

Morphology based multivariate analysis on rice [Oryza sativa]

Sudhir Deepak M

sudhir6788@gmail.com

Agricultural College and Research Institute, Madurai

Research Article

Keywords: Rice, D2, Clusters, eigen values

Posted Date: March 11th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9082657/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Morphology based multivariate analysis on rice [*Oryza sativa*]

Sudhir Deepak M¹ and Monika S²

¹Department of Plant Breeding and Genetics, Agricultural College and Research Institute,
Madurai.

²Department of Genetics and Plant Breeding, Anbil Dharmalingam Agricultural College and
Research Institute, Tiruchirappalli.

Corresponding Author: E-mail: sudhir6788@gmail.com

ORCID ID: 0009-0009-4624-1671

Abstract

Rice is grown in a variety of agroclimatic zones throughout the world, particularly in Asia. Several breeding programmes are undertaken to develop new varieties to respond to the rising needs of society. Since variation is a prerequisite for any breeding programme, it is exploited from the existing diversity within the crop. This study evaluated forty rice genotypes for nine yield-related attributes. An eigenvalue-based multivariate approach (PCA) was used to assess relationships among traits and to classify genotypes, while D² analysis identified the variability among them. PCA revealed nine components, of which three had eigenvalues greater than one and contributed most to the total variance. Traits such as panicle numbers per plant, tillers per plant, grains per panicle along with per plant grain yield revealed notable variability. D² analysis showed that plant height (39.10%), panicle numbers per plant (33.08%), per plant grain yield, and spikelet fertility (10.51%) contributed most to genetic divergence.

The genotypes were grouped into six clusters, with cluster II being the largest, followed by cluster I. The maximum inter-cluster distance occurred between clusters V and

VI, indicating substantial diversity. Hybrids derived from such diverse clusters are likely to exhibit substantial heterosis and offer useful segregants for future breeding programs.

Keywords: Rice, D², Clusters, eigen values

Introduction

Over the years, rice production has grown significantly, it is expected to reach a record level of over 561.6 million tonnes (FAOSTAT 2026). This is the result of continuous execution of plant breeding programmes to increase the yield and it often starts with the exploitation of variation. Since varietal development is a vibrant process, an uninterrupted flow of genetic sources is required. To exploit the variation necessary for a breeding programme, distant genotypes are selected and crossing programmes are executed. Furthermore, it is critical to comprehend the corresponding contribution along with correlation between the individual attributes and grain yield as selection guide (Singh and Choudary, 1981). The above-mentioned processes are achieved with the help of several multivariate analyses. Multivariate analyses provide an adequate way to quantify the extent of genetic variation prevailing in a population (Salem *et al.*, 2021). It is used to summarize and describe the intrinsic diversity across several rice genotypes (Kumari *et al.*, 2021). The principal approach for examining the inherent genetic diversity is Principal Component Analysis (PCA). Characters that are helpful in the separation of particular genotypes are analysed in this method. Thus, this study focused on studying the diversity among the forty genotypes and to elucidate the role of yield-contributing traits in determining grain yield. Generally, plant breeders handle a large data set, where the PCA helps in identifying the key traits contributing to the variation and helps in selection of traits. It tends to reduce the redundancy in the dataset (Maji and Shabu, 2012).

Genetic divergence is used to choose parents for a successful hybridization and breeding programme. When the genotypes are to be classified into distinct groups based on genetic or agronomic characteristics, cluster analysis is performed (Shabir *et al.*, 2013). A useful biometrical analysis for assessing genetic diversity is grouping analysis (Shrestha *et al.*, 2021). Therefore, the goal of the current investigation was to evaluate genetic diversity and determining variables affecting grain yield.

Methodology

The current experiment was investigated at Annamalai University farm. The study encompassed forty rice genotypes (Table 1).

Table 1. Genotypes evaluated in the Experiment

S. No	Genotypes
1	NDRK 11-21
2	NDRK 11-22
3	NDRK 11-24
4	CR2860-S-B-189-1-1-1
5	CR3882-7-1-6-2-2-1
6	CR3879-3-1-6-1-3-1
7	CR 2839-1-S-10-B2-B-43-2B
8	CR3884-244-8-5-11-1-4
9	CR3884-244-8-5-6-1-1
10	CR2838-1-S-2B-9-1
11	CR3903-161-1-3-2
12	CR3899-134-2-6-4
13	CR3881-4-1-6-3-4-1

S. No	Genotypes
14	CR3909-192-1-7-1
15	TR 09027
16	TR13069
17	TR 09030
18	PAU 3835-12-1-1-1
19	PAU 71114-3480-1-1-1-0
20	PAU 3835-36-6-3-3-4
21	CSR 7-1
22	CSR-2748-4441-22
23	CSR-2748-4441-111
24	CSRC(S)47-7-B-B-1-1
25	CSR-2748-4441-104
26	CSR-2748-4441-193
27	CSR2711-171
28	CSR-2748-4441-133
29	CSR-2748-197
30	PUSA 44
31	KS-12
32	KR 15003
33	KR15006
34	KR 15014
35	KR 15005
36	RAU 1396-14

S. No	Genotypes
37	JK-95
38	CSAR 1610
39	RP-320-4-3-2-1
40	RP-5683-101-30-2-3-1

The seedlings of genotypes were grown in raised beds before being shifted to the research field. seedlings were transplanted in a row of length 3m with distance between the rows as 30 cm and between plants as 10 cm. Research design adopted as RBD (Randomized Block Design) consists of three replicates. Recommended agricultural practices were performed for good crop establishment, crop production and protection.

Each genotype were evaluated for various yield parameters such as days to 50% flowering, plant height, tillers per plant, panicle number plant, length of panicle, spikelet fertility, number of grains per panicle, test weight, and per plant grain yield. The mean values were computed based on observations from random sample of five plants per genotype. An eigenvalue-based multivariate approach (PCA) was executed to find out proportionate contribution of each character to total variance (Singh and Chaudhury, 1981). PCA analyses were performed using the "factoextra" (Kassambara and Mundt, 2020) and "FactoMineR" (Le et al., 2008) modules of R software V 4.3.0. The diversity of the genotypes was calculated using Mahalanobis D² analysis (Mahalanobis, 1963), and The genotypes were categorized into different clusters using Tocher approach suggested by Rao (1952). The D² analysis had been conducted using the INDOSTAT (version 8.1)software.

Result and Discussion

Eigenvalue-based multivariate approach / PCA: PCA enables the determination of the relative importance of various traits in explaining population-level variation. It lowers the dimension of set of multivariate data to a finite number of fundamental axes, each yielding an

Eigenvector. "Eigenvectors" represent the relative contributions of each original variable to each primary component, whereas "eigenvalues" represent the significance of each component to overall variance (Jolliffe, 2002). The foremost advantage in Eigenvalue-based multivariate approach is it uses new components generated from linear combinations of variables to represent the bulk of variation in the dataset allowing each dimension's significance to be assessed for the purpose of describing a data set's variability in more comprehensible and visually appealing dimensions (Burman 2021). Table 2 depicts the eigen values, percentages of contribution, and cumulative percent of variation. Among nine yield attributes analysed, nine Principal Components were identified. Out of these, only three principal components were seen to have high contribution to variation (Table 3), since they had eigen values >1 (2.21, 2.00 and 1.42 respectively). Lakshmi *et al.*, (2022); Sharma *et al.*, (2014) reported three principal components in selected rice varieties. Other traits gave negligible contribution to variation, which is evident from the eigen values. These three components together represented 62.58 % of the variation, among which the first component gave 24.58%, second one 22.27% and the third component contributed 15.76% to the variation. This results was in accordance with Thakur (2023).

The maximum positive loading was noticed by the traits panicle number per plant (0.909), tillers per plant (0.907) in PC1. Attributes such as days to 50% flowering, spikelet fertility and number of grains per panicle showed negative loadings in PC1. In PC2, per plant grain yield showed the maximum positive loading (0.688). In PC3, highest positive loading was noticed in the trait spikelet fertility (0.795). Other components are neglected due to negative loading. Figure 4 flashed the traits in the quadrant I such as panicle number per plant, tillers per plant, per plant grain yield and 1000 grain weight has high PC1 and PC2 values and the genotypes in quadrant I has high values for these attributes. Similar findings were obtained in by (Gunasekaran 2023; Akilan 2023) rice. Figure 4 depicts a biplot, with

the vector length representing the contribution of each trait to the total extent of divergence. The magnitude of contribution increases with vector length (Lakshmi et al., 2022). The panicle numbers per plant and tillers per plant had the longest vectors and contributed most to divergence.

The angle formed by the trait vectors indicates which way the traits are correlated. A positive and negative correlation is represented as acute angle ($<90^\circ$), and obtuse angle ($>90^\circ$). A right angle (90°) between the vectors shows no correlation. Out of the nine traits studied, a positive association between the panicle number per plant, panicle length, plant height and tillers per plant. The variables spikelet fertility, days to 50% blooming, 1000 grain weight, and grain yield per plant all showed a positive correlation. Negative correlation was found between number of panicles / plant, tillers/ plant, per plant grain yield, 1000 grain weight, days to 50% flowering, spikelet fertility and number of grains/ panicle. Spikelet fertility, days to 50% flowering, 1000 grain weight, grain yield per plant grain yield showed negative correlation for panicle length along with plant height.

Genotypes aligned with the trait vector within a given quadrant demonstrated enhanced expression for the respective trait. In this regard, genotype 4, CR2860-S-B-189-1-1-1 is a good yielding genotype. The genotypes, CR3879-3-1-6-1-3-1 and CSR 7-1 have high spikelet fertility compared to other genotypes. The genotype, CR 2839-1-S-10-B2-B-43-2B is the best performer for tillers per plant.

Diversity analysis: ANOVA encountered a high MSS for genotypes, denoting substantial phenotypic divergence among the evaluated genotypes (Sudhir et al., 2023). A measure of group distance based on multiple characters was carried out by D^2 analysis. On assessing The partitioning of divergence among traits (Table 4) demonstrated that plant height had the greatest influence (39.10%) followed by grain numbers per panicle (33.08%). Spikelet fertility and per plant grain yield showed 10.51% contribution each followed by panicle

length (6.41%). tillers per plant and 1000 grain weight showed negligible contribution whereas days to 50% flowering and panicle number per plant showed no contribution to divergence, indicating that the genotypes did not show much variation for these traits.

Tocher's approach was used to divide the forty genotypes into six distinct clusters (Table 6), three of which are poly genotypic. Cluster III covered the majority of genotypes, with twenty-two, followed by clusters I and III, which have ten and five, respectively. Three clusters IV, V and VI are solitary clusters. From the cluster means of different traits, it can be concluded that the genotypes took minimum days for blooming and hence can be utilized in breeding program aiming to develop early maturing transgressive segregants. Cluster II has the genotypes with low mean plant height (Table 7) which can contribute to the development of dwarf or semi dwarf, non- lodging genotypes. Cluster V has the genotype, CR2860-S-B-189-1-1-1 which has high values for yield and its related attributes. This makes it suitable for the development of high yielding cultivars. Intra cluster distance (Table 5) is high for the cluster III, followed by cluster II and I. Higher intra cluster distance indicates that the genotypes in that cluster are diverse, compared to other clusters. The absence of intra-cluster distance in clusters IV, V, and VI reflects the presence of solitary genotypes possessing distinct genetic profiles. Kumari *et al.*, (2021) grouped 121 rice genotypes into 12 different clusters; Tiwari *et al.*, (2022) classified 90 genotypes into 10 groups; Sinha and Mishra (2013), where 34 landraces were grouped into five clusters. Discrete and solitary clusters result from intense natural and human selection for a variety of different and adaptable gene complexes or because their ancestors' geographic barriers impeded the gene flow (Tiwari *et al.*, 2022). Distant clusters are those which have high inter cluster distance between them. Genotypes chosen from such distant clusters can be employed in breeding programs to obtain useful transgressive segregants. In this regards, inter-cluster distance (Table 5) is high between cluster V (CR2860-S-B-189-1-1-1) and cluster VI (CR3909-192-1-7-1), which

indicates their ability to contribute to heterotic combinations. Also, crosses between the cluster III and VI can yield transgressive segregants in a wide range of frequency.

Conclusion

The study found high genetic heterogeneity across the forty rice genotypes tested for yield and other attributes. PCA identified key traits contributing significantly to overall variation, D² analysis efficiently separated genotypes into various clusters based on genetic divergence. attributes such as n grain numbers/ panicle, Plant height, spikelet fertility along with per plant grain yield all contributed significantly to genotype diversity. Pronounced divergence between clusters V and VI may be effectively harnessed for developing superior recombinants through hybridization. Crossing genotypes from genetically distant clusters is expected to result in greater heterosis and the formation of superior segregants. In order to improve yield potential and promote rice development programs meant to satisfy rising food demands, the identified diverse genotypes might thus be efficiently used in future breeding methods.

Table 2. Eigenvalues and Proportion of Total Variance Explained by Principal Components

Components	Eigenvalues	Variance explained (%)	Cumulative variance explained (%)
PC1	2.21	24.55	24.55
PC2	2.00	22.27	46.82
PC3	1.42	15.76	62.58
PC4	0.97	10.73	73.31
PC5	0.89	9.91	83.22
PC6	0.54	6.03	89.25
PC7	0.48	5.33	94.58
PC8	0.43	4.74	99.32
PC9	0.06	0.68	100.00

Table 3. PCA for nine traits for forty rice genotypes (Factor Loadings of traits on Principal Components)

Parameters	PC 1	PC 2	PC 3
DFF	-0.066	0.584	-0.409
PH	0.486	-0.449	0.354
TP	0.907	0.242	0.091
PNP	0.909	0.230	0.048
PL	0.497	-0.526	-0.401
NGP	-0.012	-0.523	0.525

SF	-0.199	0.314	0.795
TGW	0.060	0.476	0.112
PGY	0.172	0.689	0.183

DDF – days to 50% flowering, **PH**- Plant height, **TP**- Tillers per Plant, **PNP**- Panicles

Number per Plant, **PL**- Panicle Length, **NGP**- Number of Grains per Panicle, **SF**- **Spikelet**

Fertility, **TGW**- **Thousand grain Weight**, **PGY**- Per Plant Grain Yield

Table 4. Partitioning of Divergence Among Different Traits

S. No	Characters	Contribution (percent)
1	DDF	0.00
2	PH	39.10
3	TP	0.26
4	PNP	0.00
5	PL	6.41
6	NGP	33.08
7	SF	10.51
8	TGW	0.13
9	PGY	10.51

Table 5. Average intercluster and intracluster distances

Cluster Number	I	II	III	IV	V	VI

I	174.24 (13.20)	486.64 (22.06)	388.09 (19.70)	347.44 (18.64)	406.82 (20.17)	1075.18 (32.79)
II		252.49 (15.89)	757.35 (27.52)	781.20 (27.95)	1063.41 (32.61)	644.65 (25.39)
III			320.769 (17.91)	849.72 (29.15)	696.97 (26.40)	1488.41 (38.58)
IV				0.00 (0.00)	319.33 (17.87)	945.57 (30.75)
V					0.00 (0.00)	1755.61 (41.90)
VI						0.00 (0.00)

Table 6. Composition of clusters containing different genotypes calculated based on Cluster-based classification

Clusters	Number of genotypes	Name of genotypes
I	10	CSRC(S)47-7-B-B-1-1, CR2838-1-S-2B-9-1, PAU 3835-36-6-3-3-4, TR 09030, NDRK 11-24, CSR-2748-197, KR 15005, CSR2711-171, CSR-2748-4441-133, RP-320-4-3-2-1
II	22	NDRK 11-21, KS-12, JK-95, RAU 1396-14, CSR-2748-4441-104, CSR-2748-4441-111, NDRK 11-22, CSR-2748-4441-193, PAU 3835-12-1-1-1, KR 15003, CR3884-244-8-5-11-1-4, KR15006, PAU 71114-3480-1-1-1-0, TR13069, CSR-2748-4441-22, PUSA 44, CR3903-161-1-3-2, CSAR 1610, CSR 7-1, CR3879-3-1-6-1-3-1, TR 09027, RP-5683-101-30-2-3-
III	5	CR3884-244-8-5-6-1-1, KR 15014, CR3899-134-2-6-4, CR3881-4-1-6-3-4-1, CR3882-7-1-6-2-2-1
IV	1	CR 2839-1-S-10-B2-B-43-2B-1
V	1	CR2860-S-B-189-1-1-1
VI	1	CR3909-192-1-7-1

Table 7. Cluster-based mean estimates of traits in forty rice genotypes

Characters	DFF	PH	NTP	NPP	PL	NG/Pa	SF	TGW	GY/P
Clusters									
I	8.50	116.40	18.37	13.40	26.54	130.63	74.84	20.12	26.57

II	106.32	84.35	17.64	13.32	24.61	96.72	72.04	21.57	29.84
III	90.93	104.40	15.33	10.53	26.60	177.33	67.15	19.40	25.87
IV	100.67	136.00	23.33	18.33	32.27	99.00	52.67	20.40	26.48
V	126.00	151.67	21.00	16.00	25.10	146.83	61.267	20.20	34.89
VI	106.67	75.33	18.00	14.00	27.27	48.33	19.53	15.46	17.54
General mean	103.17	111.35	18.94	14.26	27.06	116.47	57.91	19.52	26.86

Figure1. Scree plot illustrating the principal components

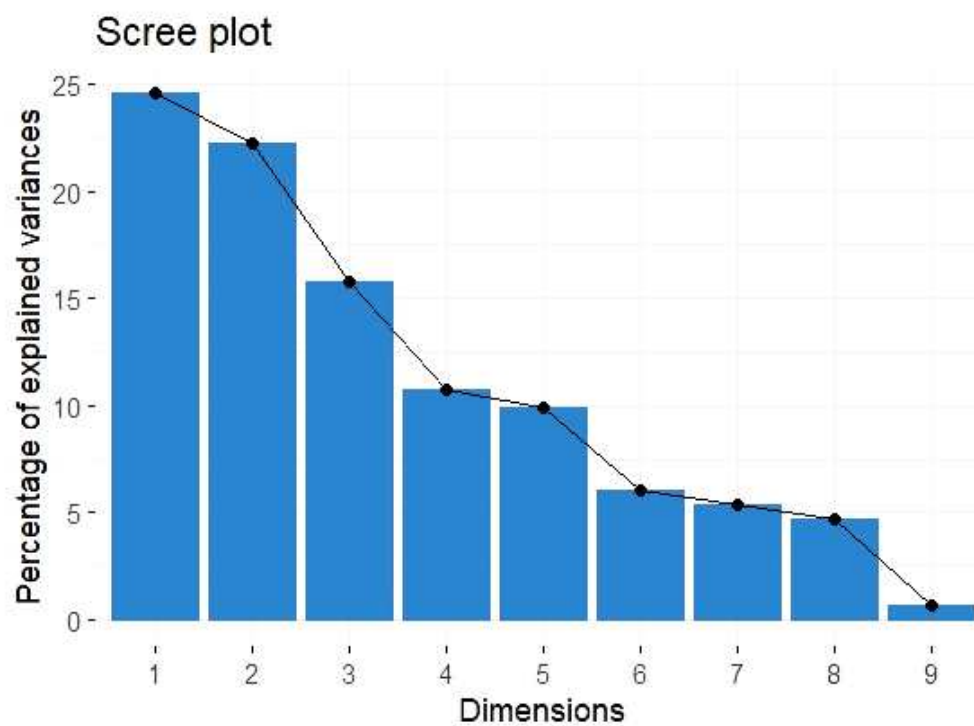


Figure 2. Two-dimensional biplot of Principal Component 1 and Principal Component 2 depicting the association among yield-related traits and the evaluated genotypes.

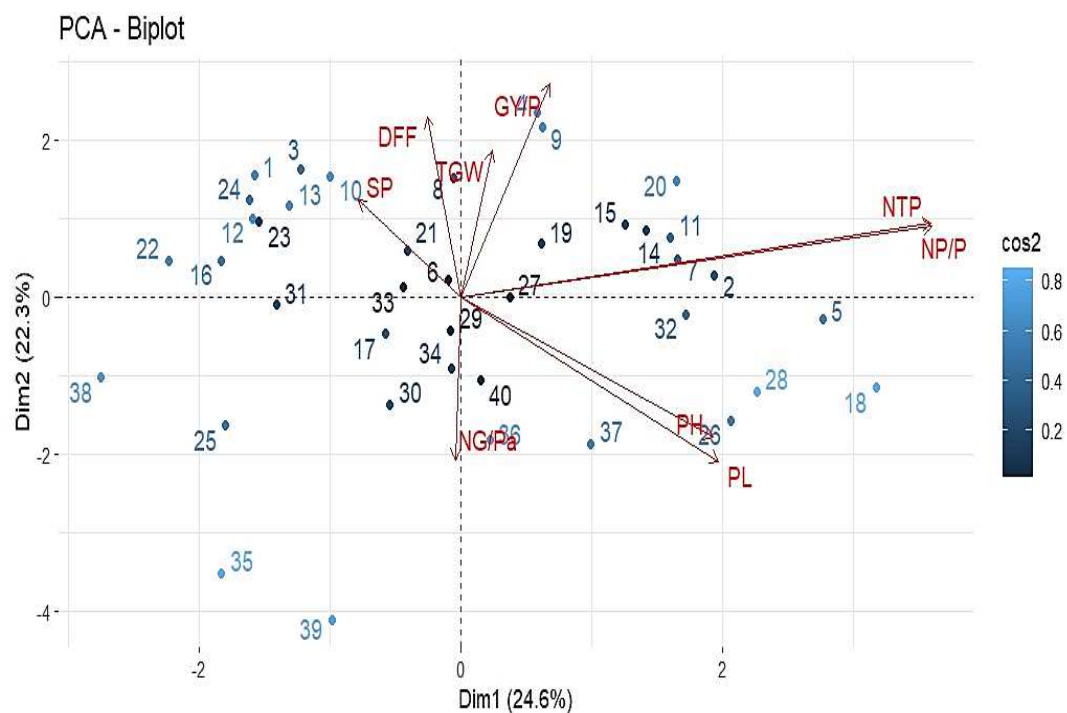


Fig 3. Percentage contribution of different traits to divergence

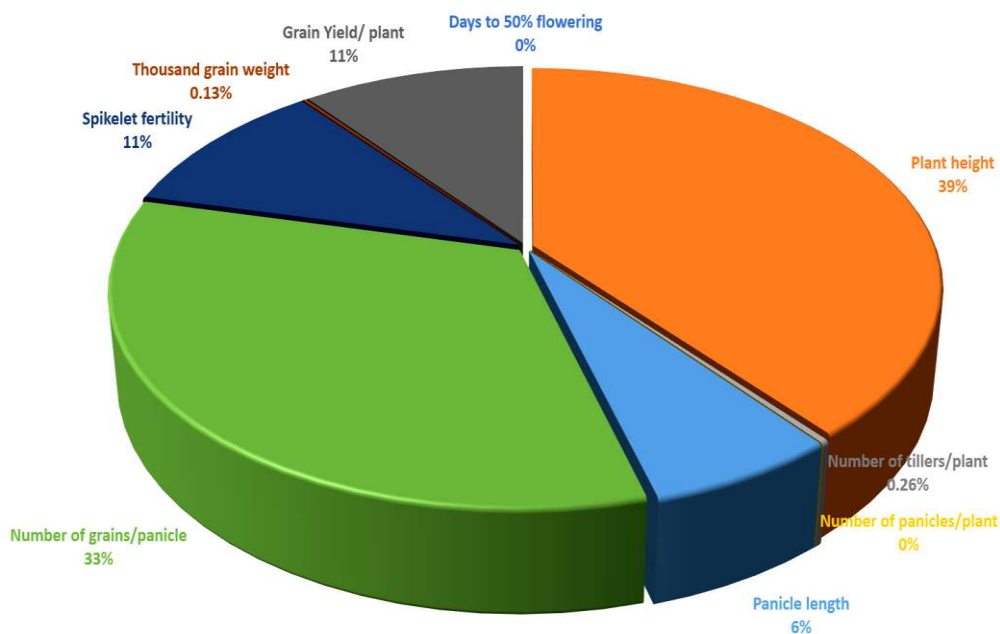
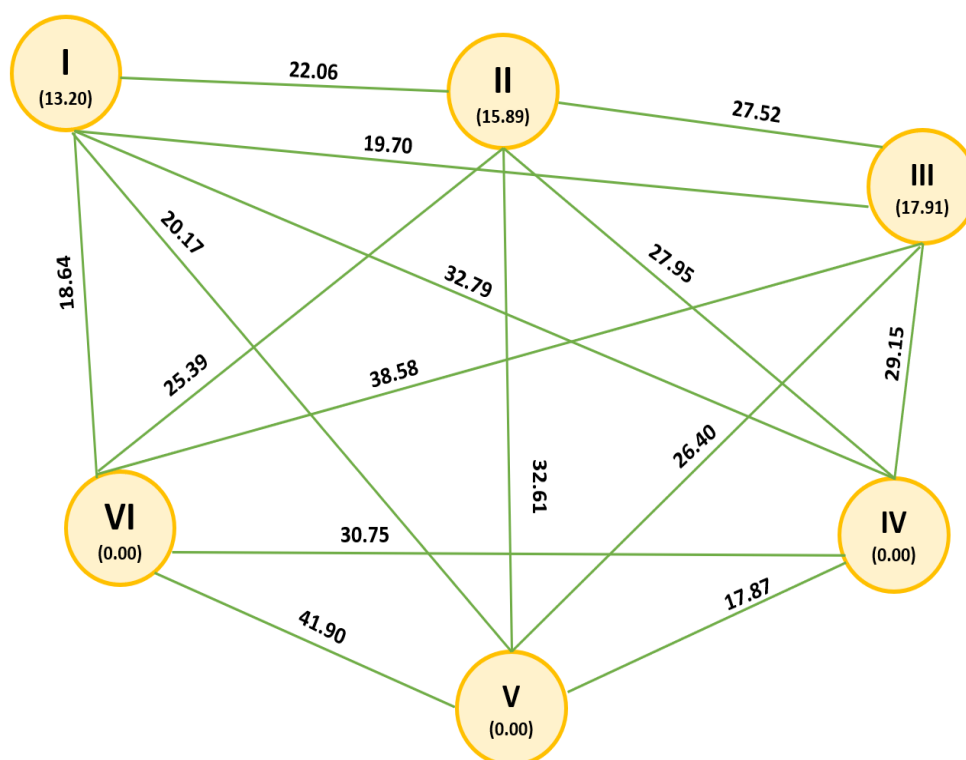


Fig 4. Cluster diagram depicting the classification of genotypes



Acknowledgement

The authors would like to sincerely thank everyone who helped make this study possible. We would especially want to thank our colleagues for their insightful conversations and helpful criticism during the course of the study. We also thank our affiliated institution for its assistance, which enabled this effort.

References

1. Akilan, M., P. Jeyaprakash, M. Shanmuganathan, S. Meena, V. Rajanbabu, and C. Vanniarajan. "Comparative patterns of principal component and cluster analysis under sodicity and normal soil conditions in rice (*Oryza sativa* L.)." *Electronic Journal of Plant Breeding* 14, no. 3 (2023): 893-901.
2. Burman, Maumita, Sunil Kumar Nair, and Arvind Kumar Sarawgi. "Principal component analysis for yield and its attributing traits in aromatic landraces of rice (*Oryza sativa*

- L.)." *International Journal of Bio-resource and Stress Management* 12, no. 4 (2021): 303-308.
3. Gunasekaran, Ariharasutharsan, Geetha Seshadri, Saraswathi Ramasamy, Raveendran Muthurajan, and Krishna Surendar Karuppasamy. "Identification of newer stable genetic sources for high grain number per panicle and understanding the gene action for important panicle traits in rice." *Plants* 12, no. 2 (2023): 250.
 4. <https://www.fao.org/>
 5. Jolliffe, Ian. "Principal component analysis." *Encyclopedia of statistics in behavioral science* (2005).
 6. Kassambara, Alboukadel, and Fabian Mundt. "Factoextra: extract and visualize the results of multivariate data analyses." *CRAN: Contributed Packages* (2016).
 7. Kumari, B. Kusuma, BNVS Ravi Kumar, D. Jyothula, and N. Mallikarjuna Rao. "Diversity analysis in rice breeding lines for yield and its components using principal component analysis." *Journal of Pharmacognosy and Phytochemistry* 10, no. 1 (2021): 905-909.
 8. Lakshmi, M., M. Shanmuganathan, P. Jeyaprakash, and T. Ramesh. "Genetic variability and diversity analysis in selected rice (*Oryza sativa* L.) varieties." *Electronic Journal of Plant Breeding* 13, no. 3 (2022): 959-966.
 9. Lê, Sébastien, Julie Josse, and François Husson. "FactoMineR: an R package for multivariate analysis." *Journal of statistical software* 25 (2008): 1-18.
 10. Mahalanobis, Prasanta Chandra. "On the generalized distance in statistics." *Sankhyā: The Indian Journal of Statistics, Series A* (2008-) 80 (2018): S1-S7.

11. Maji, A. T., and A. A. Shaibu. "Application of principal component analysis for rice germplasm characterization and evaluation." *Journal of Plant Breeding and Crop Science* 4, no. 6 (2012): 87-93.
12. Rao, C. Radhakrishna. "Advanced statistical methods in biometric research." (1952).
13. Salem, K. F., Maysoun M. Saleh, Laila Aldahak, and Abdelmoaty B. Elabd. "Assessment phenotypic diversity of rice (*Oryza sativa* L.) genotypes by multivariate analysis." *Journal of Aridland Agriculture* 7 (2021): 52-59.
14. Shabir, Ghulam, Shahzad Amir Naveed, and Muhammad Arif. "Estimation of phenotypic variability and mutual association of yield and its components in rice (*Oryza sativa* L.) germplasm using multivariate analysis." *Journal of Agricultural Research (JAR)*. 51, no. 4 (2013): 361-378.
15. Sharma, S. K., Jogendra Singh, Mahesh S. Chauhan, and S. L. Krishnamurthy. "Multivariate analysis of phenotypic diversity of rice (*Oryza sativa*) germplasm in North-West India." *Indian Journal of Agricultural Sciences* 84, no. 2 (2014): 295-299.
16. Shrestha, Jiban, Sudeep Subedi, Ujjawal Kumar Singh Kushwaha, and Bidhya Maharjan. "Evaluation of growth and yield traits in rice genotypes using multivariate analysis." *Heliyon* 7, no. 9 (2021).
17. Singh, Ram Kathin, and Bajrang D. Chaudhary. "Biometrical methods in quantitative genetic analysis." (1981): x+-304.
18. Sinha, A. K., and P. K. Mishra. "Morphology based multivariate analysis of phenotypic diversity of landraces of rice (*Oryza sativa* L.) of Bankura district of West Bengal." *Journal of Crop and Weed* 9, no. 2 (2013): 115-121.

19. Thakur, Kangkana, and M. K. Sarma. "Genetic diversity and principal component analysis in cultivated rice (*Oryza sativa*) varieties of Assam." (2023): 145-150..