

International Journal of Advanced Biochemistry Research



ISSN Print: 2617-4693
ISSN Online: 2617-4707
NAAS Rating (2026): 5.29
IJABR 2026; SP-10(9): 1196-1207
www.biochemjournal.com
Received: 18-06-2026
Accepted: 22-07-2026

K Madhu Kumar
Agricultural Research Station,
Ragolu, Acharya N G. Ranga
Agricultural University
(ANGRAU), Andhra Pradesh,
India

G Sreenivas
Agricultural Research Station,
Ragolu, Acharya N G. Ranga
Agricultural University
(ANGRAU), Andhra Pradesh,
India

Udaya Babu
Agricultural Research Station,
Ragolu, Acharya N G. Ranga
Agricultural University
(ANGRAU), Andhra Pradesh,
India

M Srinivasa Rao
Agricultural Research Station,
Ragolu, Acharya N G. Ranga
Agricultural University
(ANGRAU), Andhra Pradesh,
India

D Manojkumar
Agricultural Research Station,
Ragolu, Acharya N G. Ranga
Agricultural University
(ANGRAU), Andhra Pradesh,
India

Corresponding Author:
D Manojkumar
Agricultural Research Station,
Ragolu, Acharya N G. Ranga
Agricultural University
(ANGRAU), Andhra Pradesh,
India

Genetic Diversity Analysis of Rice (*Oryza sativa* L.) Genotypes Using Multivariate Statistical Approaches for the Identification of Divergent Parents

K Madhu Kumar, G Sreenivas, Udaya Babu, M Srinivasa Rao and D Manojkumar

DOI: <https://www.doi.org/10.33545/26174693.2026.v10.i9So.9981>

Abstract

Rice (*Oryza sativa* L.) is the principal staple food for more than half of the global population and plays a pivotal role in ensuring food and nutritional security, particularly in Asia. Broadening the genetic base through the identification of genetically diverse parents is one of the most effective approaches for developing superior rice cultivars with higher yield potential and wider adaptability. The present investigation was undertaken to assess the extent of genetic diversity among advanced rice genotypes using Mahalanobis D² statistics & Cluster analysis, factor analysis and principal component analysis. Significant multivariate differences among genotypes were confirmed through Wilks' Lambda, indicating the presence of substantial genetic variability. Analysis of variance for dispersion further confirmed significant differences among genotypes. Character contribution studies identified key traits contributing to total divergence, providing useful information for effective selection in breeding programmes. Factor analysis revealed that a few latent factors explained a major proportion of the total variability, suggesting that yield-contributing characters were primarily responsible for genetic divergence. Based on Mahalanobis D² values, the genotypes were grouped into distinct clusters using Tocher's method, demonstrating considerable genetic diversity. High inter-cluster distances indicated the possibility of obtaining greater heterosis and wider segregation through hybridization between genetically divergent parents. Cluster mean analysis facilitated the identification of superior donor parents possessing desirable combinations of agronomic traits. Principal component analysis further supported the clustering pattern and revealed the distribution of genotypes in multidimensional space. The combined application of Mahalanobis D² statistics, factor analysis and principal component analysis proved effective in identifying genetically diverse parents for future rice improvement programmes. The identified divergent genotypes may serve as valuable parental resources for developing high-yielding, climate-resilient and widely adapted rice cultivars.

Keywords: Cluster Analysis, Genetic diversity, Parent selection, Principal Component Analysis and Rice

1. Introduction

Rice (*Oryza sativa* L.) is one of the world's most important cereal crops and serves as the staple food for more than half of the global population. It contributes substantially to food security, nutritional security, employment generation and rural livelihoods, particularly across Asia where nearly 90% of global rice production and consumption occur (Tagliapietra *et al.*, 2024) [36]. India occupies the second position in global rice production after China and cultivates rice over approximately 46 million hectares with an annual production exceeding 130 million tonnes (FAO, 2024) [10]. Despite remarkable achievements in varietal development and crop management, future rice production faces serious challenges due to rapid population growth, climate change, diminishing natural resources, emerging pests and diseases and a narrowing genetic base of modern cultivars (Satyanarayana *et al.*, 2024) [28]. Sustainable enhancement of rice productivity largely depends on the availability and effective utilization of genetic variability. Genetic diversity constitutes the foundation of every crop improvement programme because it provides opportunities for breeders to identify superior parents possessing complementary alleles for economically important traits (Madhukumar *et al.*, 2025) [16]. Hybridization between genetically diverse parents generally produces wider segregation, greater recombination frequencies and increased chances of

recovering transgressive segregants in subsequent generations (Falconer & Mackay, 1996) ^[9]. Therefore, precise assessment of genetic divergence among available breeding materials is an essential prerequisite for designing efficient breeding strategies.

Yield is a complex quantitative trait governed by numerous component characters and is highly influenced by environmental conditions. Consequently, direct selection for grain yield alone often proves less effective than simultaneous selection based on multiple yield-contributing traits (Manojkumar *et al.*, 2023) ^[19]. Conventional univariate statistical methods evaluate each character independently and fail to account for the interrelationships among traits. In contrast, multivariate statistical techniques simultaneously analyze several correlated variables and provide a more realistic estimate of genetic diversity (Mohammadi & Prasanna, 2003).

Among multivariate techniques, Mahalanobis D² statistics (Mahalanobis, 1936) ^[18] has become one of the most widely accepted approaches for assessing genetic divergence in crop plants. This method combines information from several quantitative characters into a single generalized distance measure, thereby enabling objective classification of genotypes based on their overall genetic divergence. The technique has been extensively utilized in rice breeding programmes to identify genetically diverse parents suitable for hybridization (Kumar *et al.*, 2018). Genotypes belonging to clusters separated by greater inter-cluster distances are generally considered more suitable for crossing because they are expected to produce superior heterotic combinations and broad genetic variability in segregating generations (Rao, 1952) ^[26].

Tocher's method of clustering further improves the usefulness of Mahalanobis D² analysis by grouping genotypes into relatively homogeneous clusters based on genetic distances (Behera *et al.*, 2018) ^[4]. Estimation of intra- and inter-cluster distances provides valuable information regarding genetic relationships among breeding materials, while cluster mean analysis facilitates identification of donor parents possessing desirable agronomic traits. Such information greatly assists breeders in selecting appropriate parental combinations for future hybridization programmes (Singh *et al.*, 2018).

Principal Component Analysis (PCA) has emerged as another powerful multivariate technique for evaluating genetic diversity. PCA transforms correlated variables into a smaller number of independent principal components that explain maximum variability present in the experimental population (Duppala *et al.*, 2018) ^[8]. This approach not only reduces data complexity but also identifies the traits contributing most to total variation. The graphical representation of genotypes through principal component scores facilitates visualization of genetic relationships and population structure. Likewise, factor analysis summarizes highly correlated variables into a limited number of latent factors, thereby simplifying the interpretation of complex datasets and improving understanding of the underlying genetic architecture (Johnson & Wichern, 2007).

Numerous investigations have demonstrated the usefulness of Mahalanobis D² statistics, principal component analysis and factor analysis in assessing genetic diversity among rice germplasm. Earlier studies on variability, association analysis, path coefficient analysis and principal component analysis have successfully identified important yield-

contributing traits and genetically divergent parents suitable for rice improvement programmes (Satyanarayana *et al.*, 2024 and Kumar *et al.*, 2025) ^[28].

The Agricultural Research Stations of ANGRAU, has developed several advanced rice breeding lines possessing wide variation for yield and associated agronomic characters. However, comprehensive information regarding the genetic divergence among these breeding materials is limited. Therefore, systematic evaluation of genetic diversity using Mahalanobis D² statistics in combination with factor analysis and principal component analysis is essential for identifying genetically divergent parents capable of producing superior recombinants and high-yielding cultivars.

Therefore, the present investigation was undertaken with the following objectives:

1. To estimate the extent of genetic divergence among advanced rice genotypes using Mahalanobis D² statistics.
2. To classify the genotypes into distinct clusters using Tocher's method.
3. To identify the major traits contributing to total genetic divergence through factor analysis and character contribution analysis.
4. To assess the genetic relationships among genotypes using principal component analysis.
5. To identify genetically diverse and agronomically superior parents for future rice breeding programmes.

2. Materials and Methods

2.1 Experimental Site

The present investigation was carried out during the Kharif season at the Acharya N.G. Ranga Agricultural University (ANGRAU), Andhra Pradesh, Agricultural Research Station (ARS), Ragolu, located in the North Coastal Zone of Andhra Pradesh (18.240° N latitude, 83.240° E longitude) at an elevation of 27 meters above mean sea level. The region experiences an average annual rainfall of around 1111 mm over approximately 57 rainy days (Madhukumar *et al.*, 2023) ^[17]. The experiment was carried out during the Kharif 2024 season under a typical monsoon rice-growing system. The experimental farm is situated in the North Coastal Agro-climatic Zone of Andhra Pradesh and represents a typical rice-growing ecosystem. The station experiences a tropical humid climate characterized by warm temperatures, moderate to high relative humidity and seasonal monsoon rainfall, providing favourable conditions for rice cultivation. The soil of the experimental field is representative of the rice-growing tracts of the region and standard agronomic and plant protection practices recommended by ANGRAU were uniformly followed throughout the crop growth period to ensure healthy crop establishment and minimize environmental variation.

2.2 Experimental Material

The experimental material consisted of 55 advanced rice genotypes developed under the rice improvement programme of Agricultural Research Stations of ANGRAU. These breeding lines were selected based on their agronomic superiority and variability for important yield and yield-contributing traits. The genotypes represented diverse genetic backgrounds and were evaluated for multivariate genetic diversity studies.

2.3 Experimental Design and Crop Management

The experiment was conducted using a Randomized Complete Block Design (RCBD) with two replications. Each genotype was raised in uniform plots following the recommended spacing for transplanted rice. Healthy seedlings were transplanted into puddled fields after attaining suitable seedling age. Standard crop management practices including fertilizer application, irrigation scheduling, weed control and plant protection measures were adopted uniformly for all entries to minimize non-genetic sources of variation. Recommended packages of practices prescribed by Acharya N.G. Ranga Agricultural University were followed throughout the crop growth period.

2.4 Recording of Observations

Observations were recorded on five randomly selected competitive plants from each genotype in every replication for important quantitative characters associated with grain yield. Standard evaluation procedures described by the International Rice Research Institute (SES, IRRI, 2014) [11] were followed while recording observations. Mean values over replications were used for statistical analysis to minimize experimental error.

2.5 Statistical Analysis

The recorded data were subjected to multivariate statistical analysis to quantify the extent of genetic diversity among the rice genotypes.

2.5.1 Wilks' Lambda Test

The significance of multivariate differences among genotypes was evaluated using Wilks' Lambda (Λ) statistic. This multivariate test determines whether the observed differences among genotypes are statistically significant when all quantitative characters are considered simultaneously. A significant Wilks' Lambda indicates the presence of substantial genetic variability within the experimental material.

2.5.2 Analysis of Variance for Dispersion

Multivariate analysis of variance was performed to partition the total variability into genotype and residual components. The significance of genotype effects was tested using appropriate multivariate F-tests. Significant differences among genotypes confirmed the suitability of the experimental material for subsequent diversity analysis.

2.5.3 Character Contribution

The relative contribution of individual characters towards total genetic divergence was estimated by calculating the frequency with which each trait appeared first in pairwise D^2 comparisons. Traits contributing higher percentages were considered major determinants of genetic diversity and therefore useful selection criteria for breeding programmes.

2.5.4 Mahalanobis D^2 Statistics

Genetic divergence among genotypes was estimated using Mahalanobis D^2 statistics (Mahalanobis, 1936) [18]. This technique simultaneously considers several correlated quantitative traits and computes generalized squared distances between all possible pairs of genotypes. The D^2 values provide an objective measure of genetic divergence and have been extensively employed in plant breeding

programmes for identifying genetically diverse parents suitable for hybridization.

2.5.5 Tocher's Cluster Analysis

Based on the computed D^2 values, the genotypes were grouped into distinct clusters using Tocher's optimization method as described by Rao (1952) [26]. The clustering procedure minimizes intra-cluster divergence while maximizing inter-cluster divergence, thereby facilitating identification of genetically distinct parental groups. Intra- and inter-cluster distances were subsequently estimated to evaluate the magnitude of genetic divergence among the identified clusters.

2.5.6 Factor Analysis

Factor analysis was performed to identify the major latent factors responsible for total variability among the rice genotypes. Eigenvalues and factor loadings were estimated using the principal factor extraction method. Factors with eigenvalues greater than one were retained according to Kaiser's criterion (Kaiser, 1960) [13]. The contribution of each trait to individual factors was interpreted based on factor loadings, enabling identification of the major yield-related attributes contributing to genetic divergence.

2.5.7 Principal Component Analysis

Principal Component Analysis (PCA) was carried out to summarize the multidimensional variability present among the rice genotypes. Eigenvalues, percentage variance explained and principal component scores were computed following the procedure described by Johnson and Wichern (2007). Two-dimensional and three-dimensional scatter plots were generated using the first few principal components to visualize genetic relationships among the genotypes and to validate the clustering pattern obtained through Mahalanobis D^2 analysis.

2.5.8 Statistical Software

All multivariate analyses, including Mahalanobis D^2 statistics, Tocher's clustering, factor analysis, principal component analysis, Wilks' Lambda and character contribution analysis, were performed using WINDOSTAT 9.3.0 statistical software following standard biometrical procedures for multivariate genetic analysis.

3. Results and Discussion

3.1 Multivariate Analysis of Variance (Wilks' Lambda Test)

The quantitative data of 55 rice genotypes were subjected to multivariate analysis of variance (MANOVA) using Wilks' Lambda to test the significance of multivariate differences among the genotypes for the eight characters (Table 1), viz., days to 50 per cent flowering (DFF), plant height (PH), effective bearing tillers (EBT), panicle length (PL), number of filled grains per panicle (NFGP), test weight (TW) and grain yield per plant (GYP). The Wilks' Lambda test revealed highly significant differences among the genotypes, indicating substantial multivariate variation. Similar findings were reported by Devi *et al.* (2018) and Rathan *et al.* (2020) [27], who also observed significant differences among genotypes based on the Wilks' Lambda test.

The determinant of the error matrix was 26261.80E+2, whereas the determinant of the error plus variety matrix was

15266.04E+14. The Wilks' criterion was 17202.75E-16, with a V statistic of 2045.18. The calculated M value was 75.50 with 432 degrees of freedom and the associated probability was 0.000**, indicating a highly significant multivariate difference among the genotypes. The highly significant MANOVA result confirmed the existence of substantial multivariate variability among the 55 genotypes for the eight quantitative characters. Thus, the genotypes differed significantly when DFF, PH, EBT, PL, NFGP, NIFGP, TW and GYP were considered simultaneously, providing a statistical basis for further assessment of genetic divergence through Mahalanobis D^2 analysis, Tocher's clustering, factor analysis and principal component analysis.

3.2 ANOVA for Dispersion

The analysis of variance for dispersion further confirmed the significant differences among the genotypes (Table 2). The varieties component had 54 degrees of freedom, with a sum of squares of 1.52E18 and mean square of 2.82E16. The calculated F value was extremely high (5.70E11) and was significant at $P = 0.000^*$. The error component had 53 degrees of freedom, with a sum of squares of 2.62E06 and mean square of 4.95E04. The total variation comprised 107 degrees of freedom, with a total sum of squares of 1.52E18 and mean square of 1.42E16. The very high and significant F value demonstrated that the variation among varieties was considerably greater than the experimental error, confirming the presence of substantial variation among the genotypes for the characters included in the multivariate analysis. The analysis of variance for dispersion revealed significant differences among the genotypes, indicating considerable variability, which was in agreement with the findings of Asante *et al.* (2019) [3] and Arteaga *et al.* (2020) [1]. This significant varietal variation justified the estimation of genetic divergence using Mahalanobis D^2 statistics.

3.3 Contribution of Individual Characters towards Genetic Divergence

The relative contribution of the eight quantitative characters to genetic divergence was assessed from the number of times each character ranked first in pairwise comparisons. A marked difference was observed among the characters in their contribution to total divergence (Table 3). The contribution of individual characters towards genetic divergence revealed that different traits contributed variably to the overall genetic divergence among the genotypes, which was in accordance with the findings of Prakash *et al.* (2024) [24] and Mounika *et al.* (2020) [22]. Panicle length (PL) was the most important contributor to genetic divergence, ranking first 1186 times and accounting for 79.87 per cent of the total contribution. The second most important character was number of infertile grains per panicle (NIFGP), which ranked first 156 times and contributed 10.51 per cent. Days to 50 per cent flowering (DFF) ranked first 100 times, contributing 6.73 per cent.

The contribution of the remaining characters was comparatively low. Number of filled grains per panicle (NFGP) ranked first 26 times, contributing 1.75 per cent, while test weight (TW) ranked first 13 times, contributing 0.88 per cent. Effective bearing tillers (EBT) ranked first only once, contributing 0.07 per cent. Grain yield per plant (GYP) ranked first three times, contributing 0.20 per cent, whereas plant height (PH) did not rank first in any comparison and consequently showed 0.00 per cent

contribution. Thus, the divergence among the genotypes was predominantly associated with PL (79.87%), followed by NIFGP (10.51%) and DFF (6.73%). Together, these three characters accounted for 97.11 per cent of the total contribution, indicating that they were the major discriminatory characters responsible for differentiation among the genotypes. In contrast, PH, EBT, GYP and TW made relatively small contributions to the observed divergence. The results indicated that panicle length was by far the most influential character in differentiating the genotypes, while the contribution of plant height was negligible under the present set of observations.

3.4 Clustering of Genotypes Based on Mahalanobis D^2 Statistics

Based on Mahalanobis D^2 statistics and Tocher's optimization method, the 55 genotypes were grouped into six distinct clusters, indicating considerable genetic divergence among the experimental material (Table 4 and Figure 1). The distribution of genotypes among the clusters was highly unequal. Cluster I was the largest cluster and comprised 29 genotypes, namely NLR 3954, RGL 7053, RGL 7054, RGL 7045, BPT 3485, RGL 7051, BPT 3538, RGL 7052, BPT 3451, MTU 1282, RGL 7062, MTU 1318, RGL 7061, BPT 3543, RGL 7063, RGL 7049, BPT 3456, NLR 3955, BPT 3548, BPT 3554, MTU 1262, BPT 5204, RGL 7048, BPT 3468, NLR 3969, MTU 1224, MTU 1443, MTU 1426 and NLR 3968. Cluster II contained only two genotypes, MTU 1226 and RGL 7030.

Cluster III comprised 17 genotypes, namely RGL 7064, RGL 7065, MTU 1310, MTU 1121, RGL 7044, RGL 7058, MTU 1239, SWARNA, RGL 7060, NLR 3991, MCM 140, RGL 2537, RGL 7059, BPT 3372, NLR 3937, BPT 3746 and BPT 3454, NLR 3970. Cluster IV was a solitary cluster containing RGL 7055, while Cluster V was also solitary and contained NLR 3944. Cluster VI consisted of three fine grain genotypes, NLR 3694, RNR 15048 and BPT 3284. The occurrence of six clusters, including two solitary clusters, indicated substantial genetic differentiation among the genotypes.

Cluster I accounted for the largest proportion of genotypes followed by Cluster III, whereas Clusters IV and V contained single genotypes. The solitary clustering of RGL 7055 and NLR 3944 suggested that these genotypes possessed distinct combinations of the quantitative characters considered in the analysis. Similar clustering patterns based on Mahalanobis D^2 statistics have been reported in rice by several researchers, including Tejaswini *et al.* (2018) [37], Devasena *et al.* (2023) [5] and Aruna *et al.* (2024) [2].

3.5 Intra- and Inter-Cluster Genetic Distances

The intra- and inter-cluster Mahalanobis D^2 values showed considerable variation in the magnitude of genetic divergence among the six clusters (Table 5 and Figure 2). The intra-cluster distances ranged from 0.00 to 881.34. Cluster III recorded the highest intra-cluster distance of 881.34, followed by Cluster II (826.09), Cluster VI (757.49) and Cluster I (545.56). Clusters IV and V had intra-cluster distances of 0.00, as each contained only one genotype. The intra- and inter-cluster Mahalanobis D^2 values revealed substantial genetic divergence among the six clusters, with inter-cluster distances generally exceeding intra-cluster distances, indicating greater genetic diversity between

clusters. Similar results were reported by Dey *et al.* (2020)^[7] and Mohamud *et al.* (2022)^[21] in rice.

The inter-cluster distances showed a much wider range, from 1247.97 to 37560.64. The maximum inter-cluster distance was observed between Cluster II and Cluster VI (37560.64), indicating the greatest genetic divergence between these two groups. The next highest distances were recorded between Cluster IV and Cluster VI (21597.73) and between Cluster I and Cluster II (22258.43). A high distance was also observed between Cluster II and Cluster V (16213.88) and between Cluster II and Cluster III (8632.07). The distance between Cluster I and Cluster IV was 10421.25, while Cluster III and Cluster VI showed a distance of 11547.80. Cluster II and Cluster IV had a distance of 2682.68, whereas Cluster III and Cluster IV showed a distance of 2217.97.

The lowest inter-cluster distance was observed between Cluster I and Cluster V (1247.97), followed by Cluster III and Cluster V (1961.40) and Cluster IV and Cluster III (2217.97). Thus, Cluster II and Cluster VI were the most divergent groups, whereas Cluster I and Cluster V were comparatively closer based on the Mahalanobis D^2 values. The large inter-cluster distances, particularly between Clusters II and VI, indicate the possibility of obtaining greater genetic recombination by crossing suitable genotypes belonging to these groups. However, divergence should be considered together with the agronomic performance and desirable cluster means of the genotypes.

3.6 Cluster Mean Analysis

Marked differences were observed among the six clusters for all the eight quantitative characters, demonstrating considerable variation in the mean performance of the genotype groups (Table 6). For days to 50 per cent flowering (DFF), the cluster means ranged from 106.33 days in Cluster VI to 129.50 days in Cluster V. Cluster II recorded a mean of 111.00 days, Cluster IV 114.50 days, Cluster III 118.28 days and Cluster I 118.64 days. Thus, Cluster VI represented the earliest flowering group, whereas Cluster V represented the latest flowering group.

For plant height (PH), Cluster IV recorded the lowest mean value of 228.00 cm, while Cluster V recorded the highest mean of 313.50 cm. The means of Clusters I, II, III and VI were 272.14, 275.50, 278.36 and 257.50 cm, respectively. Therefore, Cluster IV represented the shortest plant type, whereas Cluster V represented the tallest plant type. For effective bearing tillers (EBT), the highest cluster mean was recorded in Cluster III (271.16), followed by Cluster I (269.38) and Cluster VI (239.60). Cluster V recorded 224.20, whereas Cluster IV and Cluster II had means of 177.60 and 165.80, respectively. Thus, Cluster III was characterized by the highest number of effective bearing tillers, while Cluster II recorded the lowest.

For panicle length (PL), considerable variation was observed among the clusters. The highest mean was recorded in Cluster II (27.91 cm), followed by Cluster IV (23.80 cm) and Cluster III (20.39 cm). Cluster V recorded 17.40 cm, Cluster I 15.57 cm, while the lowest mean was recorded in Cluster VI (11.69 cm). The wide difference between Cluster II and Cluster VI indicated substantial divergence for panicle length. For number of filled grains per panicle (NFGP), Cluster IV recorded the highest mean value of 136.70, closely followed by Cluster V (136.00). Cluster III recorded 125.45, Cluster VI 121.27 and Cluster I

120.31, while the lowest value was observed in Cluster II (112.20). For number of infertile grains per panicle (NIFGP), the highest mean was recorded in Cluster V (29.15), followed by Cluster IV (26.28), Cluster III (25.41), Cluster I (25.33) and Cluster VI (24.95). Cluster II had the lowest mean value of 23.35.

For test weight (TW), the highest cluster mean was recorded in Cluster VI (45.72 g). Cluster I and Cluster V recorded means of 37.44 g and 36.50 g, respectively. Cluster IV had a mean of 30.20 g, while Cluster III recorded 21.44 g. The lowest test weight was recorded in Cluster II (20.63 g). For grain yield per plant (GYP), Cluster VI recorded the highest mean value of 5.16, followed by Cluster II (4.97) and Cluster I (4.67). Cluster IV and Cluster III recorded 4.58 and 4.52, respectively, while Cluster V had the lowest mean grain yield of 3.78. The cluster mean analysis demonstrated that no single cluster was superior for all the characters. Cluster VI was characterized by the earliest flowering (106.33 days), highest test weight (45.72 g) and highest grain yield per plant (5.16), whereas Cluster II possessed the highest panicle length (27.91 cm) but the lowest EBT (165.80), NFGP (112.20) and TW (20.63 g). Cluster III recorded the highest EBT (271.16), while Cluster IV had the highest NFGP (136.70) and shortest plant height (228.00 cm). Cluster V showed the highest DFF (129.50 days), PH (313.50 cm) and NIFGP (29.15), but the lowest GYP (3.78). These contrasting cluster means indicate the availability of complementary character combinations that could be exploited in hybridization programmes. Such differences in cluster means indicate the potential of diverse clusters for identifying promising genotypes and selecting suitable parents in rice breeding programmes, as also reported by Shoba *et al.* (2019)^[32] and Sruthi *et al.* (2024)^[35].

Based on the genetic divergence, cluster means and trait contributions, Cluster II $\times$ Cluster VI ($D^2 = 37560.64$) was identified as the most divergent and promising combination, with Cluster II contributing the highest panicle length (27.91 cm) and Cluster VI contributing early flowering (106.33 days), high test weight (45.72 g) and highest grain yield per plant (5.16 g). Cluster IV $\times$ Cluster VI ($D^2 = 21597.73$) was another desirable combination, where Cluster IV contributed shorter plant height (228.00 cm) and highest filled grains per panicle (136.70), while Cluster VI contributed earliness, high test weight and yield. Cluster III $\times$ Cluster VI ($D^2 = 11547.80$) may also be utilized to combine high effective bearing tillers (271.16) and filled grains (125.45) of Cluster III with the high-test weight and grain yield of Cluster VI. Therefore, crosses involving MTU 1226/ RGL 7030 $\times$ BPT 3284, RGL 7055 $\times$ BPT 3284/ NLR 3694, and superior genotypes from Cluster III $\times$ BPT 3284/ NLR 3694 are suggested for generating segregating populations with complementary traits and enhanced genetic variability.

3.7 Factor Analysis with Varimax Rotation

Factor analysis with Varimax rotation was performed to identify the major groups of characters contributing to variation among the rice genotypes. Three factors were extracted and together they accounted for 100.00 per cent of the variation, with Factor I, Factor II and Factor III accounting cumulatively for 33.33, 66.66 and 100.00 per cent, respectively (Table 7). Factor I, accounting for 33.33 per cent of the variation, showed negative loadings for DFF (−0.190), PL (−0.423), NFGP (−0.504) and NIFGP (−0.569), while PH (0.419), EBT (−0.047), TW (0.057) and

GYP (0.164) showed comparatively smaller associations. The highest absolute loading on Factor I was observed for NIFGP (−0.569), followed by NFGP (−0.504) and PL (−0.423). Hence, Factor I was predominantly associated with the grain-number and panicle-related traits, particularly NIFGP, NFGP and PL.

Factor II accounted for an additional 33.33 per cent, increasing the cumulative variation explained to 66.66 per cent. The highest loading on this factor was recorded for EBT (0.657), followed by DFF (0.539) and PH (0.479). PL had almost no association with this factor (0.008), while NFGP (0.069), NIFGP (0.093), TW (0.164) and GYP (0.084) showed relatively low loadings. Thus, Factor II was primarily characterized by EBT, DFF and PH. Factor III accounted for the remaining 33.34 per cent of variation, resulting in 100.00 per cent cumulative variance. Plant height showed the highest positive loading (0.560), whereas TW (−0.526) and GYP (−0.491) showed relatively high negative loadings. PL showed a loading of 0.304, while DFF (−0.211), EBT (−0.052), NFGP (0.156) and NIFGP (−0.071) had relatively smaller associations. The factor analysis therefore indicated that the variation among the genotypes was distributed among different character complexes. The first factor was mainly associated with NIFGP, NFGP and PL, the second with EBT, DFF and PH and the third with PH, TW and GYP. Similar applications of factor analysis for identifying important traits and underlying sources of variation in rice have been reported by Sesmeskandi *et al.* (2023) [30] and Rahman *et al.* (2024) [25].

3.8 Principal Component Analysis

Principal Component Analysis (PCA) was carried out to further identify the major characters contributing to genetic variability among the 55 rice genotypes. The first three principal components accounted for 67.76 per cent of the total variation (Table 8 and Figure 3&4). PC1 explained 33.75 per cent, PC2 explained an additional 19.88 per cent, resulting in a cumulative contribution of 53.63 per cent, while PC3 contributed a further 14.13 per cent, giving a cumulative contribution of 67.76 per cent. Principal Component Analysis (PCA) was employed to identify the major components contributing to the total variation among the rice genotypes, and similar findings have been reported by Satyanarayana *et al.* (2023) [29] and Prakash *et al.* (2024) [24] in rice.

PC1 accounted for the largest proportion of variation (33.75%). The largest absolute loading on PC1 was recorded for panicle length PL (−0.527), followed by NFGP (−0.447), PH (−0.427), DFF (−0.378) and NIFGP (−0.343). EBT (0.185), TW (0.178) and GYP (0.118) showed comparatively lower loadings. Therefore, PC1 was primarily influenced by PL, NFGP, PH, DFF and NIFGP, with PL being the most influential character. PC2 explained 19.88 per cent of the total variation, taking the cumulative contribution of the first two components to 53.63 per cent. Test weight (TW) recorded the highest absolute loading on PC2 (−0.591). This was followed by PH (−0.439), NIFGP (0.404), DFF (−0.381) and NFGP (0.303). EBT (−0.173), PL (−0.161) and GYP (−0.042) showed relatively lower loadings. Thus, PC2 was predominantly associated with TW, PH, NIFGP, DFF and NFGP.

PC3 accounted for 14.13 per cent of the total variation and increased the cumulative contribution to 67.76 per cent. The

largest loading was recorded for EBT (−0.773), indicating that effective bearing tillers was the principal contributor to this component. GYP showed a positive loading of 0.369, while DFF had a loading of −0.353, TW 0.266, PH 0.146, PL 0.117 and NIFGP 0.090. NFGP had a loading of −0.163. Thus, PC3 was strongly influenced by EBT, with additional contributions from GYP, DFF and TW. The PCA results demonstrated that PL, NFGP, PH, DFF and NIFGP were important contributors to PC1, TW, PH, NIFGP, DFF and NFGP contributed substantially to PC2, while EBT was the dominant character associated with PC3. The first three components together explained 67.76 per cent of the total variability, indicating that a substantial proportion of the multivariate variation among the genotypes could be represented by these major components.

3.9 Overall Results of Multivariate Analysis

The combined results of MANOVA, character contribution, Mahalanobis D^2 clustering, cluster mean analysis, factor analysis and PCA clearly demonstrated substantial genetic divergence among the 55 rice genotypes for the eight quantitative characters studied. The MANOVA was highly significant ($P = 0.000$), while the ANOVA for dispersion also showed an extremely high and significant F value (5.70E11). Among the individual characters, panicle length (PL) was the predominant contributor to genetic divergence (79.87%), followed by NIFGP (10.51%) and DFF (6.73%). The genotypes were grouped into six clusters, with Cluster I containing 29 genotypes and Cluster III containing 17 genotypes. Clusters IV and V were solitary clusters containing RGL7055 and NLR 3944 respectively.

The greatest inter-cluster divergence was observed between Cluster II and Cluster VI ($D^2 = 37560.64$), followed by Cluster IV and Cluster VI (21597.73) and Cluster I and Cluster II (22258.43). Cluster VI was particularly notable for its combination of early flowering (106.33 days), high test weight (45.72 g) and highest grain yield per plant (5.16). Cluster II possessed the longest panicles (27.91 cm), whereas Cluster IV had the highest NFGP (136.70) and Cluster III had the highest EBT (271.16). Cluster II × Cluster VI ($D^2 = 37560.64$) is the most promising cross, combining long panicle length (27.91 cm) with early flowering, high test weight (45.72 g) and grain yield (5.16). Cluster IV × Cluster VI ($D^2 = 21597.73$) can combine short plant height (228.00 cm) and high filled grains (136.70) with high yield and test weight. Cluster III × Cluster VI ($D^2 = 11547.80$) can combine high tillering (271.16) and filled grains (125.45) with superior yield attributes. Thus, MTU 1226/ RGL7030 × BPT 3284 and RGL 7055 × BPT 3284/ NLR 3694 are recommended for generating promising segregants.

Factor analysis identified NIFGP, NFGP and PL as major variables associated with Factor I, EBT, DFF and PH with Factor II and PH, TW and GYP with Factor III. PCA similarly identified PL as the largest contributor to PC1, TW as the largest contributor to PC2 and EBT as the dominant contributor to PC3. The first three principal components together accounted for 67.76 per cent of the total variation. Overall, the results demonstrate that the genetic variability present in the material was substantial and was expressed through different combinations of flowering duration, plant height, tillering ability, panicle characteristics, grain number, grain sterility, grain weight and yield. The contrasting cluster means and large inter-cluster distances,

particularly between Clusters II and VI, provide useful information for identifying genetically divergent and agronomically complementary parents for future rice breeding programmes. Similar applications of multivariate

techniques for characterizing genetic divergence and identifying promising genotypes in rice have been reported by Shrestha *et al.* (2021) ^[33], Sharma *et al.* (2024) ^[31] and Munna *et al.* (2025) ^[23].

Table 1: Wilks' Lambda test

Statistic	Value
Determinant of Error Matrix	26261.80524E+2
Determinant of Error + Variety Matrix	15266.04399E+14
Wilks' Criterion	17202.75748E-16
V statistic	2045.18400
M value	75.5
Degrees of freedom	432
Probability	0.00000

Table 2: ANOVA for dispersion

Source	df	SS	MS	F	P
Varieties	54	1.5266E18	2.8270E16	5.705E11	0.000***
Error	53	2.6262E06	4.9551E04		
Total	107	1.5266E18	1.4267E16		

Table 3: Contribution of traits towards divergence

Trait	Times ranked first	Contribution (%)
DFF	100	6.73
PH	0	0.00
EBT	1	0.07
PL	1186	79.87
NFGP	26	1.75
NIFGP	156	10.51
TW	13	0.88
GYP	3	0.20

Table 4: Clustering of Genotypes

Cluster Number	Number of genotypes	Name of the genotypes
Cluster I	29	NLR 3954, RGL 7053, RGL 7054, RGL 7045, BPT 3485, RGL 7051, BPT 3538, RGL 7052, BPT 3451, MTU 1282, RGL 7062, MTU 1318, RGL 7061, BPT 3543, RGL 7063, RGL 7049, BPT 3456, NLR 3955, BPT 3548, BPT 3554, MTU 1262, BPT 5204, RGL 7048, BPT 3468, NLR 3969, MTU 1224, MTU 1443, MTU 1426, NLR 3968.
Cluster II	2	MTU 1226, RGL 7030.
Cluster III	17	RGL 7064, RGL 7065, MTU 1310, MTU 1121, RGL 7044, RGL 7058, MTU 1239, SWARNA, RGL 7060, NLR 3991, MCM 140, RGL 2537, RGL 7059, BPT 3372, NLR 3937, BPT 3746, BPT 3454, NLR 3970.
Cluster IV	1	RGL 7055.
Cluster V	1	NLR 3944.
Cluster VI	3	NLR 3694, RNR 15048, BPT 3284.

Table 5: Inter- and intra-cluster distances

	Cluster I	Cluster II	Cluster III	Cluster IV	Cluster V	Cluster VI
Cluster I	545.56	22258.43	4115.70	10421.25	1247.97	2669.90
Cluster II	22258.43	826.09	8632.07	2682.68	16213.88	37560.64
Cluster III	4115.70	8632.07	881.34	2217.97	1961.40	11547.80
Cluster IV	10421.25	2682.68	2217.97	0.00	6096.18	21597.73
Cluster V	1247.97	16213.88	1961.40	6096.18	0.00	5840.39
Cluster VI	2669.90	37560.64	11547.80	21597.73	5840.39	757.49

Table 6: Cluster means

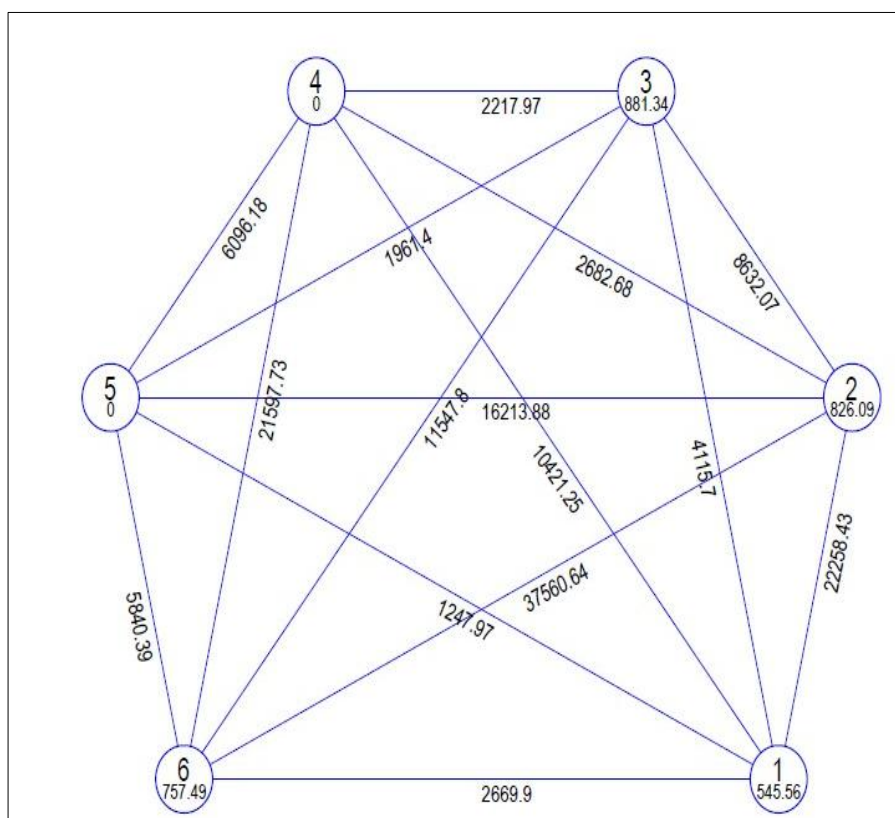
Cluster	DFF	PH	EBT	PL	NFGP	NIFGP	TW	GYP
Cluster I	118.64	272.14	269.38	15.57	120.31	25.33	37.44	4.67
Cluster II	111.00	275.50	165.80	27.91	112.20	23.35	20.63	4.97
Cluster III	118.28	278.36	271.16	20.39	125.45	25.41	21.44	4.52
Cluster IV	114.50	228.00	177.60	23.80	136.70	26.28	30.20	4.58
Cluster V	129.50	313.50	224.20	17.40	136.00	29.15	36.50	3.78
Cluster VI	106.33	257.50	239.60	11.69	121.27	24.95	45.72	5.16

Table 7: Factor analysis (Varimax rotation)

Character	Factor 1	Factor 2	Factor 3
DFF	-0.190	0.539	-0.211
PH	0.419	0.479	0.560
EBT	-0.047	0.657	-0.052
PL	-0.423	0.008	0.304
NFGP	-0.504	0.069	0.156
NIFGP	-0.569	0.093	-0.071
TW	0.057	0.164	-0.526
GYP	0.164	0.084	-0.491
Variance explained	33.33	66.66	100.00

Table 8: Principal Component Loadings

Character	PC 1	PC 2	PC 3
DFF	-0.378	-0.381	-0.353
PH	-0.427	-0.439	0.146
EBT	0.185	-0.173	-0.773
PL	-0.527	-0.161	0.117
NFGP	-0.447	0.303	-0.163
NIFGP	-0.343	0.404	0.090
TW	0.178	-0.591	0.266
GYP	0.118	-0.042	0.369
Variance explained	33.75	53.63	67.76

**Fig 1:** Mahalanobis D² Euclidean distances between the clusters

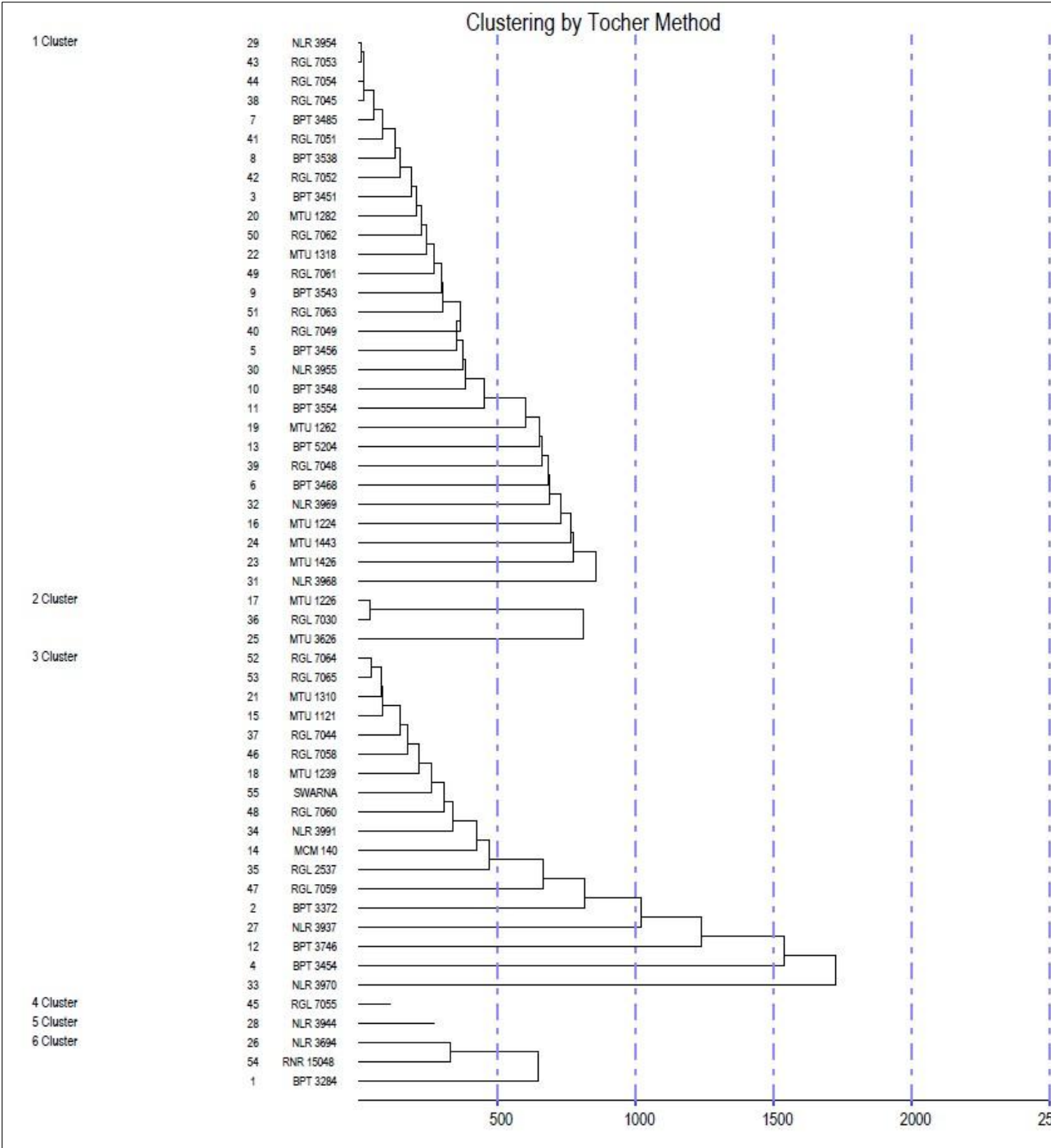


Fig 2: Clustering by Tocher's optimization method

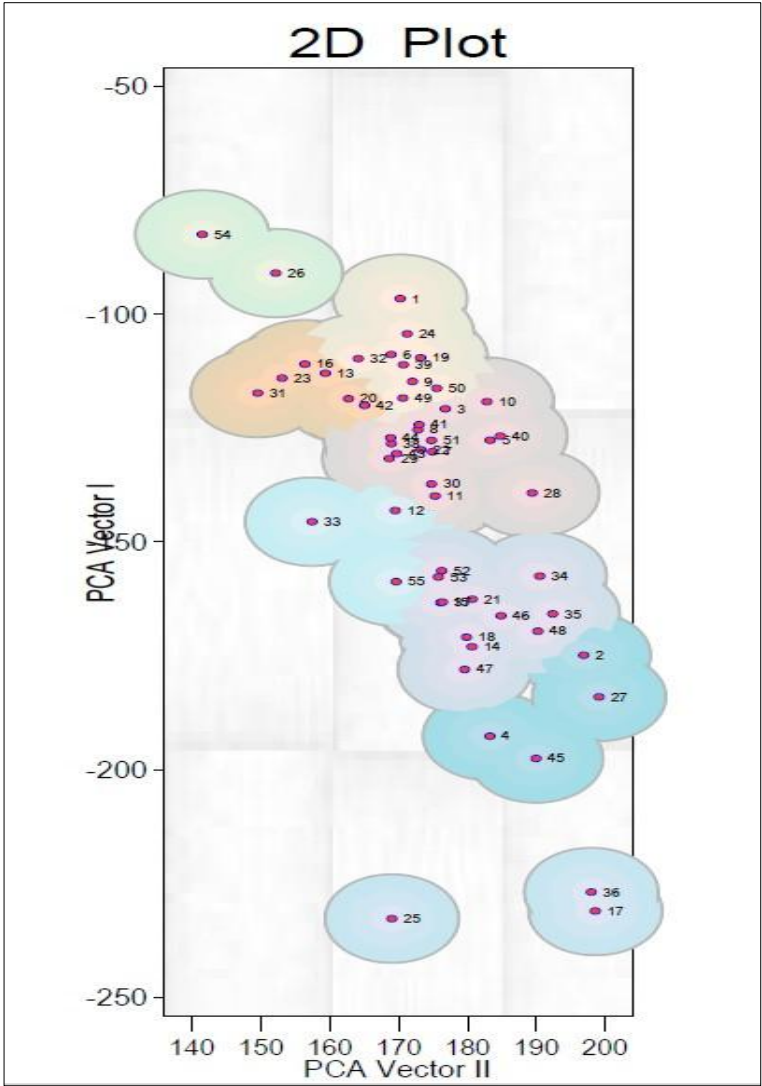


Fig 3: Loadings of genotypes on principal components

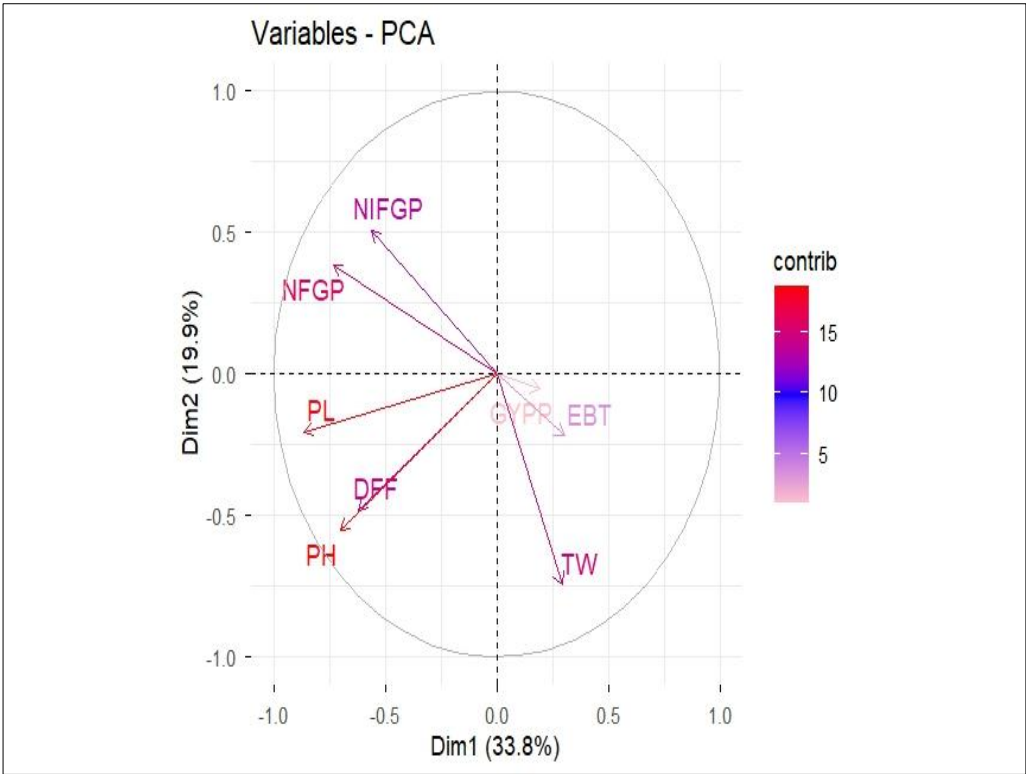


Fig 4: Loadings of characters on principal components

4. Conclusion

The present investigation revealed substantial genetic diversity among the advanced rice genotypes, as demonstrated by Mahalanobis D² statistics, factor analysis, principal component analysis (PCA) and Tocher's clustering. The highly significant Wilks' Lambda confirmed considerable multivariate variability, while factor analysis and character contribution analysis indicated that a limited number of quantitative and economically important traits accounted for a major share of the observed divergence. Mahalanobis D² analysis further classified the genotypes into distinct clusters, including solitary clusters representing potentially unique genetic resources. The presence of high inter-cluster distances indicated considerable genetic divergence among certain groups, providing opportunities for selecting diverse parents for hybridization.

Cluster mean analysis showed that different clusters were superior for different agronomic traits, emphasizing the need to consider both genetic divergence and desirable performance during parent selection. MTU 1226/ RGL 7030 × BPT 3284 was identified as a promising cross for combining long panicles with earliness, high test weight and grain yield.

RGL 7055 × BPT 3284/ NLR 3694 was also suggested to combine short plant stature and high filled grains with superior yield attributes. PCA effectively summarized the multivariate variation and supported the clustering pattern, confirming the reliability of the analyses. Overall, the results indicate that the evaluated advanced rice breeding lines possess a broad and exploitable genetic base. Selection of parents based on a combination of high inter-cluster distance, superior cluster mean performance and desirable individual genotype characteristics is likely to enhance genetic variability and selection efficiency, thereby supporting the development of high-yielding, climate-resilient and widely adapted rice cultivars.

Future Research Perspectives

Based on the findings of the present investigation, future rice improvement programmes should prioritize genetically divergent genotypes from clusters with maximum inter-cluster distances, while giving greater emphasis to major traits identified through factor and character contribution analyses in selection strategies. The identified diverse parents may be further evaluated for combining ability through diallel or line × tester designs and their genetic relationships can be validated using molecular markers such as SSRs and SNPs. Multi-location testing should be undertaken to assess genotype × environment interactions and identify stable, widely adapted materials. In addition, these diverse genotypes may be integrated into marker-assisted breeding and genomic selection programmes to accelerate the development of high-yielding, climate-resilient and widely adapted rice varieties.

References

- Artega S, Yabor L, Díez MJ, Prohens J, Boscaiu M, Vicente O. The use of proline in screening for tolerance to drought and salinity in common bean (*Phaseolus vulgaris* L.) genotypes. *Agronomy*. 2020;10(6):817.
- Aruna N, Elangaimannan R, Sriramachandrasekharan MV, Vennila S. Unravelling the genetic distance among traditional and commercial rice (*Oryza sativa* L.) genotypes using Mahalanobis D² statistics. *Biochem Cell Arch*. 2024;24(1):1357-1363.
- Asante MD, Adjah KL, Annan-Afful E. Assessment of genetic diversity for grain yield and yield component traits in some genotypes of rice (*Oryza sativa* L.). *J Crop Sci Biotechnol*. 2019;22(2):123-130.
- Behera PP, Singh SK, Singh DK, Reddy YS, Habde S, Khaire A, Ashrutha MA. Genetic diversity analysis of rice (*Oryza sativa* L.) genotypes with high grain zinc content for yield and yield traits. *J Pharmacogn Phytochem*. 2018;7(4):1319-1323.
- Devasena N, Sharmili K, Wilson D. Principal component and cluster analysis on eating and cooking quality parameters in rice (*Oryza sativa* L.) germplasm. *Biol Forum Int J*. 2023;15(5):326-332.
- Devi SV, Singh V, Lal GM. Studies on genetic divergence in rice germplasm (*Oryza sativa* L.). *Electron J Plant Breed*. 2019;10(04):1593-1599.
- Dey S, Badri J, Eswari KB, Prakasam V. Diversity analysis for yield traits and sheath blight resistance in rice genotypes. *Electron J Plant Breed*. 2020;11(01):66-74.
- Duppala MK, Beena N, Gowtham KS, Rathod A. Evaluation of advanced breeding lines of Indian mustard using principal component analysis. *J Oilseed Brassica*. 2018;9(1):45-48.
- Falconer DS, Mackay TFC. Introduction to quantitative genetics. 4th ed. Longman; 1996.
- Food and Agriculture Organization. FAOSTAT statistical database. Food and Agriculture Organization of the United Nations; 2024.
- International Rice Research Institute. Standard evaluation system for rice. 5th ed. IRRI; 2014.
- Johnson RA, Wichern DW. Applied multivariate statistical analysis. 6th ed. Pearson Education; 2007.
- Kaiser HF. The application of electronic computers to factor analysis. *Educ Psychol Meas*. 1960;20(1):141-151.
- Kumar DM, Beena N, Gowtham KS. Evaluation of advanced lines of mustard spp. using hierarchical cluster analysis. *Bioscan*. 2017;10(1):299-307.
- Kumar KM, Sreenivas G, Udayababu P, Srinivasa Rao M, Manojkumar D. Assessment of genetic variability and trait associations in advanced rice lines. *J Sci Res Rep*. 2025;31(11):699-709.
- Madhukumar K, Sreenivas G, Udayababu P, Srinivasa Rao M, Manojkumar D. Exploring genetic diversity analysis of advanced rice lines for selection and generation of high-yielding genotypes using enhanced breeding strategies. *Plant Arch*. 2026;26(Suppl 1):967-978.
- Madhukumar M, Satyanarayana PV, Babu PU, Sreenivas G, Manojkumar D. Studies on genetic parameters and trait association analysis of yield and yield factors in the advanced breeding lines of greengram. *Int J Theor Appl Sci*. 2023;15(1):75-80.
- Mahalanobis PC. On the generalized distance in statistics. *Proc Natl Inst Sci India*. 1936;2:49-55.
- Manojkumar D, Srinivas T, Rao LS, Suneetha Y, Sundaram RM, Kumari VP, Ratnam TV. Generation mean analysis for yield and yield component traits in inter-specific cross of rice (*Oryza sativa* L.). *Agric Sci Dig*. 2023;7(1):1-7.

20. Mohammadi SA, Prasanna BM. Analysis of genetic diversity in crop plants: Salient statistical tools and considerations. *Crop Sci.* 2003;43(4):1235-1248.
21. Mohamud MA, Hoque A, Hassan AH, Hossain MM, Monira S, Muse MI. The study of genetic divergence in rice (*Oryza sativa* L.) genotypes with grain yield and different morpho physiological traits using Mahalanobis' D2 analysis. *Int J Adv Agric Sci Technol.* 2022;9:20-33.
22. Mounika K, Singh SK, Pandey V, Singh DK, Khaire AR, Habde S, Majhi PK. Study of genetic divergence in rice (*Oryza sativa* L.) genotypes with high grain zinc using Mahalanobis' D2 analysis. *Electron J Plant Breed.* 2020;11(02):367-372.
23. Munna NA, Dewan MMR, Rana MM, Sajjadul AKM. Multivariate analysis of rice (*Oryza sativa* L.) genotypes. *J Biosci Agric Res.* 2025;35:2799-2809.
24. Prakash S, Reddy SS, Chaudhary S, Vimal SC, Kumar A. Multivariate analysis in rice (*Oryza sativa* L.) germplasms for yield attributing traits. *Plant Sci Today.* 2024;11(1):1-10.
25. Rahman MA, Al Mamun MA, Islam MR, Quddus MR, Munna NA, Azim MA, *et al.* Identifying key determinants of rice yield potential through multiple statistical techniques. *J Bangladesh Agric Univ.* 2024;22(1):17-28.
26. Rao CR. Advanced statistical methods in biometric research. John Wiley & Sons; 1952.
27. Rathana ND, Singh SK, Singh RK, Singh DK. Estimating genetic divergence in rice (*Oryza sativa* L.) using D² analysis. *Electron J Plant Breed.* 2020;11(03):836-840.
28. Satyanarayana PV, Madhu Kumar K, Uday Babu P, Srinivas T, Manojkumar D. Genetic architecture studies to identify the selection criteria for yield in long duration rice lines (*Oryza sativa* L.). *Int J Environ Clim Chang.* 2024;14(4):842-851.
29. Satyanarayana PV, Madhukumar K, Udayababu P, Srinivas T, Manojkumar D. Association studies for identifying the selection criteria among early varieties of rice in North Coastal Zone of Andhra Pradesh. *Biol Forum Int J.* 2023;15(11):329-335.
30. Semeskandi MN, Mazloom P, Arabzadeh B, Moghadam MN, Ahmadi T. Application of correlation coefficients and principal components analysis in stability of quantitative and qualitative traits on rice improvement cultivation. *Braz J Biol.* 2023;84:e268981.
31. Sharma A, Kumar J, Redhu M, Kumar P, Godara M, Ghiyal P, *et al.* Estimation of rice yield using multivariate analysis techniques based on meteorological parameters. *Sci Rep.* 2024;14(1):12626.
32. 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(03):1095-1104.
33. Shrestha J, Subedi S, Kushwaha UKS, Maharjan B. Evaluation of growth and yield traits in rice genotypes using multivariate analysis. *Heliyon.* 2021;7(9):e07981.
34. Singh HP, Raigar OP, Chahota RK. Estimation of genetic diversity and its exploitation in plant breeding. *Bot Rev.* 2022;88(3):413-435.
35. Sruthi SR, Kumar NL, Kishore K, Johnny SI, Anbuselvam Y. 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.* 2024;11(4):1-9.
36. Tagliapietra BL, Soares CF, Clerici MTPS. Rice (*Oryza sativa* L.) and its products for human consumption: General characteristics, nutritional properties, and types of processing. *Food Sci Technol.* 2024;44:e112423.
37. Tejaswini KLY, Srinivas Manukonda, Ravi Kumar BNVS, Ramana Rao PV, Krishnam Raju S. Application of principal component analysis for rice F5 families characterization and evaluation. *Emerg Life Sci Res.* 2018;4:72-84.