Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Agro-morphological variability coupling trait-specific SNP genotyping guides identification of suitable candidate germplasm for improving nutritional quality in rice

Shuvendu Shekhar Mohapatra
Central Rice Research Institute
C Anilkumar
Central Rice Research Institute
Torit Baran Bagchi
Central Rice Research Institute
Shaswati Rout
Central Rice Research Institute
Manoj Kumar Panda
Berhampur University
Krishnendu Chattopadhyay
krishnenducrri@gmail.com
Central Rice Research Institute

Research Article

Keywords: Rice germplasm, genetic diversity, principal component analysis, molecular markers, biofortification

Posted Date: October 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10450265/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Genetic variation is the cornerstone of genetic improvement of quantitative traits in crops. Improvement of grain quality traits is challenging because intensive selection for quality often results in trade-offs with yield. Therefore, simultaneous improvement of yield and quality requires broad genetic variation and a strategic breeding approach. Our objective in this study was to identify potential donor genotypes combining high grain protein content with improved yield potential from a diverse genetic base. A panel of 435 rice accessions, including landraces, released varieties, and breeding lines from a nutritional breeding program, was evaluated for yield and related agro-morphological traits. Principal component analysis and cluster analysis were used to identify a representative subset while retaining genetic diversity. The selected genotypes were genotyped using trait-specific single nucleotide polymorphism markers and profiled for alleles associated with yield, grain quality, and stress tolerance. Substantial allelic diversity was detected, complementing phenotypic variation and enabling identification of genotypes carrying favorable alleles at key loci such as *Gn1a*, *GS3*, and *NAL1*. Genotypes with superior phenotypic performance and favorable allele combinations were identified as potential donors for biofortification and trait improvement. Elite genotypes including Dhinkiasali, Salibahan, CR Dhan 313, and Ratna showed high favorable allele loads, while Malabati, Jasmine, and CR Dhan 310 exhibited superior grain protein and quality traits. These results provide a comprehensive framework for strategic selection and utilization of rice germplasm as donors in enhancing productivity, resilience, and nutritional quality.

Introduction

The development of nutrient-rich crop varieties is the need of the hour. However, a major limitation is the trade-off between nutritional quality and productivity. Striking a balance between these two targets is essential to develop cultivars with both high nutritional value and high yield potential. In rice, biofortified high-protein varieties such as CR Dhan 310 have been released for cultivation in India (Chattopadhyay et al. 2019) demonstrating the feasibility of genetic enhancement of nutritional traits. However, further improvement in yield potential and stress tolerance is still required. The amalgamation of enhanced nutritional quality, high productivity, and climate resilience is crucial, as rice is highly vulnerable to climate change. Polygenic inheritance of these traits, their interactions with the environment, and the frequent occurrence of negative correlations among them complicates the simultaneous improvement. Achieving these goals necessitates the strategic evaluation and utilization of novel germplasm resources.

Genetic resources in rice form the backbone of genetic improvement programs. Agro-morphological characterization serving as a critical first step in harnessing their potential. Systematic evaluation and characterization of germplasm for agro-morphological, nutritional, and stress-responsive traits are critical for identifying superior donor lines and trait-specific genetic variation. Population improvement combining alleles from various superior donors for multiple traits and yield is essential to ensure the sustainability of production and commercial viability of new varieties. Such an approach also

strengthens the genetic foundation of breeding populations, ensuring sustained genetic improvement and gain (Awad-Allah et al. 2022).

Molecular resources including functional genetic markers available in rice help identification of genotypes with favorable allele combinations for multiple traits (Nachimuthu et al. 2015; Choudhury et al. 2023; Anjali et al. 2022; Zhang et al. 2023). Integration of molecular diversity information with phenotypic and nutritional data enhances parent selection, enables pyramiding of favorable alleles, and helps overcome negative correlations among complex traits. Genotyping with 1K RiCA genotyping array includes trait-specific single nucleotide polymorphism (SNP) for quality, yield and stress tolerance traits (Arbelaez et al. 2019). The cheaper cost of 1K RiCA array genotyping facilitates effective utilization in regular breeding programs for improvement of stress tolerance (Maniruzzaman et al. 2022) and quality traits (Cruz et al. 2021).

In the last decade, numerous studies emphasized the importance of agro-morphological and phenotypic diversity in rice germplasm for the development of high-yielding and resilient cultivars (Sar et al. 2024). The application of multivariate statistical tools to agro-morphological and molecular datasets enables robust interpretation of complex variation and supports informed decision-making in breeding programs (Bhattarai and Subudhi, 2019; John et al. 2023). Clustering and classification of genotypes based on combined phenotypic and molecular diversity help eliminate genetically redundant accessions and allow efficient rationalization of germplasm collections, thereby optimizing the choice of parental lines. Identification of genetically divergent and complementary clusters facilitates the strategic selection of parents with maximum recombination potential, enhancing the likelihood of transgressive segregants. Moreover, allelic profiling of representative genotypes from distinct clusters for known functional genes enables the targeted identification of superior donor lines carrying favorable alleles for nutritional quality, yield, and stress tolerance. Such an integrated approach directly strengthens parental selection, accelerates trait introgression, and improves the efficiency and predictability of developing elite rice cultivars with multiple desirable traits.

The objectives of the present study were to: (i) characterize genetically diverse rice germplasm through comprehensive agro-morphological evaluation and narrow it down to a representative set of genotypes; (ii) assess grain protein content in the selected representative genotypes; and (iii) elucidate allele profiles of known functional genes in these genotypes. The outcomes of this study are expected to facilitate the identification of genetically diverse and complementary parents for multi-trait mapping and breeding, as well as to identify superior donor lines combining high grain quality, stress tolerance, and yield potential. This integrated approach aims to address the trade-off between nutritional quality and productivity and to accelerate the development of improved rice cultivars.

Material and Methods

Plant Materials and experimental layout

A diverse panel of 435 rice germplasm consisting of landraces (284), released varieties (116) and breeding lines (35) were considered for this study (Supplementary Table 1). Varieties and landraces, collected from all over India and maintained in ICAR-Central Rice Research Institute (CRRI) Gene Bank were considered for this study. Breeding lines were drawn from breeding programme for improvement of grain nutritional quality. Genetic material represents geographical regions and different cultivation ecologies. Seedlings of 21–25 days old were transplanted to the experimental plot of ICAR-CRRI, Cuttack, Odisha (latitude- 20.448165° N and longitude- 85.932977° E) in wet-season 2023. Two rows for each genotype were planted in two replications with the spacing of 20 cm (row to row) × 15 cm (plant to plant) and grown using standard package of practices. Observation are recorded on days to 50% flowering, plant height (cm), number of tillers/plant, number of panicles/plant, panicle length (cm), 100 grain weight (g) and grain yield/plant (g) were recorded from randomly selected 5 plants from each plot following standard evaluation procedures (IRRI 2014).

Identification of a representative panel of 150 genotypes

Agro-morphological data of 450 genotypes was subjected to preliminary analysis. Data were first subjected to descriptive statistical analysis to determine the population mean, range, standard deviation, and coefficient of variation (CV%). Pearson's correlation coefficients were computed among all trait pairs to evaluate the direction and strength of linear associations. Principle component analysis, a multivariate statistic, was performed to reduce the dimensionality of variation in the population. Major contributing traits to total genetic variance were identified based on trait loadings on the principle components. Furthermore, genotypes in the panel were clustered using K-means method based on Euclidian distance.

Based on the genetic diversity and variance in different clusters a representative set of 150 genotypes (Supplementary Table 2) was selected from the panel. This representative set was used for further experimentation.

Estimation of grain protein content

The representative set was raised in two replicated plots in wet-season 2024 using same package of practices. Grains of selected 150 genotypes from wet-season 2023 and wet-season 2024 were analysed for the grain protein content. The dehusked rice samples (20g) were subjected to estimation of protein content using calibrated NIR spectroscopy (Bagchi et al. 2016). Protein yield/plant (g) was estimated by multiplying grain protein content (GPC) (in 1 g) and grain yield/plant (g).

SNP Genotyping using gene-specific markers

The representative set of 150 genotypes comprising of 22 breading lines, 49 landraces and 79 varieties was subjected to trait-specific SNP genotyping (Supplementary Table 3). Young and healthy leaves were collected from each plant for DNA extraction. Leaf tissues were punched into uniform discs using a

sterile leaf puncher to ensure consistency in sample size. The leaf discs were shipped to service provider for DNA extraction and genotyping. Extracted and purified DNA was subjected to genotyping by using 128 SNP markers of genes related to abiotic, biotic stress tolerance, grain yield and quality traits within the 1K RiCA SNP genotyping panel (Arbelaez et al. 2019). The genotyping service was carried out by AgriPlex Genomics Cleveland, USA by the service provider INTERTEK India Pvt. Ltd., Hyderabad, Telangana, India. The genotypes in the representative set were sorted based on favorable alleles and genotypes with more favorable alleles for target traits were identified.

Results

Genetic variation for yield traits in the population

The population showed a significant genetic variation for all the quantitative traits. The analysis of variance (ANOVA) was performed for quantitative traits such as days to 50% flowering (DFF), plant height (cm) (PH), panicle length (cm) (PL), tiller number/plant (TN), panicle number/plant (PN), hundred-grain weight (HGW), and single plant yield (SPY). ANOVA revealed that all traits were found significant differences ($p < 0.05$) among rice genotypes (Supplementary Table 4). A wide range of values of each of the traits and higher genotypic and phenotypic coefficient of variances indicated the significant level of variability in the studied germplasm.

Correlation matrix showing the degree of association among the yield and yield attributing traits is presented in Fig. 1. Single plant yield (SPY) exhibited significantly high positive correlations with the number of productive tillers/plant ($r = 0.906$, $p < 0.01$), number of tillers/plant ($r = 0.872$, $p < 0.01$), panicle length ($r = 0.275$, $p < 0.01$) and hundred-grain weight ($r = 0.467$, $p < 0.01$). These findings highlighted that grain yield in this germplasm panel was primarily driven by enhanced tiller productivity, panicle architecture and grain weight. Days to 50% flowering (DFF) showed a significant negative association with panicle length ($r = -0.204$, $p < 0.01$), hundred-grain weight ($r = -0.345$, $p < 0.01$) and single plant yield ($r = -0.139$, $p < 0.01$). Panicle length (PL) exhibited positive correlations with tiller number ($r = 0.208$, $p < 0.01$) productive tiller number ($r = 0.207$, $p < 0.01$), hundred-grain weight ($r = 0.255$, $p < 0.01$) and single plant yield ($r = 0.275$, $p < 0.01$), highlighting its contribution to grain yield.

Agro-morphological diversity and trait-contributions to variation

Principal Component Analysis (PCA) was performed to investigate the contribution of agro-morphological traits to the total variation among 435 rice germplasm. The scree plot revealed that the first two principal components, PC1 and PC2 accounted for 43.63% and 20.99% of the total variance, respectively, cumulatively explaining 64.62% of the overall variability among the accessions. Eigenvalues of rest of the principal components (Supplementary Fig. 1) indicated that these two components captured the majority of meaningful variation, justifying their use for subsequent biplot interpretation.

The PCA biplot (Fig. 2) illustrated the contribution and interrelationship of the studied traits to the overall variability. Traits such as days to fifty percent flowering (DFF) and plant height (PH) showed strong loadings along PC1, whereas panicle length (PL) and hundred-grain weight (HGW) were positively associated with PC2. The PCA scatter plot (Supplementary Fig. 2) demonstrated the distribution pattern of breeding lines, landraces, and released varieties based on the first two principal components. Although a considerable overlap was observed among the three groups. Landraces exhibited wider dispersion, reflecting greater phenotypic diversity, whereas breeding lines and varieties clustered more closely, suggesting selective narrowing of variability through genetic improvement. Released varieties were mainly located in the outer portion of 2D plot while landrace located inside the plot. Overall, the PCA results highlight that agro-morphological variability in the studied rice germplasm is largely influenced by flowering behavior, plant stature, and yield-contributing traits. The distinct yet partially overlapping group distributions reflect both the genetic differentiation and shared ancestry among breeding lines, landraces, and improved varieties.

Euclidean K-means clustering analysis grouped the genotypes into distinct clusters based on similarity in trait profiles. The clustering pattern indicates the presence of both closely related and highly distinct genotypes, offering opportunities for selecting diverse parental combinations to enhance transgressive segregants and genetic gain. The resulting dendrogram, visualised in a circular layout, revealed substantial genetic diversity within the panel. Euclidean distance as the similarity measure, the 435 genotypes were grouped into five distinct clusters (Fig. 3). Cluster I being the largest with 203 genotypes was followed by Cluster II (88 genotypes), Cluster V (79 genotypes), Cluster III (49 genotypes), and Cluster IV (16 genotypes). The clustering pattern aligned with the PCA groupings, where TN, PN, PL, and DFF emerged as the primary traits contributing to the discrimination among clusters. No clear demarcation in 5 clusters based on types of genotypes and all types of germplasm, i.e. landraces, varieties and breeding lines were distributed among clusters.

A sub-set of 150 representative genotypes were selected from all five clusters and grown in the next year (Supplementary Table 2). The across-year mean grain protein content ranged from 4.89% to 13.33% with a mean of 8.50% and standard deviation 1.58%, indicating significant phenotypic diversity for GPC. Correlation coefficient matrix (Fig. 4) derived from association among yield attributing traits and grain protein content of 150 selected genotypes revealed that grain yield had significant ($p < 0.01$) positive association with number of panicles/plant, grain weight and panicle length. But GPC did not show any significant association with yield and yield attributing traits. Standard deviation was highest in cluster III (~ 1.85) and lowest in cluster V (~ 0.44), indicating that Cluster-III captured the larger diversity for GPC and was particularly useful for selecting both high and low protein types. The across-year mean of grain yield/plant ranged from 6.38g to 38.07g with a mean of 18.7g and standard deviation of 7.02, indicating moderate to high variability among selected germplasm for grain yield. Cluster III followed by Cluster II, characterized by the high mean GPC (8.75% and 8.61%), and protein yield (1.71 g/plant and 1.77 g/plant) and elite genotypes belonged to theses clusters containing both high protein and yield could be used as donor for improving both grain productivity and nutritional value. Cluster I, despite moderate GPC and protein yield, contained the highest number of genotypes with diverse nature, valuable for broadening the

genetic base and introducing variability for recombination breeding. Cluster IV also exhibited intermediate GPC (8.49%) and protein yield (1.48 g/plant). In contract, Cluster V had lowest mean GPC and protein yield among all clusters.

Allelic diversity of germplasm for trait-specific gene based markers

Selected 150 rice germplasm were genotyped with 128 trait-specific SNP markers distributed across the 12 chromosomes (Supplementary Table 3). These SNPs were specific to genes or QTLs linked with grain yield, flowering time, panicle architecture, grain quality, biotic or abiotic stress tolerance. The genomic location and distribution of these SNPs are shown in Supplementary Fig. 3. The genotype-wise allele profiles and frequency of each marker allele in the population are presented in Fig. 5 and Supplementary Fig. 4A, respectively. Percentage of favourable alleles of SNPs associated with genes for economically important traits in rice genotypes is presented in Supplementary Fig. B. The distribution of allelic classes (favorable, unfavorable, and heterozygous alleles) across the tested markers revealed substantial genetic variability among the rice germplasm for target traits (Fig. 5). Several markers of the QTLs such as *DTY1*-1, *DTY2*-2, and *DTY3*-2 displayed a distribution of favorable and unfavorable alleles, indicating greater opportunity for accumulating favorable alleles through combinatorial breeding. Conversely, several loci associated with yield contributing traits such as *Gn1a*, *GS3*, *GW5*, and *NAL1* showed a higher proportion of favorable alleles, suggesting prior selection of these loci in breeding programs. In contrast, genes such as *Hd1*, *Hd3a*, and *Wx* exhibited a greater frequency of unfavorable or mixed alleles, pointing toward untapped opportunity for genetic improvement.

Notably, the presence of heterozygous alleles was relatively low across most markers, reflecting the predominantly inbred nature of the genotypes tested. However, certain loci (*PRR37*, *Ghd7*, and *DTY12*-1) showed appreciable heterozygosity, suggesting potential introgressions or residual heterozygosity in breeding lines and landraces. The allelic composition of each genotype for the 128 SNPs was visualized to evaluate cumulative allele stacking (Fig. 5). Considerable variation was observed across the 150 genotypes, with some lines harbouring predominantly favorable alleles across multiple yield QTLs. Elite cultivars such as IR64, Swarna, and N22 derivatives displayed mixed allele profiles, with favorable alleles at major yield loci but unfavorable alleles at grain quality (Wx) or flowering-time (Hd) related markers.

Identification of donors with favorable allele for functional genes

Several genotypes exhibited a markedly higher density of advantageous loci linked to biotic stress resistance, abiotic stress tolerance, and grain quality enhancement (Supplementary Table 5), positioning them as elite genomic resources for rice improvement. Notably, Dhinkiasali (M97), Salibahan (M71), CR Dhan 313 (M200), Ratna (M57), Gayatri (M123), Nuakalazeera (M43), CR Dhan 317 (M375), and Banskathi (M50) carried the highest favorable allele loads (79–85 alleles), reflecting broad genetic potential and a high likelihood of stable agronomic performance across environments. Dhinkiasali (M97)

emerged as the most enriched genotype, closely followed by Salibahan (M71) and CR Dhan 313 (M200), all of which showed strong representation of loci governing multi-stress resilience. These genotypes harbored key biotic stress resistance genes, including *Pi2, Pi9, Pid2, Pid3, Xa7, Xa21, BPH17, BPH32*, and RYMV resistance alleles, conferring broad-spectrum protection against rice blast, bacterial blight, brown planthopper, and viral diseases. Such genetic configurations make them highly suitable donors for developing durable pest- and disease-resistant cultivars. In parallel, several elite lines possessed critical abiotic stress tolerance loci, particularly Saltol (salinity tolerance), Sub1 (submergence tolerance), and major drought-yield QTLs such as *DTY1*.1, *DTY2*.2, and *DTY3*. The co-occurrence of multiple abiotic stress alleles in genotypes including CR Dhan 203 (M374), CR Dhan 210 (M201), Kamini (M128), and MP-1-114 (M70) suggests enhanced adaptability to variable and stress-prone agro-ecosystems, reinforcing their value in climate-resilient breeding pipelines.

Grain quality allele profiling further identified genotypes with superior end-use and nutritional attributes. Lines such as Naveen (M34), ARC-10063 (M14), Ratna (M57), and Satabdi (M32) carried favorable alleles for major quality-determining genes, including *Wx* (amylose content), *GS3* and *GW5* (grain size and shape), *Chalk5* (chalkiness reduction), *SSIIa* (gelatinization temperature), and *TGW6* (grain weight). Combining agro-morphological and nutritional traits such as grain yield, grain protein and protein yield with the presence of favourable alleles of important biotic and abiotic stress tolerance and superior plant type has been utilized to identify donors in breeding for nutritional quality (Table 1). Importantly, genotypes like Malabati (M103), Jasmine (M64), MP-1-114 (M70), and CR Dhan 310 (M39) combined high mean GPC (≥ 10.6%) with superior grain yield and a rich complement of yield-, nutrition-, quality-, and stress-associated SNPs, underscoring their potential as multi-trait donor parents.

Discussion

Agro-morphological diversity among germplasm

In India and Asia as a whole, rice is the key to the food and nutritional security. Due to the overemphasis on improving grain yield, often at the expense of stress tolerance and grain quality, both growers and consumers of present-day high-yielding varieties suffer substantially as a result of climate change and malnutrition. To address the problem, breeding scheme has been developed to simultaneously improve grain yield and nutritional traits such as protein content (Chattopadhyay et al. 2018). Similarly, many near isogenic lines (NILs) has been released with biotic and abiotic stress tolerance in the high yielding background (Dixit et al. 2020). Identification and deployment of donors for quality breeding programme to broadening genetic base and development of biofortified climate resilient varieties are gaining importance. Biofortified rice will only be widely adopted if high micronutrient content is combined with competitive yields, multiple stress tolerance, and grain quality that matches current farmer and market favourites (Sanjeeva Rao et al. 2020). A large-scale diversity studies using national and international germplasm repositories has potentiality to add significant value to breeding programs. The use of multivariate statistics continues to aid in reducing data dimensionality and identifying key traits for selection, while correlation studies enhance understanding of yield determinants and selection of desired traits for varietal development. In the present study, the phenotypic evaluation of 435 *indica* rice

genotypes revealed substantial variation in yield and yield contributing traits. Correlation coefficient analyses also supported these findings, providing additional insight into the relationships between traits. The correlation coefficients confirmed positive association of plant yield with number of productive tiller, panicle length and hundred-grain weight. These results were in agreements with reported results in diverse rice populations (Meena et al., 2020; Subramanian et al. 2025).

The cluster analysis grouped the genotypes into five distinct clusters with clear phenotypic differentiation, largely consistent with PCA patterns. Classification of genotypes into different clusters which endorse the significant morphological differences indicate the substantial genetic diversity in the germplasm panel (Verma et al., 2023). This genetic diversity is suitable for identifying novel allelic combinations contributing to yield and yield attributing traits.

Identification of diverse genotypes with differential protein yield

Quantitative differences between grain yield and micronutrient content traits indicate exploitable QTLs are positively associated, allowing simultaneous improvement through careful QTL pyramiding and selection (Prasad et al. 2025). From a breeding perspective, the observed clustering reveals clear opportunities for strategic parental selection to improve grain yield, grain protein content (GPC), and protein yield simultaneously. This approach of parental selection was reported earlier by Chattopadhyay et al. (2018) for improvement of protein content and gain yield in rice. The genotypes with high protein yield such as M 26 (Pandari Gurmatiya) (2.42 g/plant), M39 (CR Dhan 310) (2.85 g/plant) belonged to Cluster II and M 70 (MP-1-114) (2.55g/plant), M-80 (PB-140) (3.08 g/plant), M-165 (Kalinga-II) (2.65 g/plant) belonged to Cluster III could be hybridized with high yielding genotypes such as M318 (CR Dhan 318) with 1.00 g/plant protein yield of Cluster V having low grain protein and protein yield. This type cross combinations allow superior transgressive segregants to produce high grain protein content and yield. These combinations help accumulate gene combination to overcome trade-off between protein content and yield.

Allele profiling in diverse germplasm sub-set

Marker-assisted breeding are now used to stack micronutrient QTLs with major genes/QTLs for abiotic and biotic stress tolerance while maintaining an elite, high-yield background (Mohanavel et al. 2024). The agro-morphological diversity has been complemented by molecular diversity in the current investigation. Trait-specific SNP markers play a crucial role in rapid identification of genotypes for different traits based on molecular markers and deploy in breeding activities (Shanmugam et al., 2025). 1K SNP chip has been instrumental with trait-specific SNPs for grain yield, biotic, abiotic stress tolerance and other economical important traits (Arbelaez et al. 2019). The distribution of favorable and unfavorable alleles at loci such as *DTY1*-1, *DTY2*-2, and *DTY3*-2 indicates moderate frequency of yield-enhancing variants within the studied panel. In contrast, the subset was enriched with favorable allele at loci including *Gn1a, GS3, GW5*, and *NAL1*, suggesting prior selection for yield improvement in breeding programs. The low frequency of favorable alleles at key loci such as *Hd1, Hd3a*, and *Wx* highlights the importance of identifying suitable

donor lines for improving quality traits. A few landraces and breeding lines, particularly those carrying favorable alleles for *Gn1a*, *GS3*, *NAL1*, and *DTY* loci, and favorable alleles for quality traits emerged as promising donors for allele pyramiding. The relatively low heterozygosity across loci reflects the inbred nature of most genotypes, although loci such as *PRR37*, *Ghd7*, and *DTY12*-1 showed evidence of introgression or residual heterozygosity, indicating possible utility in recombination breeding.

High yielding nutrient-rich genotypes with positive alleles of biotic-abiotic stress tolerant and yield -related genes

The identification of landraces and breeding lines carrying favorable alleles for major yield loci highlights their potential as pre-breeding resources for allele pyramiding and genomic-assisted selection. Furthermore, enhancing grain protein content is in staple food like rice is significant, as recent studies emphasize its importance in improving rice nutritional quality (Jukanti et al. 2025; Chen et al. 2022). These results reinforce the importance of combining SNP-based diversity analysis with phenotypic and biochemical data to accelerate the development of rice varieties that ensure yield stability while addressing nutritional challenges. Overall, the integration of favorable allele mapping enabled precise identification of multi-dimensional elite genotypes with strong potential for use in breeding programs aimed at enhancing stress resilience, productivity, and grain quality. The top-performing genotypes, enriched with advantageous alleles across biotic, abiotic, and quality-related pathways, provide valuable genetic resources for accelerating varietal improvement and boosting genetic gains in rice.

Conclusion

The integration of agro-morphological evaluation and trait-specific SNP genotyping provided a robust framework to elucidate the extent and structure of genetic diversity among rice germplasm. The results underscore the coexistence of wide phenotypic variability and substantial allelic diversity, reflecting a rich genetic base that can be strategically harnessed for improving yield, nutritional quality, and stress resilience. The clustering patterns and allele distribution across functional loci identified specific genotypes as valuable donors for future biofortification and climate-resilient breeding programs.

The findings emphasize that combining morphological characterization with targeted genomic profiling enables precise selection of parental lines for gene pyramiding and genomic-assisted breeding. The diverse set of donors identified in this study including landraces and pre-breeding lines offers significant promise to strengthen the genetic foundation of rice improvement efforts, supporting the development of high-yielding, nutrient-enriched, and climate-resilient varieties suited to diverse agro-ecological regions.

Declarations

Authors contribution

Conceptualize, design the study, Supervision: KC, MKP; Writing original dtaft: SSM, KC; Review and editing the draft: KC, AKC; Methodology, formal analysis, data curation, visualization, software: TBB, SR, AKC, KC, SSM

Acknowledgements

Authors acknowledge the logistic support provided by the Director, ICAR-CRRI, Cuttack and for the financial support extended by the CRP in Biofortification project, ICAR, India.

Funding

This research was supported by the project 'CRP in Biofortification in selected crops', Indian Council of Agricultural Research (ICAR)

Ethics statement

The authors have nothing to report

Conflicts of interest

The authors declare that they have no conflicts of interest or competing interests.

Data availability statement

The datasets generated and analyzed during the current study are available in supplementary file and also from the corresponding author on reasonable request.

References

1. Anjali N, Krishnan SG, Vinod KK, Bhowmick PK, Atla G, Ellur RK, Mondal S et al (2022) Genetic diversity and population structure of rice landraces from Kerala, India analysed through genotyping-by-sequencing. Plant Genet Resource 20(1):41–50. https://doi.org/10.1017/S1479262121000345
2. Arbelaez JD, Dwiyanti MS, Tandayu E, Llantada K, Jarana A, Ignacio JC, Platten JD, Cobb J, Rutkoski JE, Thomson MJ, Kretzschmar T (2019) 1k-RiCA (1K-Rice Custom Amplicon) a novel genotyping amplicon-based SNP assay for genetics and breeding applications in rice. Rice 12:55. https://doi.org/10.1186/s12284-019-0311-0
3. Awad-Allah MMA, Shafie WWM, Alsubeie MS, Alatawi A, Safhi FA, ALshamrani SM, Albalawi DA, Al-Amrah H, Alshehri D, Alshegaihi RM, Basahi MA, Masrahi AS (2022) Utilization of genetic resources, genetic diversity and genetic variability for selecting new restorer lines of rice (Oryza sativa L). Genes 13(12):2227. https://doi.org/10.3390/genes13122227
4. Bagchi TB, Sharma SG, Chattopadhyay K (2015) Development of NIRS models to predict protein and amylose content of brown rice and proximate compositions of rice bran. Food Chem 191:21–27. https://doi.org/10.1016/j.foodchem.2015.05.038

5. Bhattarai U, Subudhi PK (2019) Genetic diversity, population structure, and marker-trait association for drought tolerance in US rice germplasm. Plants 8:530. https://doi.org/10.3390/plants8120530
6. Chattopadhyay K, Sharma SG, Bagchi TB, Molla KA, Sarkar S, Marndi BC, Sarkar A, Dash SK, Singh ON (2018) Development of recombinant high yielding lines with improved protein content in rice (Oryza sativa L). J Agric Sci 1–17. https://doi.org/10.1017/S0021859618000230
7. Chattopadhyay K, Sharma S, Bagchi TB, Mohanty B, Sardar SS, Sarkar S, Singh ON (2019) High-protein rice in high-yielding background, cv. Naveen Curr Sci 117(10):1722–1726
8. Chen P, Lou G, Wang Y, Chen J, Chen W, Fan Z, Liu Q et al (2022) The genetic basis of grain protein content in rice by genome-wide association analysis. Mol Breed 43(1):1. https://doi.org/10.1007/s11032-022-01347-z
9. Choudhury DR, Kumar R, Maurya A, Semwal DP, Rathi RS, Gautam RK, Trivedi AK et al (2023) SSR and SNP marker-based investigation of Indian rice landraces in relation to their genetic diversity, population structure, and geographical isolation. Agriculture 13(4):823. https://doi.org/10.3390/agriculture13040823
10. Cruz M, Arbelaez JD, Loaiza K, Cuasquer J, Rosas J, Graterol E (2021) Genetic and phenotypic characterization of rice grain quality traits to define research strategies for improving rice milling, appearance, and cooking qualities in Latin America and the Caribbean. Plant Genome 14:e20134. https://doi.org/10.1002/tpg2.20134
11. Dixit S, Singh UM, Singh AK, Alam S, Venkateshwarlu C, Nachimuthu VV et al (2020) Marker assisted forward breeding to combine multiple biotic–abiotic stress resistance/tolerance in rice. Rice 13(1):29. https://doi.org/10.1186/s12284-020-00391-7
12. International Rice Research Institute (IRRI) (2014) Standard evaluation system for rice, 5th edn. IRRI, Los Baños
13. John R, Bollinedi H, Jeyaseelan C, Padhi SR, Sajwan N, Nath D, Singh R et al (2023) Mining nutri-dense accessions from rice landraces of Assam, India. Heliyon 9(7):e17524. https://doi.org/10.1016/j.heliyon.2023.e17524
14. Jukanti AK, Karapati D, Bharali V, Gudla M, Thati S, Yadla YS, Kumar M, Sundaram RM (2025) From gene to plate: Molecular insights into and health implications of rice grain protein. Int J Mol Sci 26(7):3163. https://doi.org/10.3390/ijms26073163
15. Maniruzzaman S, Rahman MA, Hasan M, Rasul MG, Molla AH, Khatun H, Akter S (2022) Genetic mapping to detect stringent QTLs using 1k-RiCA SNP genotyping platform from the new landrace associated with salt tolerance at the seedling stage in rice. Plants 11(11):1409. https://doi.org/10.3390/plants11111409
16. Meena D et al (2020) Assessment of correlation and path coefficient analysis for grain yield and its attributing traits in rice (Oryza sativa L). Int J Curr Microbiol Appl Sci 9(7):64–71
17. Mohanavel V, Muthu V, Kambale R, Palaniswamy R, Seeli P, Ayyenar B et al (2024) Marker-assisted breeding accelerates the development of multiple-stress-tolerant rice genotypes adapted to wider environments. Front Plant Sci 15:1402368. https://doi.org/10.3389/fpls.2024.1402368

18. Nachimuthu VV, Muthurajan R, Duraialaguraja S, Sivakami R, Pandian BA, Ponniah G et al (2015) Analysis of population structure and genetic diversity in rice germplasm using SSR markers. Rice 8:30. https://doi.org/10.1186/s12284-015-0062-5
19. Prasad GR, Rao CS, Muralidharan K, Sridhar R, Siddiq EA (2025) Assessing two decades of breeding for biofortified rice with zinc, iron or protein. J Agric Food Res 24:102317. https://doi.org/10.1016/j.jafr.2025.102317
20. Sanjeeva Rao D, Neeraja CN, Madhu Babu P, Nirmala B, Suman K, Rao LVS et al (2020) Zinc biofortified rice varieties: Challenges, possibilities, and progress in India. Front Nutr 7:26. https://doi.org/10.3389/fnut.2020.00026
21. Sar P, Gupta S, Behera M, Lenka S, Panda A, Raghuwanshi R, Kar MK, Behera TK, Kariali E, Sarla N (2024) Exploring genetic diversity within aus rice germplasm. Rice 17:20. https://doi.org/10.1186/s12284-024-00700-4
22. Shanmugam S, Samonte SOPB, Sanchez DL et al (2025) Genetic assessment of diverse rice germplasm for yield-related and stress tolerance traits. Genet Resour Crop Evol 72:885–904. https://doi.org/10.1007/s10722-024-02004-4
23. Subramanian S, Ramamoorthy P, Alagesan S, Amalraj JJ, Muthurajan R, Alagarsamy S et al (2025) Unraveling the genetic potential of Indian rice germplasm for reproductive stage drought tolerance. Front Plant Sci 16:1454299. https://doi.org/10.3389/fpls.2025.1454299
24. Verma H, Theunuo S, Das SP, Kumar A, Aochen C, Devi L et al (2023) Genetic diversity and marker-trait association analysis for grain quality, yield and yield attributes in hill rice. Indian J Genet Plant Breed 83(1):1–12. https://doi.org/10.31742/IJGPB.83.1.1
25. Zhang C, Li M, Liang L, Xiang J, Zhang Y (2023) Rice3K56 is a high-quality SNP array for genome-based genetic studies and breeding in rice (Oryza sativa L). Crop J 11(2):800–810. https://doi.org/10.1016/j.cj.2023.02.006

Tables

Tables are available in the Supplementary Files section.

Figures

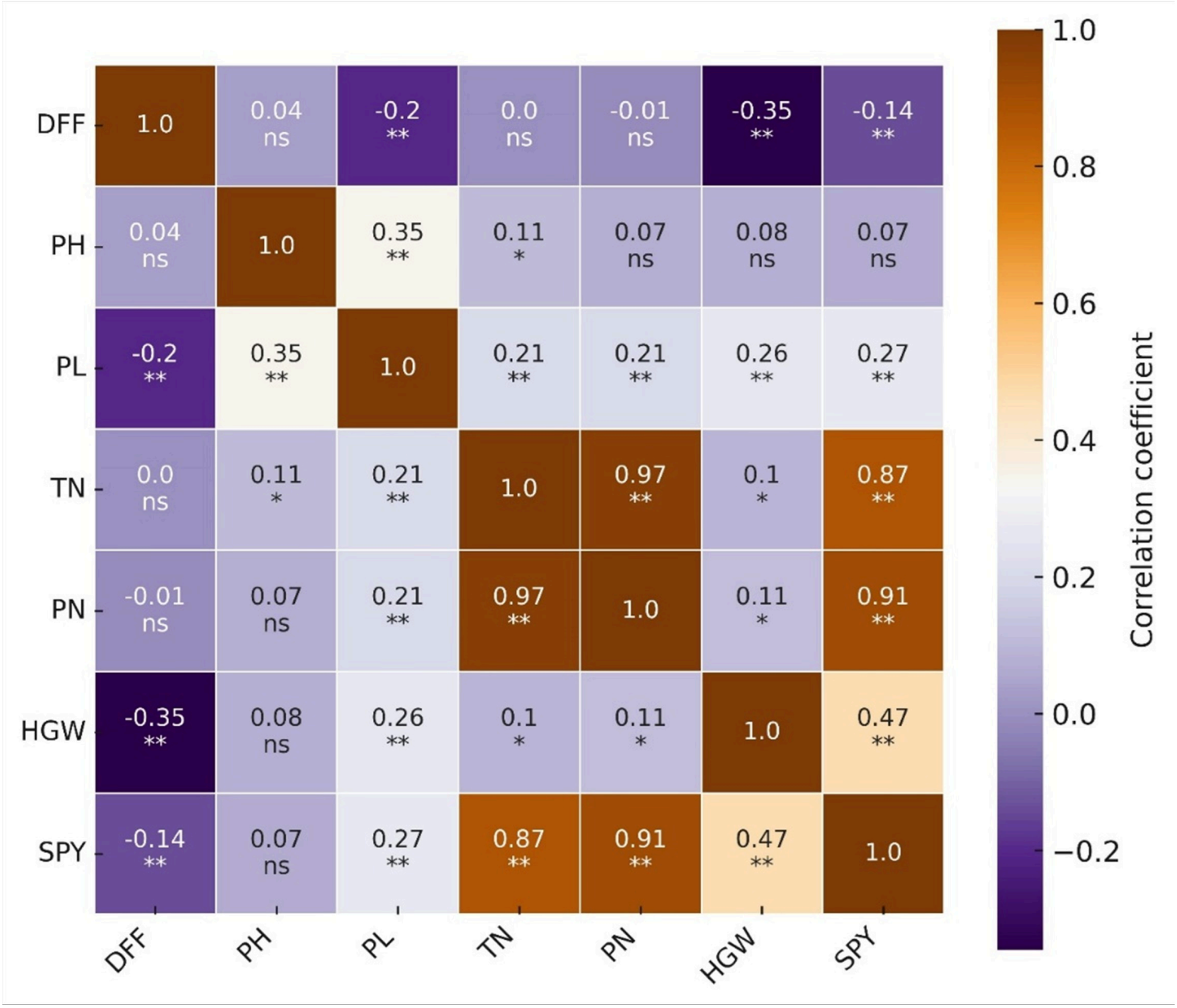


Figure 1

Correlation coefficient matrix of yield and yield-attributing traits in rice derived from 435 rice genotypes. Values inside the boxes represent Pearson's correlation coefficients, with significance levels indicated as $p < 0.05$ (*), $p < 0.01$ (**), and non-significant (NS). Traits include days to 50% flowering (DFF), plant height (PH), panicle length (PL), tiller number/plant (TN), panicle number/plant (PN), hundred grain weight (HGW) and single plant yield (SPY).

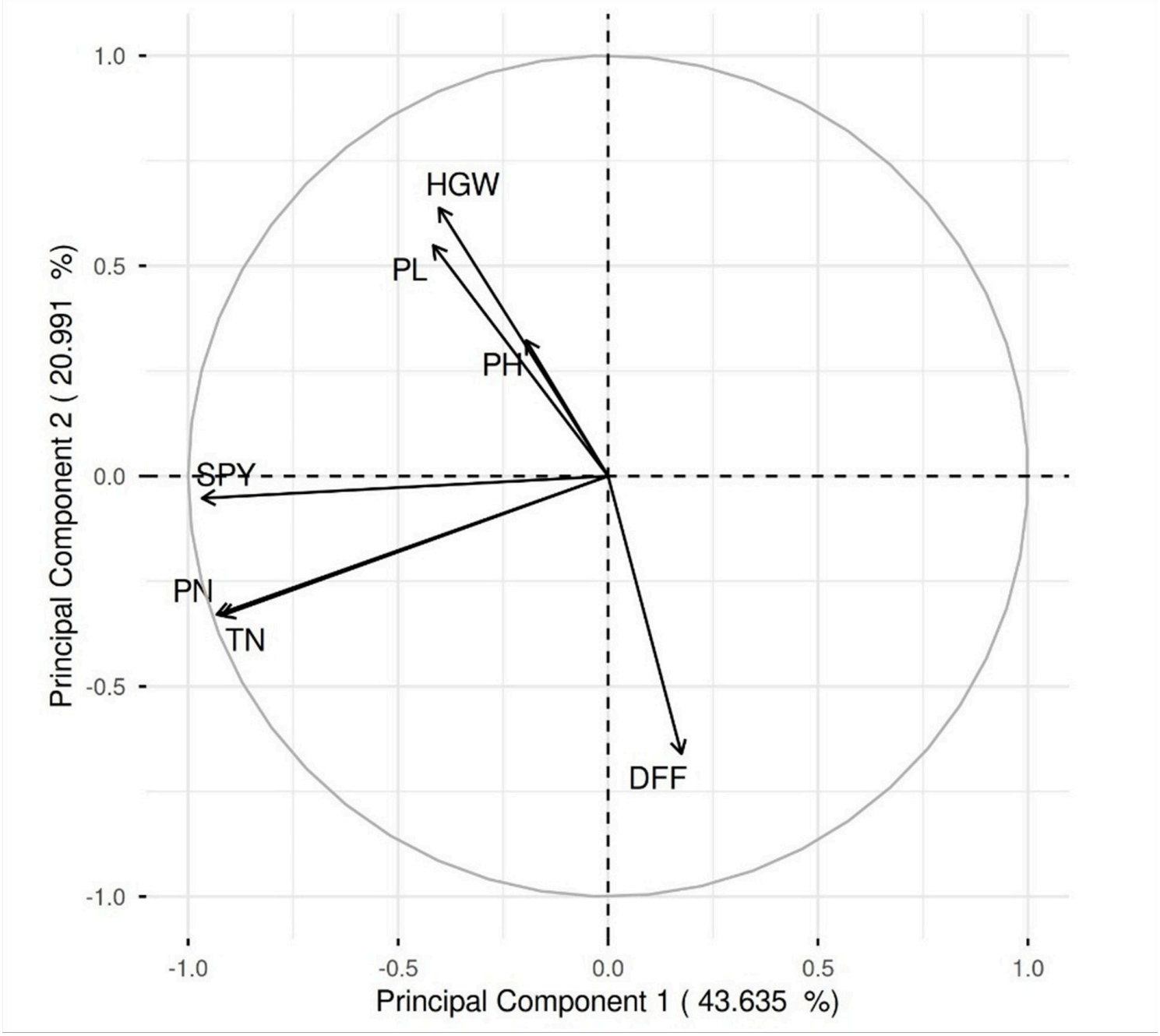


Figure 2

Principal Component Analysis (PCA) of 435 rice germplasm accessions based on agro-morphological traits; Location of traits in 2-D graph generated based on principal component analysis (PCA) taking first two principal components, PC1 (43.63%) and PC2 (20.99%) of the total variation

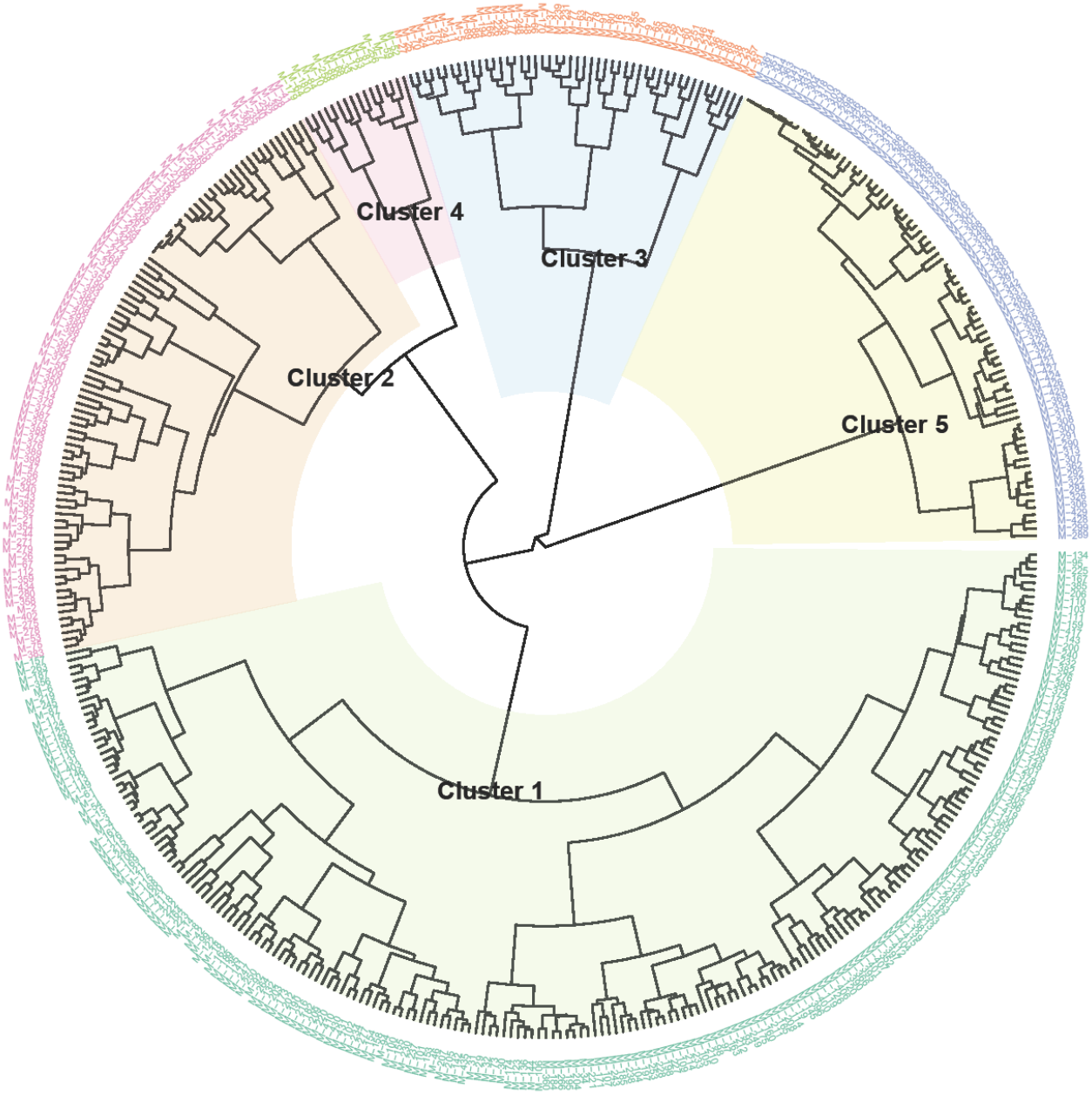


Figure 3

Phylogenetic tree based on Hierarchical clustering using Euclidean distance taking yield and yield attributing traits in rice

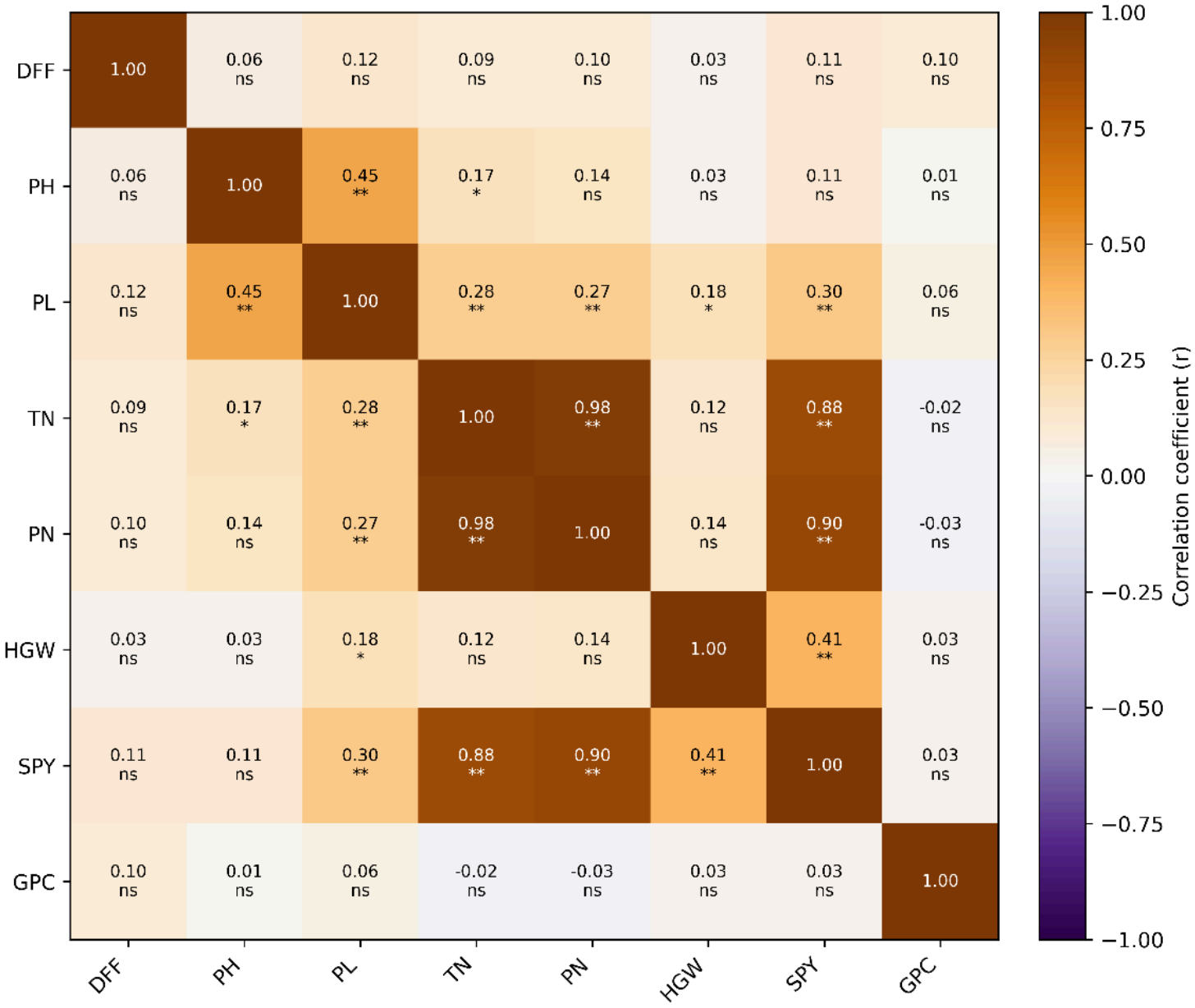


Figure 4

Correlation coefficient matrix of yield, yield-attributing traits and grain protein content derived from the selected 150 rice genotypes. Values inside the boxes represent Pearson's correlation coefficients, with significance levels indicated as $p < 0.05$ (*), $p < 0.01$ (**), and non-significant (NS). Traits include days to 50% flowering (DFF), plant height (PH), panicle length (PL), tiller number/plant (TN), panicle number/plant (PN), hundred grain weight (HGW), single plant yield (SPY) and grain protein content (GPC).

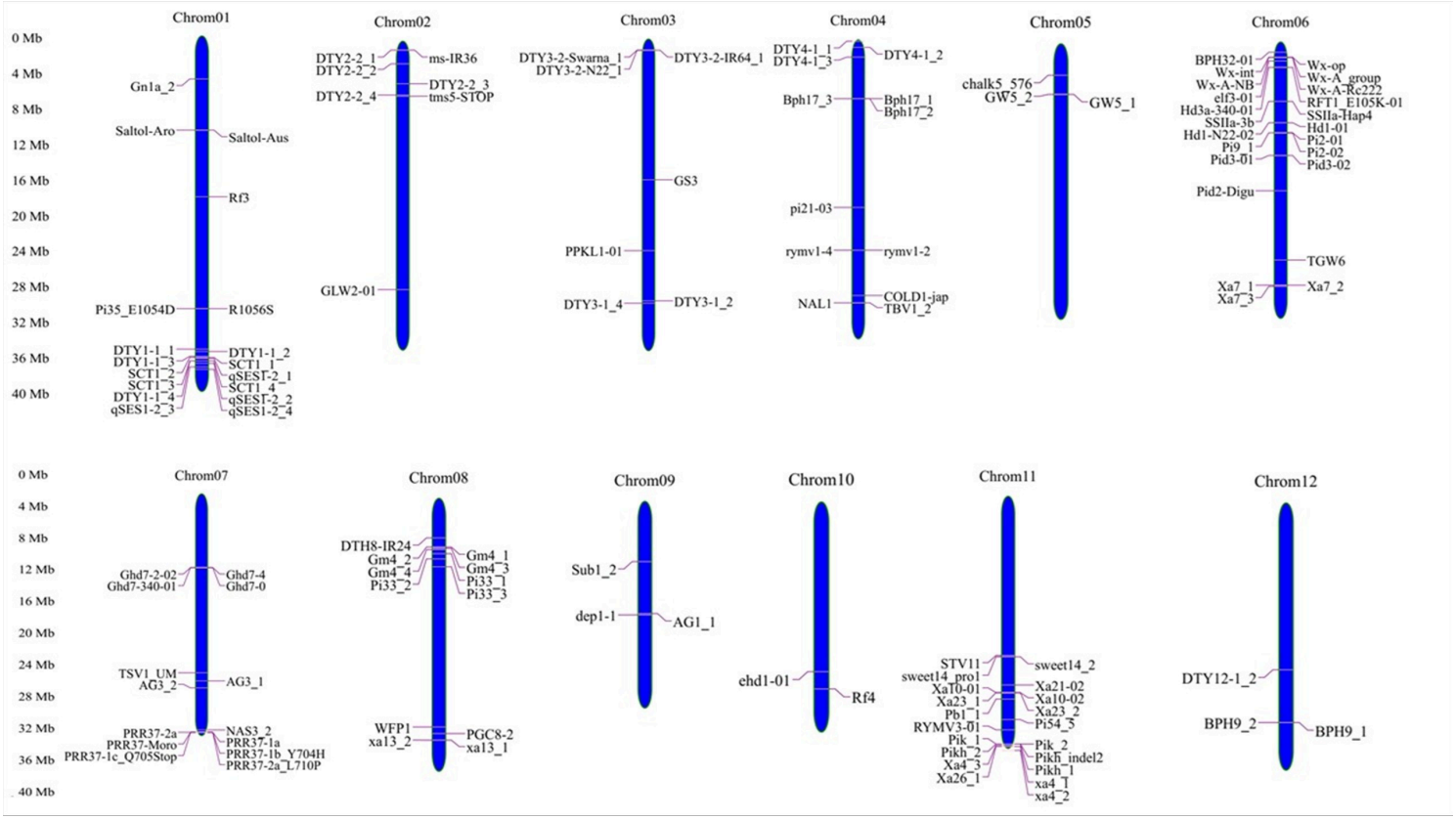


Figure 5

Chromosomal distribution of 128 SNPs markers associated with important traits in rice across all 12 chromosomes. Each chromosome is represented as a vertical bar, with markers indicated at their respective positions in mega base pairs (Mb). (Details of genes and markers are available in Table S3)

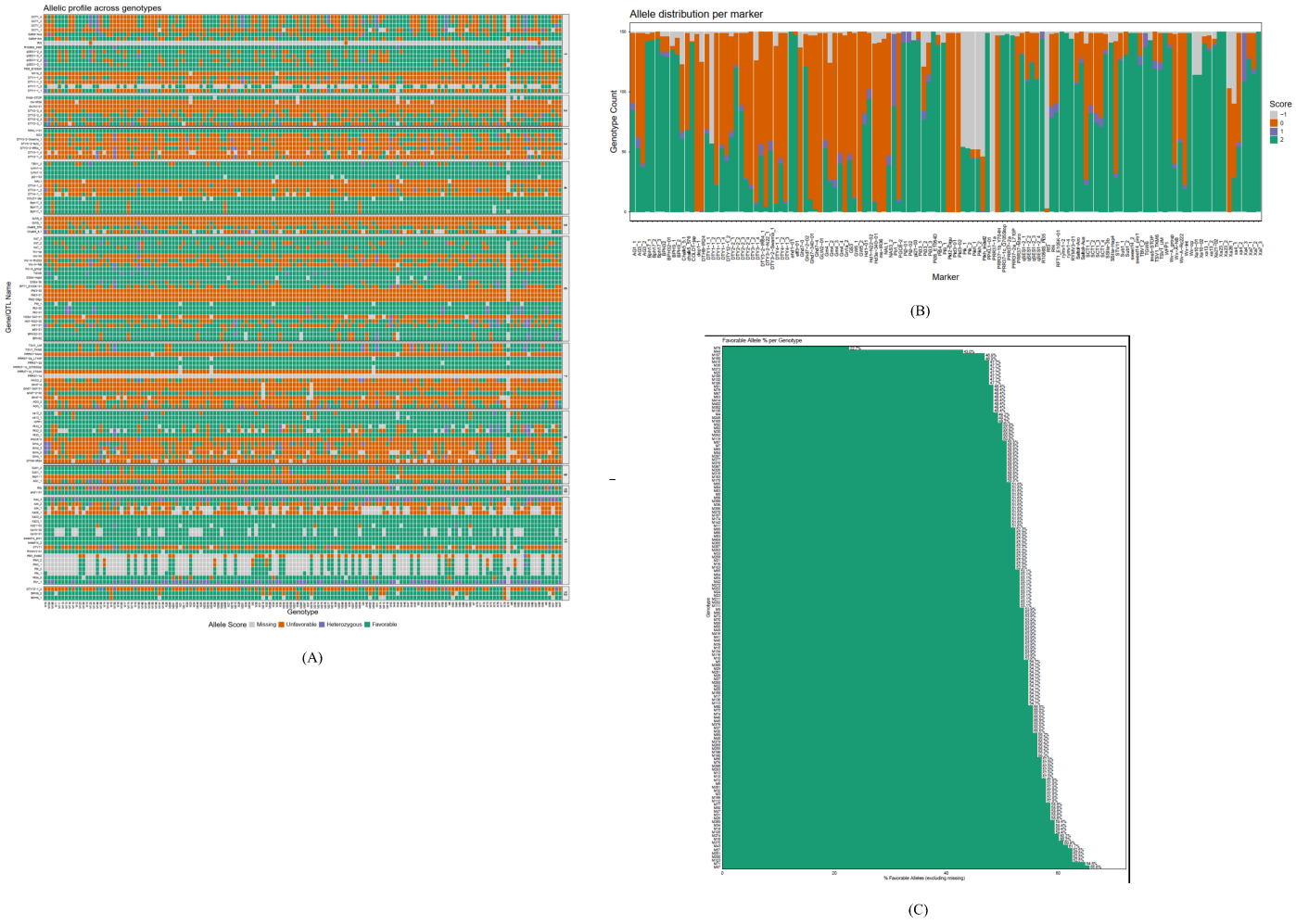


Figure 6

Allelic profiles across rice genotypes; (A) Graphical presentation of allelic variation across genotypes using 128 trait-based SNP markers, where each cell shows the allelic status: green for favorable alleles, orange for unfavorable alleles, purple for heterozygous alleles, and grey for missing data, (B) Graphical presentation of genotype count versus favourable allele frequency and, (C) Percentage of favourable alleles of SNPs associated with genes for economically important traits in rice genotypes (allelic status: green for favorable alleles, orange for unfavorable alleles, purple for heterozygous alleles, and grey for missing data)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfiles22726.docx
- Tables.docx