

Research Note

Principal component analysis (PCA) as a genetic diversity tool to understand the variation of rice mutant culture

S. Palaniyappan¹, P. Arunachalam^{1*}, S. Banumathy¹ and S. Muthuramu²

¹Department of Plant Breeding and Genetics, Agricultural College and Research Institute, Tamil Nadu Agricultural University, Madurai – 625 104, Tamil Nadu.

²Agricultural Research Station, Tamil Nadu Agricultural University, Paramakudi – 623 707, Tamil Nadu.

*E-Mail: arunachalp@gmail.com

Abstract

The current research experiment was carried out to analyze the genetic variations among rice mutants and homozygous lines from cross derivatives by principal component analysis (PCA). The PCA results had shown the first five PCs accounted 83.20 per cent of total genetic variation. The thirty six rice genotypes were formed into four groups. Among the variables studied, hundred grain weight (0.907), grain length (0.837), grain width (0.799) and plant height on 50th day (0.617) positively contributed to 29.15 % variation of first PC. Whereas, days to maturity (0.708), days to fifty percent flowering (0.643) and flag leaf breadth (0.609) contributed mostly for 21.60 percent variation of second PC. More precisely, the grouping revealed that the EMS mutants of Anna (R) 4 were potentially mutated not only for the grain type and also for yield related traits such as flag leaf length and breadth, panicle length, plant height. The genotypes grouped 1 and 4 had potential to select higher grain yield with improved grain quality suitable for rainfed condition.

Keyword: Rice, mutants, PCA, variation.

Rice (*Oryza sativa*. L) construed as the primary food crop for over 50% of the global population, with its consumption providing up to 80% of their daily caloric intake (Rasheed *et al.*, 2020; Sahebi *et al.*, 2018). Additionally, rice cultivation serves as a fundamental practice and a significant source of income in various regions worldwide (Das *et al.*, 2017). The total rice production for the 2021-22 period reached a record high of 130.29 million tonnes. This figure represented an increase of 13.85 million tonnes compared to the average production of the last five years, which stood at 116.44 million tonnes. (Annual Report, 2022). Notably, the rice crop had contributed to around 30% of the total caloric requirements for the country's population (Ratnam *et al.*, 2022). While the human population continues to grow, crop production is struggling to keep pace. To address the increasing demand for food, it is crucial to focus on developing high-yielding varieties. The effectiveness of any breeding program depends on the careful

selection of diverse parent plants. Therefore, assessing diversity within the population becomes imperative (Kumari *et al.*, 2021)

India's rich diversity and variability offer a valuable resource for expanding the genetic base of rice. Tamil Nadu, recognized for its extensive history of rice cultivation, boasts a complex ecosystem and a diverse genetic pool. Landraces are crucial as they serve as valuable source, containing a diverse array of genes that contribute to adaptation. The inclusion of landraces holds the potential to ensure a broad genetic base (Mondal *et al.*, 2016).

Multivariate techniques are widely employed to summarize and describe the inherent variability among diverse rice genotypes. Key multivariate statistical tools encompass Principal Component Analysis (PCA), Cluster Analysis, and Discriminate Analysis

(Oyelola, 2004). Principal Component Analysis (PCA) stands out as a primary tool for examining the inherent genetic variation and it is mainly estimated from correlation or covariance matrix (Sheela *et al.*, 2020). The principal component technique, known for its orthogonal transformation of original variables, is widely used in genetic variability studies (Mani Lohani *et al.*, 2012). To assess the variability of a crop under study, it is essential to first identify the significant or principal components responsible for this variability. This analytical approach aids in pinpointing specific characteristics that facilitate the differentiation of selected genotypes based on similarities in one or more traits, enabling the classification of genotypes into distinct groups (Ariyo, 1987, Nair *et al.*, 1998 and Ravi Kumar *et al.*, 2015,). According to Chandraker *et al.* (2024) PCA is the technique used to identify plant traits that distinguish between selected genotypes. It is often used to classify a population into distinct groups based on similarities in one or more characteristics, thereby aiding in the selection of parents for hybridization. The analysis of genetic diversity is instrumental for breeders as it guides the selection of divergent parents, ultimately contributing to the enhancement of various traits. Principal Component Analysis (PCA) has been a commonly employed method for studying genetic diversity in various crops, including rice (Bharadwaj *et al.*, 2001; Francis and Packiaraj, 2020; Shanmugam *et al.*, 2023) and wheat (Hailu *et al.*, 2006).

The current study was mainly focused on the development of short slender grain mutants from the long slender drought tolerant and short duration rice variety of Anna (R) 4. Further more the suitable mutants identification and the traits contribution assessment was studied by grouping of genotypes based on principal component analysis for further crop improvement.

The current research experiment was conducted during *kharif* 2019 at Agricultural College and Research Institute, Madurai. The experiment includes the three different population including mutants from Anna (R) 4 rice variety and homozygous lines from cross derivatives. The Anna (R) 4 is a commercial drought tolerant rice variety and long slender in nature. To increase the farmers preference and market value, there was a need for identification of drought tolerant short slender grain type. So this experiment was carried out to development and grouping of rice mutants lines based on principal component analysis. The homozygous mutant families includes 12 lines from EMS mutants in M_7 generation, 9 lines from gamma irradiation and 6 lines from electron beam irradiations in M_4 generation. The 7 homozygous lines were derived from MDU 6 X *Jaldi Dhan* 6 and one homozygous line from TKM 6 X *Jaldi Dhan* 6 in F_5 generation. All the genotypes were raised in three meter rows of Randomized Block Design (RBD) with the spacing of 25 x 15 cm. A total of 17 quantitative traits namely Plant height 50th day (cm), Days to fifty percent flowering (days), SPAD value (%), Plant

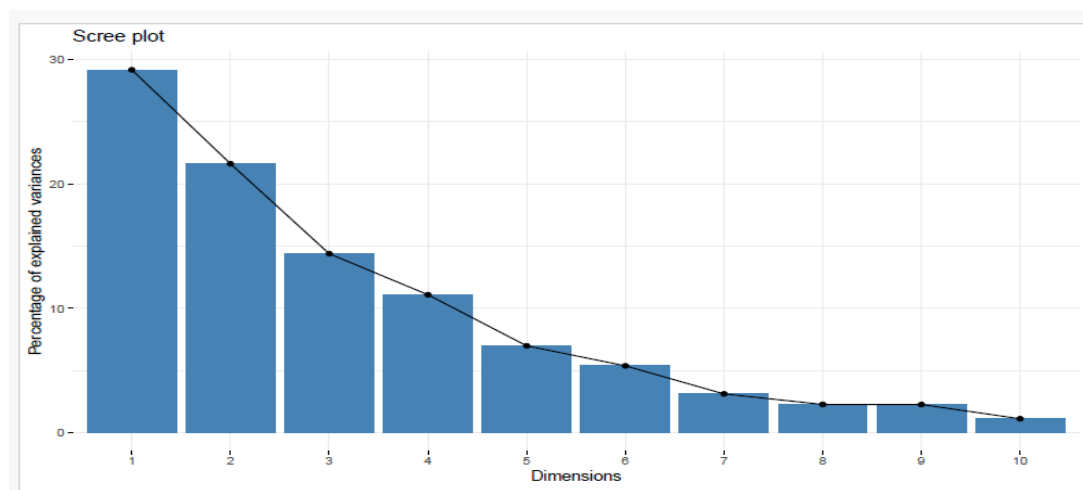
height (cm), Panicle length (cm), Flag leaf length (cm), Flag leaf breadth (cm), Total number of tillers (numbers), Productive tillers (numbers) , Days to maturity (days), Panicle weight (grams), Number of grains per panicle (numbers) , Filled grains (numbers), Grain length (cm), Grain width (cm), 100 grain weight (grams), Single plant yield (grams) were observed from five randomly selected plants for each genotype in two replications. The GL and GW were measured in centimetre by using ten hulled rice. The SPY was measured in gram by taking entire grain weight from a single plant. The Principal Component Analysis (PCA) was analysed by using “factoextra” and “FactoMineR” packages of R software.

Principal Component Analysis is an important technique for identifying the plant traits that distinguish different promising genotypes (Sheela *et al.*, 2020). PCA is the statistical analysis which gives relative importance and independent contribution of particular component to the total variance. It also provides the information about independent effect of particular trait to the total variance and the degree of contribution of each variable to the total variance is determined by coefficient vectors of that particular component. The resulted eigen values, percent variance, cumulative variance and coefficient vectors of 17 variables are given in **Table 1**. According to Kaiser (1960), principal component with eigen value greater than 1 accounts more variance contributed by every individual variable associated with that specific component. The scree plot method supports the eigen value approach. The scree plot was obtained between eigen values of different principal components (**Fig. 1**) and it explains the variation associated with each components. In this scree plot shows the first two components had greater than 3 eigen values and explains about 50.75 per cent of cumulative variance. So, while the eigen value of components reduces progressively the corresponding per cent variance of each components were reduced and it plotted as steep curve. After six components the curve bend and tends to horizontal line, revealed the contribution of maximum variability by first five components and the traits associated with these components contributed maximum variation towards total variance. Pathak *et al.* (2018) also explained the first two PCs occupied 44.65% of total variation and 75.2% of variation was accompanying within first six PCs was resulted from Tiruneh *et al.* (2019) study in rice. Shanmugam *et al.* (2023) reported 74.41 percent cumulative variation contributed by first five PCs in sixty south Indian land races of rice.

Variables from first two components solely contributed 50.80% of total variation (**Fig. 2**) and it also gives positive and negative association of different variables contributed to the variability. The biplot also explain the contribution of different variables on different principal component. The loadings are ranged from -1 to +1. If the loadings are close to -1 or +1 indicate very strong influence of variable towards that component whereas those variables close to

Table 1. Principal component analysis and vector loadings of 17 traits in rice genotypes

Eigen value/PCs	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Eigenvalue	4.96	3.67	2.44	1.88	1.19	0.91	0.53
Percentage of variance	29.15	21.60	14.38	11.09	6.98	5.37	3.12
Cumulative percentage	29.15	50.75	65.13	76.22	83.20	88.58	91.70
Eigen vectors							
Plant height 50 th day	0.617	0.323	0.414	-0.329	-0.030	0.181	-0.130
Days to fifty percent flowering	-0.196	0.643	-0.651	-0.197	0.139	0.175	0.070
SPAD value	-0.389	-0.190	0.163	0.598	0.008	0.508	0.367
Plant height (PH)	-0.016	0.576	0.358	-0.328	0.053	0.607	-0.126
Panicle length (PL)	0.490	0.423	0.612	0.024	0.203	-0.144	0.133
Flag leaf length (FL)	0.383	0.416	0.555	-0.442	0.081	-0.182	0.234
Flag leaf breadth (FB)	0.377	0.609	-0.464	0.261	0.180	-0.085	0.188
Total number of tillers (TT)	0.234	-0.790	-0.051	-0.143	0.500	0.137	-0.071
Productive tillers (PT)	0.281	-0.752	-0.075	-0.106	0.532	0.158	-0.046
Days to maturity (DM)	-0.235	0.708	-0.575	-0.108	0.179	0.130	-0.070
Panicle weight (PW)	0.078	0.506	0.084	0.601	0.380	-0.162	-0.384
Number of grains per panicle (NGPP)	-0.715	0.227	0.394	0.373	-0.008	0.027	-0.081
Filled grains (FG)	-0.841	0.014	0.426	0.247	-0.054	-0.014	-0.100
Grain length (GL)	0.837	0.074	0.169	0.344	-0.046	0.216	-0.106
Grain width (GW)	0.799	-0.013	-0.148	0.405	-0.110	0.024	0.119
100 grain weight (100 GW)	0.907	0.043	-0.011	0.348	-0.013	-0.014	0.022
Single plant yield (SPY)	-0.529	0.240	0.252	0.023	0.598	-0.165	0.220

**Fig. 1. Scree plot representing principal components and their variations**

0 designate feeble influence on that component. Among the 17 quantitative traits, yield attributing traits had higher coefficient vector values (**Table 1**). Among this, hundred grain weight (0.907), grain length (0.837), grain width (0.799) and plant height on 50th day (0.617) positively contributed to the first component. However, the number of filled grains (-0.841) and number of grains per panicle (-0.715) had higher contribution with negative magnitude.

It indicates that these variables are negatively correlated with the PC1. Nachimuthu *et al.* (2014) also accepted the yield attributing traits like grain length, breadth and filled grains contributed towards first two components. The higher and positive coefficient values of variables like days to maturity (0.708), days to fifty per cent flowering (0.643), flag leaf breadth (0.649) and plant height (0.576) were contributed to second component and it

exploited 21.60% of total variation. The 14.4 percent of total variability of PC3 was greatly contributed by panicle length (0.612) and flag leaf length (0.555) with positive coefficient vector. So this PC showed the significance for the yield related traits and it may play a major role in selection of genotypes based on yield contributing variables. Similarly, the PC4 had contributed by panicle weight (0.601) and single plant yield in PC5 for greater variations. The greater the magnitude of the coefficients, whether positive or negative, the more effectively they will differentiate between accessions (Chandraker *et al.*, 2024). So, the variables with greater coefficient values identified as most discriminating traits contributing to the existing variability (Salunkhe *et al.*, 2023)

Based on first two principal components of 17 quantitative traits was formed (Fig.2) which explore 50.8% of variance, which groups the 36 genotypes into four different groups. The genotypes are interestingly grouped in different quarters of the biplot. It showed the irradiated mutants of Anna (R) 4 in M_4 generation which falls under group 1 in first quarter except M_4 -10. Most of the genotypes associated with first group were derived from 100 Gy gamma rays and electron beam except M_4 -6, M_4 -7 and M_4 -8 which were derived from 150 Gy irradiation and M_4 -13, M_4 -14 and M_4 -15 was derived from 150 Gy electron beam irradiation. The genotypes were presented in the third group plotted over second and third quarter was EMS mutant lines. It revealed that the EMS mutant lines not only alters the grain type and also create variations in some other traits like grain yield, seed weight, days to maturity. So it was grouped separately from Anna (R) 4 rice. But gamma ray mutants showed variation only

in grain type ranged from 6.28 to 6.98 mm with a mean value of 6.49 mm for grain length compared to Anna (R) 4 (7.10 mm) check variety. Arunachalam *et al.* (2022) conducted a comparative analysis of electron beam and gamma ray-induced mutations to improve grain type, discovering that electron beam mutations produced smaller grain size mutants than gamma ray.

All the homozygous lines were grouped as a single cluster in fourth quarter. Seven homozygous lines of MDU 6 x *Jaldi Dhan* 6 and one homozygous line of TKM 6 x MDU 6 were selected based on very early maturity because the parents are distinctly differ for days to maturity. Days to maturity of MDU 6 was 114 days and 75 days for *Jaldi Dhan* 6, whereas the cross derivatives ranged from 99.50 to 103 days for maturity. So it was formed as separate group.

The biplot between genotypes and traits useful for understanding the relationship between traits and association among the traits (Fig. 2). It also useful to determine the trait profile towards particular genotype (Yan and Kang, 2002). The angle between two traits discovered the positive and negative association between traits. Yan and Tinkers (2006) also reported acute angle between two traits indicated the positive correlation between those traits, whereas obtuse angle between the traits revealed negative correlation of those traits.

The characters such as flag leaf breadth, flag leaf length, panicle length, plant height on 50th day, grain length, grain breadth and hundred seed weight fall under first quarter were highly correlated with each other because these

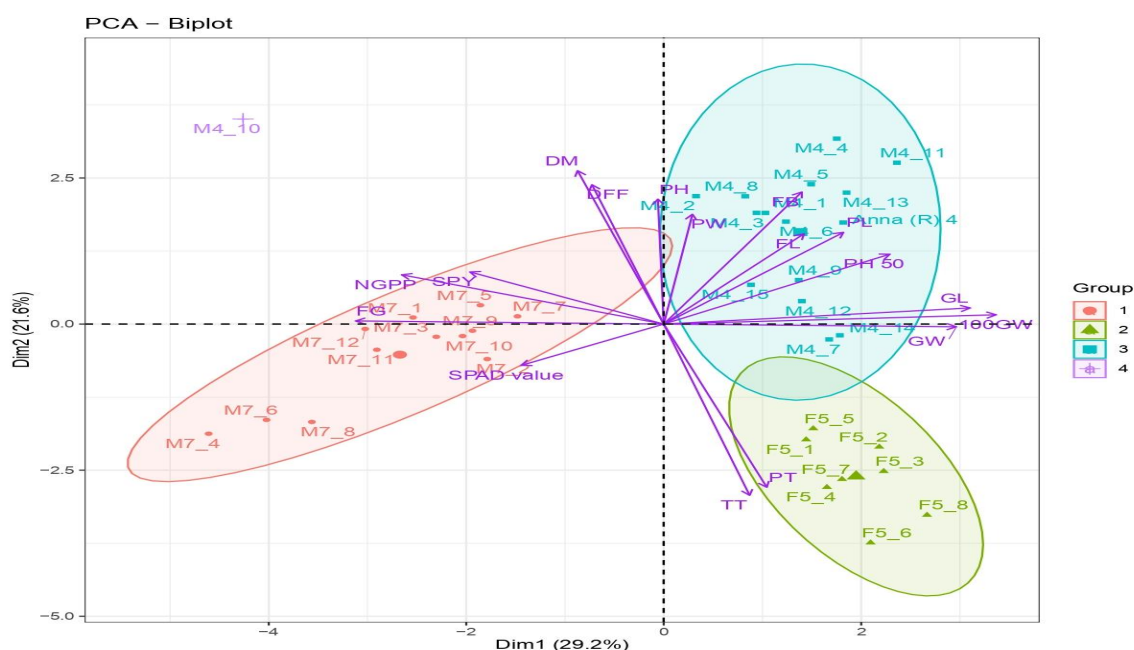


Fig. 2. Principal component analysis of 36 rice genotypes

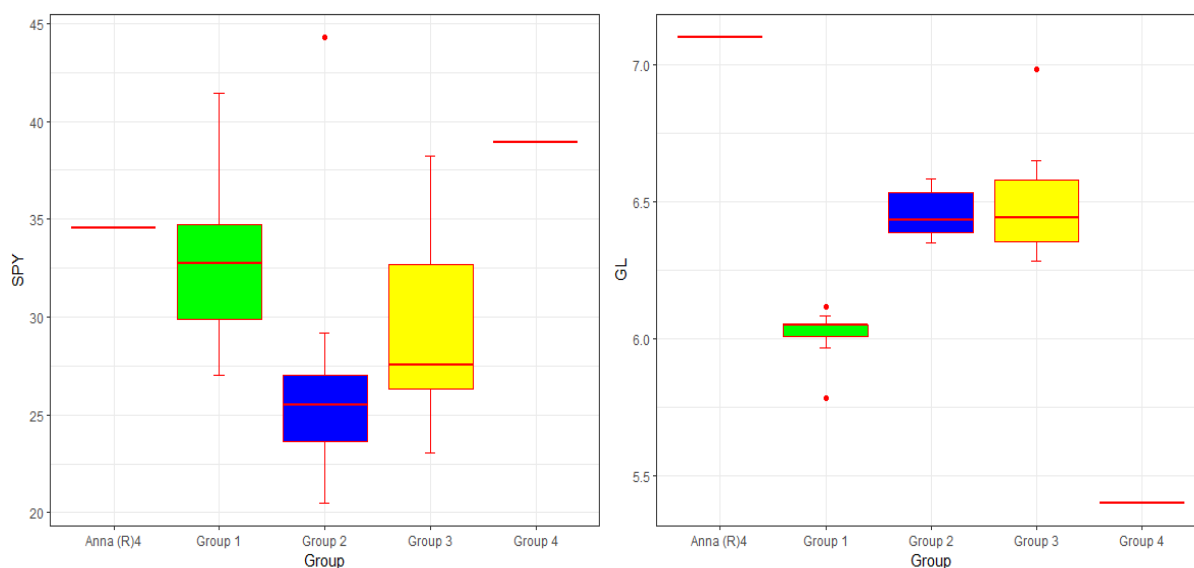


Fig. 3. Mean comparison of single plant yield and grain length of four groups and Anna (R) 4 rice variety

traits had acute angle for these traits. The genotypes located in the first quarter had better performance for these characters compared to other genotypes located in the other quarters.

All the recombinant homozygous lines (fourth quarter) located in right angle to the genotypes located in first quarter. So, the genotypes existing in first and fourth quarter independently exhibited same range of genetic variation. But the recombinant lines had better performance for the traits of total tillers and productive tillers compared with the genotypes present in first quarter. The angle between these two variables indicated that existence of strong positive association between them and Priyanga and Amudha (2022) also reported the same results in landraces of rice.

Angle between the traits in the second quarter such as filled grains, number of grains per panicle and single plant yield had positive association with each other. Shanmugam *et al.* (2023) also reported that positive association between NGPP and SPY. But the obtuse angle between the traits present in the first quarter such as grain length, breadth, hundred seed weight and the traits of second quarter such as filled grains, number of grains per panicle and single plant yield revealed presence of negative correlation between them. So the genotypes fall in the first group exhibited better performance for filled grains, number of grains per panicle and single plant yield compared with second and third group.

The mean comparison of four groups with Anna (R) 4 variety for single plant yield and grain length were presented in the **Fig 3**. All the four groups showed the reduction in grain length of medium slender type with more reduction in the genotypes of group 1 (6 mm) and group

4 (5.5 mm). The grain yield performance of group 1 and 4 were comparable with commercial variety compared to less grain yield performance in group 2 and 4. This offers scope to select the genotypes for group 1 and 4 for higher grain yield with improved grain quality suitable for rainfed condition. Salunkhe *et al.* (2023) also studied the diversity by using the PCA in thirty rice varieties based on yield and related variables.

The current investigation was revealed the existence of genetic divergence among the mutants genotypes analysed by PCA. The grouping patterns of genotypes revealed that differences between EMS, gamma ray, electron beam and cross derivative homozygous lines from Anna (R) 4 parental rice variety. Based on eigen vectors the traits like hundred grain weight, grain length, grain width and plant height on 50th day were the most critical variables contributing for more divergence. Moreover the grouped genotypes and their associated trait relationship with that specific group were more helpful to further crop improvement.

REFERENCES

- Ariyo, O.J. 1987. Multivariate analysis and the choice of parents for hybridization in okra (*Abelmoschus esculentus* (L.) Moench). *Theoretical and Applied Genetics*, **74**: 361-363. [Cross Ref]
- Arunachalam, P., Lalitha, R., Vanniarajan, C. and Souframanien, J. 2022. Comparative analysis of gamma rays and electron beam in altering rice (*Oryza sativa* L.) grain size. *Indian Journal of Genetics and Plant Breeding*, **82**(03): 355-358. [Cross Ref]

- Bharadwaj, C., Satyavathi, C.T. and Subramanyam, D. 2001. Evaluation of different classificatory analysis methods in some rice (*Oryza sativa*) collections. *Indian Journal of Agricultural Science*, **71**(2): 123-125.
- Chandraker, P., Sharma, B., Parikh, M. and Saxena, R.R. 2024. Assessment of genetic diversity in aromatic short grain rice (*Oryza sativa* L.) genotypes using pca and cluster analysis. *International Journal of Plant & Soil Science*, **36**(5): 82-94. [Cross Ref]
- Francis, N. and Packiaraj, D. 2020. Genetic uniformity of varieties and an assessment on the diversity among the elite varieties of rice (*Oryza sativa* L.). *Electronic Journal of Plant Breeding*, **11**(3): 896-900. [Cross Ref]
- Hailu, F., Merker, A., Belay, G. and Johansson, E. 2006. Multivariate analysis of diversity of tetraploid wheat germplasm from Ethiopia. *Genetic Resources and Crop Evolution*, **53**: 1089-1098. [Cross Ref]
- Kaiser, H.F. 1960. The application of electronic computers to factor analysis. *Educational and psychological measurement*, **20**(1): 141-151. [Cross Ref]
- Kumari, B.K., Kumar, B.R., Jyothula, D.P.B. 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]
- Lohani, M., Singh, D. and Singh, J.P. 2012. Genetic diversity assessment through principal component analysis in potato (*Salamun tuberosun* L.). *Vegetable Science*, **39**(02): 207-209.
- Mondal, S., Rutkoski, J.E., Velu, G., Singh, P.K., Crespo-Herrera, L.A., Guzman, C., Bhavani, S., Lan, C., He, X. and Singh, R.P. 2016. Harnessing diversity in wheat to enhance grain yield, climate resilience, disease and insect pest resistance and nutrition through conventional and modern breeding approaches. *Frontiers in plant science*, **7**: 991. [Cross Ref]
- Nachimuthu, V.V., Robin, S., Sudhakar, D., Raveendran, M., Rajeswari, S. and Manonmani, S. 2014. Evaluation of rice genetic diversity and variability in a population panel by principal component analysis. *Indian journal of science and technology*, **7**(10): 1555-1562. [Cross Ref]
- Nair, N.V., Balakrishnan, R. and Sreenivasan, T.V. 1998. Variability for quantitative traits in exotic hybrid germplasm of sugarcane. *Genetic Resources and Crop Evolution*, **45**: 459-463. [Cross Ref]
- Oyelola, B.A. 2004. The Nigerian Statistical Association preconference workshop, 20-21 September, 2004. In Conference centre, University of Ibadan.
- Pathak, S.K., Lavanya, G.R., Babu, G.S. and Srivastava, N. 2018. Evaluation of rice germplasm for genetic diversity on yield characters by principal component analysis. *Pharma Innovat J*, **7**: 661-664.
- Priyanga, R.S. and Amudha, S. 2022. Studies on variability, association and path analysis in rice landraces of Tamil Nadu. *Madras Agricultural Journal*, **109**(June (4-6)): 1. [Cross Ref]
- Rasheed, A., Hassan, M.U., Aamer, M., Bian, J.M., Xu, Z.R., He, X.F., Yan, G. and Wu, Z.M. 2020. Iron toxicity, tolerance and quantitative trait loci mapping in rice; a review. *Applied Ecology & Environmental Research*, **18**(6): 7483-7498. [Cross Ref]
- Ratnam, T.V., Kumar, B.R., Rao, L.V., Srinivas, T. and Kumar, A.A. 2022. Principal component analysis of yield and quality traits in Zinc rich landraces of rice (*Oryza sativa* L.). *Electronic Journal of Plant Breeding*, **13**(4): 1162-1169. [Cross Ref]
- Ravikumar, B.N.V.S.R., Kumari, P.N., Rao, P.V.R., Rani, M.G., Satyanarayana, P.V., Chamundeswari, N., Vishnuvardhan, K.M., Suryanarayana, Y., Bharatalakdhmi, M. and Vishnuvardhan, R. 2015. Principal component analysis and character association for yield components in rice (*Oryza sativa* L.) cultivars suitable for irrigated ecosystem. *Current Biotica*, **9**(1): 25-35.
- Sahebi, M., Hanafi, M.M., Rafii, M.Y., Mahmud, T.M.M. Azizi, P., Osman, M., Abiri R., Taheri, S., Kalhori, N., Shabanimofrad, M., Miah, G. and Atabaki, N. 2018. Improvement of drought tolerance in rice (*Oryza sativa* L.): genetics, genomic tools, and the WRKY gene family. *BioMed research international*, **2018**(1): 3158474. [Cross Ref]
- Salunkhe, H., Kumar, A., Krishna, B., Talekar, N. and Pawar, P. 2023. Genetic diversity and principal component analysis for yield and its component trait in rice. In *Biological Forum—An International Journal*, **15**: 102-107.
- Shanmugam, A., Suresh, R., Ramanathan, A., Anandhi, P. and Sassikumar, D. 2023. Unravelling genetic diversity of South Indian rice landraces based on yield and its components. *Electronic Journal of Plant Breeding*, **14**(1): 160-169. [Cross Ref]
- Sheela, K. S., Robin, S. and Manonmani, S. 2020. Principal component analysis for grain quality characters in rice germplasm. *Electronic Journal of Plant Breeding*, **11**(1): 127-131. [Cross Ref]
- Tiruneh, A., Gebrselassie, W. and Tesfaye, A. 2019. Genetic diversity study on upland rice (*Oryza sativa*

L.) genotypes based on morphological traits in Southwestern Ethiopia. *Asian Journal of Crop Science*, **11**(1): 17-24. [[Cross Ref](#)]

Yan, W. and Kang, M.S. 2002. *GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists*. CRC press. [[Cross Ref](#)]

Yan, W. and Tinker, N.A. 2006. Biplot analysis of multi-environment trial data: Principles and applications. *Canadian journal of plant science*, **86**(3): 623-645. [[Cross Ref](#)]