Plant Archives Vol. 26, Supplement 1, 2026 pp. 2353-2360 e-ISSN:2581-6063 (online), ISSN:0972-5210

Plant Archives

Journal homepage: http://www.plantarchives.org
DOI Url : https://doi.org/10.51470/PLANTARCHIVES.2026.v26.supplement-1.318

IDENTIFICATION OF GENETICALLY DIVERSE AND HIGH-YIELDING RICE BREEDING LINES THROUGH MULTIVARIATE TRAIT ANALYSIS

Alok Kumar Singh[1,2*], Devendra Pratap Singh[3], Avantika Maurya[2], Ashutosh Singh[4], Saurabh Dixit[5], N.A. Khan[1] and D.K. Dwivedi[1]

[1*]Department of Plant Molecular Biology and Genetic Engineering, Acharya Narendra Deva University of Agriculture & Technology, Ayodhya -224229, U.P, India.
[2]Division of Genomic Resources, ICAR - National Bureau of Plant Genetic Resources, New Delhi, India-110012.
[3]Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi-110012, India.
[4]Department of Molecular Biology & Biotechnology college of Agriculture, Rani Lakshmi Bai Central Agricultural University, Jhasi-284003, U.P, India.
[5]Crop Research Station Masodha, Acharya Narendra Deva University of Agriculture & Technology, Ayodhya -224229, U.P, India.
*Corresponding author E-mail: singhak8483@gmail.com
(Date of Receiving : 28-11-2025; Date of Acceptance : 30-01-2026)

ABSTRACT

Genetic variability among breeding lines is essential for effective rice improvement. The present study evaluated agronomic trait variability among 114 rice genotypes, including five national and global checks, using multivariate statistical analysis. The experiment was conducted during the *kharif* season of 2021–22 under an alpha lattice design with three replications, and thirteen yield and yield-related traits were recorded. Principal component analysis revealed that the first six principal components with eigenvalues greater than one explained 69.56% of the total variation among the genotypes. Traits such as number of spikelets per panicle, grain yield per plant, biological yield, harvest index, panicle length and number of grains per panicle contributed substantially to genetic variation. Correlation analysis showed positive associations of grain yield with biological yield, harvest index and number of grains per panicle. Cluster analysis based on Euclidean distance grouped the breeding lines into distinct clusters, indicating the presence of considerable genetic divergence. The study highlights the usefulness of multivariate approaches in identifying diverse and promising rice breeding lines for yield improvement.
Keywords **:** Rice breeding lines; Agronomic traits; Genetic diversity; Principal component analysis; Cluster analysis

Introduction

Rice (*Oryza sativa* L.) is a primary staple food for an enormous percentage of the global population (Sapna *et al*., 2024). The continuing rise in population has put significant pressure on rice-producing processes. At the same time, cultivable land is shrinking while climatic variability is expanding (Sabar *et al*., 2024). These concerns urge the establishment of high-yielding and stable rice cultivars with superior agronomic performance (Haggag *et al*., 2015).

The availability of genetic variety in breeding material is a key factor in crop development (Swarup *et al*., 2021). The presence of sufficient variability allows breeders to select superior genotypes and design effective hybridization programs (Sabar *et al*., 2024). Rice yield is a complex trait and is influenced by several agronomic characters (Luzikihupi *et al*., 1998). Traits such as plant height, number of tillers, panicle length, number of grains per panicle, and grain weight collectively determine yield potential. Therefore, the evaluation of agronomic traits is essential for identifying promising breeding lines for further improvement (Shrestha *et al*., 2021).

A crucial component of every successful breeding effort is the selection of genetically diverse and

productive parental lines. In segregating generations, genotypes with a high degree of genetic divergence are more likely to provide better recombinants (Bose *et al*., 2005). Therefore, it is crucial to recognize the degree and pattern of diversity among breeding lines. The combined impact of several factors on genetic diversity cannot be explained by traditional univariate statistical approaches, which only offer information on individual traits.

Multivariate statistical techniques offer an effective approach to analyze complex datasets involving several correlated traits simultaneously (Word *et al*., 1987). Principal component analysis (PCA) is a widely used multivariate technique for reducing data dimensionality (Salem *et al*., 2019). It helps identify traits that contribute most to total variation among genotypes (Hasan *et al*., 2021). Cluster analysis complements PCA by grouping genotypes based on their similarity and divergence (Evgenidis *et al*., 2011; Gewers *et al*., 2021). The combined use of these methods enables effective characterization of genetic diversity and supports informed parental selection in breeding programs (Azam *et al*., 2023).

The present study aimed to evaluate agronomic trait variability among 114 rice breeding lines using PCA and cluster analysis. The objectives were to identify key traits contributing to genetic variation and to select genetically diverse and high-performing genotypes for future rice improvement programs.

MATERIALS AND METHODS

Experimental site and plant material

The study was conducted during the kharif season of 2021–22 at the Crop Research Station, Masodha, Acharya Narendra Deva University of Agriculture and Technology, Kumarganj, Ayodhya (U.P.), India, involving 114 rice breeding lines, including five global and national checks (IRR154, IR-64, BPT-5204, IRRI-148, and IRRI-119).

Experimental Design & Data Recording

The experiment was conducted in an Alpha Lattice design (ALD) with three replications (Kashif *et al*., 2010). Each genotype was grown in a single plot under uniform field conditions. Yield and yield-related traits were recorded at crop maturity. These included number of spikelets per panicle (NSP), number of grains per panicle (NGP), spikelet fertility (%) (SF), biological yield per plant (g) (BYP), harvest index (%) (HI), 1000-grain weight (g) (TW), and grain yield per plant (g) (GYP). Mean values were calculated for each genotype before statistical analysis.

Statistical Analysis

The pooled mean data were subjected to multivariate statistical analysis. All traits were standardized to minimize scale effects.

Principal component analysis (PCA) was performed to identify major traits contributing to total variation among the rice breeding lines. Principal components (PCs) with eigenvalues greater than one were considered significant.

Cluster analysis was carried out using Euclidean distance to assess genetic divergence among genotypes. Hierarchical clustering was performed using the Neighbour-Joining (NJ) method. A dendrogram was constructed to depict the grouping pattern of the rice breeding lines using PAST v5.3 software (Hammer *et al*., 2001).

All statistical analyses were performed using the R statistical software.

Result and Discussion

Phenotypic variability in 13 yield-related traits across 114 breeding lines

The boxplot depicts the distribution of 13 agro-morphological and yield-related traits across 114 rice breeding lines, revealing significant phenotypic variability for all the traits studied (Figure 1). Wide interquartile ranges and the presence of extreme values for key yield-attributing traits such as number of NSP, NGP, BYP, HI, TW and GYP indicated the existence of broad genetic diversity within the 114 lines. Traits directly contributing to yield, particularly GYP and BYP, showed distinct dispersion, suggesting differential assimilate production and partitioning efficiency among genotypes. Moderate variability was also observed for phenological traits such as Days to flowering (DTF), spikelet number (SV), Plant height (PH), Panicle length (PL) and flag leaf area (FLA), reflecting differences in growth duration. Such wide variability is desirable in breeding populations as it enhances the scope for effective selection of superior genotypes and trait recombination for crop yield improvement.

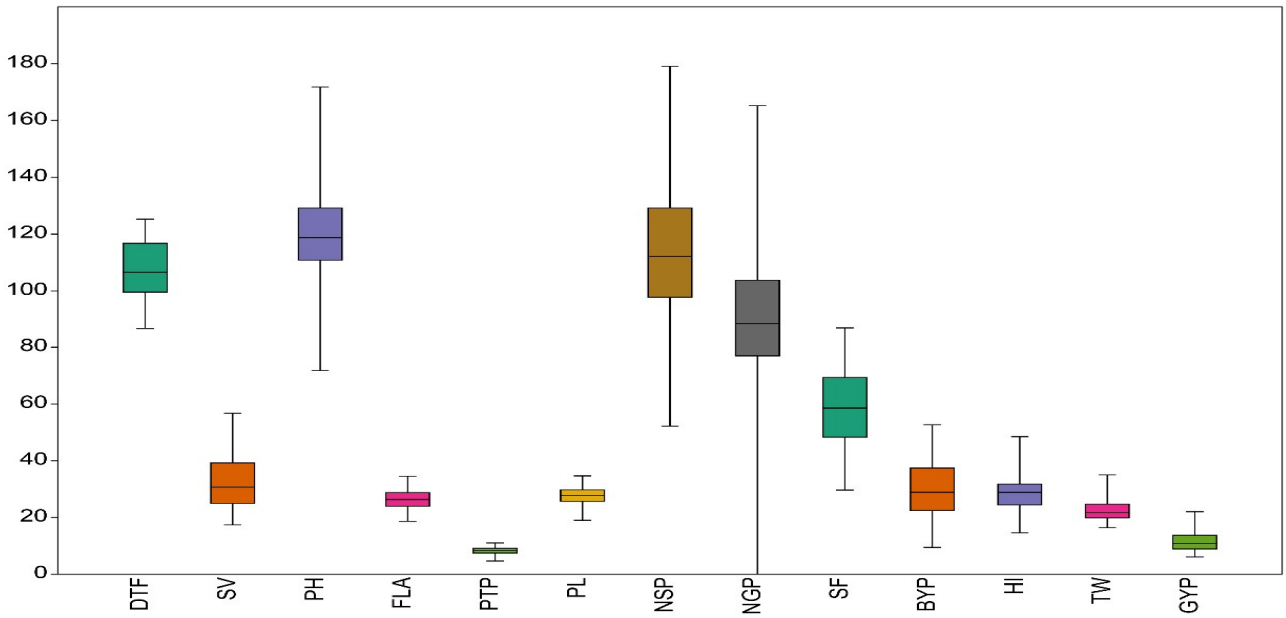

Fig. 1: Boxplot depicting the distribution of 13 yield related traits across 114 breeding lines.

Correlation among yield and yield-related traits

Pearson correlation analysis among yield and yield-related attributes revealed several significant associations highlighting the interrelationships governing grain yield in rice (Figure 2). GYP showed positive correlations with BYP, HI, NGP and PL indicating that genotypes with higher biomass production and efficient assimilate partitioning tended to exhibit superior yield performance. Positive associations between NSP and NGP further emphasize the importance of sink size in determining final yield. In contrast, SF showed a negative association with certain sink-related traits suggesting possible relation between spikelet number and fertility under the given environmental conditions. These relationships collectively indicate that selection for grain yield should be based on a combination of traits such as biomass accumulation, panicle architecture and grain filling efficiency rather than on a single trait.

Fig. 2: Pearson Correlation coefficient among the yield and yield related attributes.

Principal component analysis

PCA reduces the dimensionality of the dataset and identifies major sources of variation among the breeding lines (Edukondalu *et al.*, 2024). The first six PCs had eigenvalues greater than one and together explained 69.56% of the total variation, indicating that

these components captured most of the information present in the original variables. PC1 accounted for 15.74% of the total variance, followed by PC2 (14.15%), PC3 (11.83%), PC4 (11.21%), PC5 (8.62%) and PC6 (8%) (Table 1). The scree plot showed a gradual decline in eigenvalues after PC6, confirming the adequacy of the first few PCs in representing overall variability (Figure 3). The significant variance explained by the initial components reflects the polygenic nature of yield and its associated traits and justifies the use of PCA for effective discrimination among genotypes.

Table 1: Eigenvalue, variance explained and cumulative variance for 13 PCs.

PC	Eigenvalue	Variance_Explained	Cumulative_Variance
PC1	2.045871	0.15737	0.15737
PC2	1.838833	0.14145	0.29882
PC3	1.53786	0.1183	0.41712
PC4	1.457815	0.11214	0.52926
PC5	1.121454	0.08627	0.61553
PC6	1.040695	0.08005	0.69558
PC7	0.987848	0.07599	0.77157
PC8	0.931919	0.07169	0.84325
PC9	0.82471	0.06344	0.90669
PC10	0.652121	0.05016	0.95686
PC11	0.395497	0.03042	0.98728
PC12	0.146686	0.01128	0.99856
PC13	0.01869	0.00144	1

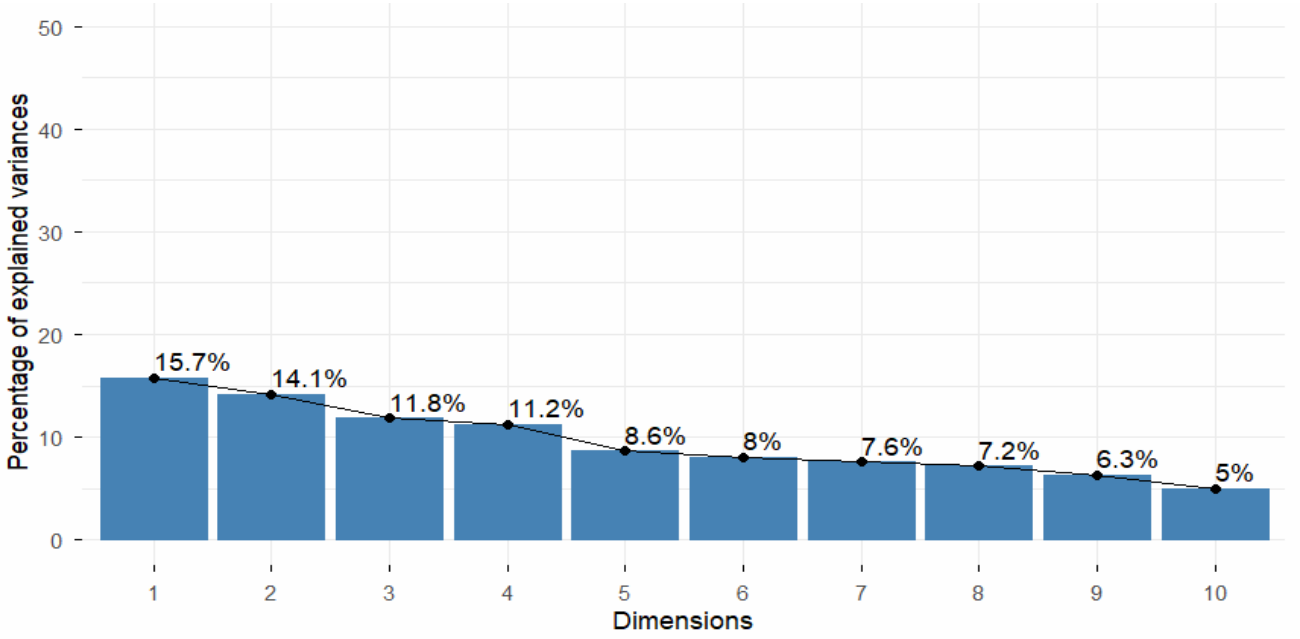

Fig. 3: Scree plot showing the percentage of variance.

The loading matrix provided insights into the contribution of individual traits to each principal component. PC1 was predominantly influenced by NSP (positive loading) and SF (negative loading) indicating that variation in spikelet production and fertility played a major role in differentiating genotypes along this axis (Table 2). PC2 was strongly associated with PH, PL, and GYP suggesting that plant architecture and yield performance together contributed to variability captured by this component. PC3 showed high positive loadings for HI, TW and GYP emphasizing the importance of assimilate partitioning and grain weight in yield determination.

PC4 was linked to BYP and GYP reflecting differences in total biomass production.

The loading plot illustrating the contributions of variables to the first two PCs further confirmed that yield and yield-related traits such as GYP, BYP, NSP, NGP and PL were the major contributors of genetic divergence (Figure 4). Traits positioned far from the origin (HI, GYP, PL, PH, BYP, NSP, SV and SF) contributed more strongly to variability, whereas closely clustered (FLA, TW, Productive tillers per plant (PTP), DTF and NGP) traits exhibited correlated behaviour. This distribution highlights the relevance of these key traits in distinguishing among rice breeding lines.

Table 2: Loading scores for 13 yield and yield-related traits.

Traits	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
DTF	0.04	0.08	-0.25	-0.12	0.68	-0.09	-0.15	-0.10	-0.51	0.40	-0.01	0.03	0.03
SV	0.31	-0.21	-0.07	0.02	-0.04	0.34	0.48	0.37	0.14	0.57	0.14	-0.01	0.01
PH	-0.22	0.41	-0.30	-0.34	-0.04	0.16	0.17	0.11	0.16	0.04	-0.69	-0.03	-0.03
FLA	-0.12	-0.10	0.14	0.15	0.35	-0.51	0.36	0.57	0.00	-0.28	-0.13	0.09	-0.02
PTP	-0.03	0.16	-0.13	0.02	-0.58	-0.23	-0.27	0.46	-0.50	0.19	0.03	0.03	0.02
PL	-0.22	0.48	-0.22	-0.22	0.11	0.20	0.14	0.20	0.04	-0.26	0.67	-0.05	0.01
NSP	0.57	0.22	0.21	-0.20	0.03	-0.11	0.01	0.06	-0.09	-0.15	-0.06	-0.70	-0.04
NGP	0.06	-0.02	0.23	-0.04	0.26	0.33	-0.65	0.49	0.29	-0.01	-0.07	0.09	0.00
SF	-0.55	-0.29	-0.18	0.21	0.04	0.02	-0.10	0.07	0.08	0.15	0.03	-0.70	-0.04
BYP	0.20	0.31	-0.26	0.65	0.06	0.05	-0.05	-0.02	0.07	-0.02	-0.05	0.02	-0.60
HI	-0.30	0.19	0.61	-0.23	-0.02	-0.11	0.06	-0.07	0.00	0.36	0.08	0.02	-0.55
TW	-0.14	-0.04	0.31	0.20	0.03	0.60	0.20	0.05	-0.58	-0.28	-0.17	-0.03	0.00
GYP	-0.10	0.50	0.31	0.44	0.04	-0.06	0.04	-0.07	0.13	0.28	-0.03	-0.05	0.58

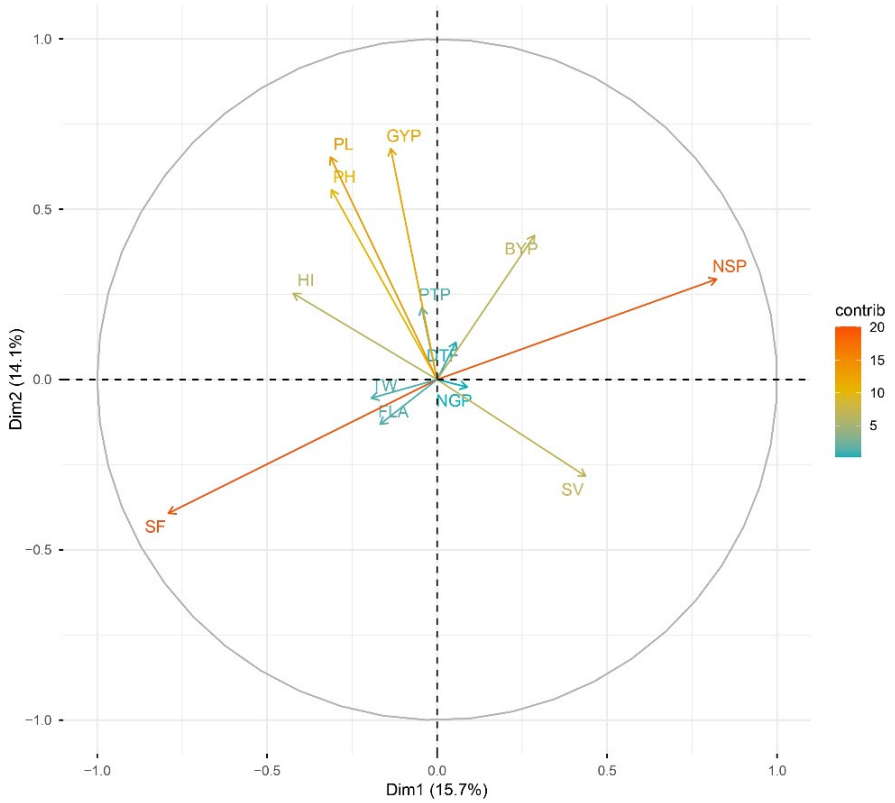

Fig. 4 : Loading plot illustrating the variables contributions to the first two PCs.

The biplot depicting the distribution of 114 rice breeding lines along the first two principal components revealed clear dispersion of genotypes across the multivariate space (Figure 5). This wide distribution reflects substantial genetic divergence among the breeding lines. Several genotypes were positioned in directions associated with favourable yield-attributing traits, indicating their potential as superior performers. Genotypes located near the vectors of GYP, BYP, HI and NGP can be considered promising candidates for yield improvement, while those occupying extreme positions in opposite quadrants represent genetically diverse lines suitable for use as parents in hybridization programs. The clear separation among genotypes in the biplot validates the effectiveness of PCA in capturing underlying variability and proves to be effective in simplifying complex multivariate datasets and revealing key relationships and variations among traits. These insights are particularly significant for rice breeding as they enhance the understanding of trait interactions and facilitate the efficient identification and selection of superior genotypes.

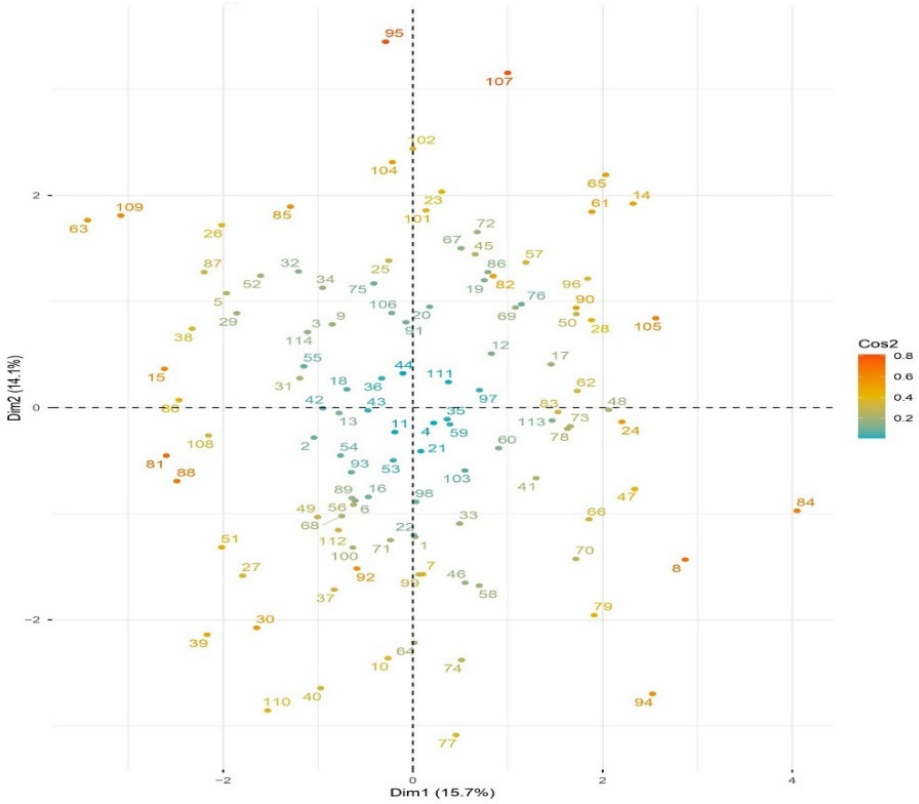

Fig. 5: Biplot depicting the distribution of 114 breeding line in the first two PCs.

NJ method-based cluster analysis of 114 breeding lines

Hierarchical cluster analysis based on Euclidean distance using phenotypic trait data grouped the 114 rice breeding lines into distinct and divergent clusters (Figure 6). The dendrogram revealed wide genetic diversity and differential combinations of yield and yield-related traits. The checks were distributed across different clusters indicating that they represent diverse genetic backgrounds. BPT 5204 and IRRI 154 grouped with moderate-yielding, tall to medium-tall genotypes with relatively lower grain yield, whereas IRRI 148, IRRI 119 and IR64 clustered with high-biomass and high-yielding genotypes. High-yielding genotypes such as IR17A3003, IR18A1027, IR18A1197, IR18A1156, IR17A3050 and IR18A1135 were positioned in distinct sub-clusters with BPT 5204 and IRRI 154. The long branch lengths separating these superior genotypes from low-yielding checks indicated significant genetic divergence, suggesting that they possess unique and complementary allelic combinations for yield traits.

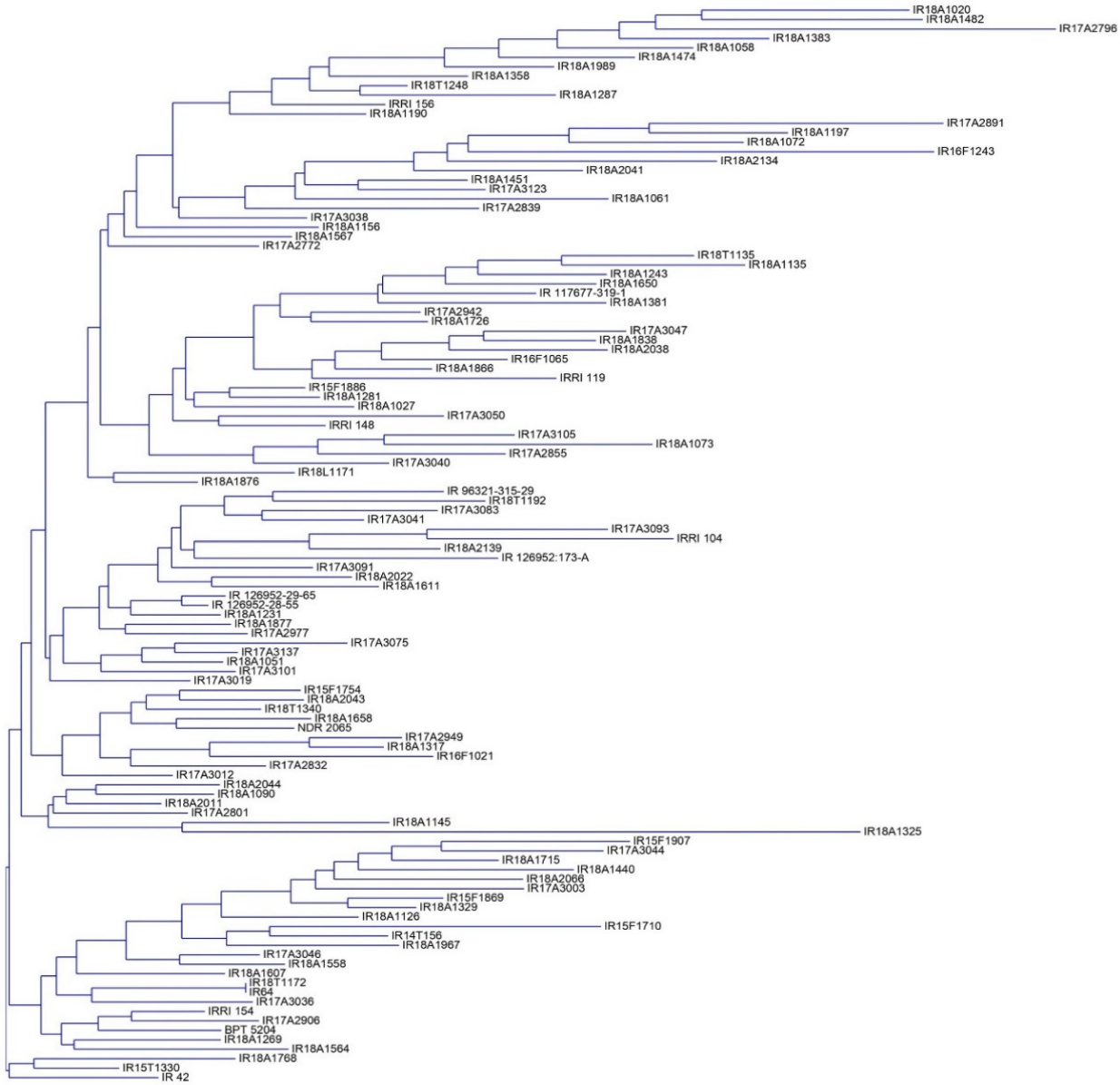

Fig. 6: Cluster analysis of 114 rice breeding lines based on Euclidean distance and NJ method.

Conclusion

The present study revealed substantial phenotypic variability among the 114 rice breeding lines including 5 checks for yield and yield-related traits, indicating the availability of useful genetic diversity for selection. Principal component analysis showed that a limited number of components explained a major proportion of the total variation, with traits such as number of spikelets per panicle, grain yield per plant, biological yield, harvest index, panicle length and number of grains per panicle contributing most to genetic divergence. Correlation analysis confirmed the importance of biomass production and assimilate partitioning in determining grain yield. Cluster analysis grouped the breeding lines into distinct and divergent clusters, with high-yielding genotypes distributed separately from popular checks, reflecting their diverse genetic backgrounds. The identified genetically diverse and high-performing breeding lines can be effectively utilized as potential parents in future rice improvement programs aimed at yield enhancement.

Acknowledgement- The authors sincerely acknowledge the Department of Plant Molecular Biology and Genetic Engineering, Acharya Narendra Deva University of Agriculture and Technology, Kumar Ganj, Ayodhya (U.P.), India, for providing the necessary field, laboratory, and experimental facilities to carry out the present investigation.

Conflict of Interest- The authors declare that there is no conflict of interest.

Generative AI Disclosure statement- The authors used generative AI tools solely for language editing and to improve the clarity and organization of the manuscript. All scientific content, data analysis, and interpretations were performed by the authors, who take full responsibility for the accuracy and integrity of the work

References

Azam, M. G., Hossain, M. A., Sarker, U., Alam, A. M., Nair, R. M., Roychowdhury, R., & Golokhvast, K. S. (2023). Genetic analyses of mungbean [*Vigna radiata* (L.) Wilczek] breeding traits for selecting superior genotype(s) using multivariate and multi-traits indexing approaches. *Plants*, **12**(10), 1984.

Bose, K., & Pradhan, K. (2005). Genetic divergence in deepwater rice genotypes. *Journal of Central European Agriculture*.

Edukondalu, B., Reddy, V. R., Rani, T. S., Kumari, A., & Soundharya, B. (2024). Assessment of variation in rice maintainer lines using principal component analysis. *Electronic Journal of Plant Breeding*, **15**(1), 270–276.

Evgenidis, G., Traka-Mavrona, E., & Koutsika-Sotiriou, M. (2011). Principal component and cluster analysis as a tool in the assessment of tomato hybrids and cultivars. *International Journal of Agronomy*, **2011**(1), 697879.

Gewers, F. L., Ferreira, G. R., Arruda, H. F. D., Silva, F. N., Comin, C. H., Amancio, D. R., & Costa, L. D. F. (2021). Principal component analysis: A natural approach to data exploration. *ACM Computing Surveys*, **54**(4), 1–34.

Haggag, W. M., Abouziena, H. F., Abd-El-Kreem, F., & El Habbasha, S. (2015). Agricultural biotechnology for management of multiple biotic and abiotic environmental stress in crops. *Journal of Chemical and Pharmaceutical Research*, **7**(10), 882–889.

Hammer, Ø., & Harper, D. A. T. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, **4**(1), 1.

Hasan, B. M. S., & Abdulazeez, A. M. (2021). A review of principal component analysis algorithm for dimensionality reduction. *Journal of Soft Computing and Data Mining*, **2**(1), 20–30.

Kashif, M. (2010). Efficiency of alpha lattice design in rice field trials in Pakistan. *Journal of Scientific Research*.

Luzikihupi, A. (1998). Interrelationship between yield and some selected agronomic characters in rice. *African Crop Science Journal*, **6**(3), 323–328.

Sabar, M., Mustafa, S. E., Ijaz, M., Khan, R. A. R., Shahzadi, F., Saher, H., & Sabir, A. M. (2024). Rice breeding for yield improvement through traditional and modern genetic tools. *European Journal of Ecology, Biology and Agriculture*, **1**(1), 14–19.

Sapna, I., & Jayadeep, A. (2024). Enzyme-treated red rice (*Oryza sativa* L.) bran extracts mitigate inflammatory markers in RAW 264.7 macrophage cells and exhibit anti-inflammatory efficacy comparable to bioactive compounds. *Journal of Ethnopharmacology*, **323**, 117616.

Salem, N., & Hussein, S. (2019). Data dimensional reduction and principal components analysis. *Procedia Computer Science*, **163**, 292–299.

Shrestha, J., Subedi, S., Kushwaha, U. K. S., & Maharjan, B. (2021). Evaluation of growth and yield traits in rice genotypes using multivariate analysis. *Heliyon*, **7**(9).

Swarup, S., Cargill, E. J., Crosby, K., Flagel, L., Kniskern, J., & Glenn, K. C. (2021). Genetic diversity is indispensable for plant breeding to improve crops. *Crop Science*, **61**(2), 839–852.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, **2**(1–3), 37–52.