Indian J. Genet. Plant Breed., (2022) 82(1): 1-6

doi: 10.31742/IJGPB.82.1.1

ISSN: 0975-6906

www.isgpg.org

RESEARCH ARTICLE

Identification of genomic loci governing pericarp colour through GWAS in rice (*Oryza sativa* L.)

Nitika Rana[$], Surbhi Kumawat[$], Uma Maheshwar Singh[1], Vikas Kumar Singh[2], Rupesh Deshmukh, Tilak Raj Sharma[3] and Humira Sonah*

Abstract

Rice pericarp colour is one of the nutritional traits that is now gaining attention worldwide. In the present investigation, genome-wide association GWAS) was performed to identify loci governing pericarp colour in rice. A set of 1,349,269 SNPs and precise phenotyping across 325 diverse accessions of rice were used for the GWAS. The accessions belong to five rice isozyme classification groups viz., *indica, japonica, aromatic, aus,* and *admix*. The GWAS identified two significant loci *gPC5-1*and *gPC7-1* on chromosomes, 5 and 7, respectively, associated with the pericarp colour in rice. The SNPs on chromosome 7 co-localized with the functionally characterized *Os07g0211500* (*Rc* gene) known to control pericarp colour and *Os07g0214900* which is similar to the *Chalcone synthase 2(OsCHS2)* gene involved in flavonoid synthesis pathway. Linkage disequilibrium analysis across 0.25 Mbp upstream and downstream of these markers suggested three strong linkage blocks on chromosome 7. More interestingly, the novel locus identified on chromosome 5 *gPC5*-1 does not harbor any homolog of previously reported genes. Therefore, the locus can serve as a basis for identifying a new gene for rice pericarp colour. The results presented here will be helpful to understand the genetic regulation of pericarp colour and for genomic-assisted breeding in rice.

Keywords: GWAS, haplotype diversity, linkage disequilibrium, pericarp colour, rice

Introduction

Pericarp colour is an important nutritional quality trait in rice. The characteristic red/purple colour in rice grain is governed by its flavonoid, pro-anthocyanin, and anthocyanin content which has antioxidant, anti-cancer, and anti-inflammatory properties (Wang et al. 2020). The commercial varieties which are cultivated in the majority of the rice growing countries are white rice. The coloured black, purple and red rice varieties are grown in restricted areas or only as specialty rice as these possess immense health benefits. The unpolished brown rice with bran layer is higher in nutrient content than its counterpart white milled rice grains. Rice bran is also rich in dietary fiber content and is an important part of the functional foods spectrum of rice eating countries (Rana et al. 2019).

Previous studies have deciphered the genetic control of rice pigmentation to some extent (Abeysiriwardena and Gunasekara 2020). The genes *Rc* and *Rd* are known to affect the pro-anthocyanin content and hence the red pericarp colour in rice (Furukawa et al. 2007). Similarly, the purple pericarp colour is governed by *Pb* on chromosome 4 and *Pp* gene on chromosome 1 (Rahman et al. 2013). Both *Pb* and *Pp* genes are necessary for the phenotype of the purple pericarp. The absence of *Pb* gene results in white grains, whereas the presence of *Pb* gene complemented with the absence of *Pp* gene results in brown pigment in grains

Department of Agriculture Biotechnology, National Agri-Food Biotechnology Institute (NABI), Mohali 140306, Chandigarh, India;

[1]International Rice Research Institute, South Asia Regional Centre, Varanasi 221 106, Uttar Pradesh, India;

[2]International Rice Research Institute, South-Asia Hub, ICRISAT Campus, Patencheru 502 324, Hyderabad, Telengana, India;

[3]Department of Crop Science, Indian Council of Agriculture Research (ICAR), Krishi Bhavan, New Delhi 110 001, India;

[$]Present address: Department of Biotechnology, Panjab University, Chandigarh 160 014, India

Corresponding Author: Humira Sonah, Department of Agriculture Biotechnology, National Agri-Food Biotechnology Institute (NABI), Mohali 140 306, Chandigarh, India, E-Mail: biohuma@gmail.com; humira@nabi.res.in

How to cite this article: Rana N., Kumawat S., Singh U. M., Singh V. K., Deshmukh R., Sharma T. R. and Sonah H. 2022. Identification of genomic loci governing pericarp color through GWAS in rice (*Oryza sativa* L.). Indian J. Genet. Plant Breed., **82**(1): 1-6.

Source of support: Nil

Conflict of interest: None.

Received: Nov. 2021 **Revised:** Feb. 2022 **Accepted:** Feb. 2022

(Wang and Shu 2007). In addition, the sequence variations which affect pericarp colour have also been studied. A 14 bp deletion within the *Rc* gene has been related to a white pericarp colour (Furukawa et al. 2007). Similarly, two SNPs result in a loss of function mutation in gene *Rd* (Sweeney et al. 2007). Rice pericarp colour is known to have been affected by domestication events between wild and cultivated rice varieties. It is evident from the observation that most of the coloured rice varieties are either wild or weedy cultivars whereas the slightly pigmented or white rice accessions are primarily cultivated varieties (Cui et al. 2016).

The advances in high-throughput sequencing techniques have provided an impetus for genome-wide identification of loci governing grain quality-related traits in cereals such as rice. GWAS analysis on 1378 accessions belonging to *indica*, *japonica,* and *aus* subpopulations has enabled the identification of 66 SNPs associated with traits such as awn colour, seed shattering and seed coat colour in weedy rice varieties (Lin et al. 2021). GWAS has been used to identify 13 loci associated with pericarp colour across 297 rice varieties which included 46 weedy rice accessions (Wang et al. 2020). Similarly, haplotype analysis has been conducted for the exon 6 of *Rc* gene in weedy rice, which confirmed the 14-bp deletion in white and brown rice and the SNP (C/A) in exon 6 in some weedy rice varieties (Prathepha 2009).

GWAS has been used to identify novel loci and genes governing grain pericarp colour in 419 rice accessions of *indica* and *japonica* type landraces, which resulted in the identification of eight and seven QTLs in the whole and *indica* subpanel, respectively (Yang et al. 2018). A similar analysis in 421 purified homozygous varieties of *indica, japonica, aus* and *aromatic* type subgroups identified seven QTLs to be associated with pericarp colour. GWAS analysis has revealed convergent evolution during domestication, by utilizing 203 rice varieties for amylose content, seed length, and pericarp colour trait. As a result, a novel major-effect loci was identified for pericarp colour; which has been suggested to have undergone convergent evolution in *aus* type of rice varieties independent of other sub-varieties (Wang et al. 2016). Similarly, haplotype analysis has provided important evolutionary insights across rice classification groups. The SNP-based analysis for genes governing pericarp colour (*Rc*), grain size (*GS3*) and starch synthases has revealed that the *Aus* cultivars have earlier origins and share haplotypes with wild-type accessions (Singh et al. 2017). The same study showed varying phylogenetic results for *indica* and *japonica* accessions. Hence, the availability of resequencing data has helped the identification of loci associated with important nutritional quality-related traits in cultivated and weedy rice varieties. The present study aims to utilize whole genome resequencing data from a diverse set of rice accessions to identify the novel loci for pericarp colour.

Materials and methods

Selection of core plant accessions

The Core Hunter 3 tool (http://www.corehunter.org/index), was utilized to generate a core subset from ~3024 rice genotypes (De Beukelaer et al. 2018). The method of core selection used was Growers distance (100% weight), which intends to compute distances from phenotypes. The high diversity subset was obtained through average entry-to-nearest-entry distance measure, which was representative of the original data while including extreme data values as well. The intensity of the core subset as compared to the input data was set at 10.74%.

Phenotyping for pericarp colour

Rice grains for each genotype were dried and de-husked. Four grains were visualized for 325 rice varieties each under Digimizer image analysis software (https://www.medcalcsoftware.com). Grain colours were ranked for white and coloured. The latter included genotypes which were red, purple, and brown/speckled in colour.

Genotypic data

The Single Nucleotide Polymorphisms (SNPs)were retrieved from the SNP-Seek Database (Mansueto et al. 2017). The SNPs were filtered for minor allele frequency (MAF) at 0.05 and minimum site count using Trait Analysis by aAssociation, Evolution and Linkage (TASSEL v5.2.67) statistical software (Bradbury et al. 2007)

Population structure analysis, kinship, phylogenetic tree and Principal Component Analysis

Population Structure analysis was performed by using FastStructure (v1.0) (Raj et al. 2014). The K-values for structure analysis were varied between two and ten. The phylogenetic tree was constructed for using filtered SNPs from 325 rice varieties by using TASSEL v5.2.67 statistical software (Bradbury et al. 2007). The clustering method used for phylogenetic tree construction was 'Neighbor Joining' and the resulting Newick tree was visualized with the 'Archaeopteryx tree' tool of TASSEL v5.2.67. The kinship matrix was constructed by using Van Raden method within the Genome Association and Prediction Integrated Tool (GAPIT) package (version 3). Principal Component Analysis (PCA) was conducted by using GAPIT (Lipka et al. 2012).

GWAS and Linkage disequilibrium analysis for grain pericarp colour

The GWAS was performed using GAPIT and TASSEL v5.2.67 statistical software. Multidimensional scaling (MDS) output was used as a covariate for running Mixed Linear Model (MLM) to account for the population structure. The kinship matrix was constructed by using the 'Centered_IBS' kinship

method. The MLM analysis was conducted by using the model

$$y = Si + Q + K + e$$

where *Si* represents the marker data; *Q* stands for the population structure and *K* for the Kinship or the variance-covariance matrix among individuals. Herein *y* represents the observed phenotypes and *e* accounts for any residuals. MLM analysis was further conducted with no PCA, PCA 3, 4, and 5 and multiple_analysis = TRUE, model =c "GLM", "MLM", "MLMM", "FarmCPU", "Blink". Once, TASSEL was utilized with 'no compression' compression level and 'Variance Estimate Composition' as P3D estimates. The results were obtained and visualized using Manhattan and QQ-Plots.

Linkage disequilibrium (LD) analysis was conducted by using the software Haploview v4.2 (Barrett et al. 2005). LD analysis was performed for the candidate peak regions spanning 0.25 Mb upstream and downstream genomic regions of the most significantly associated SNPs obtained from the GWAS analysis. The analysis was undertaken for 3602 variants on chromosome 5 and 1807 variants on chromosome 7. Pair-wise comparison of markers greater than 500kb apart was ignored and individuals with more than 50% missing genotypes were excluded from the analysis. Further, the Hardy Weinberg *p-value* cutoff was set at 0.001 and the maximum Mendel errors at 1 for obtaining the LD plot. The LD results were visualized for D'/LOD values.

Haplotype network for Os07g0211500 (Rc) gene

Haplotypic diversity was analyzed by using the rice resequencing dataset for ~4000 accessions available at RiceVarMap2.0 (Zhao et al. 2015). Sequence variations were obtained for the gene's open reading frame (ORF). Functionally important SNPs which caused non-synonymous, splice region, frame shift, start lost, stop-gained or stop-gained variations were used for analysis. For haplotype analysis, the rice accessions were classified into nine categories including *Indica I, Indica II, Indica III, Indica Intermediate, Aus,* Temperate *Japonica,* Tropical *Japonica, Japonica Intermediate*, and *Intermediate* (including aromatic and rest of the accessions) based on the rice isozyme classification system.

Results

Population structure of diverse rice accessions

A total of 325 rice accessions were extracted as a core subset and these rice accessions from 44 countries belong to *indica*, *japonica*, *admix*, *aromatic*, and *aus* isozyme classification groups (Fig.1). Population structure analysis of 325 diverse rice accessions was performed using filtered 1,349,269 SNPs. The FastSTRUCTURE tool revealed distinct populations for K-value ranging between 2 and 10. The subpopulations viz. *indica, japonica, aromatic, aus* and *admix* were grouped separately (Fig. 2a). The phylogenetic tree using the 'Neighbor Joining' clustering method sub-grouped the 325 accessions into five groups. The *indica* type rice accessions clustered separately, *aus* sub-varieties grouped closer to *indica* genotypes in a separate category. However, the *aromatic* and *admix* type of varieties clustered along with *japonica* isozyme rice classification (Fig. 2b). The VanRaden kinship matrix (Fig. 2c) and the principal component analysis (Fig. 2d) showed the distribution of rice accessions indicated diversity among the accessions. The PCA plot showed that the first principal component explained 16% of the genetic variation while the second and third explained 5% and 4% of the observed variation (Fig. 2d).

Phenotypic characterization for pericarp colour

Phenotypic of 325 rice accessions showed variation in pericarp colour as white, brown, red, and black (Fig 3). A

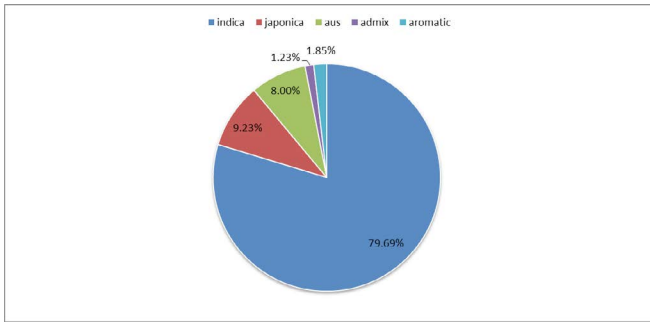

Fig. 1. Diversity of 325 rice accessions selected for genome-wide association study (GWAS) analysis. The set of accessions represents different rice groups like *Indica (Ind), Aus, Japonica (Jap),* **and** *Admixture (Admix)*

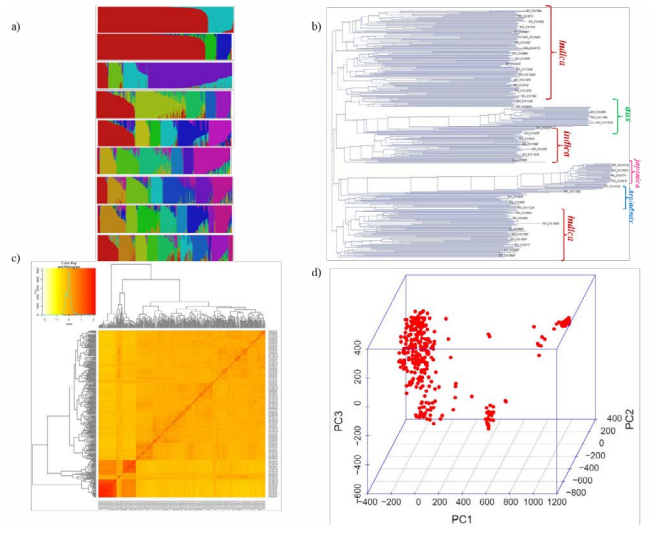

Fig. 2. Genetic diversity analysis of 325 core subset of rice. a) Population structure showing distinct populations for different values of K (between 2 and 10), b) Phylogenetic tree generated using 'neighbor joining' clustering method grouped *indica* **and** *japonica* **subpopulations separately. The** *indica* **subpopulation is labeled in maroon,** *aus* **in green,** *japonica* **in pink, and** *aromatic/admix* **in blue colour and, c-d) The kinship matrix and PCA indicate the relatedness and diversity among the rice genotypes**

total of 256 rice accessions were white and 69 rice accessions were coloured.

Genome-wide association analysis and linkage disequilibrium analysis for pericarp colour

GWAS analysis was performed using 1,349,269 MAF filtered SNPs to identify the markers associated with the pericarp colour. In rice, the most significant GWAS loci *gPC5-1* and *gPC7-1* were identified on chromosomes 5 and 7, respectively (Fig. 3). On chromosome 7, several SNPs above the threshold value were observed at genomic positions such as chr7_6128022 (*p-value*2.18E-07), chr7_6133017 (*p-value*4.80E-07), chr7_6379301(*p-value*3.18E-07) and chr7_6237231(*p-value* 5.43E-07) were found to be significantly associated with the pericarp colour (Fig. 4). The r^2 value computed for chr7_6128022 was 21.5%. While the SNP chr7_6128022 and chr7_6133017 were located intergenic between genes *Os07g0211900* and *Os07g0212200*, the SNP chr7_6379301 was positioned at downstream of*Os07g0216100*. Similarly, SNP chr7_6237231 was located intergenic between genes, *Os07g0213500* and *Os07g0213600*. Interestingly, the SNP marker chr7_6062503 with the *p-value 3.26e-05* is located 386 bp upstream of the *Rc* gene, which is known to govern the regulation of pro-anthocyanin pigmentation owing to a loss-of-function mutation in rice. Similarly, the SNP chr7_6128022 was located 58kb downstream of *Os07g0211500* (*Rc*) gene. It is noteworthy that gene *Rc* also showed a sequence polymorphism at genomic location chr7_6062909 (*p-value 0.000203*). Similarly, SNP chr7_6237231 was located 57248 bp upstream of *Os07g0214900* gene, which is similar to the *Chalcone synthase C2* (Naringenin-chalcone synthase C2) gene involved in the plant flavonoid biosynthetic pathways. The second GWAS loci *gPC5-1* was identified on chromosome 5 of rice with the significant peak of SNPs located at genomic position

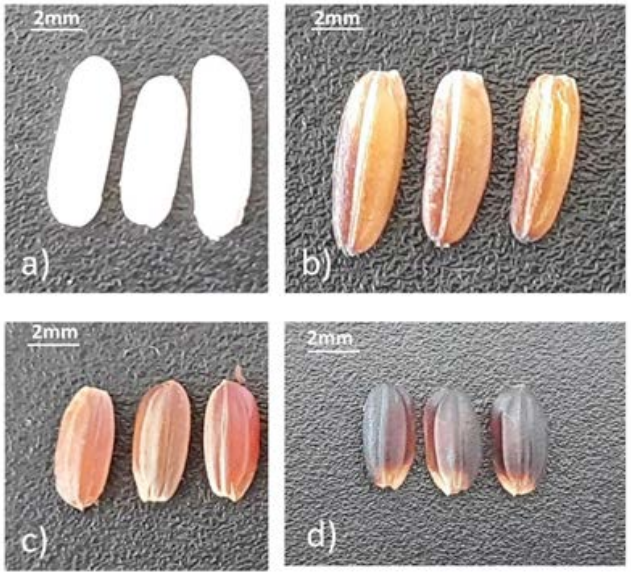

Fig. 3. Representative images of different rice genotypes with a) white b) brown c) red and d) black pericarp color.

chr5_14159818 with *p-value* of *3.90E-06*. A total of 86 genes were identified within the 1 Mb region of the significant SNP on chromosome 5 of rice.

Linkage disequilibrium analysis showed that several significantly associated SNPs on chromosome 7 were located within respective LD blocks. The SNPs chr7_6128022, chr7_6237231 and chr7_6379301 were located within separate LD blocks each. The SNP chr7_6062503 was localized within a single LD block along with chr7_6062909 SNP within the *Rc* gene. The location of the SNP marker at chr7_6062503 with *p-value 3.26e-05* is shown with a red dot in Fig. 5a. The SNP marker chr7_6062503 is located within the highly-linked block along with the *Rc* gene suggesting significant disequilibrium. The gene *Os07g0211500 (Rc)* is located within this block along with a single SNPs within this gene at genomic position 6062909 (Fig. 5b).

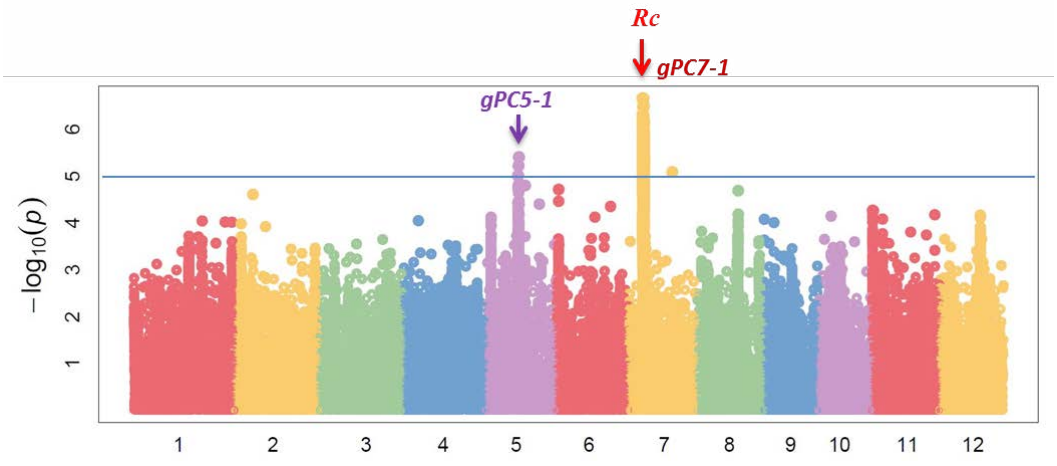

Fig. 4. Manhattan plot showing the SNPs significantly associated with pericarp color in rice. The x-axis denotes the genomic position and the y-axis is the negative log10 of *p*-values. Different chromosomes are marked by different colours. The Previously identified *gPC7-1* loci harboring *Rc* gene is marked in red colour and the new association signals identified on chromosomes 5 *gPC5-1* is shown in purple bold arrow

Haplotype network for Os07g0211500 (Rc) gene

The sequence variations such as SNPs and In/Dels within the *Rc* gene were retrieved from resequencing data at RiceVarMap2.0. As a result, we obtained 60 SNPs and 18 In/Dels. Among the 60 SNPs, sequence variation at chr7_ 6062909 resulted in a frame shift mutation. Similarly, the SNP at chr7_6063303

was a splice region variant, while chr7_6068017 introduced a stop gained mutation. In addition, four SNPs at positions chr7_6062952, chr7_6063359, chr7_6067391 and chr7_6067523 result in mis-sense mutations. Haplotype network for the functionally important SNPs within the *Os07g0211500* (*Rc*) gene across ~4,000 rice accessions revealed significant haplotypic diversity. The number of haplotypes obtained from seven functionally important SNPs at positions chr7_6062909, chr7_6062952, chr7_6063303, chr7_6063359, chr7_6067391, chr7_6067523, and chr7_6068017 was six. Among these, the highest frequency of accessions i.e., (2917) was observed in haplotype group I (TAGCAAC). The maximum accessions belonged to Temperate *Japonica* (765) followed by *Indica III* (634). In contrast, the lowest haplotypic frequency of 12 was observed for group VI (TAGCGAC).

Discussion

Compounds such as anthocyanins and pro-anthocyanins govern the pericarp colour in rice especially red and black. Several genes have been characterized which are responsible for the pericarp colour trait like *OsCHS1-2* (*chalcone synthases*), *OsCHI* (*chalcone isomerases*), *OsF3H, Rd* (*proanthocyanin synthesis),* and *OsANS* (*flavonoid pathway gene*) (Xia et al. 2021). A few regulatory transcription factors viz. *Rc* gene on chromosome 7 have also been identified (Furukawa et al. 2007). Similarly, the ectopic expression of *Kala4* gene resulted in the black pericarp colour of grain (Oikawa et al. 2015). Another gene *Rc* on chromosome 7 encodes a regulatory factor and dictates rice's purple, brown, white, and red colour (Furukawa et al. 2007). Our results are consistent with these observations as the primary SNP peak obtained in our study corresponds to the *Rc* gene on chromosome 7. In addition to a few structural genes, several regulatory transcription factors are yet to be identified and characterized for their function in pericarp colour in rice.

GWAS analysis for these diverse rice accessions revealed several significantly associated SNPs on chromosomes 7 and 5 for the pericarp colour. Out of the total SNPs on chromosome 7, chr7_6062503 was found to be in strong disequilibrium with gene *Os07g0211500 (Rc)* which is known to govern the pro-anthocyanin synthesis in rice. Sequence variation within the *Rc* gene at genomic position chr7_6062909 was located in LD block. Further, haplotypic diversity analysis for the *Rc* gene revealed the highest haplotypic frequency for group I (TAGCAAC), where most of the accessions belonged to the Temperate *japonica* group followed by *Indica III*. In the present study, we have identified another gene similar to Chalcone synthase C2 (Naringenin-*chalcone synthase C2*) at *gPC7-1* loci, which is the first dedicated step in flavonoid biosynthesis pathway. There is a possibility of having more than one causal gene at *gPC7-1*. The bi-parental or multi-parental populations

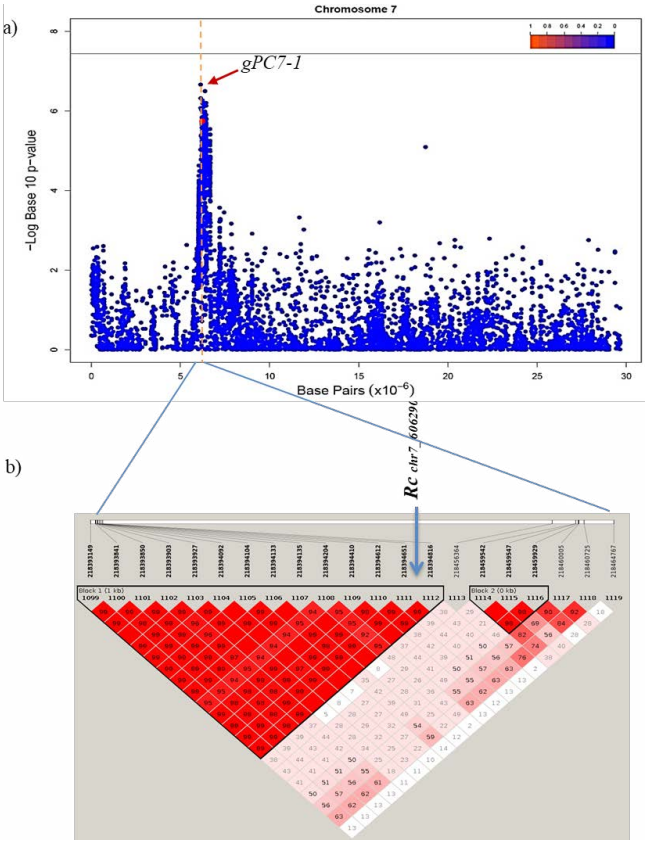

Fig. 5. Chromosome-wide Manhattan plot for significantly associated SNPs on chromosome 7. a) In the Manhattan plot, the x-axis represents the genomic location and the y-axis is the negative log10 of the *p-values* b) Linkage disequilibrium (LD) structure for the region containing the significant SNPs and the Rc gene. The *Rc gene* was observed within the same LD block

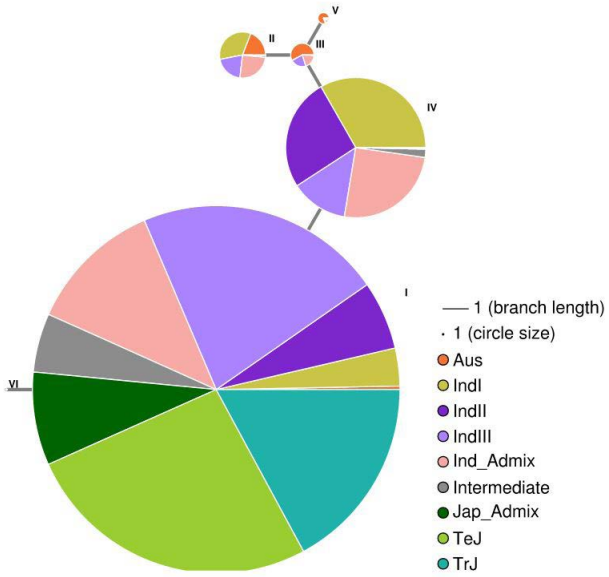

Fig. 6.: Haplotype network for gene *Os07g0211500* (*Rc*) gene which regulates the pro-anthocyanin pigmentation in rice grains. The size of circles suggests the haplotypic frequency and the rice isozyme subgroups are color-coded. The highest haplotypic frequency was observed for haplotype group I, while the lowest haplotype frequency was observed for group VI.

segregating different combinations of haplotypes at *gPC7-1* will be helpful to verify the possibility of another gene for instance, candidate gene *chalcone synthase C2*. In the case of the *gPC5-1* locus, about 86 predicted gene models are present at the significantly associated interval however, most of the genes encode for retrotransposons or proteins with unknown function. There is very limited scope for prioritizing any genes as candidates for further functional validation. Therefore, haplotypic information of the locus will be helpful to confirm the association in the breeding population.

Author's contributions

Conceptualization of research (HS); Designing of the experiments (HS); Contribution of experimental materials(NR); Execution of field/lab experiments and data collection (NR, SK); Analysis of data and interpretation (NR and SK); Preparation of the manuscript (NR and HS).

Acknowledgments

The authors are thankful to the Department of Biotechnology (DBT), Government of India (GoI) for the Ramalingaswami Fellowship Award to HS; and Grant BT/PR32853/AGIII/103/1159/2019 to HS, RD, and TRS; The Science and Engineering ResearchBoard (SERB), India, Department of Science and Technology (DST), Go, for JC Bose Fellowship to TRS, and Research grant CRG/2019/006599 awarded to HS, RD and TRS.

References

Abeysiriwardena D.S.d.Z. and Gunasekara D.C.S. 2020. Development of a red rice variety with excellent health properties and attractive grain qualities. Indian J. Genet. Plant Breed., **80**: 115-117.

Barrett J.C., Fry B., Maller J. and Daly M.J. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics, **21**: 263-265.

Bradbury P.J., Zhang Z., Kroon D.E., Casstevens T.M., Ramdoss Y. and Buckler E.S. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics, **23**: 2633-2635.

Cui Y., Song B.K., Li L.-F., Li Y.-L., Huang Z., Caicedo A.L., Jia Y. and Olsen K.M. 2016. Little white lies: pericarp colour provides insights into the origins and evolution of Southeast Asian weedy rice. G3: Genes, Genom, Genet., **6**: 4105-4114.

De Beukelaer H., Davenport G.F. and Fack V. 2018. Core Hunter 3: flexible core subset selection. BMC Bioinform., **19**: 1-12.

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

Lin Y.-L., Wu D.-H., Wu C.-C. and Huang Y.-F. 2021. Explore the genetics of weedy traits using rice 3K database. Bot. Stud., **62**: 1-16.

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

Mansueto L., Fuentes R.R., Borja F.N., Detras J., Abriol-Santos J.M., Chebotarov D., Sanciangco M., Palis K., Copetti D. and Poliakov A.J.N.a.r. 2017. Rice SNP-seek database update: new SNPs, indels, and queries. Nucleic Acids Res., **45**: D1075-D1081.

Oikawa T., Maeda H., Oguchi T., Yamaguchi T., Tanabe N., Ebana K., Yano M., Ebitani T. and Izawa T. 2015. The birth of a black rice gene and its local spread by introgression. The Plant Cell, **27**: 2401-2414.

Prathepha P. 2009. Pericarp colour and haplotype diversity in weedy rice (*O. sativa* f. spontanea) from Thailand. Pak.J.Biol. Sci., **12**: 1075-1079.

Rahman M.M., Lee K.E., Lee E.S., Matin M.N., Lee D.S., Yun J.S., Kim J.B. and Kang S.G. 2013. The genetic constitutions of complementary genes Pp and Pb determine the purple colour variation in pericarps with cyanidin-3-O-glucoside depositions in black rice. J. Plant Biol., **56**: 24-31.

Raj A., Stephens M. and Pritchard J.K. 2014. fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics, **197**: 573-589.

Rana N., Rahim M.S., Kaur G., Bansal R., Kumawat S., Roy J., Deshmukh R., Sonah H. and Sharma T.R. 2019. Applications and challenges for efficient exploration of omics interventions for the enhancement of nutritional quality in rice (*Oryza sativa* L.). Crit. Rev, Food Sci. Nutrit., 1-17. DOI: DOI: 10.1080/10408398.2019.1685454.

Singh N., Singh B., Rai V., Sidhu S., Singh A.K. and Singh N.K. 2017. Evolutionary insights based on SNP haplotypes of red pericarp, grain size and starch synthase genes in wild and cultivated rice. Front.Plant Sci., **8**: 972.

Sweeney M.T., Thomson M.J., Cho Y.G., Park Y.J., Williamson S.H., Bustamante C.D. and McCouch S.R. 2007. Global dissemination of a single mutation conferring white pericarp in rice. PLoS Genet., **3**: e133.

Wang C. and Shu Q. 2007. Fine mapping and candidate gene analysis of purple pericarp gene Pb in rice (*Oryza sativa* L.). Chin. Sci. Bull., **52**: 3097-3104.

Wang H., Xu X., Vieira F.G., Xiao Y., Li Z., Wang J., Nielsen R. and Chu C. 2016. The power of inbreeding: NGS-based GWAS of rice reveals convergent evolution during rice domestication. Mol. Plant, **9**: 975-985.

Wang W., Zhao M., Zhang G., Liu Z., Hua Y., Jia X., Song J., Ma D. and Sun J. 2020. Weedy rice as a novel gene resource: a genome-wide association study of anthocyanin biosynthesis and an evaluation of nutritional quality. Front. Plant Sci., **11**: 878.

Xia D., Zhou H., Wang Y., Li P., Fu P., Wu B. and He Y. 2021. How rice organs are coloured: the genetic basis of anthocyanin biosynthesis in rice. Crop J., **9**: 598-608.

Yang X., Xia X., Zeng Y., Nong B., Zhang Z., Wu Y., Xiong F., Zhang Y., Liang H. and Deng G. 2018. Identification of candidate genes for gelatinization temperature, gel consistency and pericarp colour by GWAS in rice based on SLAF-sequencing. PloS one **13**: e0196690.

Zhao H., Yao W., Ouyang Y., Yang W., Wang G., Lian X., Xing Y., Chen L. and Xie W.J.N.a.r. 2015. RiceVarMap: a comprehensive database of rice genomic variations. Nucleic Acids Res., **43**: D1018-D1022.