



RESEARCH ARTICLE

Deciphering the genetic diversity among a subset of rice (*Oryza sativa*) germplasm accessions through principal component analysis

Spoorthi S R^{1*}, Sanketh Ashok U¹, Pallavi Muniraja¹, Deepak C A², Sandesh G M¹, Shama P¹, Anil Kumar R¹ & Sandeep Naik¹

¹Department of Genetics and Plant Breeding, University of Agricultural Sciences, Bangalore 560 065, Karnataka, India

²Zonal Agricultural Research Station, Vishweshwaraiah Canal Farm, Mandya, University of Agricultural Sciences, Bangalore 560 065, Karnataka, India

*Correspondence email - spoorthisr648@gmail.com

Received: 27 July 2025; Accepted: 26 January 2026; Available online: Version 1.0: 22 April 2026

Cite this article: Spoorthi SR, Sanketh AU, Pallavi M, Deepak CA, Sandesh GM, Shama P, Anil KR, Sandeep N. Deciphering the genetic diversity among a subset of rice (*Oryza sativa*) germplasm accessions through principal component analysis. Plant Science Today. 2026; 13(sp2): 1-7. <https://doi.org/10.14719/pst.10802>

Abstract

Rice (*Oryza sativa* L.) is a vital staple crop that nourishes more than half of the global population, especially in Asia. To meet the rising food demand, improving rice productivity and adaptability by effectively utilizing the genetic diversity present in germplasm accessions is essential. In this study, 300 rice germplasm accessions representing diverse sub-groups of *O. sativa* were evaluated during kharif 2023 at two locations, Mandya and Gangavathi, using an alpha lattice design. Phenotypic observations were recorded for 9 agro-morphological traits. Principal component analysis (PCA) was used to assess genetic diversity and identify key traits contributing to total variation. At Mandya and Gangavathi, the first 4 principal components (PCs) with eigenvalues greater than one accounted for 64.35 % and 69.01 % of the total variability, respectively. Traits such as days to 50 % flowering, productive tillers (PT), spikelet fertility (SF) and panicle weight (PW) were major contributors to variability across locations. Principal component analysis biplots effectively visualized trait interrelationships and enabled the identification of superior-performing germplasm accessions. Germplasms located near a trait vector within the same quadrant of a PCA biplot are considered superior performers for those specific traits. Accessions such as IRGC 898, IRGC 1841 and IRGC 1022 were ranked higher based on their position and alignment with key trait vectors in the biplot. These accessions can be utilised as donor parental lines for combination breeding programs and for developing biparental mapping populations to identify Quantitative Trait Loci (QTLs) linked to specific traits.

Keywords: combination breeding; eigenvalues; genetic diversity; germplasm accessions; principal component analysis

Introduction

Rice (*Oryza sativa* L.) belongs to the tribe *Oryzeae*, family Poaceae, with a chromosome number of $2n = 24$. There are 24 species, of which two (*Oryza sativa* L. and *Oryza glaberrima* L.) are domesticated species and the remaining 22 are wild species. Rice plays an indispensable role in global food security due to its affordability, versatility and high caloric value, making it a vital part of the human diet. The global population continues to rise rapidly, placing immense pressure on crop production. By 2050, the population is expected to increase by 10 billion and we need to increase the yield of the staple crops by 70 % to meet the population demand (1). Among cereals, rice stands out as a primary staple, nourishing over half of the world population and serving as a pillar of food security. In the coming decades, the expansion of rice cultivation will face significant constraints due to the attenuation of land and water resources, driven by climate change, rapid urbanization and population growth (2). To meet imminent needs, enhancing rice yield potential and improving its resilience to climate change are decisive for ensuring sustainable development and adequate nutrition worldwide (3, 4).

Grain yield (GY) exhibits low heritability and is highly influenced by key independent traits. Therefore, it may be effectively improved through indirect selection of yield-attributing traits that are strongly correlated with yield, thereby refining overall crop productivity. The development of high-yielding, resilient varieties is possible by harnessing the genetic diversity within germplasm accessions. Since rice germplasm collections from diverse cultivated and wild species have made notable contributions to rice breeding (5) and serve as the fundamental resource for crop improvement, providing a rich reservoir of genetic variation essential for breeding programs. Exploiting the genetic variation present in diverse germplasm collections enables breeders to refine crucial agronomic traits such as height of the plant, tiller number, panicle architecture and grain filling efficiency, which directly influence productivity in low-yielding cultivars and enable breeders to develop superior cultivars that can adapt to changing environments and meet the rising demand for food (6, 7). Ultimately, the effective use of germplasm is vital for sustainably enhancing yields of the crop and securing global food security. Therefore, as the most important

staple crop, rice productivity can be enhanced by exploiting the genetic diversity present in existing germplasm and landraces using multivariate techniques (8).

Multivariate methods play a decisive role in assessing genetic diversity across diverse germplasm collections, as they enable the analysis and interpretation of complex, multidimensional genetic variable data. Principal component analysis (PCA), cluster analysis and discriminant analysis are key multivariate statistical methods used to explore genetic diversity and classify genotypes based on multiple traits (9). Principal component analysis is the true eigenvector-based multivariate data analysis through projection methods and represents different axes of differentiation through canonical roots or vectors. The original, large dataset, which includes many variables, can be summarized in just a few variables called principal components (PCs). It identifies the minimum number of components that can explain the maximum variability within the total variation of the population (10). Its main goal is to minimise the complexity of a multivariate dataset by converting it into a minimal number of variables, commonly known as summary indices (11). In light of these aspects, the current study aimed to assess genetic variability among a subset of germplasm accessions using PCA at two locations.

Materials and Methods

The material for the study entails 300 germplasm accessions, which is a subset of Three Thousand Rice Panel (3KRG) of the International Rice Research Institute, Philippines, was chosen based on visual observations on traits such as tillering ability, culm strength, lodging resistance, leaf senescence, spikelet fertility (SF) and phenotypic acceptance. These accessions represent different sub-groups of *O. sativa* i.e., *indica*, *japonica*, *tropical*, *temperate*, *aus* and *aromatic* of different countries of origin. A comprehensive list of germplasm accessions evaluated in this study is provided in Supplementary Table 1. The experiment was conducted during kharif 2023 using an alpha lattice design with two replications at two locations: the Zonal Agricultural Research Station, V.C. Farm, Mandya and the Agricultural Research Station, Gangavathi. Geographically, Mandya is situated in the Southern Dry Zone of Karnataka (Zone 6) at approximately 12°31' N latitude and 76°54' E longitude with an elevation of about 695 m above mean sea level. It received an average rainfall of around 297.4 mm during kharif 2023 (Supplementary Table 2). The Gangavathi lies in the Northern Transitional Zone of Karnataka (Zone 8) at 15°26' N latitude and 75°07' E longitude with an altitude of 678 m above mean sea level and it received about 309.9 mm of rainfall during kharif 2023 (Supplementary Table 3). These contrasting agro-climatic conditions provided a suitable framework for evaluating the genotypes across diverse environments. Standard agronomic practices recommended by the University of Agricultural Sciences, Bangalore, were followed to ensure a healthy crop stand. Twenty-

one-day-old seedlings were transplanted into the main field at a spacing of 20 cm × 15 cm (row-to-row × plant-to-plant), following the alpha lattice experimental design. The phenotypic observations were registered on five randomly selected plants on yield component traits like days to 50 % flowering, height of the plant, panicle length (PL), panicle weight (PW), productive tillers (PT), SF, test weight (TW), GY per plot and grain zinc (Zn) content. The grain samples were dehusked, then polished. A representative polished rice sample from each accession, collected from both environments, was analysed for grain Zn concentration using Energy Dispersive X-ray Fluorescence Spectrometry (ED-XRF) (X-Supreme8000, Oxford Instruments, CA, USA) at IIRR, Hyderabad. This non-destructive analytical method estimated Zn content in parts per million (mg kg⁻¹). From the mean data of 300 accessions for nine quantitative traits, PCA was performed using R-statistical software (4.5.1) with the packages '*prcomp*' and '*factoextra*'. The norms followed for selecting the PCs were based on eigenvalues greater than 1.

Results and Discussion

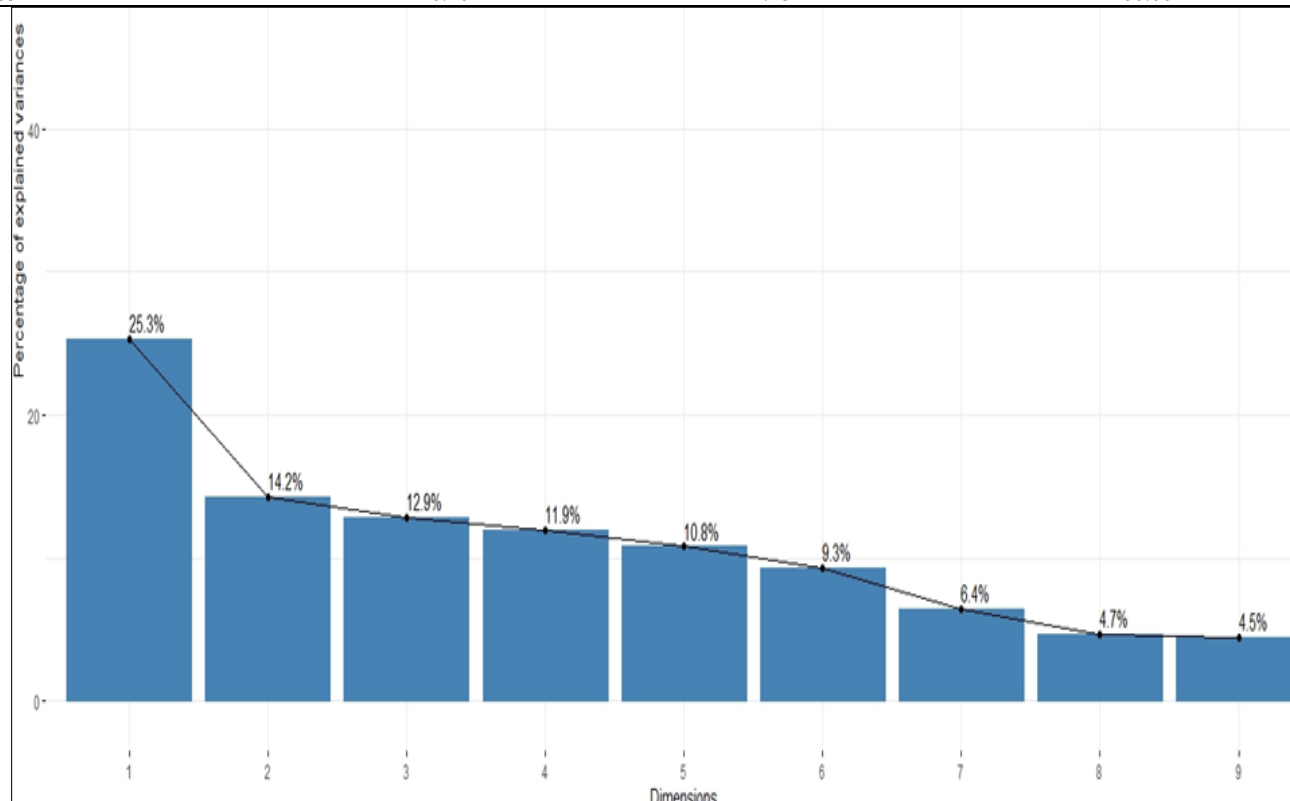
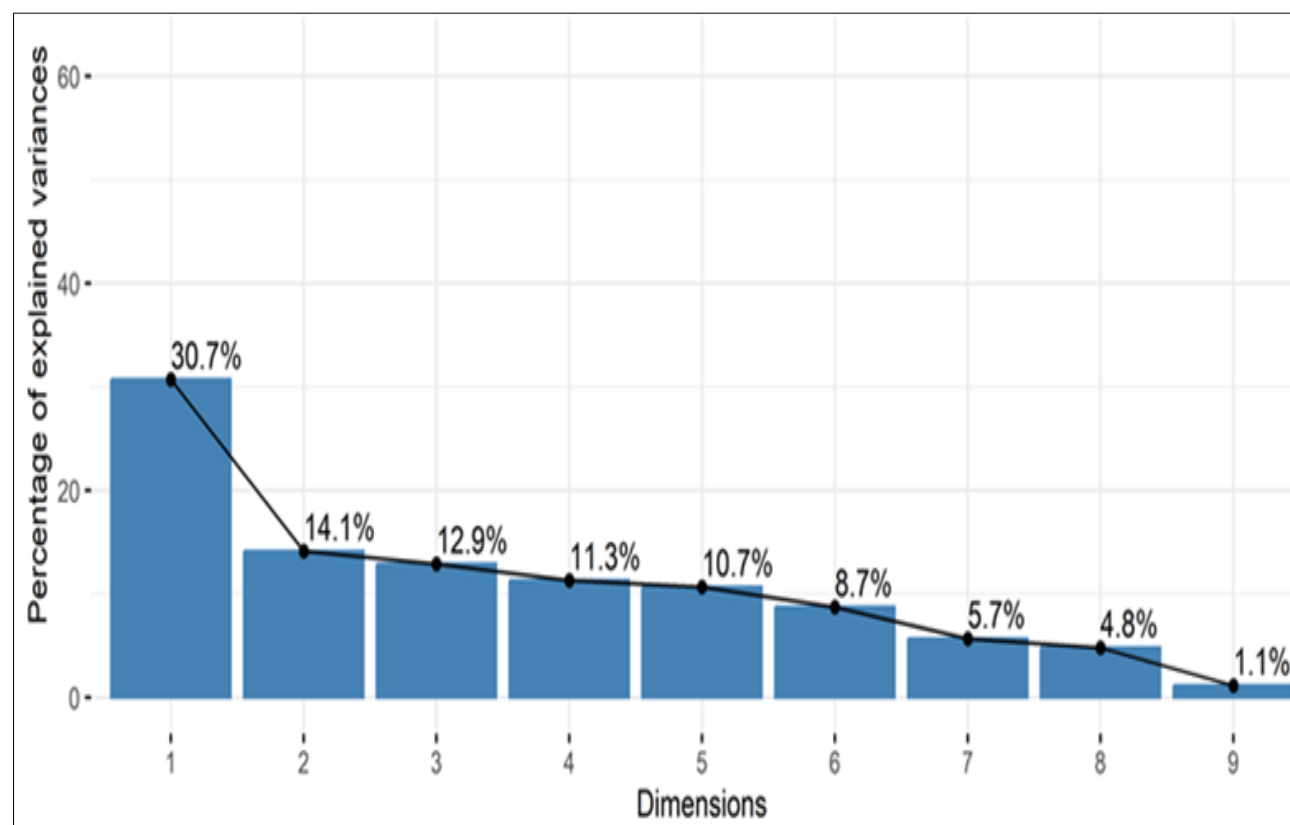
Principal component analysis is performed by eigenvalue decomposition of either the symmetric covariance or correlation matrix. The outcomes are expressed in terms of component scores, also known as factor scores and loadings. Principal component analysis reduces the dimensionality of the data without losing information from the original data. Thus, it retains original information as much as possible and each component accounts for a certain percentage of the total variability present in the dataset (12). In this regard, the first PC captures the largest proportion of the dataset's total variation, with each subsequent component accounting for progressively less. Based on the criteria, the first 4 PCs are generally considered crucial for capturing the major patterns of variation among different accessions (13, 14). The traits associated with these PCs are particularly valuable for effectively distinguishing among the accessions. In a PCA biplot, the cosine of the angle between trait vectors approximates the correlation between the corresponding traits (15). The PCs formed were equal to the number of traits under study. In the present study, the eigenvalues, percentage of variation and cumulative variance explained by the first 9 PCs were analysed for both the Mandya (Table 1) and Gangavathi (Table 2) locations. Among the nine components extracted, only four had eigenvalues greater than one, suggesting that these components accounted for the majority of the variation among the germplasm accessions. This selection criterion aligns with the rule, which recommends retaining PCs with eigenvalues greater than one (16). The percentage of variation explained by the nine PCs was illustrated using scree plots for Mandya (Fig. 1a) and Gangavathi (Fig. 1b) locations separately. The PC1 accounted for 25.33 % of the total variability with an eigenvalue of 2.28 at the Mandya location

Table 1. Eigenvalues, percent variation and cumulative variance percentage of principal components at Mandya

Principle Components	Eigenvalues	Variance percentage	Cumulative variance percentage
PC1	2.28	25.33	25.33
PC2	1.28	14.24	39.56
PC3	1.16	12.85	52.42
PC4	1.07	11.93	64.35
PC5	0.97	10.80	75.15
PC6	0.83	9.27	84.42
PC7	0.58	6.42	90.83
PC8	0.42	4.70	95.54
PC9	0.40	4.46	100.00

Table 2. Eigenvalues, percent variation and cumulative variance percentage of principal components at Gangavathi

Principle Components	Eigenvalues	Variance percentage	Cumulative variance percentage
PC1	2.77	30.72	30.72
PC2	1.27	14.11	44.83
PC3	1.16	12.88	57.71
PC4	1.02	11.29	69.01
PC5	0.96	10.67	79.68
PC6	0.79	8.73	88.40
PC7	0.51	5.66	94.07
PC8	0.43	4.79	98.85
PC9	0.10	1.15	100.00

**Fig. 1a.** Scree plot illustrating the variance explained by principal components at Mandya.**Fig. 1b.** Scree plot illustrating the variance explained by principal components at Gangavathi.

and 30.72 % of the variability with an eigenvalue of 2.77 at the Gangavathi location. Similarly, a maximum variation of 20.82 % explained by PC1 in the study involving 67 rice lines (17). Therefore, the germplasm accessions identified under PC1 could serve as valuable resources in future breeding programs, particularly for improving traits that contributed most to the observed variability, such as days to 50 % flowering and PT at the Mandya location and SF along with days to 50 % flowering at the Gangavathi location. The first four PCs (PC1, PC2, PC3 and PC4) together accounted for 64.35 % of the total variability at Mandya and 69.01 % at Gangavathi, indicating that these components captured the majority of the variation among the germplasm accessions at both locations. These results are consistent with previous research, in which four of nine PCs with eigenvalues greater than one accounted for 70.14 % of the total variability among the traits studied (18). Similar findings were reported in other studies, where the first four PCs with eigenvalues exceeding one were also identified as significant contributors to trait variation (19). At the Mandya location, PC1 accounted for 25.33 % of total variability with significant factor score of days to 50 % flowering (0.48), PT (0.47), PW (0.44), PL (0.36), SF (0.34), TW (0.22) and GY (0.15) which are positively correlated (Table 3). Whereas, at Gangavathi location, PC1 added 30.72 % to the total variability with significant loading of days to 50 % flowering (0.52), SF (0.52), PT (0.42), PW (0.36), PL (0.26), GY (0.16) and TW (0.14), which are highly correlated (Table 4). These results indicate that the listed traits contributed most significantly to the total variation at both locations.

The PCA biplot offers a comprehensive visual representation of trait interrelationships and enables the assessment of germplasm performance based on its association with key contributing traits. The length of each trait vector indicates its contribution to the total variation; the longer the vector length, the greater the contribution of the concerned traits towards the economic yield. At the Mandya location, the trait days to 50 % flowering exhibited the longest vector length in the PCA biplot, indicating its major contribution to total genetic diversity, followed by PT and PW. In contrast, at the Gangavathi location, SF %, measured over days to 50 % flowering, showed the longest vectors, followed by PL, highlighting their prominent roles in contributing to genetic variation among the

accessions. The distribution and pattern of diversity among genotypes and quantitative traits are effectively illustrated in biplot diagrams of PC1 and PC2 for Mandya (Fig. 2a) and Gangavathi (Fig. 2b). The angle between trait vectors in the biplot indicates the nature of their association. An acute angle ($< 90^\circ$) suggests a positive correlation, an obtuse angle ($> 90^\circ$) indicates a negative correlation and a right angle (90°) implies no correlation between the traits (20). Among the nine traits studied, days to 50 % flowering, PT, PW, PL, SF and TW showed highly significant correlations with GY. In contrast, plant height (PH) and grain Zn concentration showed negative correlations with GY at both locations. Germplasms located near a trait vector within the same quadrant are considered superior performers for those specific traits (20). Therefore, such germplasms can be effectively utilized in combination breeding programs to improve specific traits that are lacking in popular high-yielding varieties.

In the PCA biplot, germplasm accessions positioned farther from the origin in the direction of positive and desirable traits are considered superior and thus ranked higher and germplasms in the opposite direction of desirable traits are ranked lower. At the Mandya location (Fig. 3a), the germplasm IRGC 898, followed by IRGC 1841 and IRGC 1022, was located furthest from the origin along the vectors of key characters such as days to 50 % flowering, PT and PW, indicating superior performance. Similarly, at the Gangavathi location (Fig. 3b), the same germplasms were positioned away from the origin toward desirable attributes such as SF, days to 50 % flowering and PL, suggesting consistent superior performance across all key traits across environments.

Conclusion

The current study revealed substantial genetic variability among the evaluated rice germplasms. Principal component analysis identified key traits, such as days to 50 % flowering, SF, PL and PT, as major contributors to the observed variation among accessions. Selection of germplasms based on these traits would be highly valuable for targeted breeding programs. Germplasms that ranked higher in the PCA can serve as potential parental lines in biparental mapping populations to identify Quantitative Trait Loci (QTLs) associated with these agronomic traits. The identified QTLs can

Table 3. Proportion of variance explained by the first four principal components among rice germplasm accessions, based on eigenvectors at Mandya

Traits	PC1	PC2	PC3	PC4
Days to 50 % flowering	0.48	-0.16	0.09	0.10
Plant height	-0.05	-0.18	-0.63	-0.10
Productive tillers	0.47	-0.18	-0.21	0.26
Panicle length	0.36	0.36	-0.26	-0.04
Panicle weight	0.44	-0.19	0.32	-0.30
Spikelet fertility	0.34	0.26	0.20	-0.43
Test weight	0.22	0.46	-0.43	0.11
Grain yield	0.15	-0.66	-0.20	0.04
Grain zinc	-0.11	-0.11	-0.28	-0.78

Table 4. Proportion of variance explained by the first four principal components in rice germplasm accessions, based on eigenvectors at Gangavathi

Traits	PC1	PC2	PC3	PC4
Days to 50 % flowering	0.52	0.12	0.10	-0.05
Plant height	-0.02	0.03	-0.60	-0.07
Productive tillers	0.42	-0.08	-0.19	-0.08
Panicle length	0.26	-0.55	-0.17	0.38
Panicle weight	0.36	0.18	0.09	0.55
Spikelet fertility	0.52	0.11	0.10	-0.11
Test weight	0.14	-0.57	-0.25	-0.40
Grain yield	0.16	0.52	-0.40	-0.33
Grain zinc	-0.13	0.12	-0.55	0.48

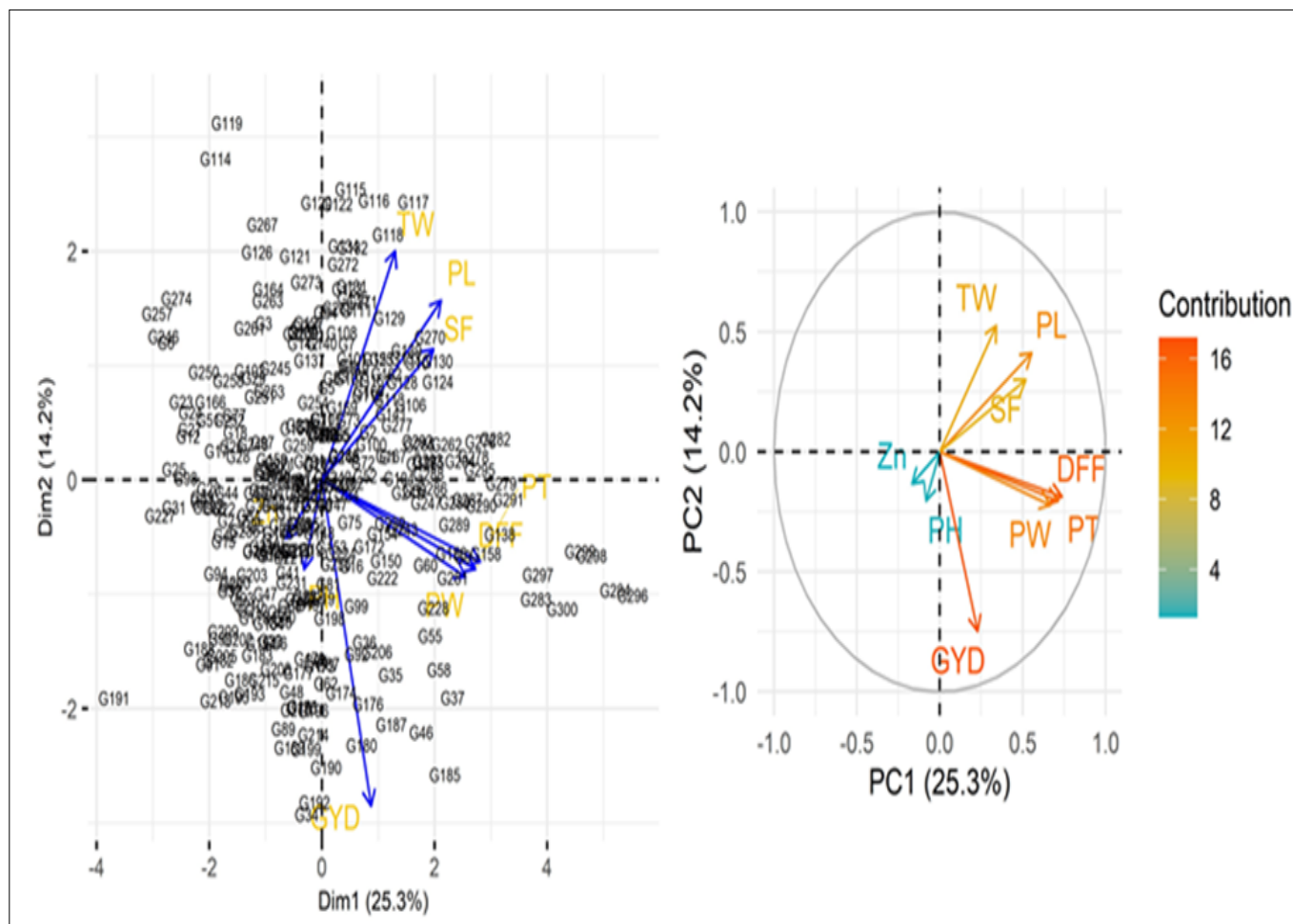


Fig. 2a. PCA-based depiction of the variables and distribution of genotypes based on the first two principal components at Mandya.

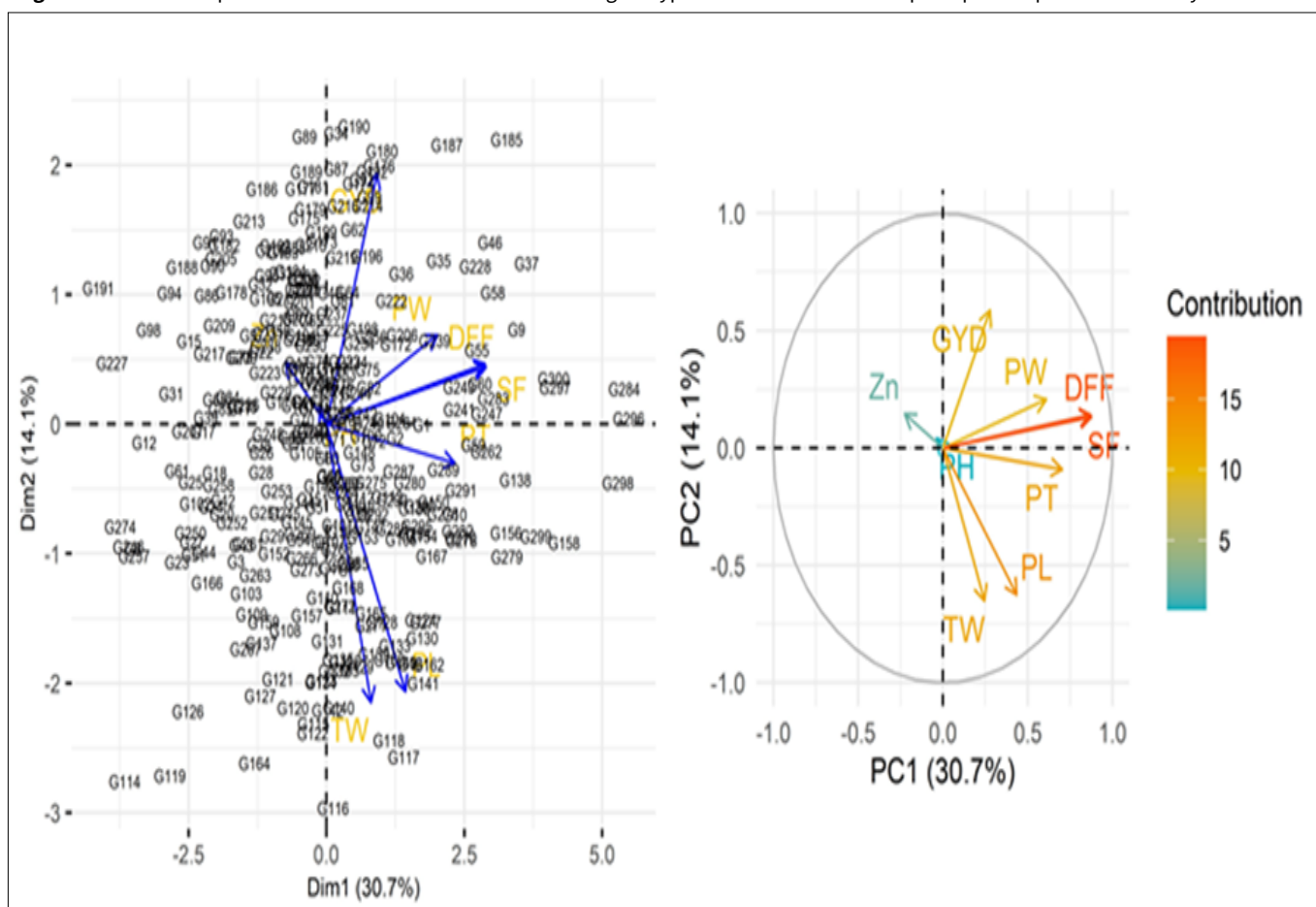


Fig. 2b. Principal component analysis-based depiction of the variables and distribution of genotypes based on the first two principal components at Gangavathi.

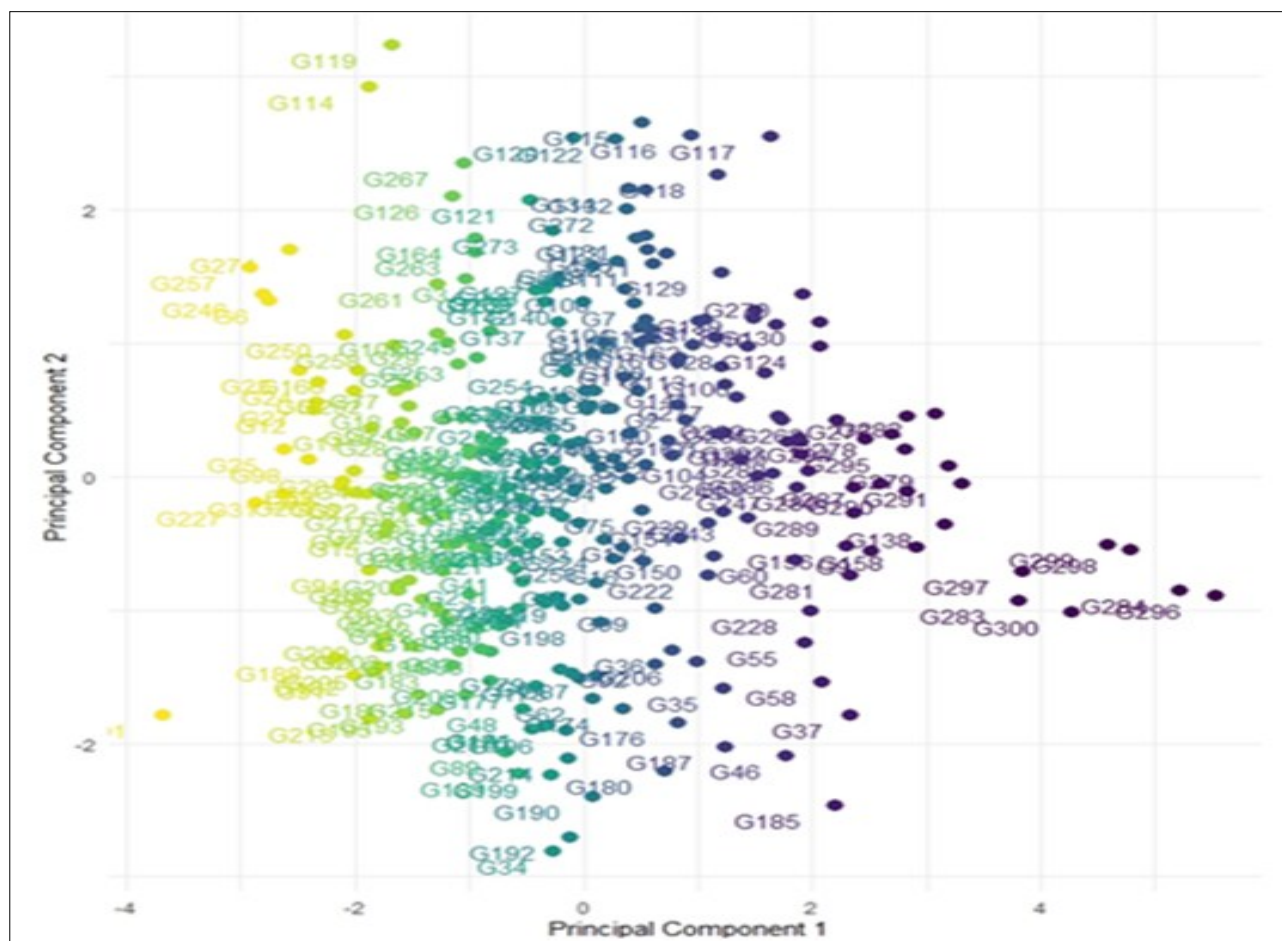


Fig. 3a. Principal component analysis of genotypes based on ranking at Mandya.

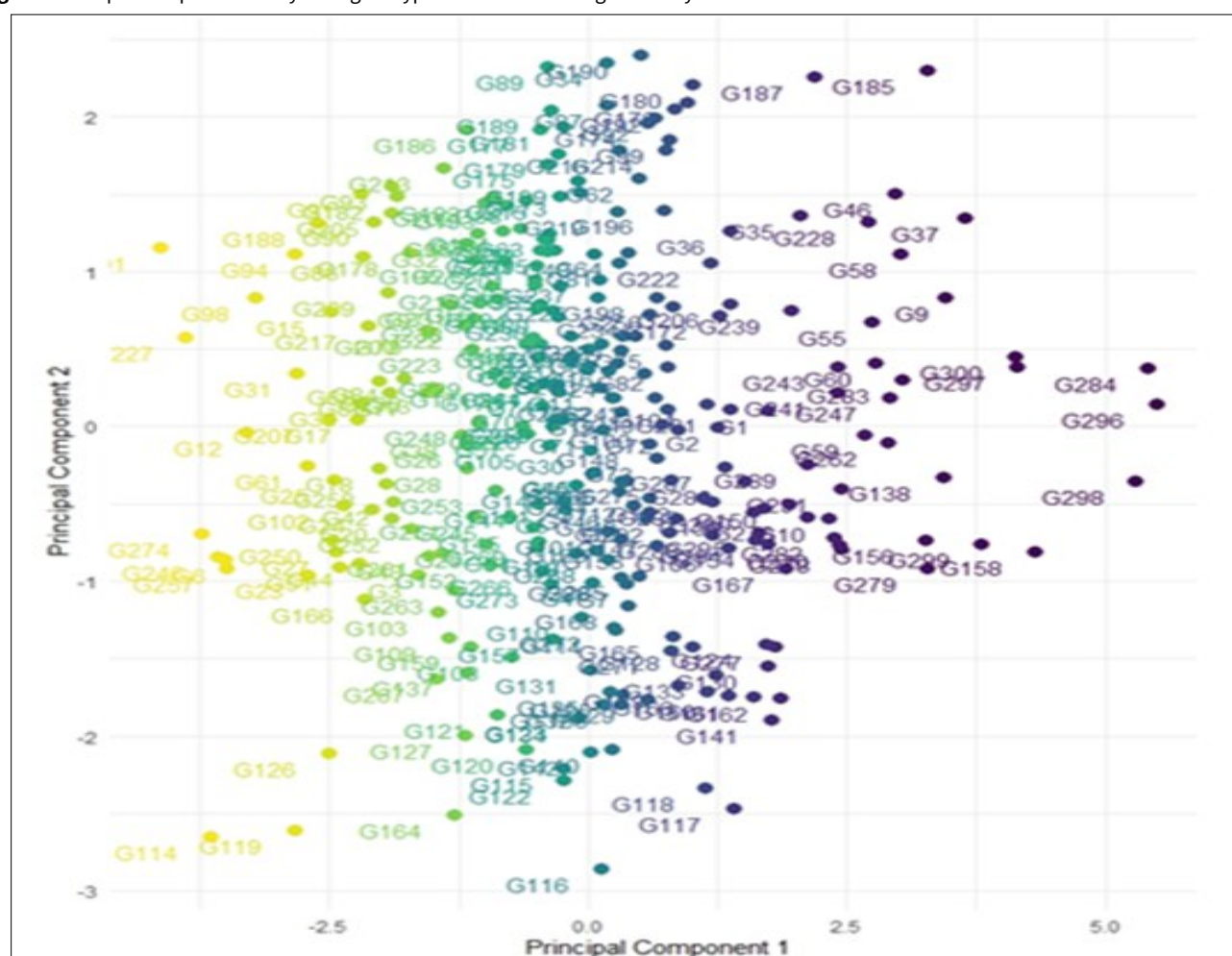


Fig. 3b. Principal component analysis of genotypes based on ranking at Gangavathi.

subsequently be introgressed into elite but trait-deficient cultivars through marker-assisted backcrossing, thereby accelerating genetic gain and improving specific trait performance.

Acknowledgements

The authors sincerely acknowledge the College of Agriculture, Mandya and the Agricultural Research Station (ARS), Gangavathi, for providing the necessary field facilities and support to carry out the research work.

Authors' contributions

SSR conducted the experiment and wrote the manuscript. SAU and PM recorded phenotypic observations. SGM and SP handled data analysis. AKR and SN prepared tables and figures. DCA supervised the study and provided critical revisions. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interest to declare.

Ethical issues: None

References

- Rahman MH. Exploring sustainability to feed the world in 2050. *J Food Microbiol.* 2016;1(1):7–16.
- Soe I, Tamu A, Asante M, Nyadanu D, Akromah R. Genetic diversity analyses of rice germplasm using morphological traits. *J Plant Breed Crop Sci.* 2019;11(4):128–36. <https://doi.org/10.5897/JPBCS2018.0786>
- Kingra PK, Kaur R, Kaur S. Climate change impacts on rice (*Oryza sativa*) productivity and strategies for its sustainable management. *Indian J Agric Sci.* 2019;89(2):171–80. <https://doi.org/10.56093/ijas.v89i2.86981>
- Ashkra, Khan AA, Jadaun KK. Estimating the potential effect of climate change on rice yield in India by considering the combined effects of temperature and rainfall. *Bhartiya Krishi Anusandhan Patrika.* 2023;38(3):284–9. <https://doi.org/10.18805/BKAP649>
- Sheela KRVS, Robin S, Manonmani S. Principal component analysis for grain quality characters in rice germplasm. *Electron J Plant Breed.* 2020;11(1):127–31. <https://doi.org/10.37992/2020.1101.023>
- Zhu Q, Zhang XL, Ni Naing NNZ, Li JQ, Chen LJ, Lee DS. Strategies for rice improvement: utilizing genetic resources from wild and cultivated *Oryza* species. *Rice Genomics Genet.* 2024;15(3):106–20. <https://doi.org/10.5376/rgg.2024.15.0012>
- Shanmugam S, Samonte SOPB, Sanchez DL, Yabes JC. Genetic assessment of diverse rice germplasm for yield-related and stress tolerance traits. *Genet Resour Crop Evol.* 2025;72(4):885–904. <https://doi.org/10.1007/s10722-024-02004-4>
- Ashfaq M, Rasheed A, Zhu R, Ali M, Javed MA, Anwar A, et al. Genome-wide association mapping for yield and yield-related traits in rice (*Oryza sativa* L.) using SNP markers. *Genes.* 2023;14(5):1089. <https://doi.org/10.3390/genes14051089>
- Oyelola BA. The Nigerian Statistical Association preconference workshop. Ibadan: University of Ibadan; 2012. p. 20–1.
- Anderson TW. An introduction to multivariate analysis. New Delhi: Wiley Eastern Pvt Ltd; 1972.
- Adams MW. An estimate of homogeneity in crop plants with special reference to genetic vulnerability in the dry bean, *Phaseolus vulgaris* L. *Euphytica.* 1977;26:665–79.
- Das S, Das SS, Chakraborty I, Roy N, Nath MK, Sarma D. Principal component analysis in plant breeding. *Biomolecules.* 2017;7:1–3.
- Clifford HT, Stephenson W. An introduction to numerical classification. New York: Academic Press; 1975. p. 229.
- Guei RG, Sanni KA, Abamu FJ, Fawole I. Genetic diversity of rice (*Oryza sativa* L.). *Agronomie Africaine.* 2005;5:17–28.
- Abdi H. Partial least squares regression and projection on latent structure regression (PLS regression). *Wiley Interdiscip Rev Comput Stat.* 2010;2(1):97–106. <https://doi.org/10.1002/wics.51>
- Jeffers JNR. Two case studies in the application of principal component analysis. *Appl Stat.* 1967;16:225–36. <https://doi.org/10.2307/2985919>
- Rahangdale S, Singh Y, Upadhyay P, Koutu G. Principal component analysis of JNPT lines of rice for the important traits responsible for yield and quality. *Indian J Genet.* 2021;81(1):127–31. <https://doi.org/10.31742/IJGPB.81.1.14>
- Shoba D, Vijayan R, Robin S, Manivannan N, Iyanar K, Arunachalam P, et al. Assessment of genetic diversity in aromatic rice (*Oryza sativa* L.) germplasm using PCA and cluster analysis. *Electron J Plant Breed.* 2019;10(3):1095–104. <https://doi.org/10.5958/0975-928X.2019.00140.6>
- Gour L, Maurya SB, Koutu GK, Singh SK, Shukla SS, Mishra DK. Characterization of rice (*Oryza sativa* L.) genotypes using principal component analysis including scree plot and rotated component matrix. *Int J Chem Stud.* 2017;5(4):975–83.
- Christina GR, Thirumurugan T, Jeyaprakash P, Rajanbabu V. Principal component analysis of yield and yield related traits in rice (*Oryza sativa* L.) landraces. *Electron J Plant Breed.* 2021;12(3):907–11.

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonpublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc. See https://horizonpublishing.com/journals/index.php/PST/indexing_abstracting

Copyright: © The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited (<https://creativecommons.org/licenses/by/4.0/>)

Publisher information: Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.