Received: 11 Apr 2022 | Revised: 28 Aug 2022 | Accepted: 01 Sep 2022

Electronic Journal of Plant Breeding

Research Article

Multivariate analysis based prediction of phenotypic diversity associated with yield and yield component traits in germplasm lines of rice (*Oryza sativa* L.)

Kasanaboina Krishna[1]*, Y. Chandra Mohan[2], L. Krishna[2], G. Parimala[1] and R. Jagadeeshwar[3]

[1]Department of Genetics and Plant Breeding, PJTSAU, College of Agriculture, Rajendranagar. Hyderabad, 500030, Telangana, India.
[2]Rice Research Centre, ARI, PJTSAU, Rajendranagar. Hyderabad, 500030 Telangana, India.
[3]Administrative Building, PJTSAU, Rajendranagar. Hyderabad, 500030 Telangana, India
***E-Mail**: kasanaboinakrishna@gmail.com

Abstract

An investigation was conducted with 217 germplasm lines of rice to estimate potential variation among rice genotypes. Multivariate analyses *viz*., PCA and cluster analysis to assess genetic diversity were performed on seven agronomic traits. For the evaluated traits in the lines of germplasm, box plots and normal probability plots displayed substantial estimates of variability. PCA showed that PC1 and PC2 represented 46.15 per cent of variation. PC1 was responsible for the most variance (24%) for four characters, followed by PC2 (22.15%) for three parameters. Filled grain number per panicle, single plant yield, plant height and thousand grain weight were identified as vital traits contributing to variability. Based on agglomerative hierarchical cluster analysis, the germplasm lines were grouped into five clusters, which explained a lot of variation in the traits. The study identified that plant height and 1000 grain weight have the greatest impact on variation. The genotypes *viz*., WGL 1063, RNR 26085, OR 2511-3, Swarna, KNM 7123, WGL 1063, MGD-101, IVT MS-6130, IVT IM-4231, IVT IME-3948 and IVT NPT-6303 were found to be the best performing genotypes for the traits panicle length, plant height and single plant yield, which can be used in hybridisation programme to improve these traits.

Keywords: Rice, Box plot**,** diversity, germplasm, principal components and variability

INTRODUCTION

Rice (*Oryza sativa* L.) is a widely grown cereal crop that provides energy to over half of the world's population. Since rice is a nutritionally important crop, the origin and diversity of *Oryza sativa* have attracted interest. Plants with genetic diversity have the potential to respond to rapid changes in environmental circumstances (Raza *et al*., 2019). Identifying and utilizing genetic variation from a wide variety of germplasm is critical to the success of a breeding effort. Breeders in various countries have used numerous ancient types as sources of genes of interest, and these materials serve as sources of desirable genes needed for relevant breeding programs to boost yield and productivity. Diversification of parents for harnessing the phenotypic variation contained in the germplasm is a frequently used strategy for overcoming the production plateau, which is often reflective of a confined genetic base. Multivariate statistical approaches are commonly used to summarize and characterize the intrinsic diversity among genotypes. The degree of difference between the genotypes is measured using multivariate analysis. As observations, Principal Component Analysis (PCA) data consists of numerous intercorrelated quantitative dependent variables (Mahendran *et al*., 2015). PCA collects data from a table and displays it as a collection of new orthogonal variables known as principle components. As variation occurs often in plants for yield and yield-related characteristics (Maji *et al*., 2012), PCA identifies patterns and reduces redundancy in datasets. The main advantage of PCA is that it allows you to quantify the importance of each dimension when connecting the variability of a dataset (Shoba *et al*., 2019). Cluster analysis, which is based on generalized distance, is a useful tool for determining the level of genetic divergence at the genotypic level (Mahalanobis, 1936). Multivariate analysis, which includes cluster analysis and PCA, has already been shown to be an effective biometrical tool for evaluating the degree of variability in a germplasm collection of different crop plants (Acquaah, 2012). Hence, the present investigation was undertaken to determine the level of genetic diversity in rice germplasm employing genetic distances calculated using morphological and agronomic traits as measures, as well as to select germplasm lines as parents in the development of promising high yielding lines.

MATERIALS AND METHODS

The germplasm used in this study consisted of 210 rice accessions, The trial was conducted during *Kharif*, 2018 at Rice Research Centre, Agriculture Research Institute, Rajendranagar, Hyderabad, Telangana State, in an augmented design introduced by Federer (1956, 1961, 1991) with seven checks along with 210 rice accessions. A healthy nursery was established and seedlings with 31 days old were transferred into a well-prepared main field. Each entry had two rows of 4 m length, consisting of 25 plants in each row, adopted a spacing of 20 cm between rows and 15 cm between plants. For each genotype, plant height, panicle number per plant, panicle length, filled grain number per panicle, thousand grain weight and single plant yield were determined on randomly selected five plants. On a full plot basis, days to 50% heading were recorded. ANOVA was used to compare the mean data of each character. To uncover the best connections among features, XLSTAT software was used to perform box plot analysis, principal component analysis (PCA), and agglomerative hierarchical clustering.

RESULTS AND DISCUSSION

A box and whisker plot is a graph that displays a visual representation of a data set's statistical five-number summary. The Sample Minimum, Lower Quartile or First Quartile, Median, Higher Quartile or Third Quartile, and Sample Maximum are the first five numbers in the Five Number Summary. The upper, median, and lower quartiles reflect the 75th, 50th, and 25th percentiles of the accessions, respectively. The frequency distribution for seven quantitative features (**Table 1**) across 217 rice accessions was presented in the form of box plots (**Fig. 1**) to highlight genetic diversity. Plant height and 1000 grain weight, in particular, showed higher genetic variability. Lakshmi *et al*. (2019) reported similar results. The population variation is represented by the vertical lines (whiskers) (**Fig.2)**. Outlines are drawn with dots. The other parameters, days to 50% flowering, plant height, panicle number per plant, panicle length, filled grain number per panicle and thousand grain weight, all deviated somewhat from the normal distribution, demonstrating variability.

Five of the seven components investigated in this study had eigenvalues of more than one. For the selection of various parents, the main components with more than one eigenvalue showed greater variability among the rice genotypes. According to PCA, the first five PCs explained 87.35 per cent of the variation in yield and yield component attributes among 217 genotypes (**Table 2**). The first principal component (PC1) accounted

Table 1. Frequency distribution for seven quantitative traits among 217 rice accessions

Statistic	DFF	PH	PNP	PL	FGNP	TSW	SPY
Number of observations	217	217	217	217	217	217	217
Minimum	91.000	72.300	4.000	18.460	71.000	9.906	13.500
Maximum	126.000	130.330	22.000	88.830	497.333	28.356	54.590
1st Quartile	103.000	95.660	11.330	22.860	171.000	16.047	23.852
Median	107.000	100.660	13.000	24.430	212.000	20.169	28.808
3rd Quartile	112.000	111.000	14.660	25.930	281.000	23.186	34.398
Mean	107.366	103.013	13.133	24.715	227.806	19.530	29.510
Variance (n-1)	47.012	141.362	7.267	24.925	605.332	19.622	62.973
Standard deviation	6.857	11.890	2.696	4.992	7.784	4.430	7.936

DFF: Days to 50% flowering, PH: Plant height (cm), PNP: Panicle number per plant, PL: Panicle length (cm), NFGP: Filled grain number per panicle , TW: 1000-grain weight, SPY; Single plant yield (g).

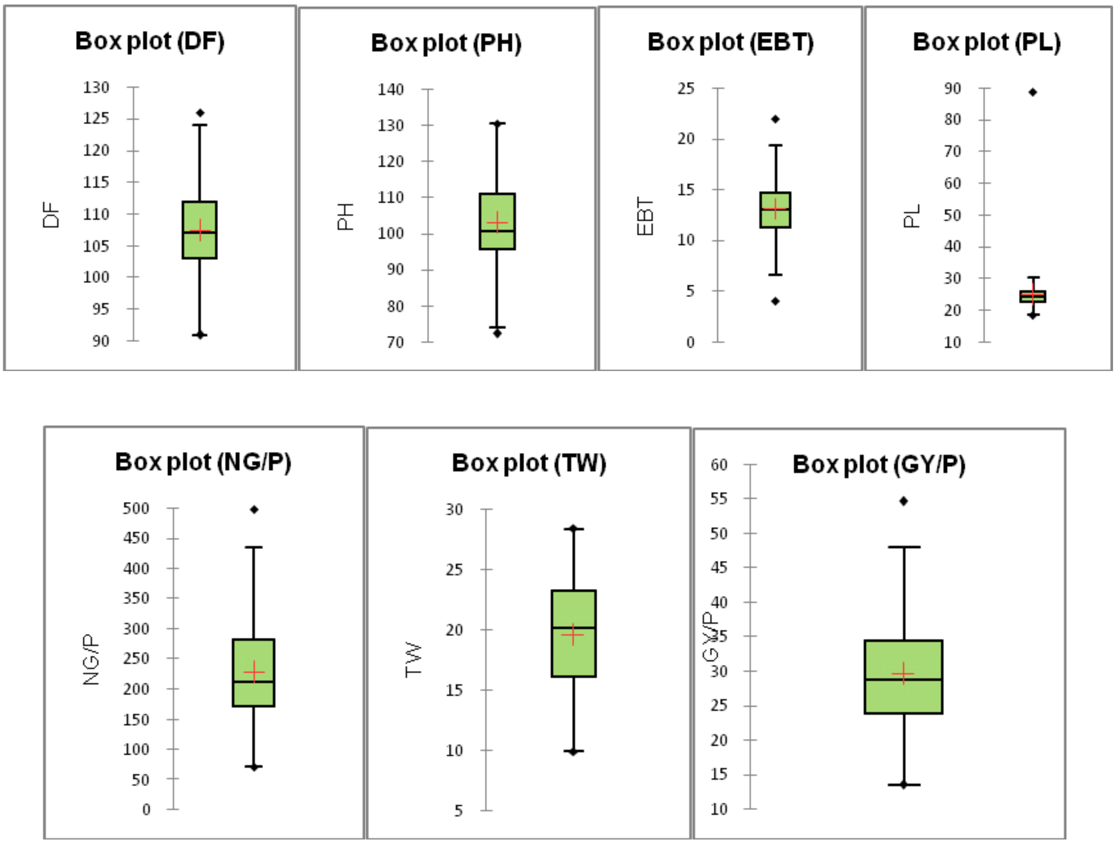

Fig. 1. Box-plots depicting the variation of the data from 217 rice accessions. Dots sketch outliers

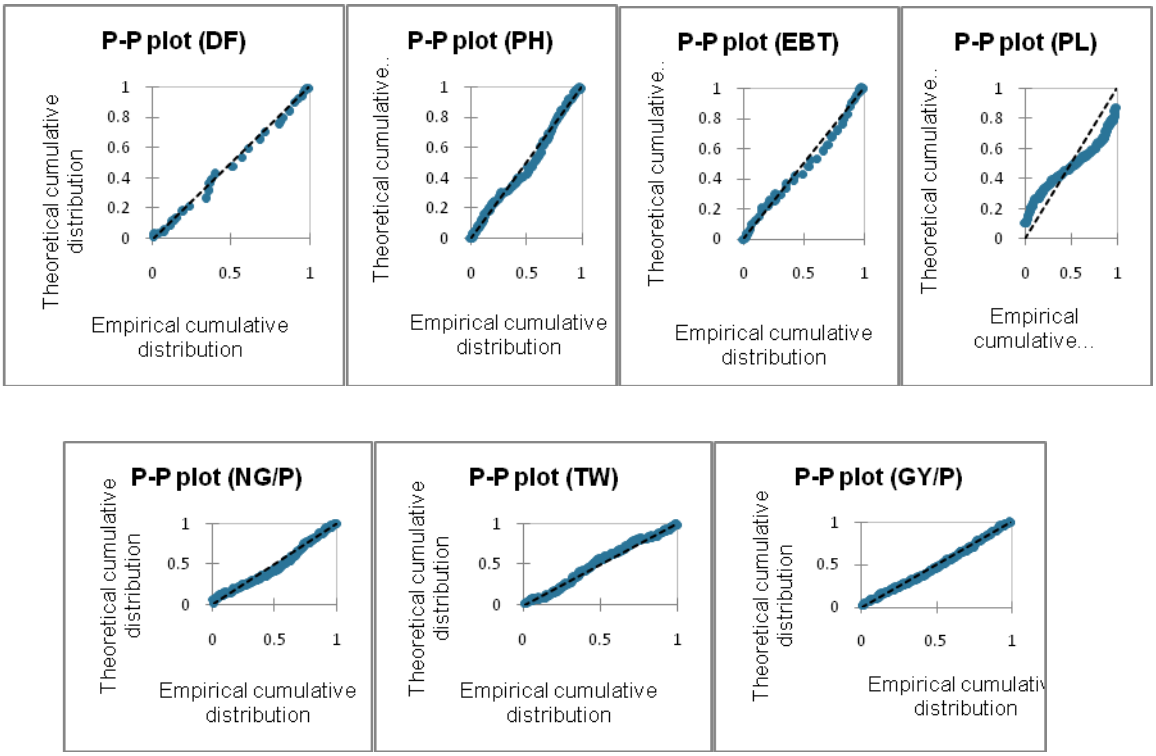

Fig. 2. Normal p-p plots depicting variation in all 7 quantitative characters of rice genotypes

for 24 per cent of total variation, with filled grain number per panicle single plant yield, plant height and thousand grain weight contributing more favorably. The PC2 which was predominantly contributed by thousand grain weight, panicle length, plant height, accounted for an additional 22.15 per cent of total variation **(Table 3)** Finally, the third, fourth, and fifth principal components (PC3, PC4, and PC5) contributed approximately 15.48, 15.29, and 10.40 per cent of the variability, respectively for the attributes, present among the germplasm accessions. PC3 described the patterns of variation in days to 50 % flowering, panicle length, single plant yield, while PC4 supplied variance through days to 50% heading, panicle number per plant and single plant yield. Finally, grain yield per plant, plant height, panicle number per plant and panicle length were major contributors to PC5 variability. These traits are largely engaged in the divergence and they also carry most of the variability. Kumari *et al*. (2021) based on their study in 119 rice genotypes along with two checks reported that the first three PCs together contributed 68.69 per cent to the total variability and that the panicle number per plant showed positive loading in PC1. Hence, the selection of traits with high variability will be rewarding for future breeding programs.

Scree plot elucidated the variation percentage between Eigenvalues and the Principal components (**Fig. 3**.). In this study, PC1 showed 24 per cent variability with an Eigenvalue of 1.64. In comparison to the other PCs, the graph clearly shows that PC1 showed the most variability. Christina *et al*. (2021) reported a maximum variation of 38.82 per cent in PC1 in their study on 49 rice lines. Hence, the genotypes selected from PC1 would be useful in future breeding programmes for the improvement of the traits contributing to maximum variability *viz.,* filled grain number per panicle, single plant yield, plant height and thousand grain weight. The biplot diagram gives the picture of interaction among the characters and also the genotypes performing better for the traits. The vector length of each trait depicts its contribution to total divergence, longer the vector length, more is the contribution of concerned traits. In this study, the distribution and nature of diversity for genotypes and quantitative traits are described in the biplot diagram (**Fig. 4**.) between PC1 and PC2. The trait thousand grain weight showed maximum vector length indicating its contribution to the total divergence followed by filled grain number per panicle, plant height and panicle length. The angle between the trait vectors indicates the direction of association between the traits. An acute angle (90º) indicates a negative correlation and a right angle (90º) indicates no correlation. Out of seven traits studied, the traits *viz.,* plant height, panicle length, and filled grain number per panicle, showed a positive correlation with single plant yield. Fathima *et al*. (2021) reported similar findings. For such traits, the genotypes that are close to the trait vector in the same quadrant will perform well. The genotypes *viz.,* NVSR 6121, HMT-Sona (BG), JGL 23835, RNR 26052 and RNR 28410 along with other genotypes in the particular quadrant perform better for the traits thousand grain weight and panicle number per plant. The genotypes like WGL 1063, RNR 26085, OR 2511-3, Swarna, KNM 7123, WGL 1063 and MGD-101, along with others are the best performing genotypes for the traits of panicle length and plant height. The genotypes like IVT MS-6130, IVT IM-4231, IVT IME-3948, IVT NPT-6303, along with others are the best performing

Table 2. Eigen valuesand contribution of variability for the principal component axis

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Eigenvalue	1.680	1.551	1.084	1.071	1.01o	0.589	0.298
Variability (%)	23.997	22.153	15.488	15.296	10.402	8.414	4.251
Cumulative %	23.997	46.150	61.637	76.933	87.335	95.749	100.000

Table 3. Contribution of variables for the principal component axis

Traits	PC1	PC2	PC3	PC4	PC5
DF	7.005	5.207	33.756	23.250	9.744
PH	16.567	20.565	4.129	0.440	18.611
PNP	1.335	2.922	11.857	62.650	16.156
PL	3.284	26.187	20.738	1.012	13.621
FGNP	40.239	5.214	6.526	0.476	2.963
TSW	12.650	33.157	9.692	1.313	5.371
SPY	18.919	6.747	13.300	10.860	33.533

DFF: Days to 50% flowering, PH: Plant height , PNP: Panicle number per plant, PL: Panicle length , NFGP: Filled grain number per panicle , TW: 1000-grain weight, SPY; Single plant yield.

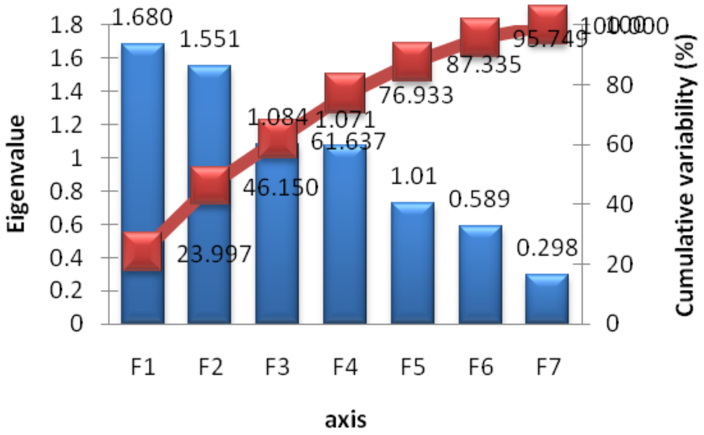

Fig. 3. Scree plot constructed using principal components which portary eigen values and cumulative variability

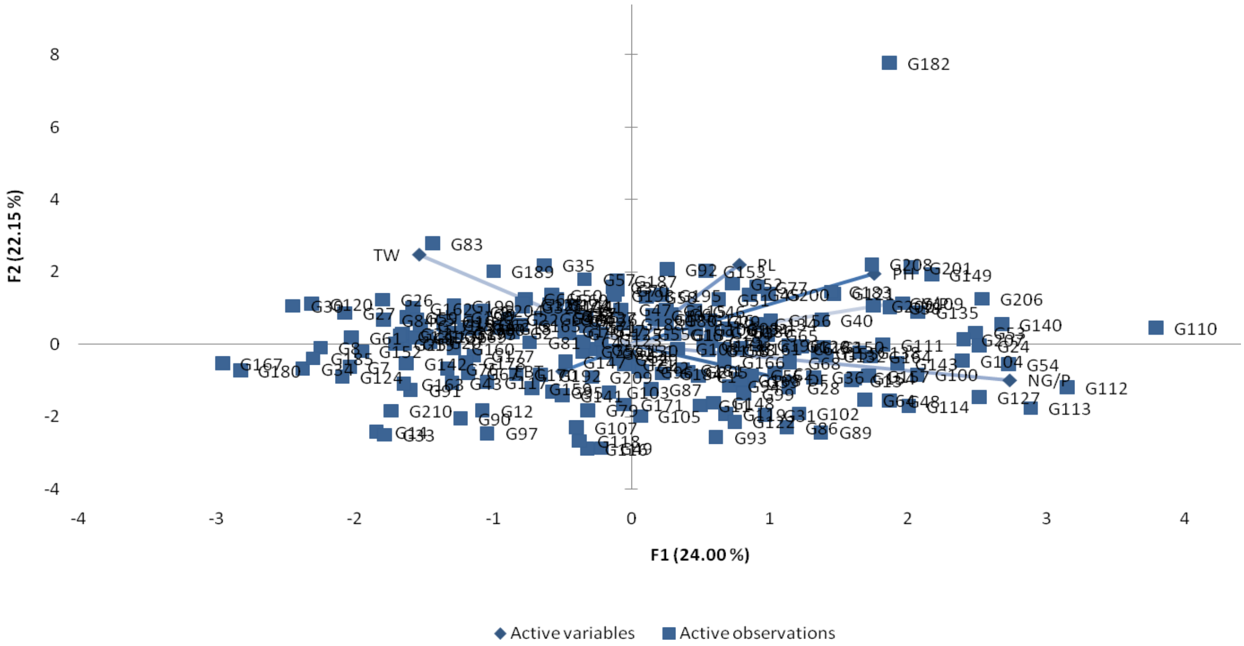

Fig. 4. Bi-plot of rice genotypes for first two principal components

genotypes for the trait single plant yield. MTU 1210, IVT MS-6157, IVT MS-6116, IVT NPT-6324, RNR 28362 are some of the poor performing genotypes for the traits under consideration as they are present in the opposite direction to the trait vector.

Agglomerative hierarchical clustering (AHC): AHC analysis was utilized to discover patterns of clustering between the rice genotypes. A total of 217 rice accessions were categorized into 5 groups using a similarity level (**Table 4 & Fig. 5**). A key conclusion drawn from AHC is that, based on composition, the variations between the accessions analyzed are still significant enough to identify them appropriately using the criteria used in this study. Cluster I had 75 genotypes in it. Accessions in Cluster I majorly contributed by parameters panicle number per plant and thousand grain weight, while cluster II had 31 genotypes. The accessions in cluster II were mostly impacted by plant height, panicle length, and thousand grain weight, while cluster III had 51 genotypes. Cluster III germplasm was mostly determined by the parameters panicle number per plant and filled grain number per panicle, while cluster IV comprised of 40 genotypes. Cluster IV accessions were primarily influenced by parameters such as panicle length, filled grain number per panicle, days to 50% flowering and seed yield per

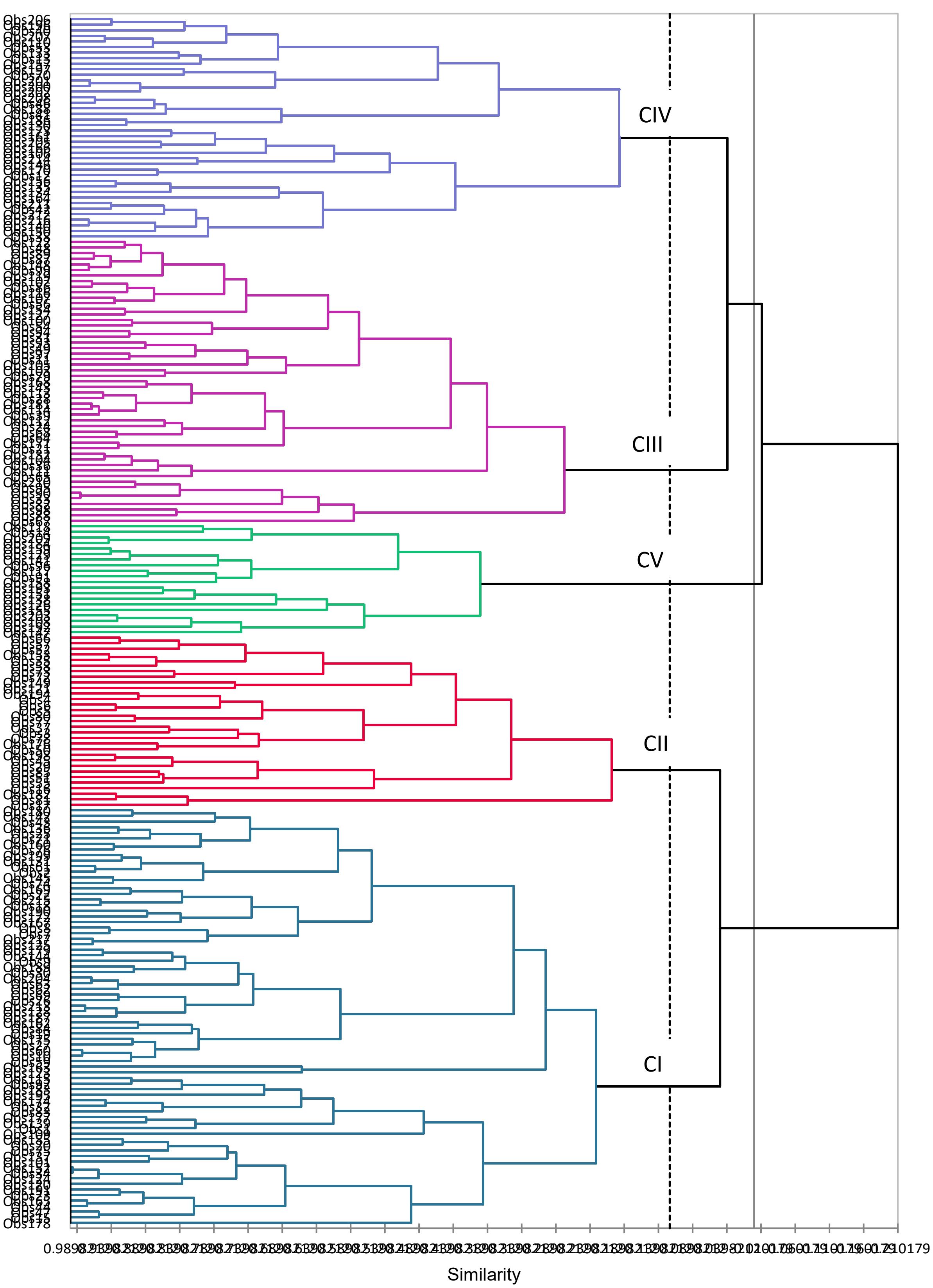

Fig. 5. Cluster analysis of 217 rice genotypes by Agglomerative hierarchical clustering (AHC) method

Table 4. Relative contribution of different traits to genetic diversity in rice

Cluster	Cluster I (75 genotypes)	Cluster II (31 genotypes)	Cluster III (51 genotypes)	Cluster IV (40 genotypes)	Cluster V (20 genotypes)
DF	-0.423	-0.138	-0.008	0.196	1.427
PH	-0.248	1.121	-0.263	0.013	-0.164
PNP	0.343	-0.525	0.372	-0.599	-0.224
PL	-0.105	0.689	-0.191	0.118	-0.423
FGNP	-0.741	-0.310	1.003	0.505	-0.308
TSW	0.825	0.359	-1.286	-0.084	-0.204
SPY	-0.122	-0.334	-0.303	1.104	-0.460
Sum of weights	75.000	31.000	51.000	40.000	20.000
Within-cluster variance	3.826	8.303	5.109	4.424	4.009

DFF: Days to 50% flowering, PH: Plant height , PNP: Panicle number per plant, PL: Panicle length , NFGP: Filled grain number per panicle , TW: 1000-grain weight, SPY; Single plant yield.

plant. Cluster V had 20 genotypes. The parameter days to 50 per cent flowering in cluster V contributed the most. In this context, it may be desirable to cross genotypes in cluster I, which has a high tiller number per plant coupled with dwarf height and high test weight, with genotypes in cluster IV to develop superior lines with a high panicle length, a high filled grain number per panicle, early maturity and a higher seed yield per plant.

Both multivariate statistical approaches revealed that the germplasm used in the study had tremendous of genetic diversity. Five of the seven principal components were significant (eigenvalue >1) and accounted for 87.35 per cent of the variance, according to PCA. The yield contributing characteristics dominated in PC1. The yield parameter was dominant in PC2. As a result, selecting germplasm with a high PC1 and PC2 score may result in increased yield and yield characteristics, by harnessing heterosis. High-performing germplasm from each cluster can be employed in a hybrid breeding programme to generate superior high yielding lines. As a result, breeders will benefit greatly from the research in terms of selecting promising parental lines from germplasm in the study.

REFERENCES

Maji, A. T. 2012. Application of principal component analysis for rice germplasm characterization and evaluation. *Journal of Plant Breeding and Crop Science*, **4**(6): 87–93. [Cross Ref]

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. *Electronic Journal of Plant Breeding*, **12**(3): .907-911. [Cross Ref]

Shoba, D., Vijayan, R., Robin, S., *et al*., 2019. Assessment of genetic diversity in aromatic rice (*Oryza sativa* L.) germplasm using PCA and cluster Analysis. *Electronic Journal of Plant Breeding*, **10**(3): 1095–1104. [Cross Ref]

Fathima, M.A., Geetha, S., Amudha, K. and Uma, D. 2021. Genetic variability, frequency distribution and association analysis in ADT (R) 48 x Kavuni derived F_2 population of rice (*Oryza sativa* L.). *Electronic Journal of Plant Breeding*, 12(3):659-666. [Cross Ref]

Federer, W.T. 1956. Augmented designs. Hawaiian Planter's Rec.55:191–208.

Federer, W.T. 1961. Augmented designs with one-way elimination of heterogeneity. Biometrics 20:540–552. [Cross Ref]

Federer, W.T. 1991. Statistic and society. Section 7.11. 2nd ed. Marcel Dekker, New York, NY.

Acquaah, G. 2012. Principles of Plant Genetics and Breeding, John Wiley & Sons, Hoboken, NJ, USA. [Cross Ref]

Kumari, B. K., Kumar, B. R., Jyothula, D. and Rao, N. M. 2021. Diversity analysis in rice breeding lines for yield and its components using principal component analysis. *Journal of Pharmacognosy and Phytochemistry*, **10**(1): 905-909. [Cross Ref]

Lakshmi, V.I., Sreedhar, M., Vanisri, S., Anantha, M.S., Rao, L.S. and Gireesh, C. 2019. Multivariate analysis and selection criteria for identification of African rice (*Oryza Glaberrima*) for genetic improvement of indica rice cultivars. *Plant Genetic Resources*, 17(6):499-505. [Cross Ref]

Mahalanobis, P. C. 1936. On the generalized distance in statistics, in *Proceedings of the National Institute of*

Science of India, vol. 2, pp. 49–55, Calcutta, India, January.

Mahendran, R., Veerabadhiran, P., Robin, S. and Raveendran, M. 2015. Principal component analysis of rice germplasm accessions under high temperature stress, *International Journal of Agricultural Science and Research*, **5**(3): 355–360.

Raza, A., Razzaq, A., Mehmood, S.S., Zou, X., Zhang, X., Lv, Y. and Xu, J. 2019. Impact of climate change on crops adaptation and strategies to tackle its outcome: A review. Plants, 8: 34. [Cross Ref]