Indian J. Genet. Plant Breed., (2025); 85(2): 198--207
https://doi.org/10.31742/ISGPB.85.2.3
ISSN: 0975-6906
www.isgpb.org

RESEARCH ARTICLE

Allele mining and haplotype-based association studies for nitrogen stress-responsive gene *Os06g0291500* in *indica* rice (*Oryz sativa* L.) panel

Ankur Poudel[$], Jeet Roy, Sachin Phogat, Jayanthi Madhavan[1], Viswanathan Chinnusamy[2], K.K. Vinod[1], Amitha Mithra Sevanthi and Pranab Kumar Mandal[*]

Abstract

Improving nitrogen use efficiency (NUE) in rice is critical for reducing nitrogen-induced environmental pollution and enhancing yield. Identifying natural allelic variants associated with NUE-related traits and utilizing them in breeding programs is an important approach. In this study, *Os06g0291500*, a novel, unannotated candidate gene associated with NUE, was used. Allelic variations were analyzed in 91 diverse rice genotypes through Illumina NovaSeq 6000 sequencing of the gene (3,078 bp) and its promoter (2,000 bp) regions. Mutation analysis revealed ample genetic variation of the gene across a panel of rice genotypes under study. The released varieties were found to accumulate more mutations, indicating limited selection for this gene in breeding programs. The phenotypic study showed differential N-response traits among genotypes, but association analysis did not reveal significant genotype-phenotype correlations due to the low phenotypic variance explained (PVE). Consequently, haplotype-based analysis was employed, identifying seven haplotypes within the panel. Among these, Hap_5 and Hap_7 were identified as the best and poorest haplotypes, respectively, based on phenotypic performance and SSI scores. The results highlight the potential of *Os06g0291500* for NUE improvement in rice through functional validation and haplotype-based breeding. This study provides valuable insights into the genetic regulation of NUE and underscores the importance of exploring natural allelic diversity to develop nitrogen-efficient rice varieties.

Keywords: Allele mining, haplotype, NUE, *indica* rice, SNP

Introduction

Rice (*Oryza sativa* L.) is the major cereal crop that supplies dietary requirements to more than 50% of the global population (Li et al. 2018). Since the green revolution, rice production has increased from 216 million tons in 1961 to 769 million tons by 2020 (FAOSTAT 2023) despite different biotic and abiotic constraints. The use of nitrogenous fertilizer was one of the key factors in increasing rice yield by a massive 400% and sustaining it. However, inadvertent use of nitrogenous fertilizer results in nitrogen-induced pollution as rice only utilizes 20 to 50% of applied nitrogen (N) and the remaining nitrogen gets lost into the environment (Chivenge et al. 2021). Although rice has high N fertilizer consumption, it has the lowest inherent NUE among the cereal crops (Usama and Khalid 2018).

Different approaches have been used to enhance the inherent NUE in rice. Although different agronomic N management practices are used, they are within the purview of the genetic potential of a plant to utilize applied N (Jaiswal and Raghuram 2024). Thus, it is imperative to enhance the inherent NUE of rice (Neeraja et al. 2019). For this, different genes and transcription factors related

Indian Council of Agricultural Research (ICAR) National Institute for Plant Biotechnology, IARI Pusa Campus, New Delhi 110 012, India.

[1]Division of Genetics, Indian Agricultural Research Institute, Pusa Campus, New Delhi 110 012, India.

[2]Division of Plant Physiology, Indian Agricultural Research Institute, Pusa Campus, New Delhi 110 012, India.

[$]Present address: Lamjung Campus, Institute of Agriculture and Animal Sciences, Tribhuvan University, Nepal.

***Corresponding Author:** Pranab Kumar Mandal, Indian Council of Agricultural Research, National Institute for Plant Biotechnology, IARI Pusa Campus, New Delhi 110 012, India, E-Mail: pranab.mandal@icar.gov.in

How to cite this article: Poudel A., Roy J., Phogat S., Madhavan J., Chinnusamy V., Vinod K.K., Sevanthi A.M. and Mandal P.K. 2025. Allele mining and haplotype-based association studies for nitrogen stress-responsive gene Os06g0291500 in indica rice panel. Indian J. Genet. Plan Breed., **85**(2): 198--207.

Source of support: ICAR, Govt. of India under the project Biological Nitrogen Fixation- Incentivizing in Agriculture.

Conflict of interest: None.

Received: Dec. 2024 **Revised:** April 2025 **Accepted:** April 2025

to nitrogen uptake, assimilation, remobilization, and translocation have been identified and engineered either by overexpression, knockdown or knockout approaches (Wang et al., 2022). Besides these, a large number of QTLs and candidate genes have been identified by QTL mapping and GWAS approaches (Sevanthi et al. 2021; Shen et al. 2021; Li et al. 2022; Phan et al. 2023). However, very few of these candidate genes have been used in rice breeding owing to the lack of validation of these QTLs and candidate genes and the paucity of molecular markers available for QTLs associated with NUE (Li et al. 2022). Improvement of NUE of rice either by single or multiple gene-based transgenics, overexpression, or knockdown might not provide significant results as NUE is a complex trait. Since most of the studies on transgenics are limited to specific tissues, stages and N dose and the choice of the gene in them is more often than not based on expression studies rather than mapping, the genetic/allelic status of the genes remains understudied, adversely affecting their use in commercial breeding. Further, haplotype breeding is an emerging tool in crop improvement that necessitates the identification of superior haplotypes from the natural variation available for breeding (Bhat et al., 2021). Hence, a study on the identification of allelic variants/haplotypes of the gene(s) controlling the trait of interest and their individual effect on the trait is imperative. Identifying natural allelic variants linked to yield and its component traits is crucial for breeding rice for different desired traits (Vemireddy et al. 2019).

LOC_Os06g18820 (*Os06g0291500*) is an unannotated novel gene, which is a likely member of the serine-threonine kinase family. In one of the studies conducted in our laboratory (Sevanthi et al. 2021), we identified *Os06g0291500* as an N stress-responsive gene in roots which is differentially expressed in IR64 and N22 and has two isoforms. Although the specific function of this gene is not characterized in rice, its homolog in Arabidopsis (At1g53660) acts as an organic anion transmembrane transporter and is associated with seed germination, growth, and biomass accumulation (Hartanto et al. 2020). This gene may influence differential nitrate (NO_3^-) uptake in rice, potentially through direct or indirect mechanisms (Sevanthi et al. 2021). So, *Os06g0291500* could be a potential candidate gene for NUE. The increasing demand for rice with improved nitrogen use efficiency necessitates the identification of superior and novel alleles for breeding programs. To explore genetic diversity and create allele-specific molecular markers for marker-assisted selection, the allele mining approach has been widely used. This method involves sequencing different alleles of a single gene across various genotypes within the species (Dixit et al. 2024). In this context, we have studied the allelic diversity and haplotype-based association analysis of a candidate gene related to NUE *Os06g0291500* in a panel of diverse rice genotypes.

Materials and methods

Plant materials and hydroponics

The present study constituted a diverse set of 91 genotypes, which included 43 landraces, 36 released varieties pertaining to different ago-climatic regions and 12 EMS-derived mutants of N22, which showed contrasting NUE responses over four years of testing (Supplementary Table S1). This is referred to as the association panel in the entire study. The genotypes in the association panel, other than the 12 mutants, were from the mini-core collection identified earlier from a set of 7000 lines of *indica* species (Tiwari et al. 2015). All the genotypes were phenotyped for NUE-related traits under N-stress and N-optimum conditions till maturity, as described in detail elsewhere (Poudelet al. 2024a). The hydroponics setup was maintained at the phenotyping facility at ICAR-National Institute for Plant Biotechnology, New Delhi, India, during the *Kharif* season of 2022-23 without any artificial light, at 24 to 35°C temperature (maximum 35°C during daytime and minimum 24°C during nighttime) and 60 to 80% relative humidity. The modified Yoshida media was used as hydroponics growing media (Poudel et al. 2024a).

Phenotyping for N-responsive traits

The experiment was laid on a completely randomized design (CRD) with three replications under both N-optimum and N-stressed treatments. Root length (RL) and root volume (RV) were recorded at the booting stage using a non-destructive method (Burdett 1979). Chlorophyll content was estimated at the anthesis stage (Hiscox and Israelstam 1979). Shoot dry weight (SDW), root dry weight (RDW), panicle dry weight (PDW), grain yield (GY), and N content in the root (NR), shoot (NS), and panicle (NG) were measured after harvest. N in the root, shoot, and grains was measured by near-infrared spectroscopy using FOSS™ DS 2500, Denmark. For each trait, observations were recorded from three independent samples per replication per treatment.

The Stress Susceptibility Index was calculated for traits under study (Fischer and Maurer 1978). For each trait, the Mean (M), Standard Deviation (SD), 'M - SD,' and 'M + SD' values were determined. Plants with trait values less than 'M - SD' were assigned a score of 1. Scores of 0.75 were given for values between M and 'M - SD,' 0.5 for values ranging from M to 'M + SD,' and 0.25 for values exceeding 'M + SD.' The scores for all traits were summed for each genotype and were further used for comparing the haplotypes identified for the target gene *Os06g0291500*. The haplotype with the highest total score is the most favorable haplotype and, hence, can be a potential donors for enhancing NUE in rice.

Amplicon sequencing and data analysis

Genomic DNA was isolated from young leaves of 10-day-old seedlings of all the 91 rice genotypes by the CTAB method (Murray and Thompson 1980). Quality and quantity

of DNA were assessed at 0.8% agarose gel and Nanodrop spectrophotometer (ND-1000 Spectrometer, Thermo Scientific, USA), respectively.

The genomic DNA sequence of *Os06g0291500* (LOC_Os06g18820) gene, along with 2 kb promoter region was downloaded from the ENSEMBL plants database. Two pairs of overlapping sets of primers were designed FP1 (forward 5'-3',CATCTCAACGCGGACATT), RP1 (reverse 5'-3',TCCATGCAGTCAACATCAG) and FP2 (forward 5'-3',CGGTTAATTCGAAGGACTGT), RP2 (reverse 5'-3'CATGTGTGGGAGTAGTATAAGG) with expected fragment size of 3048 bp and 2541 and used for amplification. The genomic DNA of 91 isolated genotypes was used as a template in PCR amplification (Veriti™ 96 well fast thermal cycler). Both fragments were amplified separately across 91 genotypes using Phusion™ High Fidelity DNA polymerase. A total of 50 µL reaction volume was prepared. The reaction mixture was prepared with the following components: 5X Phusion™ HF buffer (10 µL), 10 mM dNTPs (1-µL), forward primer (2.5 µL), reverse primer (2.5 µL), template DNA (2.5 µL), Phusion™ High-Fidelity DNA Polymerase (0.5 µL), and nuclease-free water (31 µL). The final concentrations of the primers and template DNA were 200 µM and 20 ng/µL, respectively. PCR amplification was performed using a three-step protocol. Initial denaturation was carried out at 98°C for 30 seconds, followed by denaturation at 98°C for 10 seconds. Annealing was performed at 60°C for both primer sets, with an extension step at 72°C for 1.5 minutes. The reaction concluded with a final extension at 72°C for 7 minutes. PCR-amplified products were resolved in 1% agarose gel. PCR amplicons were purified using MagGenomeXpressPure Beads PCR Cleanup Kit and quality and quantity were checked in 1% agarose gel. The pooling of amplicons was done on the equimolar concentration of the amplicons.

Paired-end sequencing of pooled amplicons was carried out using Illumina NovaSeq 6000 (Illumina, San Diego, USA). Data obtained from sequencing was de-multiplexed by Illumina's bcl2fastq conversion software v2.20 using unique dual barcode sequences and FastQ files were generated. The good-quality trimmed data of all 91 samples were used for alignment against reference (Nipponbare) sequences using the Bwa-v0.7.17 tool (Li 2013). Alignment data was processed using the Picard v1.102 tool for the removal of PCR duplicates. Haplotype mapping data was generated for the SNP using NGSEPScore-v 4.3.1 (https://github.com/NGSEP/NGSEPcore/releases/tag/v4.3.1).

Genome-wide association analysis of *Os06g0291500*

Genome-Wide Association Analysis (GWAS) of *Os06g0291500* gene was conducted using a Mixed Linear Model (MLM) in GAPIT v3 (Wang and Zhang 2021). The high-quality SNPs of *Os06g0291500* gene, including its 2 kb promoter region, were used for association analysis along with phenotypic data in N-optimum and N-stressed conditions, respectively.

Haplotype analysis of association panel

The obtained consensus sequences of 91 genotypes were aligned to the reference Nipponbare sequence using clustalW in GeneiousPrime v 2025.0 software (Supplementary Table S2). DNAsp version 5.10 (Rozas et al. 2017) was used to analyze nucleotide diversity. A haplotype network diagram was generated using PopART version 4.8.4 (Leigh et al. 2015).

We used the rice SNP Seek database available at (snp-seek.irri.org) to obtain SNPs of *Os06g0291500* (Mansueto et al. 2017). We filtered the SNPs from the coding region of the gene and used it for haplotype analysis using DnaSP v5.10. For haplotype characterization, heterozygous SNPs were considered as missing data, while the genotypes with missing SNPs were removed. Similarly, the sites with missing data/gaps were not considered during haplotype identification.

Prediction of the 3D structure of reference and haplotypes

AlphaFold2 tool was used to generate detailed 3D structural models of proteins of all haplotypes identified under the study. The protein sequences of the reference were obtained from Ensembl plants databases. Multiple alignments of CDS sequence were done using clustalW. Aligned CDS sequences were translated using expasy (https://web.expasy.org/translate/) and protein sequences were used for the prediction of 3D structure. The 3D structures were predicted using AplhaFoldColab(Jumper et al. 2021), which was visualized using ChimeraX (https://www.rbvi.ucsf.edu/chimerax). AlphaFold2 tool outputs 3D coordinates and a confidence score called pLDDT: scores above 90 indicate high accuracy, scores between 70 and 90 suggest a reliable backbone structure, and scores from 50 to 70 represent regions that may appear unstructured, often forming loops and flexible domains. AlphaFold2 also provides a Predicted Aligned Error (PAE) map, visualizing confidence in the relative positions of residue pairs, with lower PAE values indicating greater confidence in their proximity. Similarly, to identify changes in the protein structure of haplotypes, the reference 3D protein structure of haplotypes was superimposed using chimeraX software.

Results

Phenotypic variation under N-optimum and N-stress treatments in the association panel

The panel of 91 genotypes was grown under hydroponics in N-optimum and N-stressed conditions. Traits under N-optimum condition traits like TC, PDW, RDW, SDW, GY, NS, NR, and NG were higher than the N-stressed condition. The phenotypic performance of the rice genotypes grown under N-optimum and N-stressed conditions is presented

in Poudel et al. (2024b). Similarly, traits like RL and RV were higher than in the N-optimum condition.

Nucleotide polymorphism in Os06g0291500 gene and promoter

A total of 452 mutations were observed in 91 genotypes used in the study. The nucleotide variation frequency across the entire genes is presented in Fig. 1. Of the total mutations, 285 were SNPs, while 167 were InDels, with 80 insertions and 87 deletions. Insertions ranged from 1 bp to 21 bp, whereas deletions ranged from 1 bp to 14 bp. Of the total 285 SNPs, 168 (58.9%) were transitioned, and 117 (41.1%) were transversions. On analyzing the location of mutations, 79 SNPs, 10 insertions and 35 deletions were present in the promoter region, while 5'UTR and intronic regions contained just two mutations, one transversion and one transition, respectively. The CDS region contained 6 SNPs, three deletions and two insertions. Multiple sequence alignment of the CDS region revealed that a 9 bp deletion was observed from position 203 to 211 bp in 30 genotypes, another 9 bp deletion from position 199 to 207 bp in BAM 6067 and 7 bp deletion from position 195 to 201 in genotype Shabhagi Dhan. The 3'UTR region harbored maximum variations with 198 SNPs, 70 insertions and 52 deletions.

Polymorphism in an amino acid Sequence of Os06g0291500

Across the CDS region, a total of 30 amino acid changes were observed. This included the deletion of 3 consecutive alanine residues from positions 65-67 in 31 genotypes. This was the most frequent amino acid change observed in the association panel. Insertion of 4 bases at nucleotide position 441 in seven genotypes, namely BAM 2166, BAM 2614, BAM5449, PS2, ChakhoPoiteran, Pokkali and Tetep, corresponding to 148th amino acid position, resulted in a frameshift which led to two changes: (i) insertion of arginine at position 148 and (ii) creation of a stop codon at amino acid position 150. In genotype BAM6067, the deletion of three Alanine residues was detected; however, there was no termination codon was created. In BAM2096, the insertion of 4 bases at the 508 nucleotide position resulted in frameshift mutation from the 170th amino acid onward and the abolition of the stop codon present in the reference at position 194. A total of four SNPs, at positions 11, 16, 28 and 142 resulted in non-synonymous mutations in corresponding amino acid positions 11, (G to D in the mutant 3694); position 16 (E to Q in 27 genotypes); position 28 (D to G in 29 genotypes); and position 142 (S to P in 21 genotypes).

Nucleotide diversity of Os06g0291500 gene

The nucleotide diversity analysis revealed that the average pairwise nucleotide diversity (π) was 0.01232. The sliding window analysis of pairwise nucleotide diversity in DnaSP showed significant diversity in 1000-2000 bp, 3500-4600 bp and 4600-5164 bp regions, which matched with promoter and 3'UTR regions (Fig. 1). The salient Watterson's nucleotide diversity estimator, Theta (Θw) per site from S and Θ over Eta (η), was found to be the same as 0.0008, whereas Θ (per sequence) from S was 41.514. The average number of nucleotide differences, "k", was found to be 60.132. The Tajima's S statistic was 1.46, which is positive but not significant.

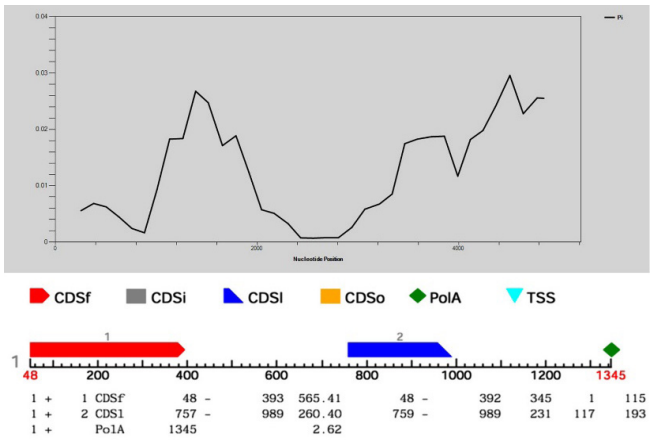


Fig. 1. Nucleotide variation at Os06g0291500 gene. A. Nucleotide diversity of *Os06g0291500* gene including 2 Kb promoter region. B. Gene structure of *Os06g0291500* gene. TSS: Transcription Start Site, CDSf: First Coding Sequence, CDSI: Last Coding Sequence, PolA: Polyadenylation Sequence

Association analysis of SNP in the Os06g0291500 gene with N-responsive traits

Mixed Linear Model (MLM) was used for association mapping using high quality 81 high-quality SNPs from a promoter and 149 SNPs fromgenic region of *Os06g0291500*. Only one association was observed from the promoter region with SDW under N-stress conditions. All the significant associations were identified in the 3'UTR region of the gene. Of the different associations observed, root-related traits (RL, RDW & NR) and pigment (TC) showed three associations under N-optimum and 5 associations under N-stress conditions (Figs. 2 and 3). The phenotypic variation explained (PVE) ranged from 1.12 to 7.62%. The SNP5148 explained the highest PVE from the UTR region for NR under N-optimum condition. This SNP also showed an association with TC (N-optimum) and NR (N-stress) but with low PVE and LOD scores.

Haplotype analysis of Os06g0291500 gene from IRRI SNP seek database

The SNP data for *Os06g0291500*, located at 10,685,619 – 10,688,696 bp positions on the forward strand on chromosome 6, extracted from IRRI SNP-Seek database containing re-sequencing data of 3,024 diverse rice genotypes, was used for this study. Out of 3024 genotypes, 310 (10.25%) could not be considered for haplotype analysis

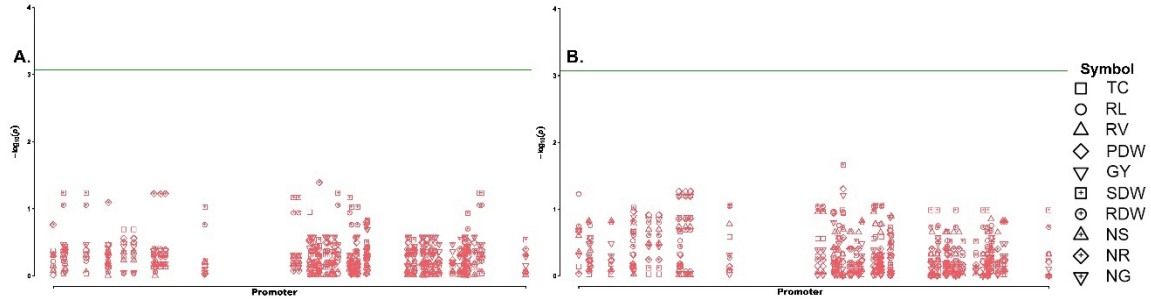


Fig. 2. Manhattan Plot derived from MLM model showing association between significant SNPs of promoter region with 10 traits under study. A. N- optimum condition. B. N-stressed condition

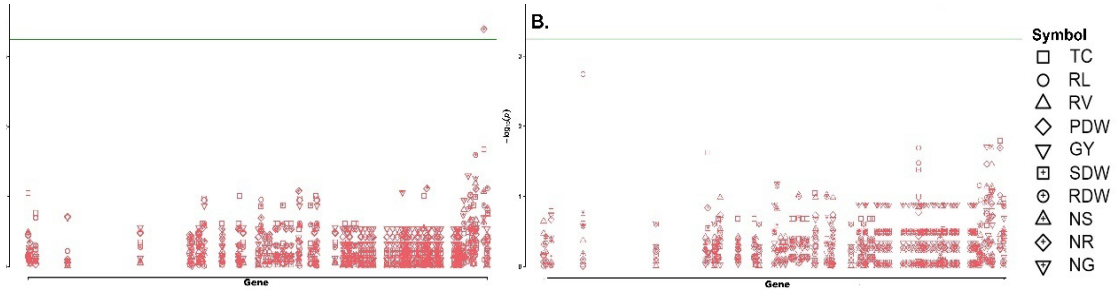


Fig. 3. Manhattan Plot derived from MLM model showing an association between significant SNPs of the gene with 10 traits under study. A. N-optimum condition. B. N-stressed condition

based on the criteria explained in data curation. A total of 2714 genotypes, harboring nine SNPs from the coding region, gave rise to 12 haplotypes. The Hap_10 was the most common haplotype (31.5% with 855 genotypes) across the 3K panel in which *indica* genotypes were the predominant ones, representing 53.10% (n = 454), followed by *japonica*, 32.63% (n = 279) and *Aus* genotypes 7.72%, (n = 66). Other haplotypes, namely Hap_7, Hap_3, Hap_4, Hap_8 and Hap_11, predominantly contained genotype with indica ancestry, whereas the less common haplotype Hap_1 predominantly contained genotype with *Aus* ancestry (75.76%).

Haplotype analysis of Os06g0291500 geneusing association panel

Haplotype analysis was done based on 5 high-quality SNPs identified in the coding region of a gene. Haplotype analysis revealed a total of 7 haplotypes with a haplotype diversity of 0.7736. Hap1 is the major haplotype consisting of 29 genotypes, followed by Hap2 consisting of 25 genotypes. Hap 5 is the smallest haplotype, consisting of only one genotype. The detail of the haplotype is presented in Table 1, and Fig. 4.

The haplotyping of the association panel revealed that most of the accessions with shared pedigree and from the same region tended to cluster within the same haplotype. Hap_1 contained genotypes that were derived as N22 mutants, which were earlier tested for NUE at our lab. Hap_2 contained a majority of commercially released varieties like ADT39, Rajendra Mahsuri, CR1009, Dhan 42, Ranjit, DRR Dhan 50, FR13A, IR64, MTU1075, NL 44, SM Sub 1 and Swarna. Similarly, Hap_3 contained a collection of elite aromatic/basmati rice varieties like Pusa Sugandh 5, Pusa Basmati 1121, Pusa Basmati 1509, etc. Hap_5 and Hap_7 contained one and two genotypes, respectively, while Hap_4 and Hap_6 mostly contained landraces.

When the haplotypes were analyzed for change in amino acid sequences, Haplotypes Hap_1 and Hap_4 did not have any changes in amino acid. This represented 40.6% (37/91), which did not have any changes in amino acid sequence. The N22-derived mutants and some released varieties like PB1401, SahabhagiDhan, CSR 30, Narendra Dhan, Dhan 51, etc., represented this group. Hap_3 and Hap_7, which represented 29.67% (27/91) of the genotypes, showed transition (G to A) on 32nd position of CDS sequence that resulted in the amino acid substitution, G to R. The released cultivars like Vandana, PB1121, PB1509, MTU1010, Lalat, IR36, HUR105, HP2216, Rasi etc. represented this group. Another transversion (G to C) on the 46th position of CDS in 24.17% (22/91) of the association panel resulted in the substitution of amino acid from Glycine to Alanine in Hap_2, Hap_5 and Hap_6. The rice accessions like Azucena, ADT39, Rajendra Mahsuri, CR1009, Dhan 42, Ranjit, DRR Dhan 50, IR64, MTU1075, SM Sub1, BAM 2096, PS2, Pokkali, Tetep, etc. belonged to this group.

Table 1. Haplotypes characterized from the data obtained from association panel genotypes under study

Haplotypes	2081	2095	2132	2346	2836	Genotypes
Hap_1	G	G	G	C	T	29
Hap_2	G	C	A	T	C	19
Hap_3	A	G	A	C	T	25
Hap_4	G	G	A	C	T	8
Hap_5	G	C	A	C	C	1
Hap_6	G	C	A	C	T	7
Hap_7	A	G	A	C	C	2

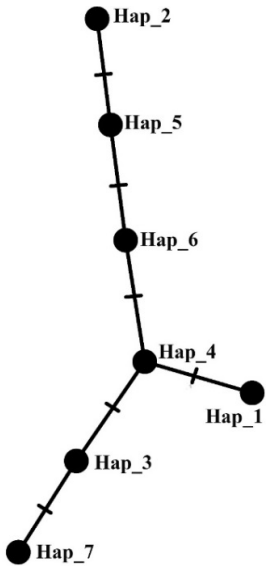


Fig. 4. Haplotypes based on Exonic SNPs of *Os06g0291500* gene

Association of haplotypes of association panel under different N regimes

The mean agro-morphological trait values were grouped as 7 different haplotype groups (Fig. 5). Among all, haplotype 5 showed the highest trait values for TC, RL, SDW, RDW and NS. Also, the PDW and GY were higher in this haplotype group under N-stress. Moreover, under N-optimum conditions, this haplotype displayed the lowest values for RL, RV, PDW, GY, SDW, RDW and NR. Haplotype 7 displayed the lowest trait values for PDW and NS under N stress. Also, the PDW and GY of this haplotype displayed lower value as compared to other haplotypes.

Classification of haplotypes based on SSI scores

The Stress Susceptibility Index (SSI) of these haplotypes was estimated. The SSI score calculated based on different trait values showed that Haplotype 5 obtained the highest score, thus representing the best group, while the lowest score was obtained for Haplotype 7, representing the poorest group. The SSI scores confirmed the results of phenotyping.

3D structure of Os06g0291500 gene and its haplotypes

AlphaFold2 was used to predict the 3D structure of the *Os06g0291500* protein based on sequence alignments (Fig. 6).

The reference protein structure of *japonica* rice *Nipponbare* revealed 4 alpha helices and 4 beta sheets. The pLDDT score was 51.5. The details of the 3D structure model, sequence coverage, pLDDT score, and PAE score are presented in Supplementary Fig 1. The 3D model of the haplotypes exhibited distinct differences as the haplotypes differed in the number of alpha helices and beta chains. The haplotype 4 did not show the presence of any alpha helix and fewer beta sheets due to premature termination of the protein. Similarly, significant changes were observed in the protein structure when the reference 3D structure was superimposed on the haplotype structure (Fig. 7). Higher RMSD scores were observed for a superimposed model with greater dissimilarity as the higher RMSD value indicates deviation between reference and haplotype model. Also, none of the superimposed models showed very low or zero RMSD value, which showed that the SNPs had a significant bearing on the 3D structure.

Discussion

The complex interplay between environment, edaphic factors, other biotic and abiotic stresses make creation and maintenance of N0 or N-sick plots challenging. Hence, studying NUE in field conditions with absolute control is almost impossible. However, phenotyping of biological material for different NUE related traits with precise control is very critical for the success of any association or mapping study. Hence, we used a hydroponics setup standardized in our laboratory to grow rice from seed to seed under optimal and low N regimes (Poudel et al. 2024a). The association panel used in the study consists of diverse set of rice germplasm, which was identified as part of minicore selected from 7000 (Tiwari et al. 2015) ecologically diverse *indica* rice collection, released varieties and EMS-induced mutants of our lab. This choice of germplasm is very

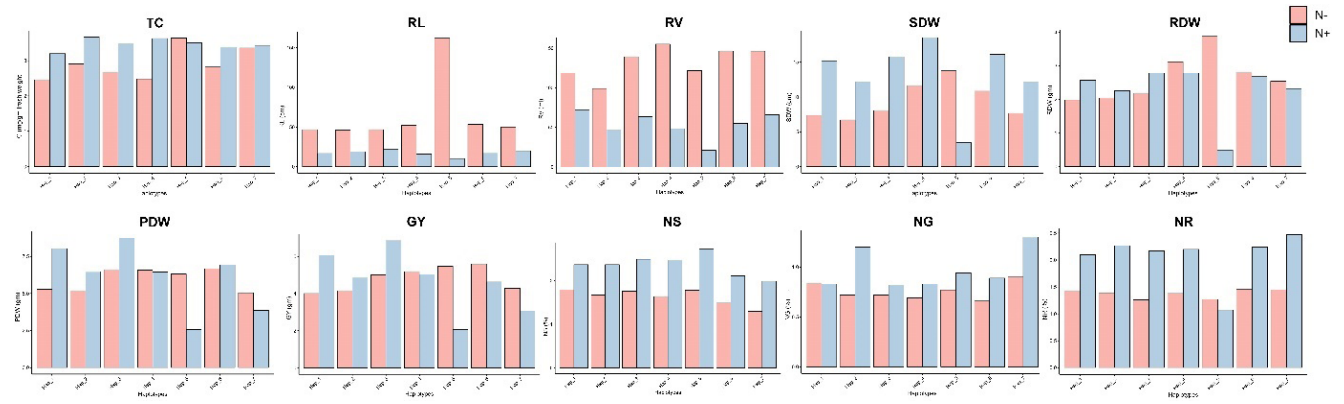


Fig. 5. Performance of haplotypes under N-optimum and N-stressed conditions. TC = Total Chlorophyll, RL = Root Length, RV: Root Volume, SDW = Shoot Dry Weight, RDW = Root Dry Weight, PDW = Panicle Dry Weight, GY = Grain Yield, NG = N content in grain, NR = N content in root, NS = N content in shoot

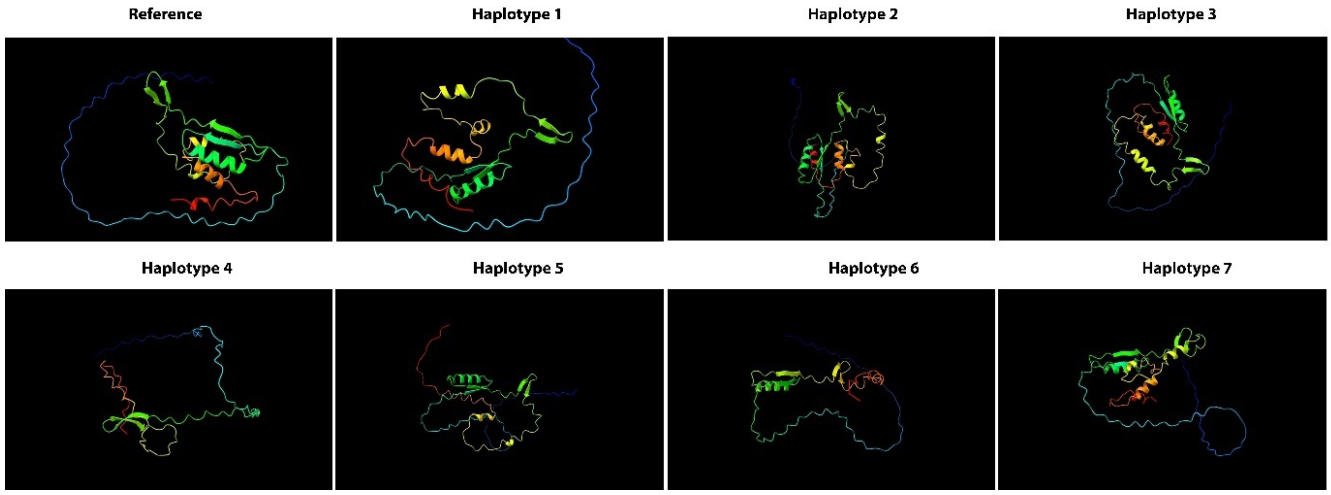


Fig. 6. 3D structure of the *Os06g0291500* gene and its haplotypes predicted using Alphafold2

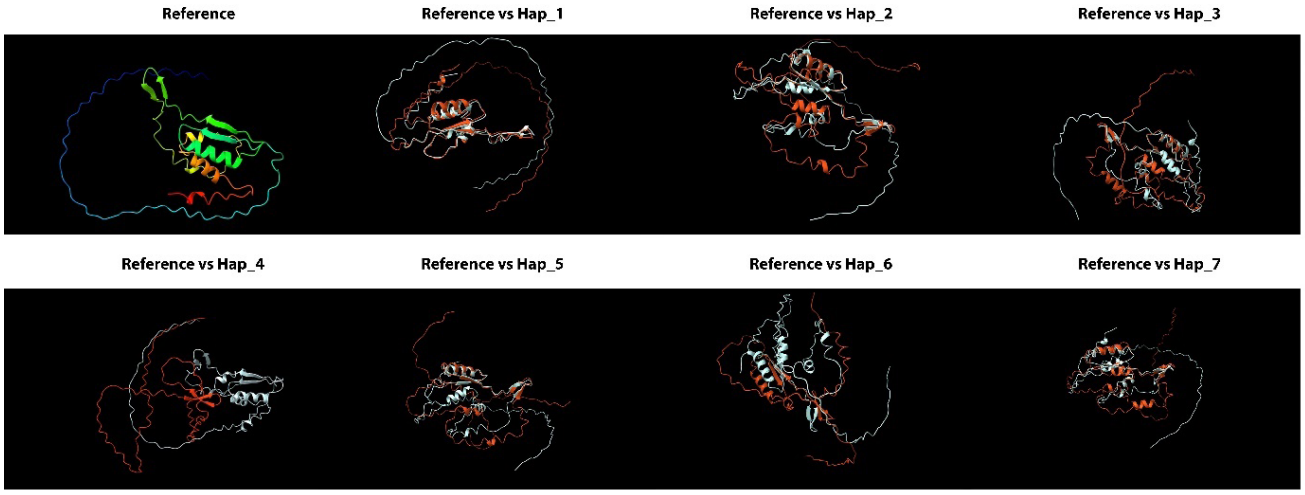


Fig. 7. The reference 3D structure of *Os06g0291500* gene superimposed to the haplotype 3D structure. I. Reference vs Haplotype 1 II. Reference vs Haplotype 2. III. Reference vs Haplotype 3. IV. Reference vs Haplotype 4. V. Reference vs Haplotype 5. VI. Reference vs Haplotype 6 VII. Reference vs Haplotype 7

critical as association analysis and the power of association mapping are directly proportional to the genetic diversity of the population (Abdurakhmonov and Abdukarimov 2008).

NUE is a complex trait and its measurement becomes more complex because of its involvement in N sensing, uptake, transport, assimilation, signaling and interaction with other nutrients (Lee 2021). Thus, important traits that depict the differences in NUE include root, pigment and yield-related traits (Sandhu et al. 2021). On this basis, 10 traits, namely, TC, RL, RV, RDW, PDW, SDW, GY, NG, NS and NR, were recorded both under N-optimum as well as N-stressed condition for phenotyping (Poudel et al. 2024b). The genotypes exhibited distinct performances under the two nitrogen regimes, highlighting their inherent genetic variability in both conditions. The higher coefficient of variation (CV) observed for grain yield (GY), number of grains (NG), and all dry weight traits (PDW, SDW, and RDW) indicated that the association panel was highly diverse and suitable for studying nitrogen use efficiency (NUE) traits. As expected, the CV for traits under nitrogen stress was lower than under optimal N conditions, since genotypes were unable to fully express their potential under stress (_ et al. 2016). However, two key root traits, RDW and NR, showed higher CV under N stress, suggesting their importance for enhanced nitrogen uptake. Variations in root length and volume under nitrogen stress are adaptive strategies to enhance nutrient acquisition (Jia et al., 2020). Therefore, the key nitrogen-responsive traits in rice included RL, RV, RDW, and GY (Poudel et al. 2024b).

The sequence analysis of the *Os06g0291500* gene consisting of 2000 bp promoter region 3078 bp genic region consisting of 1 intron and 2 exons revealed significant variation within the association panel. The exon 2 region consisted of more than 2-fold SNPs and 7-fold more insertions as compared to the promoter region. The intronic region was found to be highly conserved as only one SNP was detected. The CDS region was found to contain 4 SNPs, 1 deletion and 1 insertion event, all of which led to either non-synonymous substitutions or premature termination codons. Released varieties, namely, Purnendu, IR64, MTU 1075, ADT39, Samba Mahsuri- Sub 1, Azucena, Rajendra Mahsuri, CR1009, DRR Dhan 42, DRR Dhan 50, and landraces namely, BAM113, BAM 2064, BAM 3154, BAM 328, BAM 4168, DTY 3.2, FR 13A, IC 576938, and NL 44 contained more mutations as compared to other members of panel. The greater number of mutations, especially in the released varieties, suggests that NUE related genes were not considered during the green revolution, wherein high yielding and high N-responsive genes were primarily considered.

The sequence analysis of 5078 bp length of *Os06g0291500* gene showed significant sequence variation in the association panel under study. Exon 2 contained almost 2.5 times more number of mutations than promoter region. Since exon 1 was shorter in length (393 bp), the number of mutations was less. The intron region was found to be highly conserved with only one SNP (C to T) transition in cv. Tetep, whereas the CDS region contained multiple non-synonymous mutations. Similarly, analysis of amino acid sequence detected insertion of 4 bases at nucleotide position 441, which resulted in haplotype group 6. This mutation did show any significant improvement in SSI score, which resembled the intermediate haplotype groups in terms of stress tolerance. However, four base insertions in BAM 2096 resulted in a frameshift mutation, which removed the stop codon and showed the highest SSI score, which corresponded to the best haplotype cluster (Haplotype 5).

The average pairwise nucleotide diversity value indicated a higher genetic diversity of *Os06g0291500* in the panel. Despite potential bottlenecks during domestication, the gene has retained sufficient genetic diversity, attributing to phenotypic variability in traits associated with abiotic stress, including NUE. Maintaining such genetic variation in key functional genes is critical for providing resilience against environmental stresses, including nitrogen-limited conditions, in released rice cultivars. The higher nucleotide diversity is due to the presence of rice genotypes like released varieties which have not undergone selection pressure for the trait governed by the gene. Higher value of nucleotide diversity also represents more genetic diversity (Rana et al. 2022), which is also represented by mutation summary and haplotypes. In this study positive Tajima D value (1.46) suggested an excess of polymorphism at intermediate frequencies, which could be due to balancing selection and overdominance (Dixit et al. 2024).

The candidate gene-based GWAS identified a total of 8 associations, mainly with root traits. However, the PVE was very low, thus defined as minor effect associations. Since no significant associations were detected, we resorted to haplotype-based analysis for this study. The haplotype analysis from IRRI rice SNP seeks 3K database showed that *indica* rice genotypes were predominant ones representing the larger haplotype number and genotypes. Since Indian *indica* rice germplasm for diverse NUE is not properly represented in the IRRI rice 3K panel, haplotyping was done with the help of the association panel used in this study. The SNPs from the association panel and 3K panel were used separately, and differences in sequences were obtained in haplotypes for both cases. These haplotypes need functional validation before their use in breeding.

Thus, allele mining of the novel gene *Os06g0291500*, a potential candidate for NUE, revealed significant variation across diverse indica rice genotypes. Due to low PVE, the haplotype-based analysis identified seven haplotypes, with Hap_5 and Hap_7 as the best and poorest groups, respectively. These findings provide a foundation for functional validation and haplotype-based breeding for

NUE in rice.

Supplementary material

Supplementary Tables S1-S2 and Supplementary Fig. 1 are provided with the text, which can be accessed at www.isgpb.org

Authors' contribution

Conceptualization of research (PKM, AMS); Designing of the experiment (PKM, AMS); Collection of experimental materials (AMS); Execution of lab experiments and data collection (AP, SP, JR), Analysis of data and interpretation (AP, JM, VC, KKV, AMS, PKM); Preparation of the manuscript (AP, JR, SP, JM, VC, KKV, AMS, PKM).

Acknowledgments

The first author acknowledges the SAARC Agriculture Centre (SAC) for awarding the SAARC Agriculture PhD Scholarship to complete this work as part of PhD thesis. The authors also would like to acknowledge the financial support from the Indian Council of Agricultural Research as part of the project Biological Nitrogen Fixation under the initiative Incentivizing in Agriculture and also by the ICAR-National Institute for Plant Biotechnology.

References

Abdurakhmonov I. Y. and Abdukarimov A. 2008. Application of association mapping to understanding the genetic diversity of plant germplasm resources. Int. J. Plant Genomics, 2008, 574927. https://doi.org/10.1155/2008/574927

Bhat J. A., Yu D., Bohra A., Ganie S. A. and Varshney R. K. 2021. Features and applications of haplotypes in crop breeding. Commun. Biol., **4**(1): 1266. https://doi.org/10.1038/s42003-021-02782-y

Burdett A. 1979. A non-destructive method for measuring the volume of intact plant parts. Can. J. For. Res., **9**(1): 120–122.

Chivenge P., Sharma S., Bunquin M. A., and Hellin J. 2021. Improving Nitrogen Use Efficiency—A Key for Sustainable Rice Production Systems. Front. Sustain. Food Syst., **5**: 737412. https://doi.org/10.3389/fsufs.2021.737412

Dixit D., Siddiqu, N., Bollinedi H., Malik A., Bhowmick P., Ellur R., Nagarajan M., Vinod K., and SinghA. 2024. Allele mining for Granule Bound Starch Synthase1 (GBSS1) gene governing amylose content in aromatic rice (*Oryza sativa* L.) germplasm. Indian J. Genet. and Plant Breed., **84**(01): 38–45.

Fischer R. and Maurer R. 1978. Drought resistance in spring wheat cultivars. I. Grain yield responses. Aust. J. Agr. Res., **29**(5): 897–912.

Hartanto M., Joosen R. V. L., Snoek B. L., Willems L. A. J., Sterken M. G., de Ridder D., Hilhorst H. W. M., Ligterink W., and Nijveen, H. 2020. Network Analysis Prioritizes DEWAX and ICE1 as the Candidate Genes for Major eQTL Hotspots in Seed Germination of *Arabidopsis thaliana*. *G3* (Bethesda, Md.), **10**(11): 4215–4226. https://doi.org/10.1534/g3.120.401477

Hiscox J. D. and Israelstam G. F. 1979. A method for the extraction of chlorophyll from leaf tissue without maceration. Can. J. Bot., **57**(12): 1332–1334. https://doi.org/10.1139/b79-163

Jaiswal D. K. and Raghuram N. 2024. Molecular interventions for improving crop nitrogen use efficiency: Trends, opportunities and challenges in rice. Improving Nitrogen Use Efficiency in Crop Production; Ladha, JK, Ed.; Burleigh Dodds Science Publishing Limited: Cambridge, UK, 67–112.

Jia Z., Giehl R. F. H. and von Wirén N. 2020. The Root Foraging Response under Low Nitrogen Depends on DWARF1-Mediated Brassinosteroid Biosynthesis. Plant Physiol., **183**(3): 998–1010. https://doi.org/10.1104/pp.20.00440

Jumper J., Evans R., Pritzel A., Green T., Figurnov M., Ronneberger O., Tunyasuvunakool K., Bates R., Žídek, A. and Potapenko A. 2021. Highly accurate protein structure prediction with AlphaFold. Nature, **596**(7873): 583–589.

Lee S. 2021. Recent Advances on Nitrogen Use Efficiency in Rice. Agronomy, **11**(4): 753. https://doi.org/10.3390/agronomy11040753

Leigh J. W., Bryant D. and Nakagawa S. 2015. POPART: full-feature software for haplotype network construction. Methods Ecol. Evol., **6**(9).

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013.

Li J., Xin W., Wang W., Zhao S., Xu L., Jiang X., Duan Y., Zheng H., Yang L., Liu H., Jia Y., Zou D. and Wang J. 2022. Mapping of Candidate Genes in Response to Low Nitrogen in Rice Seedlings. Rice, **15**(1): 51. https://doi.org/10.1186/s12284-022-00597-x

Li P., Li Z., Liu X., Zhang H., Wang Q., Li N., Ding H. and Yao, F. 2022. Development and Application of Intragenic Markers for 14 Nitrogen-Use Efficiency Genes in Rice (*Oryza sativa* L.). Front. Plant Sci., 13. https://doi.org/10.3389/fpls.2022.891860

Li Y., Xiao J., Chen L., Huang X., Cheng Z., Han B., Zhang Q. and Wu C. 2018. Rice Functional Genomics Research: Past Decade and Future. Mol. Plant, **11**(3): 359–380. https://doi.org/10.1016/j.molp.2018.01.007

Mansueto L., Fuentes R. R., Borja F. N., Detras J., Abriol-Santos J. M., Chebotarov D., Sanciangco M., Palis K., Copetti D., Poliakov A., Dubchak I., Solovyev V., Wing R. A., Hamilton R. S., Mauleon R., McNally K. L. and Alexandrov N. 2017. Rice SNP-seek database update: New SNPs, indels, and queries. Nucleic Acids Res., **45**(D1): D1075–D1081. https://doi.org/10.1093/nar/gkw1135

Murray M. G. and Thompson W. F. 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res., **8**(19): 4321–4326. https://doi.org/10.1093/nar/8.19.4321

Neeraja C. N., Voleti S. R., Subrahmanyam D., Surekha K., and Rao P. R. 2019. Breeding rice for nitrogen use efficiency. Indian J. Genet. Plant Breed., **79**(Sup-01): 208–215.

Phan N. T. H., Draye X., Van Pham C. and Bertin P. 2023. Identification of quantitative trait loci controlling nitrogen use efficiency-related traits in rice at the seedling stage under salt condition by genome-wide association study. Front. Plant Sci., **14**.

Poudel A., Saini M. R., Phogat S., Roy J., Chauhan H., Madhavan J., Sevanthi A. M. and Mandal P. K. 2024. Growing *Indica* and *Aus* Type Rice till Maturity in Hydroponics under N-Optimum and N-Stressed Condition: A Detailed Account.

Prakash C., Mithra S. V. A., Singh P. K., Mohapatra T. and Singh N. K. 2016. Unraveling the molecular basis of oxidative stress management in a drought tolerant rice genotype Nagina 22. BMC Genomics, **17**(1): 774. https://doi.org/10.1186/s12864-016-3131-2

Rana N., Kumawat S., Kumar V., Bansal R., Mandlik R., Dhiman P., Patil G. B., Deshmukh R., Sharma T. R. and Sonah H. 2022.

Deciphering Haplotypic Variation and Gene Expression Dynamics Associated with Nutritional and Cooking Quality in Rice. Cells, **11**(7): 1144. https://doi.org/10.3390/cells11071144

Rozas J., Ferrer-Mata A., Sánchez-DelBarrio J. C., Guirao-Rico S., Librado P., Ramos-Onsins S. E. and Sánchez-Gracia A. 2017. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Mol. Biol. Evol., **34**(12): 3299–3302. https://doi.org/10.1093/molbev/msx248

Sandhu N., Sethi M., Kumar A., Dang D., Singh J. and Chhuneja, P. 2021. Biochemical and Genetic Approaches Improving Nitrogen Use Efficiency in Cereal Crops: A Review. Front. Plant Sci., **12**: 657629. https://doi.org/10.3389/fpls.2021.657629

Sevanthi A. M., Sinha S. K., V S., Rani M., Saini M. R., Kumari S., Kaushik M., Prakash C., K., V., Singh G. P., Mohapatra T. and Mandal P. K. 2021. Integration of Dual Stress Transcriptomes and Major QTLs from a Pair of Genotypes Contrasting for Drought and Chronic Nitrogen Starvation Identifies Key Stress Responsive Genes in Rice. Rice, **14**(1): 49. https://doi.org/10.1186/s12284-021-00487-8

Shen C., Chen K., Cui Y., Chen J., Mi X., Zhu S., Zhu Y., Ali J., Ye G., Li Z. and Xu, J. 2021. QTL Mapping and Favorable Allele Mining of Nitrogen Deficiency Tolerance Using an Interconnected Breeding Population in Rice. Front. Genet., **12**: 616428. https://doi.org/10.3389/fgene.2021.616428

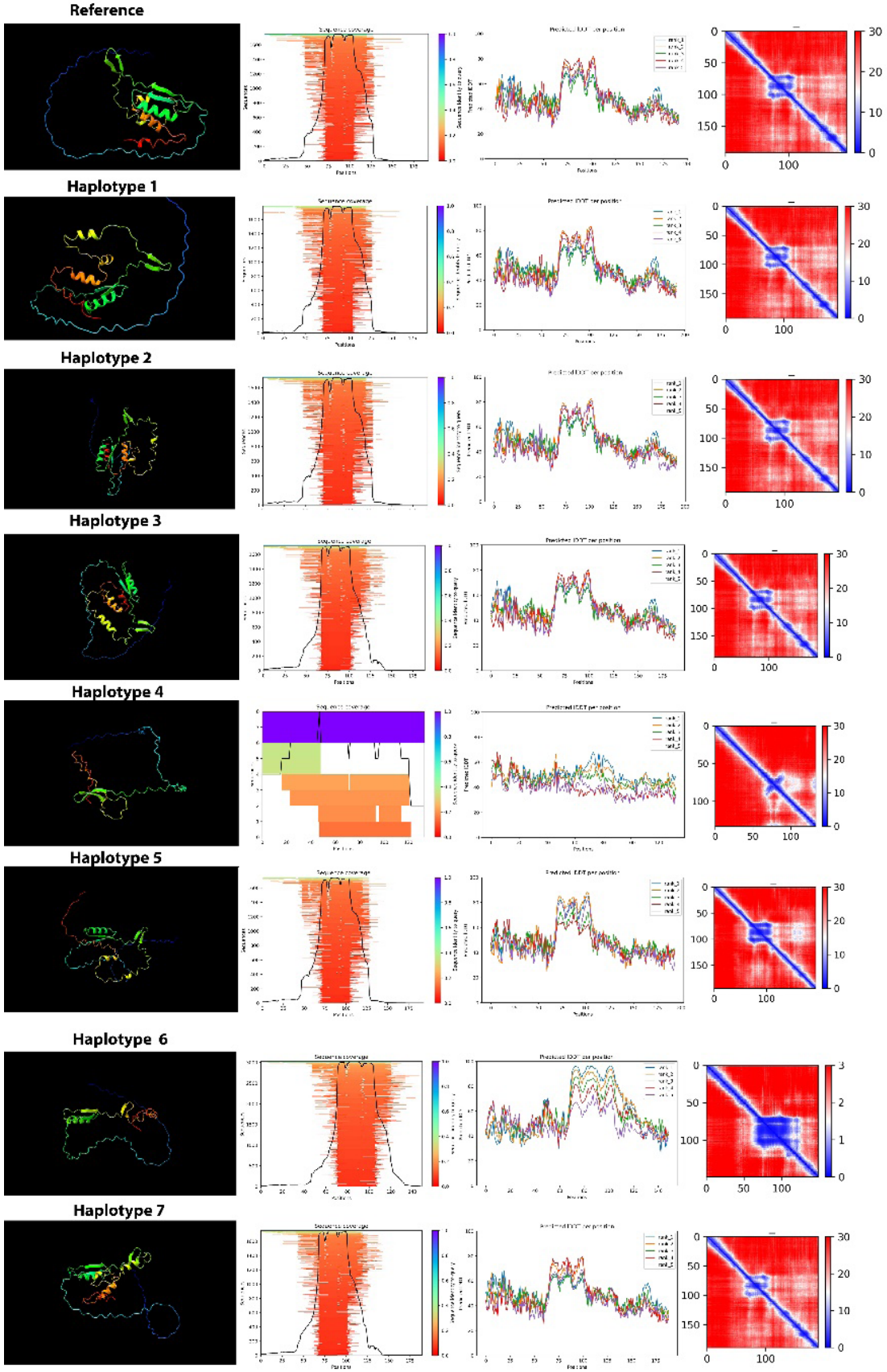
Tiwari K. K., Singh A., Pattnaik S., Sandhu M., Kaur S., Jain S., Tiwari S., Mehrotra S., Anumalla M., Samal R., Bhardwaj J., Dubey N., Sahu V., Kharshing G. A., Zeliang P. K., Sreenivasan K., Kumar P., Parida S. K., Mithra S. V. A. and Mohapatra T. 2015. Identification of a diverse mini-core panel of Indian rice germplasm based on genotyping using microsatellite markers. Plant Breed., **134**(2): 164–171. https://doi.org/10.1111/pbr.12252

Usama M. and Khalid M. 2018. Fertilizer consumption in India and need for its balanced use: A review. Indian J. Environ. Prot., **38**(7): 564–577.

Vemireddy L. R., Kadambari G., Reddy G. E., Kola V. S. R., Ramireddy E., Puram V. R. R., Badri J., Eslavath S. N., Bollineni S. N., Naik B. J., Chintala S., Pottepalem R., Akkareddy S., Nagireddy R., Reddy L. V. B., Bodanapu R., Lekkala S. P., Chakravartty N. and Siddiq E. A. 2019. Uncovering of natural allelic variants of key yield contributing genes by targeted re-sequencing in rice (*Oryza sativa* L.). Sci. Rep., **9**(1): 8192. https://doi.org/10.1038/s41598-019-44708-z

Wang B., Zhou G., Guo S., Li X., Yuan J. and Hu A. 2022. Improving Nitrogen Use Efficiency in Rice for Sustainable Agriculture: Strategies and Future Perspectives. Life (Basel, Switzerland), **12**(10): 1653. https://doi.org/10.3390/life12101653

Wang J. and Zhang, Z. 2021. GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. Genom. Proteom. Bioinform., *19*(4), 629–640. https://doi.org/10.1016/j.gpb.2021.08.005



Supplementary Fig. 1. 3D structure model, sequence coverage, pLDDT score and PAE score of reference and haplotypes

Supplementary Table S1. List of 91 rice genotypes used as association panel in the study

S.No.	Genotype	Details	S.No.	Genotype	Details	S.No.	Genotype	Details
Gen_1	669	Mutant of N22	Gen_32	BAM 2841	Gen_63	DTY 9.1		
Gen_2	734	Mutant of N22	Gen_33	BAM 3154	Gen_64	Rasi		
Gen_3	868	Mutant of N22	Gen_34	BAM 326		Gen_65	FR 13A	Derived from the Dhal puttia landrace
Gen_4	2094	Mutant of N22	Gen_35	Thururbhog	Landrace	Gen_66	Gobindabhog	Short grain aromatic local
Gen_5	2335	Mutant of N22	Gen_36	BAM 328		Gen_67	HP 2216	Released variety
Gen_6	2347	Mutant of N22	Gen_37	BAM 4138		Gen_68	HUR 105	Mutant of MPR 7-2
Gen_7 collection	2889	Mutant of N22	Gen_38	BAM 4168		Gen_69	IC 576938	Gene Bank (NBPGR)
Gen_8	3247	Mutant of N22	Gen_39	BAM 4233		Gen_70	IC 576983	Gene Bank (NBPGR) collection
Gen_9	3474	Mutant of N22	Gen_40	BAM 5449		Gen_71	IET 25701	Local variety
Gen_10	3694	Mutant of N22	Gen_41	BAM 5645		Gen_72	IET 27625	
Gen_11	7347	Mutant of N22	Gen_42	PS 2	Released basmati variety	Gen_73	IR 36	Multiple cross derivative
Gen_12	PS 5	Mutant of N22	Gen_43	BAM 6067		Gen_74	IR 64	IR 5657-33-2-1/IR 18348-36-3-3
Gen_13	CR-4-27-5	Mutant of N22	Gen_44	BAM 6921		Gen_75	Kalajoha	Local cultivar
Gen_14	SL 275	Mutant of N22	Gen_45	2432	Mutant of N22	Gen_76	Kariagora	Local cultivar
Gen_15	Purnendu	Deepwater	Variety in Eastern India	Gen_46	BAM 712	Gen_77	kittake	*Japonica* cultivar
Gen_16	SL 42	Mutant of N22	Gen_47	BAM 7488		Gen_78	Lalat	Released variety
Gen_17	ADT 39	Short duration improved variety	Gen_48	BAM 766		Gen_79	MTU 1010	Released variety
Gen_18	Sahabhagi	Early drought tolerant variety	Gen_49	BAM 8055		Gen_80	MTU 1075	Released variety
Gen_19	Swarna Prabha	Low light tolerant variety	Gen_50	Pusa 44	Improved	Gen_81	N 22	Indica variety
Gen_20	APO	Aerobic variety	Gen_51	Rajendra Mahsuri	Improved	Gen_82	Narendradhan	Released variety
Gen_21	Azucena	Japonica Land race	Gen_52	BAM 8316		Gen_83	NL 44	NERICA cultivar
Gen_22 ety	BAM 1098		Gen_53	Chakho Poiteran	Local cultivar	Gen_84	PB 1121	Released basmati vari-
Gen_23	BAM 113		Gen_54	CR 1009	Improved	Gen_85	PB 1401	Released basmati variety
Gen_24	BAM 1264		Gen_55	CSR 27	Salt tolerant variety	Gen_86	PB 1509	Released basmati variety
Gen_25	Taipei 309	Japonica variety	Gen_56	CSR 30	Salt tolerant basmati variety	Gen_87	Pokkali	Salt tolerant landrace
Gen_26	BAM 1708		Gen_57	Dhan 42	Drought tolerant	Gen_88	SM Sub 1	Released variety
Gen_27	BAM 2096		Gen_58	Ranjit	Improved	Gen_89	Swarna	Released variety
Gen_28	BAM 2166		Gen_59	Dhan 51	Improved	Gen_90	Tetep	Local cultivar
Gen_29	BAM 2199		Gen_60	DRR Dhan 50	Submergence & drought tolerant at seedling & reproductive stages	Gen_91	Vandana	Early maturing upland from Vietnam
Gen_30	BAM 2614		Gen_61	DTY 2.2				
Gen_31	BAM 265		Gen_62	DTY 3.2				

Supplementary Table S2. List of Seven Haplotypes

Hap_1	Hap_2	Hap_3	Hap_4	Hap_5	Hap_6	Hap_7
1 669	4 2094	12 7347	19 Sahabhagi	29 BAM 2096	30 BAM 2166	42 BAM 4168
2 734	18 ADT 39	13 PS 5	27 Taipei 309		34 BAM 2614	74 IC 576938
3 868	22 Azucena	20 Swarna Prabha	38 BAM 326		44 BAM 5449	
5 Ps2	25 BAM 113	21 APO	53 BAM 7		66 46 PS 2	
6 2335	37 BAM 3154	31 BAM 2199	61 CSR 30		58 Chakho Poiteran	
7 2347	40 BAM 328	39 Thurubhog	80 Kalajoha		92 Pokkali	
8 2889	49 BAM 6921	43 BAM 4233	82 kittake		95 Tetep	
9 3247	56 Rajendra Mahsuri	48 BAM 6067	87 Narendra dhan			
10 3474	59 CR 1009	54 BAM 8055				
11 3694	62 Dhan 42	55 Pusa 44				
14 CR-4-27-5	63 Ranjit	57 BAM 8316				
15 SL 275	65 DRR Dhan 50	60 CSR 27				
16 Purnendu	67 DTY 3.2	66 DTY 2.2				
17 SL 42	70 FR 13A	69 Rasi				
24 BAM 1098	79 IR 64	72 HP 2216	a			
26 BAM 1264	85 MTU 1075	73 HUR 105				
28 BAM 1708	88 NL 44	75 IC 576983				
35 BAM 265	93 SM Sub 1	76 IET 25701				
36 BAM 2841	94 Swarna	77 IET 27625				
41 BAM 4138		78 IR 36				
45 BAM 5645		83 Lalat				
50 2432		84 MTU 1010				
51 BAM 712		89 PB 1121				
52 BAM 7488		91 PB 1509				
64 Dhan 51		96 Vandana				
68 DTY 9.1						
81 Kariagora						
86 N 22						
90 PB 1401						