RESEARCH ARTICLE

Study on submergence tolerance of rice (*Oryza sativa* L*.*) in a core panel of North-East India using GWAS

Munmi Phukon, Janardan Das, R. Sruthi, Rahul K. Verma¹, Mahendra K. Modi², Ashok Bhattacharyya³ and Sanjay K. Chetia⁴*

Abstract

Rice (*Oryza sativa* L.) being the major food crop of the north-eastern region, combating floods in rice fields for higher economic yield is a major challenge. A core panel, consisting of local rice landraces that have been cultivated in the flood-prone areas of this region for years, was used for this GWAS study to uncover possible genetic resistance sources to submergence. A study on GWAS was conducted to understand the mechanism of resistance of rice under water-submerged conditions through higher expression of genes. The GWAS analysis of a core panel of 400 rice landraces generated 38,723 filtered SNPs. The result showcased nine loci across the 12 rice chromosomes, one locus each in chromosome 2 and 4, five loci in chromosome 6, and two loci in chromosome 9. The two promising loci among these nine identified loci codes for zinc fingers, C3HC4 type domain-containing proteins, with FDR adjusted *p-values* of 0.04 each and allele effect of 4.60 and 4.57, respectively. These GWAS-identified association signals are a valuable source for allele mining and can be validated and introgressed into elite germplasms to decipher submergence tolerance in future breeding programs.

Keywords: GWAS, submergence, SNPs identification, rice landraces

Introduction

Rice (*Oryza sativa* L.) requires a huge amount of water for its growth and development, but prolonged submergence causes a hypoxia environment, which affects the growth and productivity of rice plants. Approximately 15 mha area under rice cultivation is affected by flash floods in Asia alone (Sarkar et al. 2014). The change in the climate affects rainfall patterns and may result in heavy rain and causing heavy floods. About 4 lakh ha of *sali* paddy areas are affected due to floods in Assam (Bujarbaruah 2015). However, other than the natural occurrence of floods experienced in the rice growing season, intermittent flooding with various degrees of submergence is also observed for a duration ranging from 1 to 17 days (Neog et al. 2016). Understanding the tolerance mechanism of rice under submergence is an important requirement for managing flood stress in rice during the crop season. *Indica* and *japonica* ecotypes of rice confer flood tolerance by adapting to *SUB1A* gene-regulated process of quiescence. Deepwater and most of the lowland rice genotypes usually adapt different escape strategies to combat the stress. Ethylene-mediated SNORKEL1 and SNORKEL2 genes were identified, which can withstand submergence through internode elongation (Bailey-Serres et al. 2010; Nagai et al. 2010; Nishiuchi et al. 2012). The variation among the *bao* genotypes (deepwater rice) of Assam is due to elongation ability and kneeing ability as an adaptation mechanism for submergence tolerance (Neog et al. 2016). After the identification of a QTL for submergence tolerance SUBMERGENCE 1 (SUB1) from FR13A landrace, remarkable progress has been achieved in the last 10 years in the development of flood-tolerant varieties through marker-assisted backcrossing (Khalil et al. 2024).

Assam Rice Research Institute, Assam Agricultural University, Titabor, Assam 785 630, India

1 DBT-North-East Centre for Agricultural Biotechnology, Assam Agricultural University, Assam 785 013, India

2 Assam Down Town University, Guwahati 381 026, Assam, India.

3 Department of Plant Pathology, Assam Agricultural University, Assam 785 013, India

4 Directorate of Research (Agriculture), Assam Agricultural University, Assam 785 013, India

***Corresponding Author:** Sanjay K. Chetia, Directorate of Research (Agriculture), Assam Agricultural University, Assam 785 013, India, E-Mail: sanjaykumarchetia@gmail.com

How to cite this article: Phukon M., Das J., Sruthi R., Verma R.K., Modi M.K., Bhattacharyya A. and Chetia S.K. 2024. Study on submergence tolerance of rice in a core panel of North-East India using GWAS. Indian J. Genet. Plant Breed., **84**(2): 193-201.

Source of support: Department of Biotechnology, Government of India.

Conflict of interest: None.

Received: July 2023 **Revised:** April 2024 **Accepted:** May 2024

[©] The Author(s). 2024 Open Access This article is Published by the Indian Society of Genetics & Plant Breeding, NASC Complex, IARI P.O., Pusa Campus, New Delhi 110 012; Online management by www.isgpb.org

Being the secondary centre of origin, the areas in the North-East states of India is a hotspot of genetic variation and endowed with a huge diversity of rice germplasm due to the presence of various ethnolinguistic groups, variation in altitude and diverse climatic conditions (Sharma et al. 1971; Choudhury et at. 2013). The germplasms available in this region have been the sources of resistance against biotic stresses (Umakanth et al. 2017), abiotic stresses (Sharma et al. 1971; Patra et al. 2018; Verma et al. 2019) and nutrition (Vanlalsanga et al. 2019; Chattopadhyay et at. 2019). North-East India harbors diverse ecotypes of rice, namely, *sali* (coarse-grained, photosensitive), *lahi* (slender-grained), ahu (aus), *asra* (semi-deepwater rice), *bao* (deepwater rice), *bora* (waxy), *chakua* (intermediate amylase content) and aromatic (scented). Many wild species of rice are also found in the region (Choudhury et al. 2013). The rice germplasms being cultivated in this region have been selected based on its adaptation to the diverse topography and agroclimatic conditions of North-East India. The rice germplasms that do well in submergence conditions are categorized into three groups, namely, *sali, bao* and *asra*. The *sali* rice varieties are tall and are cultivated during *kharif* season under rainfed lowland conditions. The *bao* rice varieties are cultivated in deepwater areas and these varieties can escape submergence by elongating themselves with the strength of their kneeing ability as the water level rises. The *asra* rice varieties can withstand stagnant water conditions by elongating their internodes. With the available germplasm adapted to the submergence environment can be assured that these rice cultivars are excellent genetic reserves for tolerance to submergence.

Although various mechanisms of submergence tolerance have been uncovered among different germplasms, only SUB1 has been extensively used for introgression by the breeders and the SUB1 introgressed varieties have expressed tolerance for about 12 days under submergence (Nishiuchi et al. 2012). However, practically in the farmers' field, these varieties didn't perform well in varying degrees of flood duration (Neog et al. 2016). This happened due to the short height of almost all *Sub1* introgressed genotypes, which were not suitable for lowland areas with water levels of more than 100 cm and flood exits for more than 21 days. This made the researchers discover other genes and genotypes showing tolerance to submergence from the early vegetative stage to the panicle initiation stage, which could withstand intermittent flooding ranging from 1 to 21 days (Chattopadhyay et al. 2018).

Genome-wide association study (GWAS) is a powerful technique to determine genomic regions by linking phenotypic and genotypic information using a diverse set of germplasms. GWAS is useful in the identification of QTLs for specific traits in rice, such as for agronomic traits (Huang et al. 2010) and anaerobic germination of deepwater rice

(Rohila et al. 2020). The GWAS study also has the potential to understand the mechanism of resistance of rice under water-submerged conditions through higher expression of certain genes or due to specific QTL effects. Considering the above factors, the GWAS approach was used in the present study to dissect the genetic architecture and mechanisms for tolerance to submergence using a diverse rice panel to determine the genetic elements underlying the response.

Materials and methods

Plant material and experimental layout

The rice plants were grown in an experimental area at Assam Rice Research Institute, Assam Agricultural University (AAU-ARRI), Titabor, Assam, during *kharif* season in the year 2021. A panel of 400 landraces selected from over 6,000 available germplasm at Assam Rice Research Institute, Assam Agricultural University, Titabor was grown in a concrete tank for submergence treatment. These landraces consisted of 12 different groups, namely, aromatic, *asra*, *bao*, black rice, *bora*, new *bora*, *chakua*, *joha*, *lahi*, *khamti lahi*, *sali* and new *sali* (Supplementary Table S1). Augmentedrandomised complete block design was used for designing the experiment. Each germplasm was planted with seven plant hills per row, in a total of 5 rows, 35 plants per block, and with a planting density of 20 cm x 30 cm (plant x row). The submergence tank used was 2.5 m deep. The protocol described by Xu et al. (2000) was followed for screening for submergence survival percentage of the plants. About 30 days old seedlings were submerged for a continuous ten days and de-submerged. The de-submerged plants were kept undisturbed for ten days for full recovery from submergence shock. The survival percentage after recovery was recorded visually based on the rice submergence tolerance standard evaluation system (SES). The total number of lines before submergence and after de-submergence were recorded and survival was calculated in percentage.

Sampling and genome sequencing

For leaf sampling, the youngest mature leaf was collected from one-month-old plants grown in a tank before submergence in the morning and kept in an ice box during the sampling time. The genomic DNA was isolated using the CTAB method and sequencing was performed using V4 sequencing chemistry in Illumina HiSeq™ X10. The *Oryza sativa* reference genome which was downloaded from Ensembl Genomes (http://ftp.ensemblgenomes.org/ pub/ plants/release-51/fasta/oryza_sativa/). The sequence reads were mapped to the downloaded reference genome using the MEM algorithm version 0.7.5 of BWA. The mapped reads were realigned using RealignerTargetCreator and IndelRealigner inversion v3.6 of GATK (https://gatk. broadinstitute.org/hc/en-us). The vcftoolsversion0.1.17 was used for filtering the variants and removal of indels (http:// vcftools.sourceforge.net/). A minor allele frequency of 5% MAF and a missing rate of 10% were applied to filter the SNPs to target a total of 39,045 SNPs, with a minimum of 90% of samples having that particular SNPs. A total of 38,723 SNPs were mapped in the 12 rice chromosomes across the 399 samples and were further used for downstream analysis. BAM format was used in all the mapping results, and SAMtools version 0.1.18 was used to filter out the nonunique and unmapped reads.

Diversity study, population structure and PCA

The SniPlay (https://sniplay.southgreen.fr/cgi-bin/home.cgi) was used for the diversity study. With a 200 kb window in the genome, the transition-transversion ratio was calculated and plotted in VCFtools (Danecek et al. 2011). SNP density was plotted within a range of 250 kb using SNIPlay (Dereeper et al. 2011). LD plotting was done using Tomahawk software (Klarqvist 2018). LD plots were illustrated in ggplot of R. Structuring was plotted in PGDSpider, version 2.1.1.5 (Lischer and Excoffier 2018) using the 1,472 filtered SNPs with no missing data within 250 kb and filtered with the best SNP with the highest DP. STRUCTURE (version 2.3.4) was used to estimate population structure using Bayesian Markov Chain Monte Carlo model (MCMC). Each population (k) set 1-10 was run three times. The burn-in time was set to 100000 and MCMC replication number was set to 300000 for each run. Structure Harvester was used to determine the most probable K-value. PLINK (version 1.9) was used to calculate the PCA and the plot was made in R.

Genome-wide association studies

GAPIT (version 3) within R was used in the compressed mixed linear model (CMLM) to perform a genome-wide association study (GWAS) and genome prediction. Unified mixed model, EMMA, CMLM and P3D/ EMMAx were used to study statistical genetics. An association study was performed using 38,723 chromosomal SNPs by CMLM method along with kinship and population file. The significance of the SNPs was considered with a cut-off of log_{10} [P] <1e-4 threshold.

The qqman within R was used for the Manhattan plots and quantile-quantile (Q-Q) plots. A *P*-value along with a false discovery rate (FDR) was set to *P* < 0.05 to reduce any false positives and to keep the higher associations. The threshold level was set based on Q-Q plots and chromosomal positions of known loci using Bonferroni correction for submergence resistance. The polymorphism within 50kb region, flanking the highly associated SNPs with the target trait was included to identify the SNPs in intergenic regions or gene promoters (putative). Variations within the flanking region of the coding sequence were considered to keep the SNPs in open reading frames (ORFs). Nucleotide diversity was calculated and deviations from neutral equilibrium (Tajima's *D*) were studied using DnaSP (v5.10.01) (Rozas 2009) to look for the significant signals for each candidate gene.

Results

Phenotyping and GWAS overview

The plants in the submergence tank were screened for recording the degree of tolerance under submerged conditions. The submergence survival percentage of the plants was with a minimum of 10% and a maximum of 100% (Supplementary Table S1). A GWAS analysis was performed for the plants to uncover the tolerance factors in the genetic sequence. The analysis has extracted 38,723 filtered SNPs. SNPs density across the 12 chromosomes ranged from 2,252 to 4,332 SNPs with the lowest in chromosome 12 and the highest in chromosome 1 (Fig. 1). Highest number of significant SNPs was mapped in chromosome 6 (5 SNPs).

SNP identification

A number of SNPs significantly associated with submergence tolerance were detected to be nine, with 1 SNP in chromosome 2, 1 SNP in chromosome 4, 5 SNPs in chromosome 6, and 2 SNPs in chromosome 9. The SNPs have been studied for their functional role under submergence conditions. The SNP mapped to chromosome 2 is annotated as retrotransposon protein (SNP ID 2:9208802, Accession LOC_Os02g16200); in chromosome 4 as ZOS4-03 - C2H2 zinc finger protein (SNP ID 4:4296750, Accession LOC_Os04g08060); in chromosome 6 as zinc finger (SNP ID 6:20412681, Accession LOC_Os07g06560), zinc finger (SNP ID 6:20412682, Accession LOC_Os07g06560), MYB family transcription factor (SNP ID 6:20454821, Accession LOC_Os06g35140), MYB family transcription factor (SNP ID 6:20455307, Accession LOC_Os06g35140) and exostosin family protein (SNP ID 6:28365560, Accession LOC_Os06g46690); in chromosome 9 as OsFBX315 - F-box domain containing protein (SNP ID 9:9511779, Accession LOC_Os09g15570) and proteasome/ cyclosome repeat containing protein (SNP ID 9:9620587, Accession LOC_Os09g15750) (Table 1). A summary of the association mapping for submergence with data for allelic variations, minor allele frequencies (ranged 0.09–0.35), *p-*values (ranged 1.34E-05–9.58E-06), FDR Adjusted P-values $(0.04 - 0.30)$, effect $(-6.01 - 6.73)$ and R² $(0.24 - 0.25)$ are presented in Table 2.

Population structure and LD estimation

The population structure of the 400 germplasms have been analyzed using PCA and the result showed three distinct sub-groups, with group 1 containing aromatic, *joha* and *sali*; group 2 containing *asra, bao* and new *bora*; group 3 containing aromatic, *joha*, black rice, *sali*, new *sali*, *asra*, *bao*, *bora*, new *bora*, *chakua*, *lahi* and *khamti lahi*. A decrease in the LD value from its maximum has been observed within 120-350 kb across the twelve chromosomes (Fig. 2). In the present study, the pairwise squared correlation coefficient (r2) decreased to half of its maximum value, with LD decay value of 0.6 in Chromosome 1, 0.62 in Chromosome 2, 0.62 in Chromosome 3, 0.63 in Chromosome 4, 0.6 in Chromosome

Fig. 1. SNP density variation across the twelve chromosomes. 'SNPs (1)' is SNPs in chromosome

5, 0.6 in Chromosome 6, 0.58 in Chromosome 7, 0.64 in Chromosome 8, 0.59 in Chromosome 9, 0.58 in Chromosome 10, 0.6 in Chromosome 11 and 0.63 in Chromosome 12.

GWAS for submergence

To analyse genome-wise association signals for submergence tolerance, 3,808,730 SNPs were used and analysed through the SUPER method of GAPIT package within R. Ten highly

significant association signals were found which were with *p-*value of < 0.05 (Table 2). The detected loci were further studied for their gene annotation to understand their role in the physiological pathways to combat submergence. The nine candidate SNPs detected in the GWAS analysis for submergence tolerance were retrotransposon protein (LOC_Os02g16200, chromosome 2), ZOS4-03–C2H2 zinc finger protein (LOC_Os04g08060, chromosome 4), zinc $\overline{1}$

 $\overline{1}$

 \mathbf{I}

finger (LOC_Os07g06560, chromosome 6), zinc finger (LOC_ Os07g06560, chromosome 6), MYB family transcription factor (LOC_Os06g35140, chromosome 6), MYB family transcription factor (LOC_Os06g35140, chromosome 6), exostosin family protein (LOC_Os06g46690, chromosome 6), OsFBX315-F-box domain-containing protein (LOC_ Os09g15570, chromosome 9), proteasome/cyclosome repeat-containing protein (LOC_Os09g15750, chromosome 9) (Figs. 3a and 3b).

Discussion

Practical approaches through deep molecular studies such as GWAS and QTL identification are the ways to answer the challenges due to climate change (Oladosu et al. 2020). Abiotic stresses have always reduced plant yield worldwide. The plants have been fighting to grow well under adverse conditions and the need for adaptation has made a way to evolve new mechanisms to grow and develop under stressed environments. A large set of transcriptional factors and stress-responsive proteins are extensively involved in response mechanisms to stresses such as low temperature, high temperature, salinity, high light, drought, and submergence.

Based on the morpho-physiological and grain quality traits such as growth traits, yield traits, and biotic and abiotic resistance traits, the best 400 rice landraces were selected from over 6,000 landraces. This new panel of 400 landraces was used in the present GWAS analysis. Several studies on agronomically important traits in rice have reported similar findings previously (Huang et al*.* 2015; Zhang et al. 2019). In the present analysis, the SNP density was found to be higher in chromosome 1 and chromosome 8 among all the chromosomes. However, more significant SNPs were found in chromosome 6 and chromosome 9 (Fig. 1). Population structure analysis showed three sub-groups within the population. Subgroup 1 has scented aromatic and *joha,* and coarse-grained kharif *sali*; subgroup 2 has semi-deep water *asra,* deep water *bao* and glutinous new *bora*; and subgroup 3 has scented aromatic and *joha*, pigmented black rice, coarse-grained kharif *sali* and new *sali*, semi-deep water *asra*, deep water bao, glutinous bora and new bora, semiglutinous chakua, and fine-grained kharif *lahi* and *khamti lahi*. In GWAS, LD estimation is crucial in understanding the identifying genetic markers because the quality of the GWAS analysis using SNPs also relies on LD while studying population genetics (Joiret et al. 2019). In the present study, LD has decreased (Fig. 2) and the distant relatedness has increased among the landraces. The results of population structure analysis and LD decay indicate that the core panel of 400 landraces used in the present study is suitable for studies on association mapping studies.

This GWAS study was analyzed using 3,808,730 SNPs to identify any new submergence-tolerant loci using the panel of 400 rice accessions under submergence conditions.

Fig. 2. LD heatmap for twelve chromosomes

The analysis employed a mixed linear model. The highly significant loci that are associated with the submergence trait were selected with a *p-value* cut-off of 0.05. The SNPs positioned within a length of 350 kb of a gene are considered close to that particular gene (Zhang et al*.* 2019). The nine SNPs found to be associated with submergence tolerance were mapped across twelve chromosomes (Fig. 3; Table 2). The loci uncovered in the present investigation, LOC_Os02g16200, LOC_Os04g08060, LOC_Os07g06560, LOC_Os07g06560, LOC_Os06g35140, LOC_Os06g35140, LOC_Os06g46690, LOC_Os09g15570 and LOC_Os09g15750, are believed to be involved in submergence tolerance mechanism in rice (Table 1).

The identified SNP LOC_Os02g16200 (SNP ID 2:9208802, Table 1), which annotates Ty3-gypsy subclass retrotransposon protein showed significantly high expression under submergence, indicating its involvement under the stress. Retrotransposons are transposable elements that are abundant in plant genomes and they gets activated under environmental stimuli such as microbial excretion, pathogen infection, physical injury, and other environmental stresses (Galindo-González et al. 2017). The LOC_Os04g08060 with SNP ID 4:4296750 has been predicted to be a ZOS4-03- C2H2 zinc finger protein (Table 1) which belongs to C2H2 family. This protein is involved in adaptive pathways to withstand submergence stress directly or indirectly (Wang et al. 2019). The zinc finger, C3HC4 type domain-containing proteins (SNP ID 6:20412681, Accession LOC_Os07g06560; SNP ID 6:20412682, Accession LOC_Os07g06560; Table 1) helps in plant growth and development and might have a role in submergence which is yet to be validated. MYB family has four groups of transcription factors (1R-MYB, 2R-MYB, 3R-MYB and 4R-MYB) that play a role in the overall development of plants, signal transduction, secondary metabolic pathways, biotic and abiotic stress (Katiyar et al. 2012). In the present study, two highly significant loci were found (SNP ID 6:20454821, Accession LOC_Os06g35140; SNP ID 6:20455307, Accession LOC_Os06g35140; Table 1), which encodes MYB family transcription factor, suggesting that this protein might has a role in submergence tolerance mechanism. An identified SNP LOC_Os06g46690 in this study (SNP ID 6:28365560) has been annotated exostosin family protein. Exostosin protein interacts with and influences growth molecules, morphogens and proteases and thereby regulates cell to cell crosstalk events (Busse-Wicher et al. 2014). The exact function of Exostosin proteins not yet known, and the role of this important locus in rice under submergence are yet to be validated. The identification

Fig. 3. Manhattan plot of EMMAX for submergence survival in genome-wide association studies. Negative log 10 p-values from a genome-wide scan are plotted against position on each of 12 chromosomes. (a) Manhattan plots of EMMAX (MLM-SUPER model). Black horizontal solid line indicates the genome-wide significant threshold; (b) Quantile-quantile (Q-Q) plot of the EMMAX (MLM-SUPER model)

of OsFBX315 - F-box domain-containing protein (SNP ID 9:9511779, Accession LOC_Os09g15570; Table 1) in the present GWAS study suggests that this protein might have a strong role in submergence tolerance in rice. The F-box encoded proteins are involved in developmental pathways in plants such as floral development, signal transduction through hormones, secondary metabolic pathways, biological clock, and biotic and abiotic stresses (Zhang et al. 2019). Proteasome/cyclosome repeat-containing protein (SNP ID 9:9620587, Accession LOC_Os09g15750; Table 1) might be connected to the maintenance and degradation of cells during submergence. In eukaryotes, proteasomes regulate the degradation of intracellular proteins (Sorokin et al. 2009). Cyclosome or APC/C responds to proteindamaging stress (Ahlskog et al. 2010). The functions of proteasome and cyclosome under submergence are yet to be confirmed.

The escape mechanism in rice under a submerged environment includes stem elongation, modification in plant architecture, internal aeration, metabolism and control in growth (Oladosu et al. 2020). The discovery of SUB1 gene has facilitated faster development of flood-tolerant varieties. The SUB1 rice varieties cannot respond to biotic and abiotic stresses simultaneously (Sharma et al. 2013). Breeding efforts have been made in recent years to combat this challenge. A simple trait governed by a single gene can be handled effectively through marker-assisted backcrossing; it would not be possible for a complex trait. A diverse set of landraces that responds well to flood and deep water can be a good set of germplasms for gene discovery other than SUB1. The landraces expressing genes for anaerobic germination will be a good source for yield stability in areas with intermittent and stagnant floods. This kind of breeding strategy can support the development of varieties for cultivation in the rainfed rice ecosystem through higher yield by preventing loss during heavy rainfall season (Oladosu et al. 2020). Identification of nine loci given in Table 1, which are highly significant under a submerged environment, can be a valuable source for submergence tolerance in rice. Among nine, two promising loci are LOC_Os07g06560 and LOC_Os07g06560 in chromosome 6. Validation of the function of these loci in the three landraces, namely Kheron (90% submergence survival percentage), Kala birain (90%) and Dhupa Bao (100%), showing maximum tolerance along with appropriate breeding efforts, can lead to new variety development with submergence tolerance as well as increase yield.

The present study used GWAS to understand and explore genic regions significantly expressed under submergence. A total of nine strong genome-wide association signals were identified across the twelve rice chromosomes in a set of 400 landraces, including observed submergence tolerance in Kheron (90% submergence survival percentage), Kala birain (90%) and Dhupa Bao (100%). Screening for new tolerant loci other than the well-acquainted SUB1 gene treated with 10 days of submergence is an important achievement. These identified loci can be tested for different durations of flood ranging from 10 to 21 days of extended submergence. Among the nine loci, the two promising loci are LOC_Os07g06560 and LOC_Os07g06560 in chromosome 6 which encodes zinc fingers, C3HC4 type domaincontaining proteins. The identified loci can be validated and utilized through marker-assisted backcrossing with appropriate breeding strategies for submergence-tolerant variety development.

Supplementary material

Supplementary Table S1 contains the list of 400 landraces used in the GWAS, at www.isgpb.org

Authors' contribution

Conceptualization of research (SKC, MP); Designing of the experiments (SKC, MP); Contribution of experimental materials (SKC); Execution of field/lab experiments and data collection (MP, JD, PB, SR); Analysis of data and interpretation (MP, SKC); Preparation of the manuscript (MP, JD, PB,SKC, MKM, AB, RKV, SR).

Acknowledgments

This research was funded by the Department of Biotechnology, Government of India. We thank Assam Agricultural University for providing the necessary infrastructure to carry out this research work.

References

- Ahlskog J.K., Björk J.K., Elsing A.N., Aspelin C., Kallio M., Roos-Mattjus P. and Sistonen L. 2010. Anaphase-promoting complex/cyclosome participates in the acute response to protein-damaging stress. Mol. Cell Biol., **30**(24): 5608-20.
- Bailey-Serres J., Fukao T., Ronald P., Ismail A., Heuer S. and Mackill D. 2010. Submergence tolerant rice: SUB1's journey from landrace to modern cultivar. Rice, **3**: 138–147.
- Bujarbaruah K.M. 2015. Agriculture in Assam-A path forward. In Agriculture in Assam. Ed Bhattacharyya A. pp 1-21.
- Busse-Wicher M., Wicher K.B. and Kusche-Gullberg M. 2014. The exostosin family: proteins with many functions. Matrix Biol., **35**: 25-33.
- Chattopadhyay K., Reddy J., Pradhan S.K., Patnaik S., Marndi B., Swain P., Nayak A., Anandan A., Chakraborty K., Sarkar R.K., Bose L., Katara J., Parameswaram C., Mukherjee A., Mohapatra S., Poonam A., Mishra S. and Korada R.R. 2018. Genetic improvement of rice for multiple stress tolerance in unfavorable rainfed ecology. In: Rice Research for Enhancing Productivity, Profitability and Climate Resilience. (Eds: Pathak H et al.): pp 120-137.
- Chattopadhyay K., Sharma S.G., Bagchi T.B. and Mohapatra T. 2019. CR Dhan 310 and CR Dhan 311 high protein, nutrientrich rice varieties of NRRI. https://icar-nrri.in/wp-content/ uploads/2019/06/2.-leaflet_highprotein_final.pdf, accessed on 30th August 2022.
- Choudhury B., Khan M.L. and Dayanandan S. 2013. Genetic structure and diversity of indigenous rice (*Oryza sativa*) varieties in the Eastern Himalayan region of Northeast India. Springer Plus, **2**: 228.
- Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E., Lunter G., Marth G.T., Sherry S.T., McVean G. and Durbin R. 2011. 1000 Genomes Project Analysis Group, The variant call format and VCFtools. Bioinformatics, **27**(15): 2156–2158.
- Dereeper A., Nicolas S., Le Cunff L., Bacilieri R., Doligez A., Peros J.P., Ruiz M. and This P. 2011. SNiPlay: a web-based tool for detection, management and analysis of SNPs. Application to grapevine diversity projects. BMC Bioinform., **12**: 134.
- Galindo-González L., Mhiri C., Deyholos M.K. and Grandbastien M.A. 2017. LTR-retrotransposons in plants: Engines of evolution. Gene, **626**: 14-25.
- Huang X., Wei X., Sang T., Zhao Q., Feng Q., Zhao Y., Li C., Zhu C., Lu T., Zhang Z., Li M., Fan D., Guo Y., Wang A., Wang L., Deng L., Li W., Lu Y., Weng Q., Liu K., Huang T., Zhou T., Jing Y., Li W., Lin Z., Buckler E.S., Qian Q., Zhang Q.F., Li J. and Han B. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Genet., **42**(11): 961-969.
- Huang X., Yang S., Gong J., Zhao Y., Feng Q., Gong H., Li W., Zhan Q., Cheng B., Xia J., Chen N., Hao Z., Liu K., Zhu C., Huang T., Zhao Q., Zhang L., Fan D., Zhou C., Lu Y., Weng Q., Wang

Z.X., Li J. and Han B. 2015. Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. Nat. Commun., **6**: 6258.

- Joiret M., Mahachie John J.M., Gusareva E.S. and Steen K.V. 2019. Confounding of linkage disequilibrium patterns in large scale DNA based gene-gene interaction studies. BioData Mining, **12**: 11.
- Katiyar A., Smita S., Lenka S.K., Rajwanshi R., Chinnusamy V. and Bansal K.C. 2012. Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis. BMC Genomics, **3**: 544.
- Khalil M.I., Hassan M.M., Samanta S.C., Chowdhury A.K., Hassan M.Z., Ahmed N.U., Somaddar U., Ghosal S., Robin A.H.K., Nath U.K., Mostofa M.G., Burritt D.J., Ha C.V., Gupta A., Tran L.P. and Saha G. 2024. Unraveling the genetic enigma of rice submergence tolerance: Shedding light on the role of ethylene response factor-encoding gene SUB1A-1. Plant Physiol. Biochem., **206**: 108224.
- Klarqvist M.D.R. 2018. Tomahawk: Fast calculation of LD in largescale cohorts. https://github.com/mklarqvist/tomahawk/ blob/master/README.md
- Lischer H. and Excoffier L. 2018. PGDSpider version 2.1.1.5: An automated data conversion tool for connecting population genetics and genomics programs. http://cmpg.unibe.ch/ software/PGDSpider/
- Nagai K., Hattori Y. and Ashikari M. 2010. Stunt or elongate? Two opposite strategies by which rice adapts to floods. J. Plant Res., **123**: 303-309.
- Neog P., Sarma P.K., Chary G.R., Dutta S., Rajbongshi R., Sarmah K., Baruah S., Sarma D., Sarma M., Borah P., Rao C.S. and Hazarika G. 2016. Building climate resilient agriculture through traditional floating rice in flash flood affected areas of the North bank plains zone of Assam. Indian J. Tradit. Knowl., **15**(4): 632-638.
- Nishiuchi S., Yamauchi T., Takahashi H., Kotula L. and Nakazono M. 2012. Mechanisms for coping with submergence and waterlogging in rice. Rice (N Y), **5**(1): 2.
- Oladosu Y., Rafii M.Y., Arolu F., Chukwu S.C., Muhammad I., Kareem I., Salisu M.A. and Arolu I.W. 2020. Submergence tolerance in rice: Review of mechanism, breeding and, future prospects. Sustainability, **12**(4): 1632.
- Patra B., Marndi B., Sanghamitra P., Samantaray S., Umakanta N. and Katara J. 2018. Rice genetic resources: Collection, conservation, maintenance and utilization. in: rice research for enhancing productivity, profitability and climate resilience, (Eds. Pathak H. et al.): pp 18-36.
- Rohilla M., Singh N., Mazumder A., Sen P., Roy P., Chowdhury D., Singh N.K. and Mondal T.K. 2020. Genome-wide association studies using 50 K rice genic SNP chip unveil genetic architecture for anaerobic germination of deepwater rice population of Assam, India. Mol. Genet Genom., **295**(5): 1211-1226.
- Sarkar R.K., Das K.K., Panda D., Reddy J.N., Patnaik S.S.C., Patra B.C. and Singh D.P. 2014. Submergence tolerance in rice: Biophysical constraints, physiological basis and identification of donors. Central Rice Research Institute, Cuttack, India, pp 36.
- Sharma R., De Vleesschauwer D., Sharma M.K. and Ronald P.C. 2013. Recent advances in dissecting stress-regulatory crosstalk in rice. Mol. Plant, **6**(2): 250-60.
- Sharma S.D., Vellanki J.M.R., Hakim K.L. and Singh R.K. 1971. Primitive and current cultivars of rice in Assam- a rich source of valuable genes. Curr. Sci., **40**(6): 126-128.
- Sorokin A.V., Kim E.R. and Ovchinnikov L.P. 2009. Proteasome system of protein degradation and processing. Biochemistry (Mosc), **74**(13): 1411-42.
- Umakanth B., Vishalakshi B., Sathish K.P., Devi R.S.J.S., Bhadana V.P., Senguttuvel P., Kumar S., Sharma S.K., Sharma P.K., Prasad M.S. and Madhav M.S. 2017. Diverse rice landraces of North-East India enables the identification of novel genetic resources for Magnaporthe resistance. Front Plant Sci., **8**: 1500.
- Vanlalsanga S., Singh P. and Singh Y.T. 2019. Rice of Northeast India harbor rich genetic diversity as measured by SSR markers and Zn/Fe content. BMC Genet., **20**:79.
- Verma H., Borah J.L. and Sarma R.N. 2019. Variability assessment for root and drought tolerance traits and genetic diversity analysis of rice germplasm using SSR markers. Sci. Rep., **9**: 16513*.*
- Wang K., Ding Y., Cai C., Chen Z. and Zhu C. 2019. The role of C2H2 zinc finger proteins in plant responses to abiotic stresses. Physiol. Plant, **165**(4): 690-700.
- Xu K., Xu X., Ronald P.C. and Mackill D.J. 2000. A high-resolution linkage map in the vicinity of the rice submergence tolerance locus Sub1. Mol. Gen. Genet., **263**: 681-689.
- Zhang X., Gonzalez-Carranza Z.H., Zhang S., Miao Y., Liu C.J. and Roberts J.A 2019. F-Box proteins in plants. Annu. Plant Rev. Online, **2**(1).

Supplementary Table S1. A list of 400 landraces used in the GWAS study for submergence tolerance along with their submergence survival percentage $\overline{}$

