Agriculture Association of Textile Chemical and Critical Reviews Journal (2025) 672-676

AATCC Review

09 July 2025: Received
28 August 2025: Revised
06 September 2025: Accepted
04 October 2025: Available Online

https://aatcc.peerjournals.net/

Original Research Article

Open Access

Evaluation of genetic diversity in rice (*oryza sativa* L.) germplasm accessions employing multivariate analysis and association studies for yield and nutritional traits

M Madhuri[1], **Nandan Mehta**[1], **R Abdul Fiyaz***[2], **K Chaitanya**[3], **M Bharath**[3]

[1]*Indira Gandhi Krishi Vishwavidyalaya, Raipur, C.G-492012, India*
[2]*Indian Institute of Rice Research, Hyderabad, Telangana-500030, India*
[3]*Indian Institute of Oilseeds Research, Hyderabad, Telangana-500030, India*

ABSTRACT

Genetic variability and correlation of characters in rice (Oryza sativa L.) are crucial to the development of high-yielding and nutritionally improved cultivars. Principal Component Analysis (PCA) was used in the current study to determine phenotypic variation between 183 rice germplasm accessions tested over four consecutive seasons kharif 2021 and 2022 and rabi 2022 and 2023 with regard to yield and nutritional traits. Observations were recorded for eleven yield attributes like plant height, panicle length, tillers number, productive tillers number, days to reach 50% flowering, panicle weight, number of filled grains, number of unfilled grains, fertile spike lets, per plant yield and test weight, and two nutritional attributes viz., grain iron and zinc contents. PCA retained six principal components with eigenvalues greater than one, explained a combined 78.5% of the total phenotypic variance. Yield-contributing characteristics and micronutrient levels were the primary drivers of variability and discriminated among genotypes as key discriminators. Five different clusters were classified using cluster analysis, with Cluster I having the highest number of accessions and Cluster V having the least. Among all the traits, zinc content showed maximum genetic divergence. Biplot analysis also revealed a significant positive correlation of single plant yield with all the traits, except plant height, which showed a negative correlation. Similarly, these relationships were confirmed by correlation analysis, suggesting that the traits positively correlating with yield may be effective selection traits. Yield traits shown high environmental variability, whereas micronutrients have low phenotypic variation but high analytical sensitivity which was a challenge of the study and it was overcome by eliminating biasedness through proper standardization of principle component analysis. In all, the findings provide useful information about the multivariate pattern of phenotypic variation and facilitate identification of donor lines for combined improvement of yield and micronutrient use efficiency in rice.

Keywords: *Principal component analysis, genetic diversity, trait association, biplot, rice yield, micronutrient, correlation analysis and clusters.*

Introduction

Rice (*Oryza sativa* L.) is the major staple crop for more than half of the world's population, being the major source of calories for almost 3.5 billion individuals around the world (Alam *et al.*, 2024). Aside from its nutritional significance, rice cultivation is a important in maintaining national economies, especially in the developing world where it is the mainstay of agricultural livelihoods. Thus, the selection and cultivation of high-yielding rice varieties are vital to address the increasing global food demand and achieve economic stability in food-dependent countries (Meena *et al.*, 2023). Given the population growth forecast in the next few decades, increasing the productivity and nutritional value of rice has become a strategic necessity to ensure world food and nutritional security. Micronutrient inadequacies, referred to as "hidden hunger," are a common public health issue in women of reproductive age, children, and vulnerable groups. Low consumption of important micronutrients like iron and zinc may result in weakened immune function, growth retardation, diminished physical and cognitive capacity, and total health decline (Stevens *et al.*, 2022). To overcome this, the incorporation of micronutrient enrichment especially iron and zinc concentration into high-yielding rice varieties has emerged as a focal point of contemporary bio fortification as well as breeding programs, particularly in resource-poor areas. Breeding for concurrent enhancement of yield as well as nutritional quality is, however, complicated owing to the polygenic nature of most traits and their intensive association with the environment. As such, recent rice improvement schemes increasingly depend on sophisticated statistical and computational methodology for handling multidimensional data obtained from agronomic and nutritional tests. Among them, multivariate statistical methods and association studies are effective tools to examine variation in traits, comprehend the genetic architecture of complex traits, identify interrelationships between traits and select superior types with favorable associations of yield and quality components. Principal Component Analysis (PCA), especially useful in unraveling phenotypic and genotypic variation by reducing big sets of inter-related traits into fewer representative components.

---

*Corresponding Author: **R Abdul Fiyaz***

DOI: https://doi.org/10.21276/AATCCReview.2025.13.04.672

PCA helps the breeders in rice breeding programs to determine the most contributing traits to genetic differentiation and thus select genotypes that exhibit high yield potential and increased micronutrient content at the same time (Paramanik *et al*., 2025). By concentrating on previously selected elite germplasm lines, PCA can narrow down breeding plans in order to select donors with a balance of yield efficiency and nutritional supremacy. Additionally, correlation analysis is a complement to PCA since it provides quantification of the direction and magnitude of the association between traits and hence helps in designing a good selection index. Positive associations mean that increase in one attribute will likely increase another, while negative associations imply possible trade-offs among attributes (Adams and Grafius, 1971; Teja Sri *et al*., 2025). Therefore, it is important to know them for formulating breeding strategies for the simultaneous improvement of yield and nutritional value. The current study is focused on elucidating genetic variation and relationships between yield and micronutrient traits in superior rice germplasm through Principal Component Analysis and correlation analysis.

Materials and Methodology

The experimental material comprised 183 rice germplasm lines evaluated during four seasons *kharif* 2021, *kharif* 2022, *rabi* 2022 and *rabi* 2023. The trial was conducted in an augmented design consisting of three replications and five standard checks. Twenty-one-day-old seedlings were transplanted with a spacing of 15 cm between plants and 20 cm between rows. Gaps were filled within a week to ensure a uniform plant population along with standard agronomic practices for normal crop growth. Observations were recorded on eleven yield-related traits like height of the plant, length of the panicle, tiller count, productive tiller count, number of days to attain 50% flowering, panicle weight, number of filled grains, number of unfilled grains, fertile spikelet, per plant yield and test weight. Two nutritional traits, grain iron and zinc contents were also assessed. Data collection involved five randomly selected plants from each replication, following standard evaluation procedures. Grain iron and zinc concentrations were measured using Energy Dispersive X-Ray Fluorescence (ED-XRF). Using R software (v. 4.5.1), PCA was performed on the mean values of each trait to identify traits contributing most to genetic variation among the genotypes. Correlation coefficients were computed at both genotypic and phenotypic levels as suggested by Falconer (1964) to find out the association among yield and micronutrient traits.

Results and Discussion

PCA was performed using data from thirteen traits, comprising eleven yield-related traits and two nutritional parameters (iron and zinc contents) across 183 rice genotypes evaluated over four seasons. Eigenvalues were employed to measure the total principal components to retain, as values >1 indicates significant explanatory power, while those below one suggest lesser contribution to total variance (Jadhav *et al*., 2023). Among the thirteen generated components, six exhibited eigenvalues exceeding unity and were, therefore, considered significant for interpretation. In combination, these six components captured 78.5% of total phenotypic variability. The majority of variation (18.7%) was contributed by the first principal component (PC1), followed by PC2 (16%), PC3 (15%), PC4 (11%), PC5 (9.5%), and PC6 (8.3%). The percentage of variance explained by each successive component declined gradually from PC1 to PC13, while cumulative variance increased correspondingly. The (Fig. 1) showed a steady decline in eigenvalues, and the analysis confirms that the first six PCs contributed most to the observed variability among the genotypes. A summary of eigenvalues, percentage of variance, and cumulative contribution of each principal component is depicted in Table 1. Comparable results have been documented by Singh *et al*. (2020), where five principal components explained 73.84% of total variability. Trait loadings on principal components (Table 2) indicated the relative contribution of individual traits to total genetic diversity among genotypes. Positive and negative loadings signify the magnitude and direction of association of traits with each component (Sruthi *et al*., 2024). The tiller number (0.507), productive tiller number (0.509), number of unfilled grains (0.291), zinc content (0.182), and iron content (0.145) had the largest positive loadings on PC1, suggesting their greatest impact on overall variability. Contrarily, spikelet fertility (-0.384), panicle weight (-0.32), filled grains number (-0.223), and plant height (-0.157) had negative loadings, indicating lower contribution to PC1. Similar trends have been reported by Sruthi *et al*. (2024), Sandhya *et al*. (2022), and Christina *et al*. (2021). All the traits related to yield had positive loadings on PC2, while zinc (-0.087) and iron (-0.155) contents had negative loadings. PC3 was influenced by the number of spikelet fertility (0.4611), iron concentration (0.429), and zinc concentration (0.2831), whereas days to reach 50% flowering (-0.0253), plant height (-0.0457), panicle length (-0.0769), number of filled grains (-0.1222) and number of unfilled grains (-0.5166) negatively contributed.

PC4 exhibited positive correlation with days to reach 50% flowering (0.4165), spikelet fertility (0.3116), and number of filled grains (0.2914), but with negative correlation for test weight (-0.5492) and panicle length (-0.3822).

PC5 exhibited positive correlation for test weight (0.2618), number of productive tillers (0.2068), and number of tillers (0.2066), whereas negative loadings were found for zinc (-0.673) and iron (-0.4082).

PC6 had positive correlations with single plant yield (0.4026), panicle weight (0.3329), and number of unfilled grains (0.2877), but negatively correlated with panicle length (-0.4039) and plant height (-0.4925).

As shown in Fig. 3, the maximum contribution of PC1 suggests that genotypes with high scores on this component are likely to have better yield attributes like the number of tillers and productive tillers, along with increased micronutrient contents of iron and zinc. Therefore, the selection of genotypes with promising combinations of these traits would be advantageous for breeding programs for improving yield potential as well as nutritional content in rice.

Biplot Analysis

Biplot analysis has a graphic view of interrelationships among characters and assists in selecting genotypes that are better for given characters. It also depicts the direction and magnitude of variability displayed by traits and genotypes, which helps provide a better view of the general structure of the data (Yan and Kang, 2002). The nature of the correlation between characteristics is established by the angle created between their vectors; acute angles represent positive relationships, while obtuse angles represent negative associations (Yan and Tinker, 2006).

The biplot diagrams (Fig. 2.1 and Fig. 2.2) revealed that per plant yield was strongly and positively associated with several important yield and nutritional traits, including panicle weight, number of tillers, number of productive tillers, number of filled grains, spikelet fertility, days to 50% flowering, panicle length, test weight, iron content, zinc content, and number of unfilled grains (0.007), as these traits formed acute angles with yield. In contrast, plant height exhibited an obtuse angle with yield, indicating a negative association.

The length of each vector in the biplot represents the contribution of the respective trait to total variability. Longer vectors signify greater discriminatory power of the trait in differentiating genotypes. Thus, traits such as panicle weight, number of tillers, productive tillers, and micronutrient contents play a major role in distinguishing high-performing genotypes.

Table 1. Principal component, eigenvalue, percentage of variance, cumulative % of variance

Principal Component	Eigen value	% of variance	Cumulative % of variance
PC1	2.425	18.651	18.651
PC2	2.061	15.857	34.508
PC3	1.955	15.039	49.548
PC4	1.447	11.129	60.677
PC5	1.234	9.492	70.168
PC6	1.078	8.29	78.459

Table 2. Individual trait loadings on PC scores

Trait	PC1	PC2	PC3	PC4	PC5	PC6
PH	-0.157	0.286	-0.044	-0.298	-0.202	-0.516
PL	-0.09	0.313	-0.085	-0.39	-0.088	-0.398
NT	0.507	0.293	0.209	0.064	0.208	-0.141
NPT	0.509	0.288	0.211	0.07	0.209	-0.143
PW	-0.32	0.442	0.113	-0.171	0.016	0.331
NFG	-0.223	0.445	-0.146	0.284	-0.251	0.183
NUFG	0.291	0.172	-0.519	-0.174	0.21	0.277
SF	-0.384	0.055	0.451	0.321	0.067	-0.182
TW	-0.089	0.037	0.279	-0.535	0.263	0.14
DFF	0.053	0.183	0.037	0.412	-0.243	-0.259
FE	0.145	-0.146	0.438	-0.185	-0.407	0.191
ZN	0.182	-0.083	0.29	-0.08	-0.675	0.017
SPY	0.048	0.404	0.202	0.1	0.051	0.398

PH-plant height, PL-panicle length, NT- number of tillers, NPT- number of productive tillers, PW- panicle weight, NFG- number of filled grains, NUFG- number of unfilled grains, SF- spikelet fertility, TW- test weight, DFF- days to 50% flowering, FE- iron content, ZN- zinc content

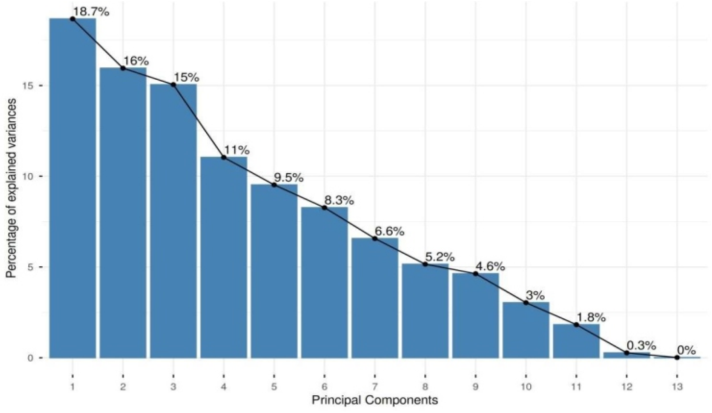

Fig 1. Scree plot illustrating percentage of variance explained by 13 principal components

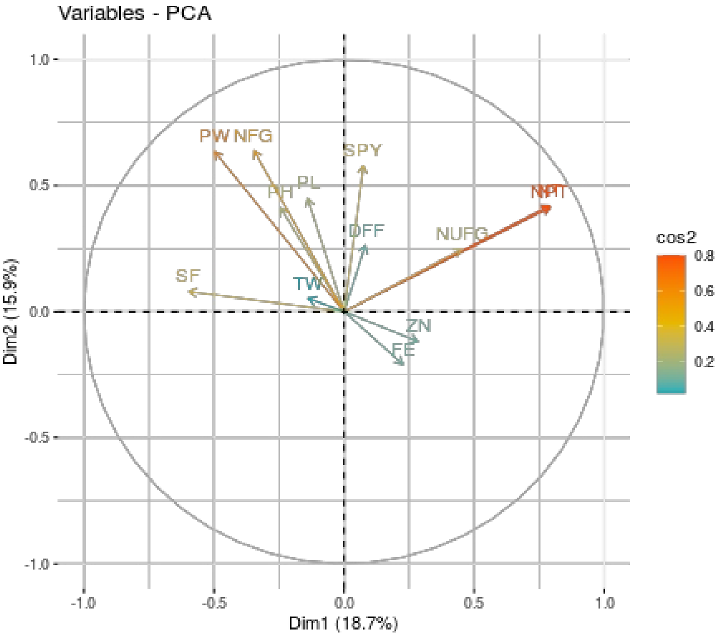

Fig 2.1 Biplot elucidating the association among 13 traits in PC1 and PC2

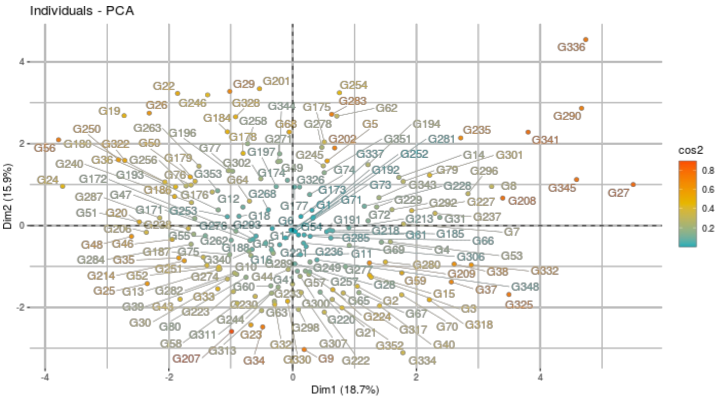

Fig 2.2 Biplot explaining the distribution of 183 rice genotypes in first two principal components

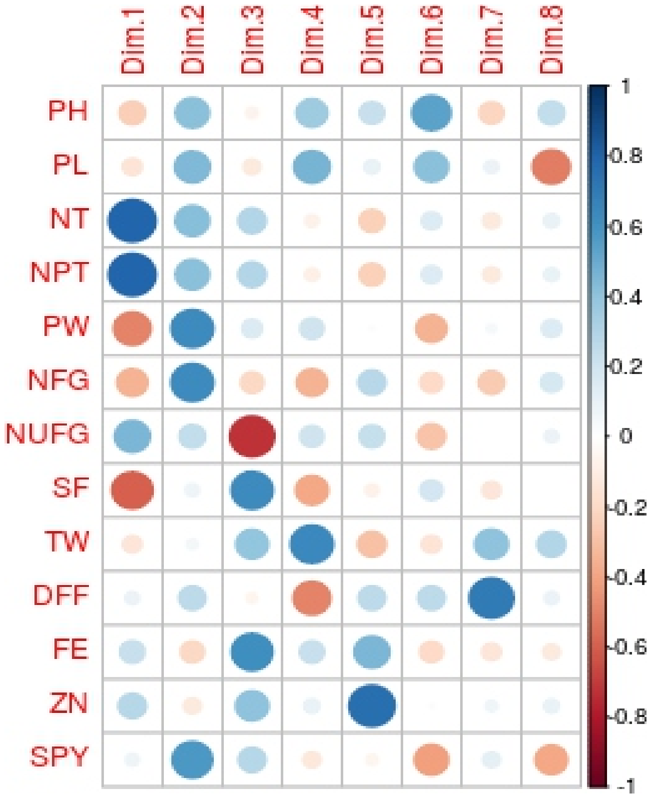

Fig 3. Correlation plot of traits vs PCs

Clustering analysis with Tocher's method and Mahalonobis D^2 distance

The genetic divergence among 183 rice genotypes was assessed through Tocher's method and Mahalonobis D^2 statistics, using the average linkage distance approach of hierarchical clustering. The genotypes were categorized into five distinct clusters (Fig. 4). Among them, Cluster I accommodated the largest proportion of genotypes (118), followed by Cluster III with 30 genotypes, Cluster II with 24 genotypes, Cluster IV with 10 genotypes and Cluster V with a single genotype. Cluster V showed superior mean performance values for most yield-related traits such as height of the plant (106.97 cm), panicle length (25.26 cm), number of tillers (7.76), productive tillers count (6.71), and single plant yield (18.5 g). Cluster IV exhibited superior performance for spikelet fertility (84.92%), days to attain 50% flowering (109.92 days), iron content (11.21 ppm) and zinc content (27.50 ppm). Cluster II registered the maximum mean value for panicle weight (3.1462 g) and filled grains count (148.94), while Cluster I showed the maximum number of unfilled grains (34.91) and Cluster III the highest test weight (21.28 g). Similar clustering trends were reported by Islam *et al*. (2024), who also observed that genotypes grouped in Cluster IV exhibited desirable combinations of yield and nutritional traits. The distinct cluster formation in this study highlights the substantial genetic variability among genotypes, which can be effectively exploited for the selection of promising donor lines in future breeding programs.

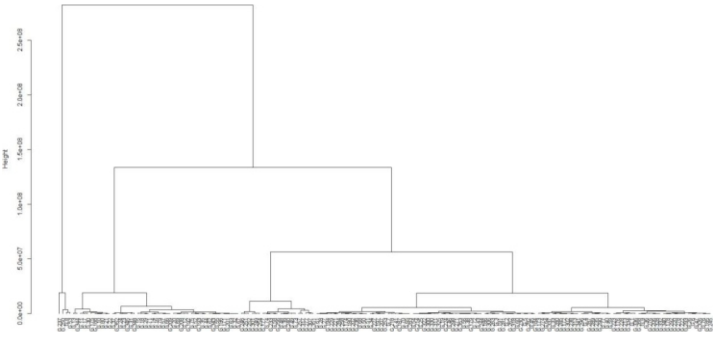

Fig 4. Dendrogram using agglomerative clustering method of 183 germplasm lines

Correlation studies for yield related traits and micronutrient content in rice

Correlation studies between various traits occur because of genetic linkage or pleiotropic effects and are critical in the comprehension of the interrelationship between agronomic and nutritional traits. The analyses are important in offering insights into how improvement in one trait could have an impact on another, thus making the selection method more efficient in breeding programs. During the current research, Pearson's correlation coefficients were estimated for all the traits in 183 diverse rice genotypes. The analysis indicated that per plant yield was positively significantly correlated with panicle weight (0.34**), tiller number (0.255**), productive tiller number (0.239**), and filled grain number (0.212**), suggesting that these characteristics may be targeted for direct selection to improve yield. This is consistent with earlier observations documented by Teja Sri *et al*. (2025) for filled grains number count. In contrast, per plant yield showed positive but statistically non-significant associations with spikelet fertility (0.107), days to 50% flowering (0.099), panicle length (0.061), test weight (0.06), iron content (0.04), zinc content (0.033), and number of unfilled grains (0.007). A non-significant negative correlation was observed with plant height (-0.004), suggesting that higher yield is often associated with shorter stature in these genotypes. These results, particularly the positive non-significant correlations, are consistent with findings from Farheen *et al*. (2023) and Teja Sri *et al*. (2025). The correlation matrix for all 13 traits across the 183 genotypes is shown in Fig. 5

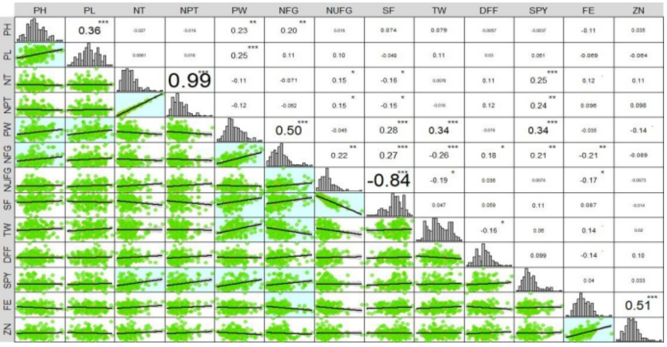

Fig 5. Correlation plot illustrating associations among yield related traits and micronutrients

Conclusion

Out of 8 principal components, the first six PCs (PC, PC2, PC3, PC4, PC5, PC6) exhibited eigen- values>1 and accounted for 78.5% of total variation. The maximum amount of total variation present in 183 rice genotypes is represented by PC1, PC2, PC3, PC4, PC5 and PC6. Traits such as number of tillers (0.507), number of productive tillers (0.509), number of unfilled grains (0.291), zinc content (0.182) and iron content (0.145) revealed the highest positive loadings on PC1, indicating the major contribution of these traits towards variation in PC1 which also displayed the highest vector lengths on biplot. Per plant yield exhibited a strong positive correlation with weight of the panicle, tiller count, productive tiller count and filled grains count implying the direct selection of these traits for yield will be effective. Biplot analysis revealed similar results with correlation studies. Plant height showed a negative correlation. Cluster I entailed the highest number of genotypes. The zinc content trait contributed towards the highest genetic diversity amongst all traits. Cluster V consisted of the maximum number of traits with the highest mean values.

Future scope

The present study provides a valuable basis for understanding genetic diversity in rice for yield and nutritional traits. In the future, the superior genotypes identified through multivariate and association analyses can be evaluated across multiple environments to confirm their stability and adaptability. These diverse accessions can be used to form core collections for more efficient breeding and conservation efforts. The high yielding and nutrient-rich lines can serve as parents in bio fortification and genetic improvement programs aimed at developing nutritionally enhanced rice varieties. Overall, the study opens pathways for genomics-assisted breeding, improved nutrient security and development of climate-resilient, high-quality rice cultivars.

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this article

Acknowledgement

The authors are thankful to the Indian Institute of Rice Research, Rajendranagar, Hyderabad and Indira Gandhi Krishi Vishwavidyalaya, Raipur, Chhattisgarh, for providing the funds and research facilities to carry out the present research work.

References

1. Adams, M. W and Grafius, J. E (1971). Yield component compensation alternative interpretations 1. Crop Science, 11(1), 33–35. Alam, M., Lou, G., Abbas, W., Osti, R., Ahmad, A., Bista, S and He, Y (2024). Improving rice grain quality through ecotype breeding for enhancing food and nutritional security in Asia–Pacific region. Rice, 17(1): 47.

2. Christina, G.R., Thirumurugan, T., Jeyaprakash P and 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.

3. Farheen, M., Murthy, K., Mohan, Y. C and Kumar, J. H (2023). Studies on correlation and path analysis for yield and morphophysiological traits in elite rice (Oryza sativa L.) genotypes under dry DSR system. International Journal of Bioresource and Stress Management, 14(4): 546–553.

4. Islam, M. A., Hasan, M. M., Hossain, M. A., Haque, M. A., Siddique, M. N. A., Shamsuddoha, M and Risha, S. S (2024). Identification and evaluation of high-performing advanced germplasm of rice through morphological and breeding value analysis. Discover Agriculture, 2(1): 126.

5. Jadhav, R. A., Mehtre, S. P., Patil D. K and Gite V. K (2023). Multivariate analysis using D2 and principal component analysis inmung bean [Vigna radiata (L.) Wilczek] for study of genetic diversity. Legume Res., 46(1): 10-17.

6. Jombart, T., Pontier, D and Dufour, A. B (2009). Genetic markers in the playground of multivariate analysis. Heredity, 102(4): 330-341.

7. Meena, D., Kumar, M and Soni, R. K (2023). Assessment of genetic diversity for grain yield in rice (Oryza sativa L.) genotypes under humid South Eastern Plain of Rajasthan, India. International Journal of Plant and Soil Science, 35(18): 971-977.

8. Paramanik, S., Panda, K. K., Rao, M. S., Jaishi, A and Pati, A. K (2025). Unlocking the Genetic Diversity and Principal Component Analysis of Selected Rice (Oryza sativa L.). Genotypes Using Physical and Biochemical Characteristics. Plant Molecular Biology Reporter, 1-15.

9. RStudio Team. RStudio: Integrated Development for R. RStudio, P. B. C, Boston, M. A (2020). http://www.rstudio.com.

10. Sandhya, Mukul, Kumar M., Kumar, P., Gaurav, A., Meena D and Soni R. K (2022). Assessment of Principal Component Analysis and Genetic diversity among linseed (Linum usitatissimum L.) germplasm under Humid South Eastern plain of Rajasthan. Agri. Mechanization in Asia., 53(2): 5809-5818.

11. Singh, K. S., Suneetha, Y., Raja, D. S and Srinivas, T (2020). Principal component analysis for yield and quality traits of coloured rice (Oryza sativa L.). The Pharma Innovation Journal, 9(7): 456-462.

12. Sruthi, S. R., Kumar, L. N., Kishore, D., Subakar J. I and Anbuselvam Y (2024). Principal Component Analysis in rice (Oryza sativa L.) varieties for three seasons in Annamalai Nagar, an east coast region of Tamil Nadu. Plant Sci. Today., 11(4): 634-639.

13. Stevens, G. A., Beal, T., Mbuya, M. N., Luo, H., Neufeld, L. M., Addo, O. Y and Young, M. F (2022). Micronutrient deficiencies among preschool-aged children and women of reproductive age worldwide: a pooled analysis of individual-level data from population-representative surveys. The Lancet Global Health, 10(11): 1590-1599.

14. Yan W and Kang M. S (2002). GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists. CRC press.