Multivariate and Genome-Wide SNP-Based Characterization of Genetic Diversity and Population Structure in a Subset of the 3K Rice (Oryza sativa L.) Genome Panel of Germplasm Accessions

Spoorthi Siddeswara Rekha

spoorthisr648@gmail.com

University of Agricultural Sciences, Bangalore

Deepak Chikkaballi Annegowda

University of Agricultural Sciences, Bangalore

Sanketh Ashok Umapathi

University of Agricultural Sciences, Bangalore

Pallavi Muniraja

University of Agricultural Sciences, Bangalore

Lohithaswa Hirenallur Chandappa

University of Agricultural Sciences, Bangalore

Sheshashayee Sreeman

University of Agricultural Sciences, Bangalore

Research Article

Keywords: Rice, Principal Component Analysis, K-means clustering, Multi-Trait Genotype–Ideotype Distance Index, STRUCTURE

Posted Date: June 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9911445/v1

Additional Declarations: No competing interests reported.

Title Page

Multivariate and Genome-Wide SNP-Based Characterization of Genetic Diversity and Population Structure in a Subset of the 3K Rice (*Oryza sativa* L.) Genome Panel of Germplasm Accessions

Spoorthi Siddeswara Rekha[a*], Deepak Chikkaballi Annegowda[b], Sanketh Ashok Umapathi[a], Pallavi Muniraja[a], Lohithaswa Hirenallur Chandappa[c], Sheshashayee Sreeman[d]

[a] Ph.D. Scholar, Department of Genetics and Plant Breeding, UAS, GKVK, Bangalore-560065, Karnataka, India.

[b] Junior Rice Breeder, AICRP on Rice, ZARS, UAS, V.C. Farm, Mandya-571405, Karnataka, India.

[c] Principal Scientist and Head, AICRP on Pigeonpea, UAS, GKVK, Bangalore-560065, Karnataka, India

[d] Emeritus Professor, Dept. of Crop Physiology, UAS, GKVK, Bangalore-560065, Karnataka, India.

[*] **Corresponding author** – spoorthisr648@gmail.com

Spoorthi S R (ORCID) - 0009-0000-4992-4683

Abstract

The narrow genetic base of cultivated rice necessitates the identification of diverse germplasm possessing superior agronomic performance and enhanced nutritional quality. The present study aimed to assess phenotypic and genetic diversity among a subset of 300 rice (*Oryza sativa* L.) germplasm accessions and identify elite accessions suitable for rice biofortification and yield improvement programmes. The accessions were evaluated during *Kharif* 2023 for nine quantitative traits using an alpha lattice design at Mandya and Gangavathi. Analysis of variance revealed significant variation among the accessions for all traits, indicating the presence of substantial phenotypic diversity. Principal Component Analysis identified the major contributors to phenotypic variation, with the first four principal components accounting for 69.3% of the total variation. Based on phenotypic performance, 50 accessions were selected and subjected to K-means clustering and Multi-Trait Genotype–Ideotype Distance Index analysis. Clustering grouped the accessions into three distinct clusters with contrasting combinations of grain yield and grain zinc concentration. MGIDI analysis identified G38 (IRGC 349) as the accession closest to the ideotype, exhibiting desirable values for grain zinc and yield-associated traits. Genetic diversity and population structure were assessed using 102,580 high-quality genome-wide SNP markers. STRUCTURE analysis, PCA and phylogenetic tree-based clustering consistently revealed the presence of two major genetic subpopulations within the germplasm panel. The combined phenotypic and genotypic analyses demonstrated substantial exploitable diversity and facilitated the identification of elite and genetically diverse accessions. These accessions represent valuable genetic resources for the development of high-yielding and zinc-biofortified rice varieties.

Keywords: Rice, Principal Component Analysis, K-means clustering, Multi-Trait Genotype–Ideotype Distance Index, STRUCTURE

1.INTRODUCTION

Rice (*Oryza sativa* L.) is the principal food crop for a large proportion of the world's population and contributes significantly to global food security (Khush, 2005). In many developing countries, particularly across Asia, rice serves as the primary source of dietary energy and sustains the livelihoods of millions of farming households (FAO, 2023). Although remarkable progress has been made in enhancing rice productivity, current breeding efforts are increasingly focused on improving both yield and nutritional quality to address the challenge of hidden hunger. Among micronutrients, zinc is of particular importance because its deficiency affects a substantial proportion of the global population and is associated with several health disorders (Cakmak, 2008).

Consequently, the development of zinc-enriched rice varieties has become a major objective in rice improvement programmes.

The success of crop improvement largely depends on the availability and effective utilization of genetic variability (Spoorthi S R *et al.,* 2025). Germplasm collections constitute an invaluable source of novel alleles and provide opportunities for identifying accessions with desirable agronomic and nutritional traits (Dwivedi *et al*., 2016). However, the large number of traits recorded in germplasm evaluation studies often makes interpretation difficult when using conventional statistical approaches. Therefore, multivariate analytical methods have gained importance for studying complex trait relationships and understanding diversity within germplasm collections (Mohammadi and Prasanna, 2003).

Principal Component Analysis (PCA) is one of the most widely used techniques for summarizing variation in large datasets. By converting correlated variables into a reduced number of independent principal components, PCA helps to identify the traits that contribute most to the observed variability and facilitates the visualization of relationships among accessions (Jolliffe, 2002). Such information is useful for identifying key traits responsible for differentiation within a germplasm panel. Another widely adopted approach is K-means clustering, which groups accessions according to their similarity across multiple traits (MacQueen, 1967). The resulting clusters provide insights into the pattern of diversity within a population and assist in identifying divergent genotypes that can be utilized as parents in hybridization programmes (Kaufman and Rousseeuw, 2005). The selection of parents from distinct clusters often increases the likelihood of obtaining superior recombinants in subsequent generations.

While PCA and clustering are useful for diversity assessment, breeders also require methods that enable simultaneous selection for several traits. The Multi-Trait Genotype–Ideotype Distance Index (MGIDI) offers an effective framework for this purpose by ranking genotypes according to their distance from an ideal genotype possessing desirable trait values (Olivoto and Nardino, 2021). Genotypes exhibiting lower MGIDI scores are considered more desirable because they combine favourable performance across multiple traits.

Alongside phenotypic characterization, molecular marker-based analyses provide a deeper understanding of genetic diversity and population structure. Single nucleotide polymorphism (SNP) markers have become the marker system of choice because of their abundance, genome-wide distribution and ability to capture genetic variation with high precision (Rafalski, 2002). Genome-wide SNP markers allow the assessment of genetic relationships and the identification of population stratification within germplasm collections.

STRUCTURE based Bayesian clustering approaches are commonly used to infer the number of genetic subpopulations and estimate the degree of admixture among accessions (Pritchard *et al*., 2000). Likewise, PCA based on SNP marker data provides a graphical representation of genetic relationships and facilitates the identification of major genetic groups within a population (Patterson *et al*., 2006). Phylogenetic analyses, particularly those based on the Neighbour-Joining method, further complement these approaches by illustrating evolutionary relationships and genetic divergence among accessions (Saitou and Nei, 1987). The combined use of STRUCTURE analysis, PCA and phylogenetic clustering therefore provides a robust framework for understanding the genetic architecture of a germplasm panel.

In view of the importance of identifying genetically diverse and agronomically superior germplasm, the present study combined phenotypic and genotypic approaches to evaluate a subset of rice accessions. Multivariate analyses, including PCA, K-means clustering and MGIDI, were employed to characterize phenotypic diversity and identify elite accessions, while genome-wide SNP markers were utilized to assess genetic diversity and population structure through STRUCTURE analysis, PCA and phylogenetic clustering. The integration of these approaches is expected to facilitate the identification of promising parental lines and accelerate the development of high-yielding and zinc-biofortified rice varieties.

2. MATERIALS AND METHODS

The plant material consists of subset of a 300 germplasm accessions selected from the 3K Rice Genome (3KRG) panel of the International Rice Research Institute (IRRI), Philippines, based on visualize observations such as, tillering ability, culm strength, leaf senescence, spikelet fertility, lodging resistance, photo-insensitivity and overall phenotypic acceptability to the local conditions. These accessions represent different subgroups of *Oryza sativa*, including indica (262), tropical japonica (21), temperate japonica (5), aus (8) and aro (4). They originate and evolve from diverse countries and capture a wide range of genetic diversity and variability.

2.1 Phenotypic evaluation

A total of 300 accessions were evaluated for yield attributing traits and grain zinc concentration using an alpha lattice design with two replications. Each accession was planted in a 1 m^2 plot consisting of six rows, with a spacing of 20 cm between rows and 15 cm between the plants, accommodating 42 plants per plot. The experimental layout comprised of 18 blocks, each accommodating 17 genotypes. The experiment was conducted during the *Kharif* 2023 at two locations: the Zonal Agricultural Research Station, V. C. Farm, Mandya, which lies in the Southern Dry Zone of Karnataka (Zone 6) at 12°31′ N and 76°54′ E and about 695 m above mean sea level,

received 297.4 mm of rainfall during the *Kharif* season 2023 (Supplementary Table 1) and the Agricultural Research Station, Gangavathi, which is located in the Northern Transitional Zone (Zone 8) at 15°26′ N and 75°07′ E at an elevation of around 678 m, recorded 309.9 mm of rainfall in the same season (Supplementary Table 2). The contrasting agro-climatic conditions at the two locations provided an opportunity to assess genotype performance across diverse environments. Observations were recorded on five randomly selected plants from each accession for days to 50% flowering (DFF), plant height (PH), panicle length (PL), panicle weight (PW), productive tillers per plant (PT), spikelet fertility SF), test weight (TW), grain zinc concentration (Zn) and grain yield (GYD).

Based on the phenotypic performance of the 300 germplasm accessions evaluated across two contrasting environments, 50 accessions were selected for further investigation. The selected set comprised accessions representing both extremes and the intermediate range of the phenotypic distribution, including 15 accessions from the lower tail, 15 accessions from the upper tail, and 20 accessions from the middle of the distribution. This sampling strategy ensured adequate representation of the entire spectrum of phenotypic variation present in the study panel.

2.2 Estimation of Grain Zinc Concentration

Grain samples from each accession were dehusked using a palm dehusker and the resulting brown rice grains were cleaned to remove residual bran particles. Zinc concentration was determined from a 5 g grain sample using an energy-dispersive X-ray fluorescence (ED-XRF) spectrometer (X-Supreme 8000, Oxford Instruments, CA, USA). Grain zinc concentration was expressed as mg/kg (ppm).

2.3 Multivariate Statistical Analysis for Diversity Assessment and Ideotype-Based Selection of Elite Germplasm Accessions

To assess the phenotypic diversity among the germplasm accessions, analysis of variance (ANOVA) was performed according to Patterson and Williams (1976) using the '*metan'* package. Genetic variability parameters, including phenotypic and genotypic coefficient of variation, were estimated using the '*psych'* package. To identify the major traits contributing to the total phenotypic variation, principal component analysis (PCA) was conducted using the '*FactoMineR'* package and the results were visualized using the '*factoextra'* package.

To identify the proportion of traits contributing to each principal components (PCs), the Circos plot was generated based on PCA Scores (Krzywinski, M. *et al* 2009). Further, to classify the germplasm accessions into

different groups, K-means clustering was performed using the '*stats'* package in R. To identify the germplasm accession posing the desirable combination of grain zinc and yield associated traits, the Multi-Trait Genotype-Ideotype Distance Index (MGIDI) was computed using the '*metan'* package to rank and assess the stability of the germplasms based on multiple traits. Hence, MGIDI analysis facilitated the selection of genotypes closest to the predefined ideotype based on multiple trait performance.

2.4 Assessment of Population Structure and Genetic Relationships among Germplasm Accessions Using STRUCTURE and Neighbour-Joining cluster Analyses

To elucidate the population structure among the 300 germplasm accessions, the high-quality filtered SNP dataset was analysed using the STRUCTURE software. The most probable number of subpopulations (K) was determined based on the ΔK method and the results were visualized through STRUCTURE bar plots and ΔK graphs. To further assess the genetic relationships among the accessions, the filtered SNP dataset was subjected to Neighbour-Joining (NJ) cluster analysis using TASSEL version 5.2.79 (Bradbury *et al*., 2007). A genetic distance matrix was generated and used to construct the NJ phylogenetic tree. The resulting tree file was exported in Newick format and visualized as a circular dendrogram using iTOL version 6.5.2. (Letunic and Bork .2024)

2.5 3KRG SNP Data set and quality SNP filtering

The SNP Dataset for the 300 accessions was retrieved from the Rice SNP-Seek Database, which is maintained by IRRI, Philippines. From the ~29 million biallelic SNPs, a base set of approximately 18 million SNPs was generated by removing those with excessive heterozygous calls. The dataset was then filtered by setting a call rate threshold of 95% was applied, retaining SNPs present in at least 285 germplasm accessions. Markers with a minor allele frequency (MAF) ≥ 0.05 were selected to remove rare alleles occurring at frequencies below 5%. A maximum heterozygosity proportion of 0.10 was imposed to exclude highly heterozygous loci. Missing genotype data were subsequently imputed using the Linkage Disequilibrium K-Nearest Neighbour Imputation **(**LD-KNNi) method. After applying these quality filtering and imputation steps, 1,02,580 high-quality SNPs were retained from approximately 18 million initial SNPs and were subsequently used for population structure analysis and phylogenetic based cluster assessment.

3. RESULTS

3.1 First- and Second-Order Genetic Statistics for Different Quantitative Traits in the Germplasm Accessions

Analysis of variance revealed highly significant differences among the germplasm accessions for grain zinc concentration and all yield-related traits (Table 1). The magnitude of variation observed among the accessions indicated the presence of a broad phenotypic range for the traits studied. The pooled analysis across environments showed that neither the location effect nor the genotype × location interaction was significant for any of the traits (Table 2). Consequently, pooled means were considered appropriate for subsequent diversity and multivariate analyses.

3.2 Per se Performance of Germplasm Accessions for Grain Zinc and Yield-Associated Traits

The frequency distribution of all quantitative traits approximated a normal distribution, indicating continuous variation among the germplasm accessions (Figure 1). Considerable variation was observed for grain zinc concentration, grain yield and their associated traits. The population means for DFF, PH, PL, PW, PT, SF, TW, GYD and Zn were 90 days, 95.83 cm, 20.43 cm, 7.25 g, 10, 86.95%, 19.26 g, 3.77 t ha^{-1} and 22.35 ppm, respectively (Figure 1). Grain zinc concentration ranged from low to high values among the accessions, while grain yield also exhibited a wide distribution, reflecting the diverse nature of the germplasm panel. The presence of accessions at both extremes of the distribution suggests the availability of contrasting genetic resources for trait improvement. The top-performing accessions identified for grain zinc concentration and grain yield are presented in Tables 3 and 4. Several accessions exhibited superior performance for individual traits, while a few accessions combined favourable values for multiple traits.

3.3 Estimates of Genetic Variability Parameters among Germplasm Accessions

The PCV estimates were slightly higher than the corresponding GCV values for all traits, indicating some influence of environmental factors on trait expression (Figure 2). However, the differences between PCV and GCV were generally narrow. Among the traits, grain zinc concentration recorded the highest variability followed by plant height, test weight and spikelet fertility. Moderate levels of variability were observed for panicle length, panicle weight, productive tillers and grain yield, whereas days to 50% flowering exhibited the lowest variability. These results indicate that substantial variability exists within the germplasm panel for several economically important traits.

3.4 Identification of Major Contributing Traits Through Principal Component Analysis

Principal component analysis reduced the nine quantitative traits into four principal components explaining 69.3% of the total variation (Figure 3D). The first principal component accounted for the largest

proportion of variation, followed by PC2, PC3 and PC4. Together, these components captured most of the diversity present within the germplasm accessions. The Circo's plot revealed differential contributions of traits to individual principal components (Figure 3C). DFF contributed predominantly to PC1, while TW contributed substantially to PC2. Grain zinc concentration contributed significantly to both PC3 and PC4. The PCA biplot clearly separated the accessions across different quadrants (Figure 3A and 3B). Accessions distributed away from the origin exhibited greater divergence than those clustered near the centre. Several accessions were associated with specific trait vectors, indicating superior expression of those traits. The accessions G248 (IRGC 349), G28 (IRGC 993), and G204 (IRGC 972) were positioned farthest from the origin in the direction of important trait vectors such as DFF, PT, and PW (Figure 3B), indicating their superior performance for these attributes and their potential value in future breeding programmes.

3.5 K-Means Clustering of Germplasm Accessions

The Elbow method indicated three optimal clusters among the 50 selected accessions (Figure 4A). Cluster I consisted of 15 accessions with high grain zinc concentration and low grain yield. Cluster II contained 16 accessions with intermediate values for both traits. Cluster III comprised 19 accessions characterized by high grain yield and comparatively low grain zinc concentration (Figure 4B and 4C).

3.6 Identification of Superior Germplasm Accessions Through MGIDI

MGIDI analysis ranked the selected germplasm accessions according to their distance from the predefined ideotype (Figure 5A). Eight accessions exhibited the lowest MGIDI values and were therefore considered superior. Among these, G38 (IRGC 647) recorded the lowest MGIDI value and ranked first. The strengths and weaknesses analysis partitioned the MGIDI values into three latent factors (Figure 5B). Factor 1 was associated with DFF, PT and PW, Factor 2 with PL, SF and TW, and Factor 3 with GYD, Zn and PH. The relative contribution of these factors varied among accessions, indicating differences in the traits limiting their proximity to the ideotype.

3.7 Population Structure Based on Genome-Wide SNP Markers

After quality filtering, 102,580 polymorphic SNP markers were retained for diversity analysis. STRUCTURE analysis identified two major genetic subpopulations, with the highest ΔK value observed at $K = 2$ (Figure 6A and 6B). Most accessions exhibited membership to one of the two groups, while a few accessions displayed admixed ancestry. The PCA based on SNP markers supported the STRUCTURE results. The first two principal components

explained approximately 93% of the total molecular variation, with PC1 and PC2 contributing 52% and 41%, respectively (Figure 7). The scree plot showed a clear elbow after the second principal component. Similarly, the neighbour-joining phylogenetic tree grouped the accessions into two major clusters (Figure 8). The clustering pattern was largely consistent with the results obtained from STRUCTURE and PCA, confirming the presence of two major genetic groups within the germplasm panel.

4. DISCUSSION

The significant differences observed among the germplasm accessions for grain zinc concentration and yield-related traits indicate the existence of considerable variability within the panel. Such variability forms the foundation of any crop improvement programme because the effectiveness of selection largely depends on the extent of variation available in the breeding material. The non-significant genotype × environment interaction further suggests that the expression of these traits was relatively stable across the two environments, thereby increasing the reliability of selection based on pooled means. Similar observations have been reported in rice germplasm evaluations, where stable trait expression across environments facilitated the identification of superior accessions (Singh *et al.,* 2022).

The high GCV and PCV values observed for grain zinc concentration indicate substantial genetic variation for this trait, suggesting that considerable progress can be achieved through selection. Since zinc concentration is a primary target in rice biofortification programmes, the availability of accessions with high zinc content provides opportunities for introgressing favourable alleles into elite breeding lines. The close correspondence between PCV and GCV estimates further indicates that environmental influence on trait expression was limited, implying that phenotypic selection would be effective for these traits.

The PCA results demonstrated that a relatively small number of principal components captured a large proportion of the total variation present in the germplasm panel. This indicates that diversity among the accessions was largely governed by a few key traits. The strong contribution of DFF, TW and Zn to the principal components suggests that these traits played a major role in differentiating the germplasm accessions. Similar findings have been reported by Gaur *et al.* (2017), who observed that flowering behaviour and grain zinc were among the major contributors to phenotypic diversity in rice. The negative association observed between grain yield and grain zinc concentration in the PCA biplot is particularly important from a breeding perspective. Such relationships have frequently been reported in rice and may arise due to dilution effects associated with increased grain production.

This suggests that simultaneous improvement of yield and grain zinc concentration remains challenging and requires the identification of genotypes that possess favourable combinations of both traits.

K-means clustering grouped the accessions into clusters with distinct trait combinations, indicating the presence of exploitable diversity within the germplasm panel. The high-zinc but low-yielding accessions grouped in Cluster I represent valuable donor sources for nutritional improvement, whereas the high-yielding accessions in Cluster III provide opportunities for enhancing productivity. The substantial divergence between these clusters suggests that crosses involving parents selected from these groups may generate wider segregation and increase the probability of recovering transgressive segregants combining high grain yield with elevated zinc concentration. The overlap observed among clusters indicates that the variation for grain zinc concentration and grain yield is quantitative and continuous rather than discrete. Such overlap is expected in diverse germplasm collections because most agronomic traits are controlled by multiple genes and influenced by environmental factors.

Conventional selection methods often focus on a few individual traits and may fail to identify genotypes with balanced performance across multiple characteristics. The MGIDI approach overcomes this limitation by simultaneously considering several traits and ranking accessions based on their similarity to a predefined ideotype. In the present study, G38 (IRGC349) emerged as the most desirable accession because it combined superior grain yield, acceptable grain zinc concentration and favourable agronomic attributes. The strengths and weaknesses analysis further revealed the specific trait groups contributing to the distance of each accession from the ideotype. Such information is particularly useful for breeders because it not only identifies superior accessions but also highlights the traits requiring improvement. Therefore, MGIDI serves as an efficient decision-support tool for multi-trait selection in rice breeding programmes.

The use of 102,580 genome-wide SNP markers enabled a comprehensive assessment of genetic diversity within the germplasm panel. The identification of two major subpopulations through STRUCTURE analysis, PCA and phylogenetic clustering indicates the presence of clear genetic stratification among the accessions. The grouping of indica, japonica, and aromatic accessions separately from aus accessions reflects their distinct evolutionary histories and breeding backgrounds. Interestingly, the phenotypic clusters did not completely correspond with the SNP-based population structure. This suggests that accessions sharing similar genetic ancestry may still differ considerably in agronomic performance due to historical selection, local adaptation and

the polygenic nature of quantitative traits. Similar observations have been reported in rice diversity studies, where genetic relatedness was not always reflected in phenotypic similarity.

4. CONCLUSION

The present study demonstrated the existence of substantial phenotypic and genetic diversity among the rice germplasm accessions for grain zinc concentration and yield-associated traits. Significant variation observed across the accessions provided opportunities for identifying superior genotypes with desirable trait combinations. Multivariate analyses, including PCA and K-means clustering, effectively characterized the diversity present within the panel and grouped the accessions into distinct clusters with contrasting agronomic and nutritional attributes. The MGIDI analysis further enabled the identification of elite accessions that closely resembled the predefined ideotype and combined favourable levels of grain yield and grain zinc concentration. The use of 1,02,580 Genome-wide polymorphic SNP markers revealed two major genetic subpopulations, which were consistently supported by STRUCTURE analysis, PCA and phylogenetic tree-based clustering. The partial correspondence between phenotypic and genotypic groupings indicated that accessions sharing a common genetic background may still exhibit considerable phenotypic variation due to environmental adaptation and the complex inheritance of quantitative traits. Overall, the integration of phenotypic and genotypic approaches facilitated the identification of elite and genetically diverse germplasm accessions with potential utility in rice improvement programmes. The superior accessions identified in this study can serve as valuable parental lines for developing high-yielding, zinc-biofortified rice varieties and contribute to future efforts aimed at enhancing both productivity and nutritional quality in rice.

REFERENCES

Bouis HE, Saltzman A (2017) Improving nutrition through biofortification: a review of evidence from HarvestPlus, 2003 through 2016. Glob Food Secur 12:49–58.

Cakmak I (2008) Enrichment of cereal grains with zinc: agronomic or genetic biofortification? Plant Soil 302:1–17.

Christina GR, Thirumurugan T, Jeyaprakash P, Rajanbabu V (2021) Principal component analysis of yield and yield related traits in rice (*Oryza sativa* L.) landraces. Electron J Plant Breed 12(3):907–911.

Dwivedi SL, Ceccarelli S, Blair MW, Upadhyaya HD, Are AK, Ortiz R (2016) Landrace germplasm for improving yield and abiotic stress adaptation. Trends Plant Sci 21(1):31–42.

FAO (2023) Rice Market Monitor. Food and Agriculture Organization of the United Nations, Rome.

Gour L, Maurya SB, Koutu GK, Singh SK, Shukla SS, Mishra DK (2017) Characterization of rice (*Oryza sativa* L.) genotypes using principal component analysis including scree plot and rotated component matrix. Int J Chem Stud 5(4):975–983.

Jolliffe IT (2002) Graphical representation of data using principal components. In: Principal Component Analysis. Springer, New York, pp 64–91.

Kaufman L, Rousseeuw PJ (2009) Finding groups in data: an introduction to cluster analysis. John Wiley & Sons, New York.

Khush GS (2005) What it will take to feed 5.0 billion rice consumers in 2030. Plant Mol Biol 59(1):1–6.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. Genome Res 19(9):1639–1645.

Letunic I, Bork P (2024) Interactive Tree Of Life (iTOL) v7: recent updates and new developments. Nucleic Acids Res. https://doi.org/10.1093/nar/gkae268

MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol 1. University of California Press, Berkeley, pp 281–297.

Mohammadi SA, Prasanna BM (2003) Analysis of genetic diversity in crop plants—salient statistical tools and considerations. Crop Sci 43(4):1235–1248.

Olivoto T, Nardino M, Meira D, Meier C, Follmann DN, de Souza VQ, Konflanz VA, Baretta D (2021) Multi-trait selection for mean performance and stability in maize. Agron J 113(5):3968–3974.

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2(12):e190.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2):945–959.

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5(2):94–100.

Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4):406–425.

Singh VK, Wahi N, Mishra SK, Singh BD, Singh NK (2022) Studies on genetic variability, correlation analysis, character association and path analysis of phenotypic characteristics of twelve mega varieties of rice and their near-isogenic lines carrying high grain number per panicle QTL qGN4.1. Curr Trends Biotechnol Pharm 16(1):35–45.

Spoorthi SR, Immadi S, Sangeetha, Bonsu SK (2025) Genetic characterization of soybean (*Glycine max* L.) genotypes for yield, maturity and rust resistance. Legume Res 48:1–7. https://doi.org/10.18805/LR-5475

Statement and Declarations

- **Funding:** This research received no specific grant from funding agencies in the public, commercial or not-for-profit sectors**.**
- **Conflict of Interest:** The author(s) declare that there are no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.
- **Ethical Approval:** Not Applicable. This study involved field evaluation of rice germplasm accessions and did not involve human participants or animals requiring ethical committee approval.
- **Informed Consent:** Not Applicable.
- **Author Contribution:**

  1. Spoorthi Siddeswara Rekha: Conceptualization, methodology, field experimentation, phenotyping, data curation, statistical analysis, interpretation of results and manuscript writing

  2**.** Deepak Chikkaballi Annegowda: Supervision, conceptual guidance, methodology refinement, manuscript review and editing

  3. Sanketh Ashok Umapathi: Data recording and statistical analysis

  4. Pallavi Muniraja: Data recording and statistical analysis

  5. Lohithaswa Hirenallur Chandappa: Research frame work and manuscript review.

  6. Sheshashayee Sreeman: Technical guidance and manuscript review.

Data Availability Statement: The phenotypic datasets generated during this study are available from the corresponding author upon reasonable request. Genotypic data used in this study were obtained from the Rice SNP-Seek Database, which is publicly accessible.

Acknowledgments

The authors sincerely express their heartfelt gratitude to Dr. Deepak C. A., chairperson, for his valuable guidance, encouragement and constant support throughout the investigation. The authors are highly thankful to Zonal Agricultural Research Station (ZARS), Mandya, and Agricultural Research Station (ARS), Gangavathi, for providing necessary facilities for the evaluation of germplasm accessions. Sincere thanks are also extended to all the members of the Advisory Committee for their constructive suggestions and technical support during the course of the study. The authors gratefully acknowledge the Department of Genetics and Plant Breeding, University of Agricultural Sciences, Bengaluru.

Table 1: Analysis of variance for grain zinc and yield associated traits among the germplasm accessions

Source of Variation	df	Mean Sum of Squares (MSS)					
		Days to 50% Flowering		Plant Height		Panicle Length	
		Mandya	Gangavathi	Mandya	Gangavathi	Mandya	Gangavathi
Genotypes	299	88.21**	85.85**	878.35**	848.79**	25.58**	22.32**
Replication	1	1.35	1.67	3.76	5.21	2.15	1.27
Blocks	34	1.02	1.10	2.98	2.94	0.99	0.03
Residual	265	8.95	6.47	11.45	15.02	6.95	9.86

Source of Variation	df	Mean Sum of Squares (MSS)					
		Panicle Weight		Productive Tillers		Spikelet fertility	
		Mandya	Gangavathi	Mandya	Gangavathi	Mandya	Gangavathi
Genotypes	299	24.93**	22.80**	31.58**	35.67**	24.94**	27.83**
Replication	1	0.56	0.46	0.35	0.19	2.95	2.25
Blocks	34	1.08	1.39	0.85	0.987	1.14	1.19
Residual	265	10.95	8.30	9.95	13.29	5.65	8.25

Source of Variation	df	Mean Sum of Squares (MSS)					
		Test Weight		Grain Yield		Grain Zinc Concentration	
		Mandya	Gangavathi	Mandya	Gangavathi	Mandya	Gangavathi
Genotypes	299	46.46**	42.32**	31.53**	28.46**	18.58**	15.22**
Replication	1	0.021	0.07	0.08	0.01	0.79	0.47
Blocks	34	0.79	0.81	0.26	0.25	0.85	1.11
Residual	265	12.95	10.57	11.30	9.41	8.17	5.16

Table 2: Pooled Analysis of Variance for grain zinc and yield associated traits among the germplasm accessions

Source of Variation	df	Mean Sum of Squares (MSS)								
		DFF	PH	PL	PW	PT	SF	TW	GYD	Zn
Genotypes	299	73.59**	96.22**	44.69**	11.71**	11.44**	39.76**	12.74**	2.98**	24.05**
Replication	2	2.01	1.68	2.71	0.618	0.19	1.14	7.64	0.04	1.03
Blocks	68	1.06	0.96	0.01	1.37	0.98	0.17	0.80	0.25	0.98
Location	1	12.9	8.59	1.59	3.75	4.05	2	0.58	3.66	0.49
Genotypes X Location	299	4.71	9.14	3.29	2.58	4.26	8.68	7.39	0.012	1.75
Residuals	576	2.164	4.02	1.33	0.605	0.359	1.06	0.3063	0.117	1.10

Table 3: *Perse* performance of germplasm accessions for yield associated traits

Germplasm accessions					
Traits	**IRGC 349**	**IRGC 993**	**IRGC 972**	**IRGC 1150**	**IRGC 1143**
DFF	99.00	101.00	98.00	102.00	98.00
PH	98.23	98.23	104.36	96.32	110.45
PL	26.23	26.89	27.15	25.78	27.23
PW	7.95	7.05	6.85	7.52	7.62
PT	14.00	13.00	14.00	12.00	13.00
SF	93.65	92.56	93.85	92.85	93.24
TW	21.65	21.12	20.63	20.45	20.23
GYD	6.23	6.13	6.02	5.94	5.85
Zn	24.23	24.96	25.87	25.45	26.68

Table 4: *Perse* performance of germplasm accessions for grain zinc concentration

Germplasm accessions					
Traits	**IRGC 647**	**IRGC 652**	**IRGC 414**	**IRGC 173**	**IRGC 940**
DFF	86.00	88.00	87.00	90.00	93.00
PH	115.21	114.25	106.54	108.32	112.65
PL	21.54	20.96	19.65	22.04	21.96
PW	5.68	6.05	5.55	6.45	6.84
PT	9.00	9.00	10.00	10.00	11.00
SF	86.21	85.24	86.96	87.65	88.65
TW	16.24	16.75	18.24	18.65	19.24
GYD	2.56	2.96	3.45	3.27	3.75
Zn	35.75	35.25	34.96	34.59	34.45

Figure 1: Normal distribution of grain zinc concentration and yield associated traits among the rice germplasm accessions

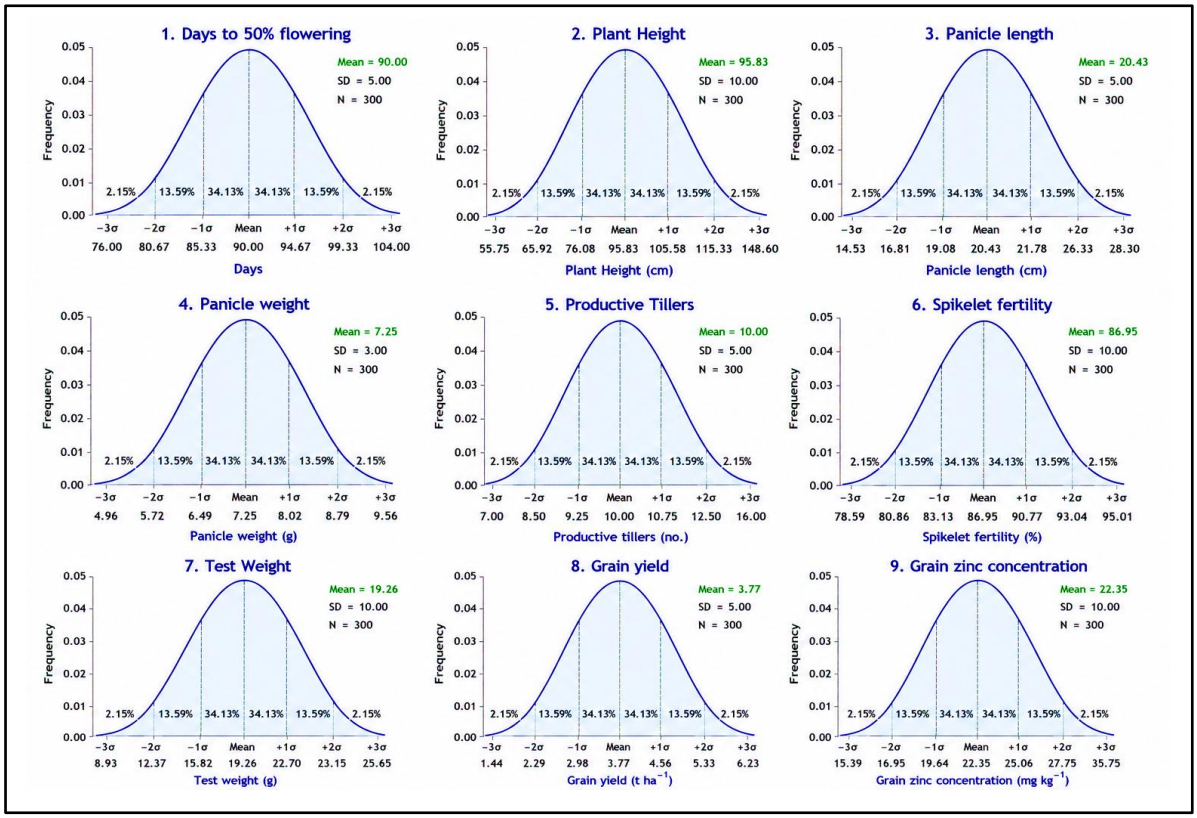


Figure 2: Estimates of genetic variability parameters for quantitative traits among rice germplasm accessions

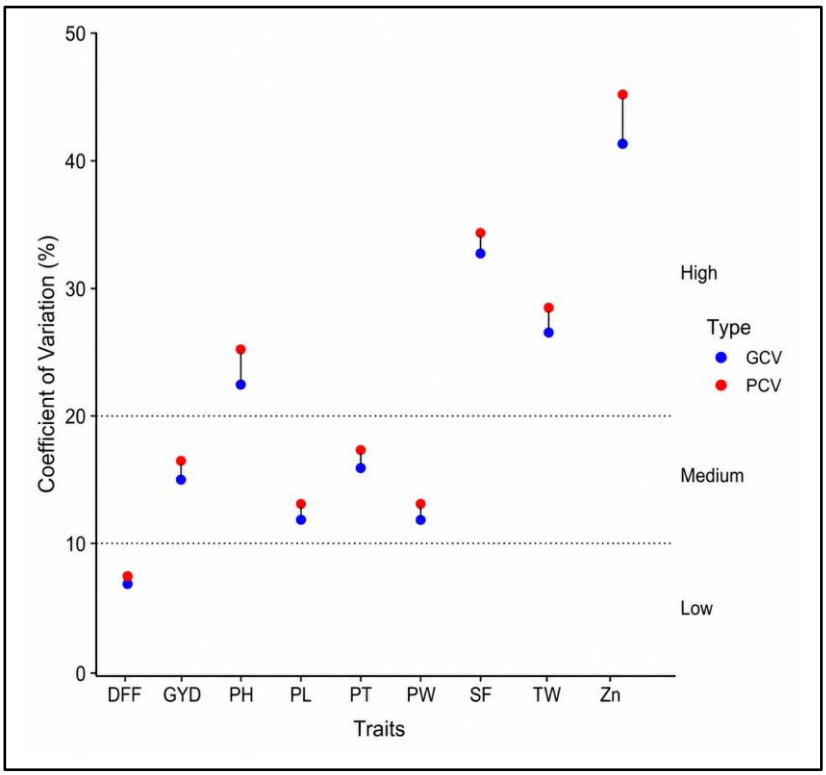


Figure 3: Principal component analysis of rice germplasm accessions: (A) PCA biplot showing the major traits contributing to genetic variation (B) Ranking of the germplasm accessions based on PCA biplot. (C) Circo's plot depicting the contribution of traits to each principal components. (D) Scree plot depicting the variance explained by the principal components

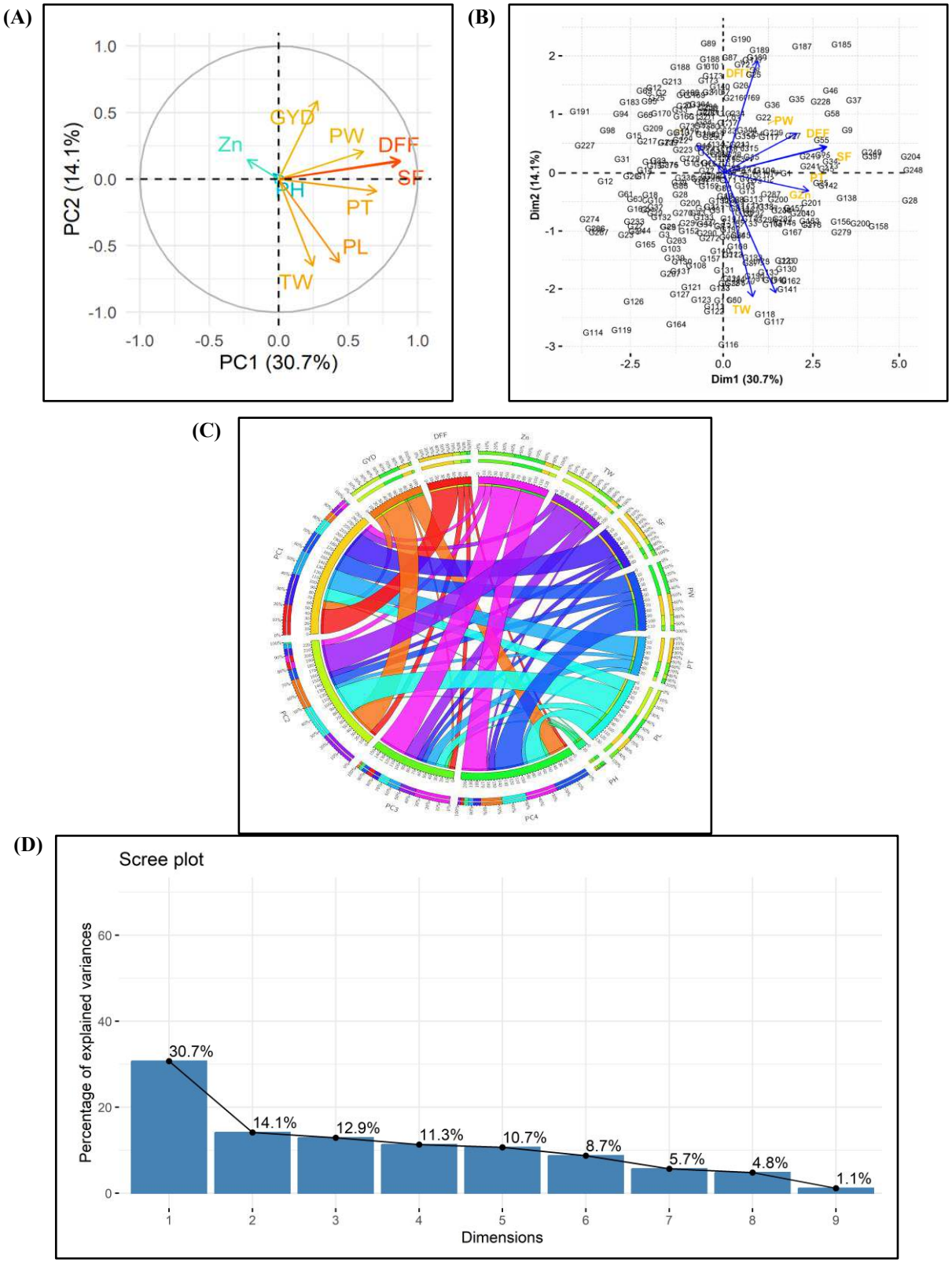
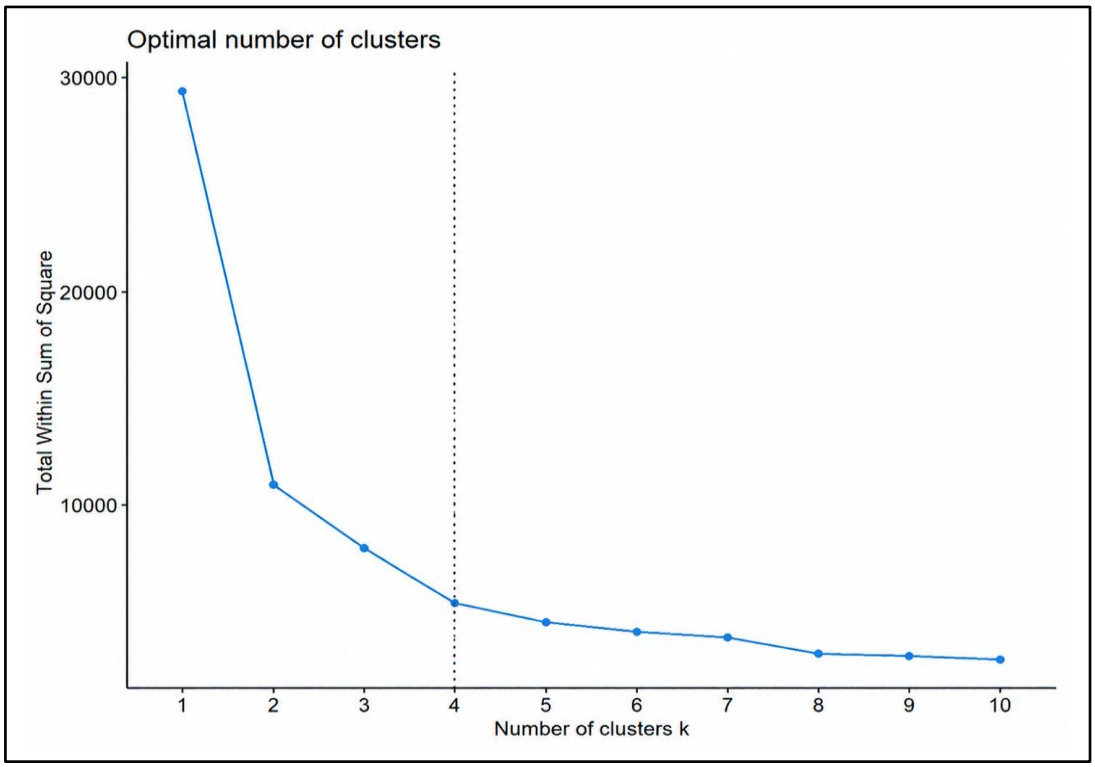
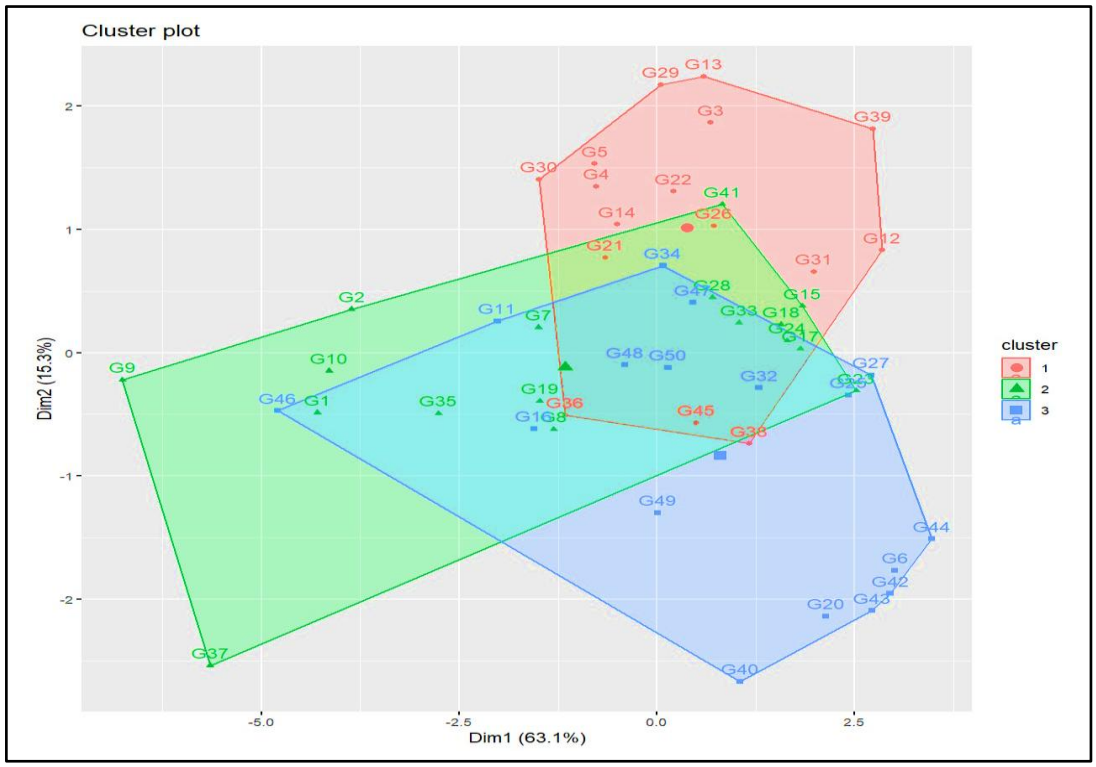


Figure 4: (A) Identification of optimum number of clusters among germplasm accession for quantitative traits. (B) K-means clustering scatter plot of germplasm accessions based on grain zinc and yield-associated traits. (C) Heat map showing the clustering of germplasm accessions and trait variation among clusters.

(A)



(B)



(C)

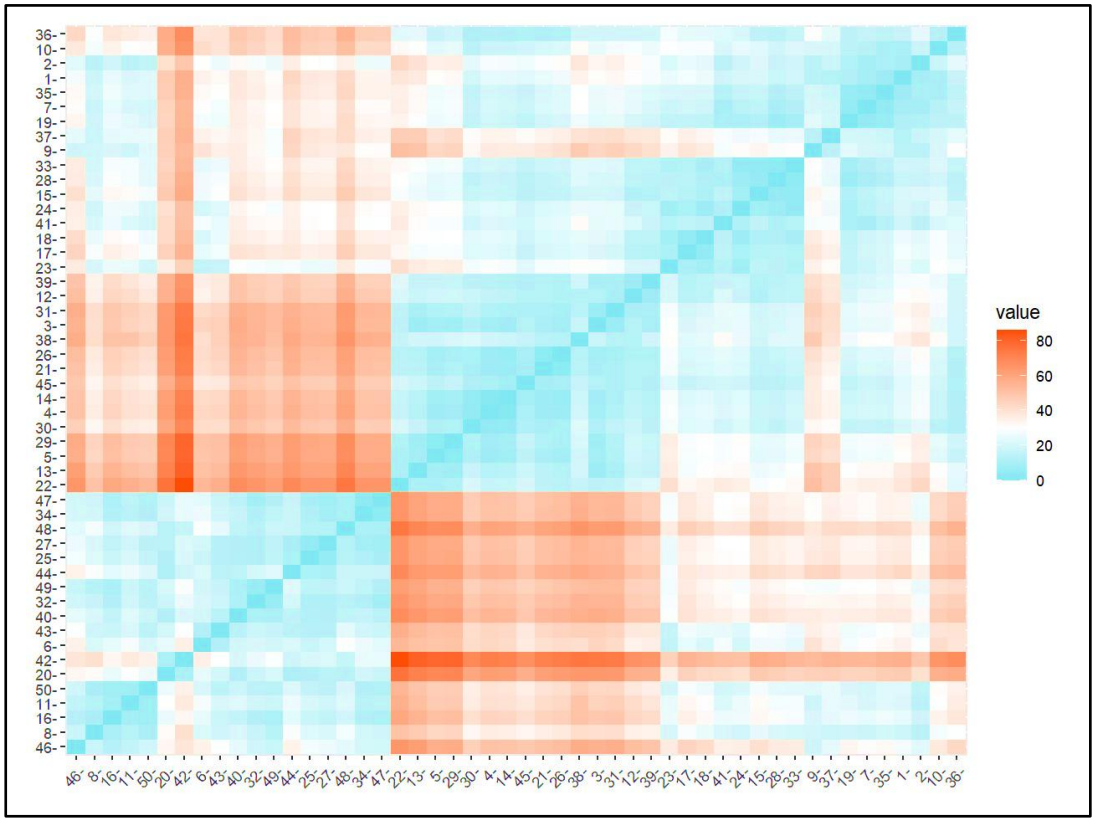
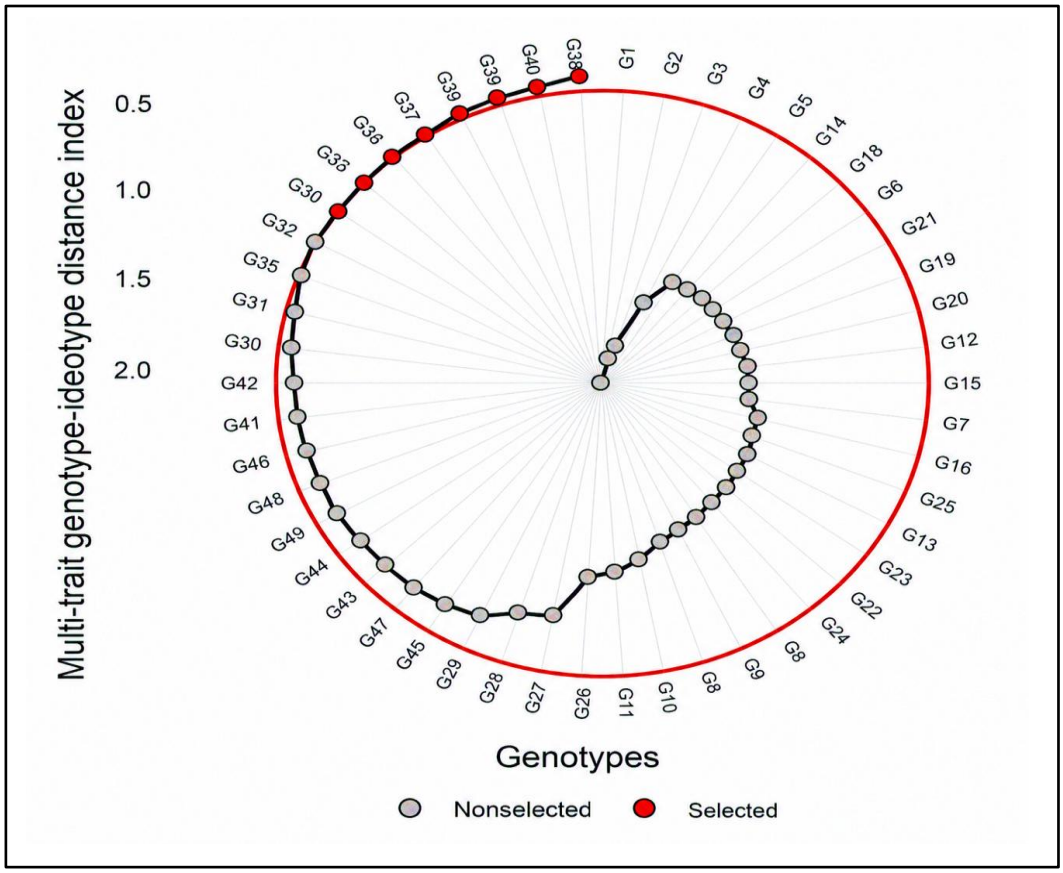


Figure 5: (A) Multi-Trait Genotype–Ideotype Distance Index (MGIDI)-based identification of superior rice germplasm accessions for grain zinc and yield-associated traits. (B) Radar chart illustrating the proportional contribution of different factors to the MGIDI values of the selected germplasm accessions

(A)



(B)

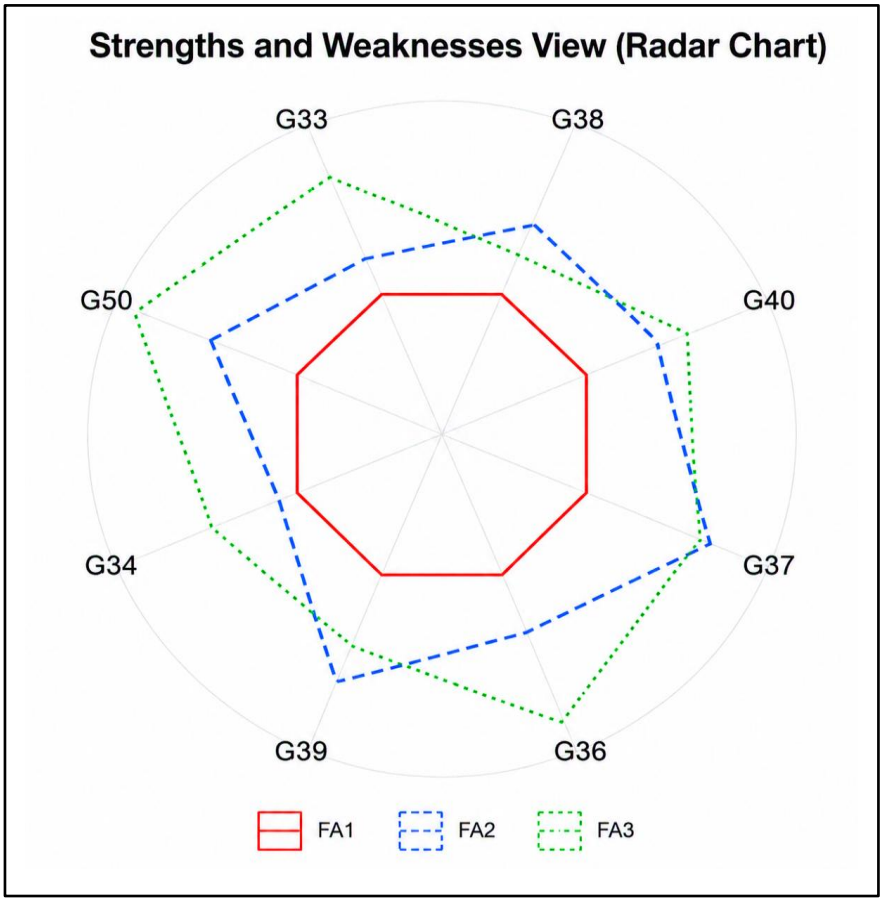
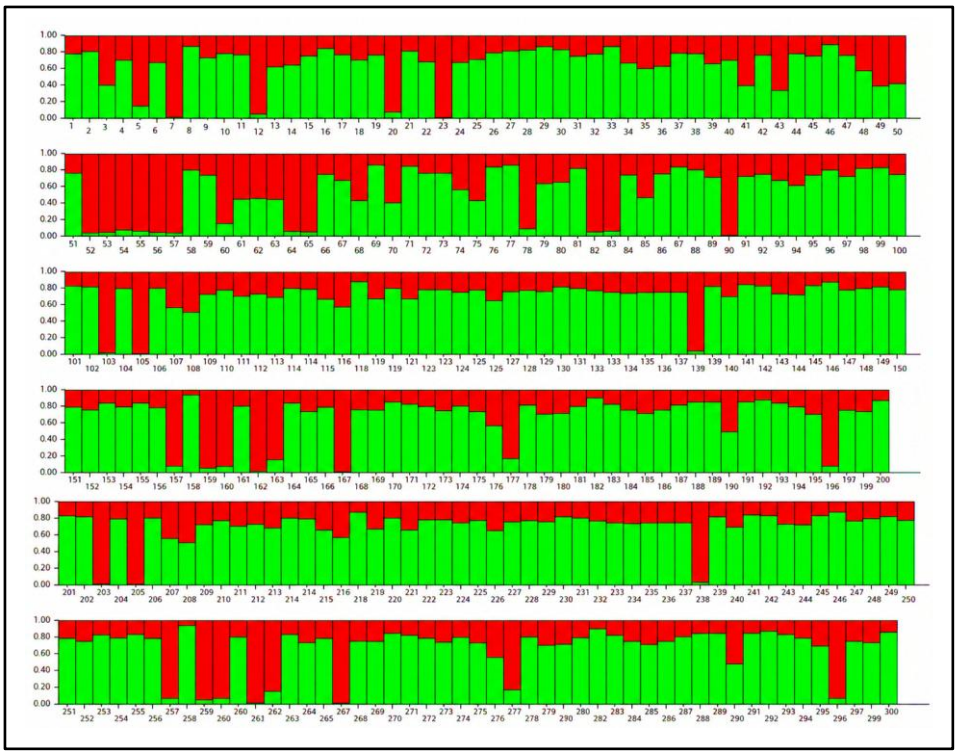


Figure 6: Population structure of 300 germplasm accessions. (A) Bar plot depicting the number of subpopulation based on colours through STRUCTURE (B) Determination of optimum number of subpopulation through Delta K graph

(A)



(B)

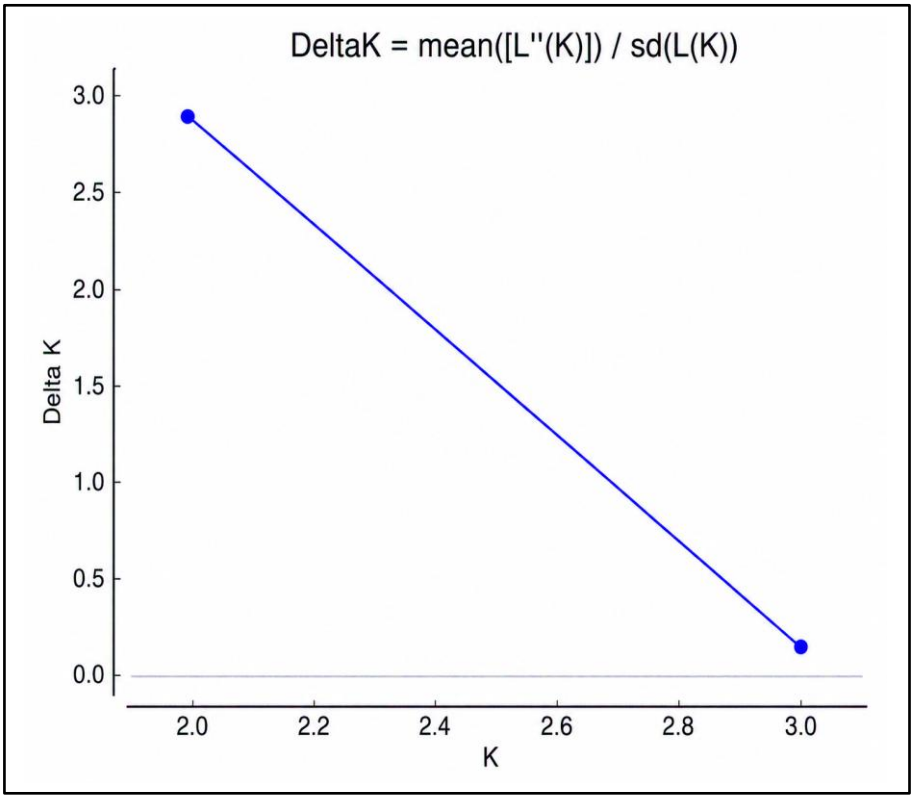


Figure 7: Scree plot-based determination of population structure among germplasm accessions through Principal Component Analysis

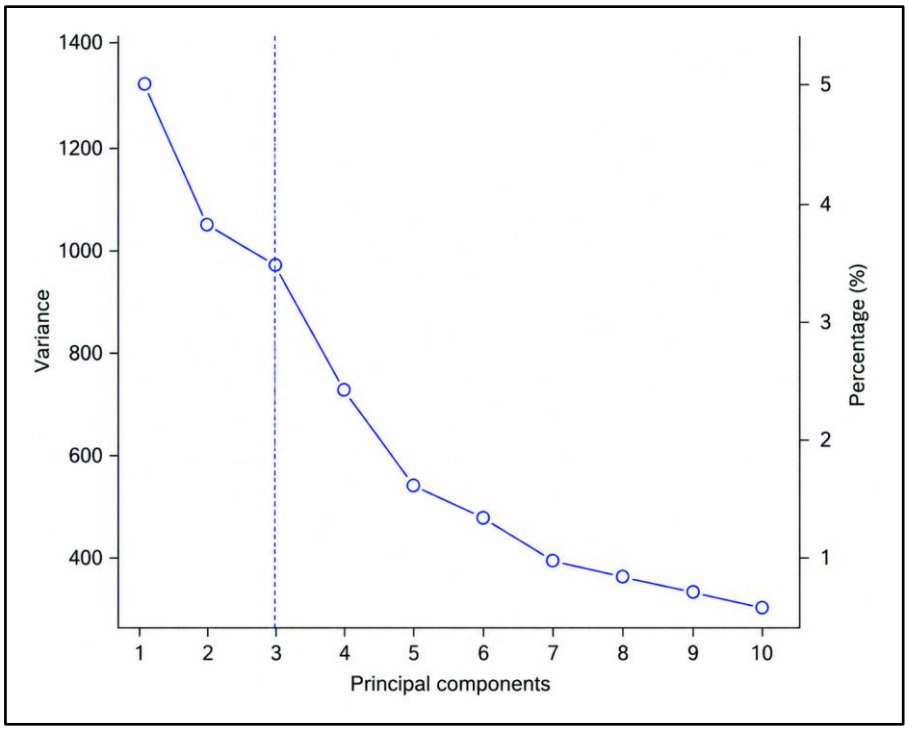


Figure 8: Neighbour-Joining cluster analysis revealing the genetic relationships among the germplasm accessions