinase OsGSK5/OsSK41 that interacts with OsARF4 to negatively regulate grain size and weight in rice. *Mol Plant*, **11**(5): 736–749.

Huang R Y, Jiang L R, Zheng J S, Wang T S, Wang H C, Huang Y M, Hong Z L. 2013. Genetic bases of rice grain shape: So many genes, so little known. *Trends Plant Sci*, **18**(4): 218–226.

Huang X H, Han B. 2014. Natural variations and genome-wide association studies in crop plants. *Annu Rev Plant Biol*, **65**: 531–551.

Hussain K, Zhang Y X, Anley W, Riaz A, Abbas A, Rani M H, Wang H, Shen X H, Cao L Y, Cheng S H. 2020. Association mapping of quantitative trait loci for grain size in introgression line derived from *Oryza rufipogon*. *Rice Sci*, **27**(3): 246–254.

Jan R, Khan M A, Asaf S, Lee I J, Kim K M. 2020. Overexpression of *OsF_3H* modulates WBPH stress by alteration of phenylpropanoid pathway at a transcriptomic and metabolomic level in *Oryza sativa*. *Sci Rep*, **10**(1): 14685.

Kaneko K, Inomata T, Masui T, Koshu T, Umezawa Y, Itoh K, Pozueta-Romero J, Mitsui T. 2014. Nucleotide pyrophosphatase/ phosphodiesterase 1 exerts a negative effect on starch accumulation and growth in rice seedlings under high temperature and CO_2 concentration conditions. *Plant Cell Physiol*, **55**(2): 320–332.

Kang H M, Zaitlen N A, Wade C M, Kirby A, Heckerman D, Daly M J, Eskin E. 2008. Efficient control of population structure in model organism association mapping. *Genetics*, **178**(3): 1709–1723.

Kassambara A, Mundt F. 2017. Factoextra: Extract and visualize the results of multivariate data analyses. [2021-7-25]. https:// mirrors.sjtug.sjtu.edu.cn/cran/web/packages/factoextra/index.html.

Katara J L, Parameswaran C, Devanna B N, Verma R L, Anil Kumar C, Patra B C, Samantaray S. 2021. Genomics assisted breeding: The need and current perspective for rice improvement in India. *Oryza*, **58**: 61–68.

Korte A, Farlow A. 2013. The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods*, **9**: 29.

Lipka A E, Tian F, Wang Q S, Peiffer J, Li M, Bradbury P J, Gore M A, Buckler E S, Zhang Z W. 2012. GAPIT: Genome association and prediction integrated tool. *Bioinformatics*, **28**(18): 2397–2399.

Liu K J, Muse S V. 2005. PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics*, **21**(9): 2128–2129.

Lu H, Redus M A, Coburn J R, Rutger J N, McCouch S R, Tai T H. 2005. Population structure and breeding patterns of 145 US rice cultivars based on SSR marker analysis. *Crop Sci*, **45**(1): 66–76.

Ma X S, Feng F J, Zhang Y, Elesawi I E, Xu K, Li T F, Mei H W, Liu H Y, Gao N N, Chen C L, Luo L J, Yu S W. 2019. A novel rice grain size gene *OsSNB* was identified by genome-wide association study in natural population. *PLoS Genet*, **15**(5): e1008191.

Mather D E, Hyes P M, Chalmers K J, Eglinton J, Matus I, Richardson K, Von Zitzewitz J, Marquez-Cedillo L, Hearnden P, Pal N. 2004. Use of SSR marker data to study linkage disequilibrium and population structure in *Hordeum vulgare*: Prospects for association mapping in barley. *In*: Jaroslav S, Jarmila J. 9th International Barley Genetics Symposium. Brno, Czech Republic: International barley genetics symposium: 302–307.

Meng L J, Zhao X Q, Ponce K, Ye G Y, Leung H. 2016. QTL mapping for agronomic traits using multi-parent advanced generation inter-cross (MAGIC) populations derived from diverse elite *indica* rice lines. *Field Crops Res*, **189**: 19–42.

Mohanty S. 2013. Trends in global rice consumption. *Rice Today*, **12**: 44–45.

Molla K A, Debnath A B, Ganie S A, Mondal T K. 2015. Identification and analysis of novel salt responsive candidate gene based SSRs (cgSSRs) from rice (*Oryza sativa* L.). *BMC Plant Biol*, **15**: 122.

Molla K A, Azharudheen T P M, Ray S, Sarkar S, Swain A, Chakraborti M, Vijayan J, Singh O N, Baig M J, Mukherjee A K. 2019. Novel biotic stress responsive candidate gene based SSR (cgSSR) markers from rice. *Euphytica*, **215**(2): 17.

Nanjo Y, Oka H, Ikarashi N, Kaneko K, Kitajima A, Mitsui T, Muñoz F J, Rodríguez-López M, Baroja-Fernández E, Pozueta-Romero J. 2006. Rice plastidial *N*-glycosylated nucleotide pyrophosphatase/phosphodiesterase is transported from the ER-Golgi to the chloroplast through the secretory pathway. *Plant Cell*, **18**(10): 2582–2592.

Norton G J, Travis A J, Douglas A, Fairley S, Alves E D, Ruang-Areerate P, Naredo M, Elizabeth B, McNally K L, Hossain M, Islam M. 2018. Genome wide association mapping of grain and straw biomass traits in the rice Bengal and Assam Aus panel (BAAP) grown under alternate wetting and drying and permanently flooded irrigation. *Front Plant Sci*, **9**: 1223.

Pahlich E, Gerlitz C. 1980. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry*, **19**: 11–13.

Patra B C, Anilkumar C, Chakraborti M. 2020. Rice breeding in India: A journey from phenotype based pure-line selection to genomics assisted breeding. *Agric Res J*, **57**(6): 816–825.

Piepho H P, Möhring J, Melchinger A E, Büchse A. 2008. BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica*, **161**: 209–228.

Ponce K, Zhang Y, Guo L B, Leng Y J, Ye G Y. 2020. Genome-wide association study of grain size traits in *indica* rice multiparent advanced generation intercross (MAGIC) population. *Front Plant Sci*, **11**: 395.

Pritchard J K, Stephens M, Donnelly P. 2000. Inference of population

structure using multilocus genotype data. *Genetics*, **155**(2): 945–959.

Qiu X J, Pang Y L, Yuan Z H, Xing D Y, Xu J L, Dingkuhn M, Li Z K, Ye G Y. 2015. Genome-wide association study of grain appearance and milling quality in a worldwide collection of *indica* rice germplasm. *PLoS One*, **10**(12): e0145577.

R Core Team. 2021. R, A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. [2021-11-25].

Rafalski J A. 2010. Association genetics in crop improvement. *Curr Opin Plant Biol*, **13**(2): 174–180.

Rahman S N, Islam M S, Alam M S, Nasiruddin K M. 2007. Genetic polymorphism in rice (*Oryza sativa* L.) through RAPD analysis. *Indian J Biotechnol*, **6**: 224–229.

Raju B R, Mohankumar M V, Sumanth K K, Rajanna M P, Udayakumar M, Prasad T G, Sheshshayee M S. 2016. Discovery of QTLs for water mining and water use efficiency traits in rice under water-limited condition through association mapping. *Mol Breeding*, **36**(3): 35.

Sahu R K, Patnaik S S C, Sah R P. 2020. Quality seed production in rice. Odisha, India: ICAR-National Rice Research Institute: 58.

Sakamoto T, Ohnishi T, Fujioka S, Watanabe B, Mizutani M. 2012. Rice CYP90D2 and CYP90D3 catalyze C-23 hydroxylation of brassinosteroids *in vitro*. *Plant Physiol Biochem*, **58**: 220–226.

Sanghamitra P, Sah R P, Bagchi T B, Sharma S G, Kumar A, Munda S, Sahu R K. 2018. Evaluation of variability and environmental stability of grain quality and agronomic parameters of pigmented rice (*O. sativa* L.). *J Food Sci Technol*, **55**(3): 879–890.

Seo H, Kim S H, Lee B D, Lim J H, Lee S J, An G, Paek N C. 2020. The rice basic helix-loop-helix 79 (OsbHLH079) determines leaf angle and grain shape. *Int J Mol Sci*, **21**(6): 2090.

Shapiro S S, Wilk M B. 1965. An analysis of variance test for normality (complete samples). *Biometrika*, **52**: 591–611.

Song X W, Li P C, Zhai J X, Zhou M, Ma L J, Liu B, Jeong D H, Nakano M, Cao S Y, Liu C Y, Chu C C, Wang X J, Green P J, Meyers B C, Cao X F. 2012. Roles of DCL4 and DCL3b in rice phased small RNA biogenesis. *Plant J*, **69**(3): 462–474.

Tan Y F, Xing Y Z, Li J X, Yu S B, Xu C G, Zhang Q F. 2000. Genetic bases of appearance quality of rice grains in Shanyou 63, an elite rice hybrid. *Theor Appl Genet*, **101**: 823–829.

Upadhyaya G, Das A, Ray S. 2021. A rice R2R3-MYB (*OsC1*) transcriptional regulator improves oxidative stress tolerance by modulating anthocyanin biosynthesis. *Physiol Plant*, **173**(4): 2334–2349.

VanRaden P M. 2008. Efficient methods to compute genomic predictions. *J Dairy Sci*, **91**(11): 4414–4423.

Varshney R K, Graner A, Sorrells M E. 2005. Genic microsatellite markers in plants: Features and applications. *Trends Biotechnol*, **23**(1): 48–55.

Vieira M L C, Santini L, Diniz A L, de Freitas Munhoz C. 2016. Microsatellite markers: What they mean and why they are so useful. *Genet Mol Biol*, **39**(3): 312–328.

Wang C H, Yang Y L, Yuan X P, Xu Q, Feng Y, Yu H Y, Wang Y P, Wei X H. 2014. Genome-wide association study of blast resistance in *indica* rice. *BMC Plant Biol*, **14**: 311.

Wang Y H, Zheng Y M, Cai Q H, Liao C J, Mao X H, Xie H G, Zhu Y S, Lian L, Luo X, Xie H A, Zhang J F. 2016. Population structure and association analysis of yield and grain quality traits in hybrid rice primal parental lines. *Euphytica*, **212**(2): 261–273.

Wei T, Simko V. 2021. R package 'corrplot': Visualization of a Correlation Matrix. Version 0.88. https://github.com/taiyun/corrplot.

Wu J H, Feng F J, Lian X M, Teng X Y, Wei H B, Yu H H, Xie W B, Yan M, Fan P Q, Li Y, Ma X S, Liu H Y, Yu S B, Wang G W, Zhou F S, Luo L J, Mei H W. 2015. Genome-wide association study (GWAS) of mesocotyl elongation based on re-sequencing approach in rice. *BMC Plant Biol*, **15**: 218.

Xing Y Z, Zhang Q F. 2010. Genetic and molecular bases of rice yield. *Annu Rev Plant Biol*, **61**: 421–442.

Xu J L, Xue Q Z, Luo L J, Li Z K. 2002. Genetic dissection of grain weight and its related traits in rice (*Oryza sativa* L.). *Chin J Rice Sci*, **16**: 6–10. (in Chinese with English abstract)

Yu J M, Pressoir G, Briggs W H, Vroh Bi I, Yamasaki M, Doebley J F, McMullen M D, Gaut B S, Nielsen D M, Holland J B, Kresovich S, Buckler E S. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet*, **38**(2): 203–208.

Yu J P, Xiong H Y, Zhu X Y, Zhang H L, Li H H, Miao J L, Wang W S, Tang Z S, Zhang Z Y, Yao G X, Zhang Q, Pan Y H, Wang X, Rashid M A R, Li J J, Gao Y M, Li Z K, Yang W C, Fu X D, Li Z C. 2017. OsLG3 contributing to rice grain length and yield was mined by Ho-LAMap. *BMC Biol*, **15**(1): 28.

Zhang D L, Zhang H L, Qi Y W, Wang M X, Sun J L, Ding L, Li Z C. 2013. Genetic structure and eco-geographical differentiation of cultivated Hsien rice (*Oryza sativa* L. subsp. *indica*) in China revealed by microsatellites. *Chin Sci Bull*, **58**(3): 344–352.

Zhang P, Liu X D, Tong H H, Lu Y G, Li J Q. 2014. Association mapping for important agronomic traits in core collection of rice (*Oryza sativa* L.) with SSR markers. *PLoS One*, **9**(10): e111508.

Zhao D S, Li Q F, Zhang C Q, Zhang C, Yang Q Q, Pan L X, Ren X Y, Lu J, Gu M H, Liu Q Q. 2018. *GS9* acts as a transcriptional activator to regulate rice grain shape and appearance quality. *Nat Commun*, **9**(1): 1240.

Zhou Q Y, An H, Zhang Y, Shen F C. 2000. Study on heredity of morphological characters of rice grain. *J Southwest Agric Univ*, **22**(2): 102–104.

(Managing Editor: FANG Hongmin)