Plant Archives Vol. 25, Supplement 2, 2025 pp. 3187-3193 e-ISSN:2581-6063 (online), ISSN:0972-5210

Plant Archives

Journal homepage: http://www.plantarchives.org
DOI Url : https://doi.org/10.51470/PLANTARCHIVES.2025.v25.supplement-2.408

PRINCIPAL COMPONENT ANALYSIS AND CORRELATION STUDIES IN RICE (*ORYZA SATIVA* L.) GENOTYPES

B. Srinivas*, P. Gonya Nayak, P. Madhukar, B. Laxmi Prasanna, D. Sreelatha and G. Sreenivas
Rice Research Scheme, Regional Agricultural Research Statio, Jagtial, PJTAU, Telangana – 505529, India
*Corresponding author E-mail: srinu.bdd@gmail.com
(Date of Receiving : 22-05-2025; Date of Acceptance : 29-07-2025)

ABSTRACT

Principal component and correlation analysis was carried out in 15 rice genotypes for eight variables at Regional Agricultural Research Station, PJTAU, Polasa, Jagtial, Telangana. ANOVA indicated significant variation for all the traits studied. PCA analysis revealed that first three principal components (PC, PC2 and PC3) were accounted for 80% of total variation, of which PC1 contributed highest (36.78%) followed by PC2 (28.87%) and PC3 (14.65%). The traits, test weight (0.530) and yield (0.395) exhibited highest positive loadings on PC1 indicating the major contribution of these traits towards variation in PC1. Biplot shows the high correlation of yield with important yield attributing traits like test weight and panicle length as yield registered acute angle with both. The variable test weight exhibited highest vector length indicating its predominant contribution towards total variance followed by yield, head rice recovery and number of grains per panicle. The genotypes, G6, G4, G7, G5 and G12 with high PC scores and high yield were found desirable. Correlation analysis indicated the strong association of yield with test weight followed by panicle length which are important yield attributing characters. Yield exhibited strong and significant negative correlation with days to 50% flowering indicating the scope for development of high yielding early maturing varieties.

Keywords **:** Rice, principal Component Analysis, Yield, Quality, Correlation

Introduction

Rice is one of the important cereal crops feeds about two third of world's population and contributing up to 50% of the daily caloric intake for Asian populations. The aspect of increase of rice area has its limitations like limited irrigation sources, market facilities, reluctant of farmers to shift from cultivating traditional crops etc. High yielding varieties plays a crucial in rising the food production to meet the requirements of growing population. It is crucial to take the sustainability of food systems into account, especially better utilizing rice production in order to feed 10 billion people in 2050 (Saha *et al.*, 2021). Parents are very important in development of high yielding varieties or hybrids and various breeding techniques like D^2 analysis, multivariate analysis, principal component analysis, estimation of general combining ability helps in choosing the appropriate parents for the development of varieties or hybrids of interest. Principal component analysis is one of the effective breeding techniques that minimizes the huge data and transform it into few variables retaining the maximum available variability. It is an orthogonal transformation process that transforms a collection of characters with potentially correlated values into a set of observations with linearly uncorrelated characters (Mohan *et al.*, 2015 and Alshugeairy *et al.*, 2023). Yield is a complex trait which in turn depends on other yield contributing characters. In rice, 1000 grain weight, number of grains per panicle, panicle length and number of productive tillers per plant are the traits that contribute directly for yield enhancement. The other characters like plant height, biomass, leaf area also plays a role in increasing the yield. Breeder must have the knowledge about the type of correlation between the yield and its important contributing traits to select the trait of interest in segregating generations for further improvement. The present investigation is aimed at selection of desirable parents through PCA analysis and finding the traits having significant correlation with yield in desirable direction.

Materials and Methods

The present investigation was carried out at Regional Agricultural Research Station, Polasa, Jagtial, Telangana located at 18^0 50’ 20.24” N latitude, $78^0$56’54.20” E longitude and 249 m above mean sea level. The experimental material constitutes 15 rice genotypes (Table 1.). The experiment was conducted during *kharif,* 2022 with plot size of 10.32 m^2 per each entry in three replications. The spacing adapted was 15 x 15 cm between the rows and between the plants within the row. Standard recommended package of practices were adapted during the crop season to raise the healthy crop. Data on various morphological, yield and quality traits was recorded. Days to 50% flowering and yield were recorded on plot basis in each replication. 5 random plants in each replication per entry were used to record plant height, panicle length and effective bearing tillers/m^2. Random seed sample from each entry in each replication was used to record the test weight, number of grains per panicle and head rice recovery. Analysis of variance (ANOVA) for the randomized complete block design was computed using the method of Panse and Sukhatme (1967). Correlation coefficients were calculated using Pearson’s simple correlations and path analysis was computed using the method suggested by Dewey and Lu (1959). Principal component analysis (PCA) was used to identify characters mainly contributing to the variation among the genotypes tested. PCA was done using quantitative traits that are correlated by converting them into uncorrelated traits known as PCs (Johnson and Wichern, 1988). Principal components were computed using R software (Version 4.3.2), derived from the mean values of eight characters across the 15 rice genotypes.

Results and Discussion

Analysis of Variance (Table 2) revealed significant variation among the genotypes for all the characters studied emphasizing much scope for these traits for further improvement.

Principal Component Analysis

The independent effects of each of eight characters studied in fifteen genotypes were determined by principal component analysis. Eigen values are generally employed to establish the number of factors to retain. If eigen value is less than one, it indicates that the explanatory efficacy of the principal component is inferior to average explanatory efficacy of the original variables (Jadhav *et al.,* 2023). The principal components having eigen values greater than one were considered as significant. The traits with high eigen value on PC1 are generally considered as key contributors to the total variance. In the present investigation, out of eight principal components the first three PCs (PC, PC2 and PC3) exhibited eigen values more than one (Table 3) and accounted for 80% of total variation, of which PC1 contributed highest (36.78%) followed by PC2 (28.87%) and PC3 (14.65%). Percent of variance contributed by each PC was observed to decrease gradually from PC1 to PC8, in turn cumulative frequency was increased. Scree plot (Fig 1) showed the percent of explained variance contributed by each of eight principal components *i.e.* PC1 (36.78%) to PC8 (0.3%) and clearly indicating that, maximum amount of total variation present in the experimental material is represented by PC1, PC2 and PC3. Highest contribution by PC1 suggests that selecting genotypes based on their scores on PC1 will be useful in improving the traits such as test weight, yield and panicle length which are important yield attributing characters. Desirable genotypes are those which are having favourable combinations of important yield contributing traits. Hence, breeders should focus on these traits with high scores on PC1 that holds maximum portion of total variability.

Individual trait PC scores (Table 3) represents the contribution of each character towards total genetic variation present in 15 rice genotypes. The positive and negative loadings of each individual trait on PCs indicate its influence on variability, association with principal component and impact on genotype performance (Sruthi *et al.,* 2024). The traits, test weight (0.530) and yield (0.395) exhibited highest positive loadings on PC1 indicating the major contribution of these traits towards variation in PC1. Palaniyappan *et al.* (2023), Vaidya *et al.* (2024) and Sruthi *et al.* (2024) reported similar findings of high positive loadings for test weight and yield on principal components. Whereas, days to 50% flowering (-0.477), head rice recovery (-0.361), number of grains per panicle (-0.343) and plant height (-0.179) with negative loadings contributed less for total variation in PC1. Sruthi *et al.* (2024), Sandhya *et al.* (2022) and Christina *et al.* (2021) reported similar results for one or other traits. In PC2, panicle length (0.549), plant height (0.466), yield (0.413), head rice recovery (0.360) and number of grains per panicle (0.197) recorded positive loadings shows the importance of these traits in enhancing yield and yield contributing traits, whereas, effective bearing tillers/m^2 (-0.370), days to 50% flowering (-0.052) and test weight (-0.032) exhibited negative influence in PC2. In PC3, only number of grains per panicle (0.594) was observed to be contributor towards total variation and remaining all the *viz.,* effective bearing tillers/m2 (-0.497), days to 50% flowering (-0.388), plant height (-0.313), test weight (-0.300), panicle length (-0.180), head rice recovery (-0.140) and yield (-0.087) contributed in negative direction towards variation.

The biplot (Fig. 2) helps in understanding the association among the traits and identifying the better performing genotypes for the particular trait. It also gives the nature and distribution of variability exhibited by both the characters and genotypes and helps to determine the trait profile towards particular genotype (Yan and Kang, 2002). The angle between the characters determines the type of correlation existing between those 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. Yield was found to be highly correlated with important yield attributing traits like test weight and panicle length as yield registered acute angle with both, whereas, remaining traits exhibited negative correlation with yield. Similarly, when the correlation between grain parameters like test weight and number of grains per panicles was observed, a high negative correlation was noticed between them, indicating that fine grain varieties accommodated more number of grains per panicle compared to coarse grain rice cultures. It was also observed from the biplot that, head rice recovery which is very important quality parameter has exhibited high positive correlation with number of grains per panicle indicating high head rice recovery in case of fine grain varieties and low in coarse varieties. The length of each trait vector also gives the amount of contribution of that particular trait towards total divergence. In the present study, highest vector length was exhibited by test weight indicating its predominant contribution for total variance followed by yield, head rice recovery and number of grains per panicle. The genotypes that are present close to the trait vector of the same quadrant would be the best performing for those traits. The phenotypic expression of each genotype was explained through the principal component (PC) scores, as delineated in the Table 4. Thus, the genotypes, G6, G4, G7, G5 and G12 with high PC scores and yield performance were found to be desirable for the traits yield and test weight whereas, G1, G2, G8 and G10 were designated as undesirable which also confirmed by the biplot. According to Singh and Chaudhary (1977), high values for variables within a specific genotype are represented by a correspondingly high PC score attributed to that genotype within the associated component. The highest PC scores of positive values >1.5 in each PC could be used them as selection indices (Das *et al.* 2024).

Correlation studies

Correlation analysis is an important part in plant breeding which gives the positive or negative association between the traits of interest and helps in identifying the group of desirable traits towards which the selection could be practiced for obtaining the genotypes with high yield and quality. Pearson's correlation analysis (Fig 3) shows the strong correlation of yield with test weight followed by panicle length which are important yield attributing characters. Saketh *et al.* (2023), Barde *et al.* (2024) and Reddy *et al.* (2024) also reported significant association of test weight and panicle length with grain yield suggesting direct selection for these traits to improve the yield. Yield exhibited strong and significant negative correlation with days to 50% flowering indicating the scope for development of high yielding early maturing varieties with the present experimental material. Prashanth *et al.* (2024) and Kalpande *et al.* (2024) also revealed the same findings from their research *i.e* negative and significant correlation of yield with flowering duration. Interestingly, head rice recovery recorded strong negative correlation with test weight suggesting the scope for development of fine grain varieties with high head rice recovery. The significant and high negative correlation between test weight and number of grains per panicle revealed that selection for high grain number could results in development of fine grain cultures and low grain number for coarse varieties.

Conclusion

Out of 8 principal components, the first three PCs (PC, PC2 and PC3) exhibited eigen values more than one and accounted for 80% of total variation. Maximum amount of total variation present in 15 rice genotypes is represented by PC1, PC2 and PC3. Test weight (0.530) and yield (0.395) with highest positive loadings on PC1 were identified as major contributors for total variation of PC1 which also exhibited highest vector lengths on biplot. From the present study, G6, G4, G7, G5 and G12 were identified as desirable genotypes and G1, G2, G8 and G10 as undesirable. Yield registered strong positive correlation with test weight and panicle length and negative correlation with days to 50% flowering.

Table 1: Details of the 15 rice genotypes

Entry designation	Entry name	Grain type	Maturity group
G1	KNM 12472	Medium slender	Long
G2	BPT 5204	Medium slender	Long
G3	KNM 12469	Medium slender	Long
G4	JGL 38206	Long slender	Early
G5	MTU 1001	Medium bold	Medium
G6	RNR 31753	Long slender	Early
G7	JGL 38889	Long slender	Early
G8	JGL 11470	Short slender	Medium
G9	JGL 38903	Short slender	Medium
G10	HMT SONA	Short slender	Medium
G11	KNM 12510	Short slender	Medium
G12	WGL 1380	Medium slender	Medium
G13	KRISHNA	Medium slender	Medium
G14	WGL 1590	Medium slender	Medium
G15	RNR 35197	Medium slender	Medium

Table 2: Analysis of variance (ANOVA) for 8 characters in 15 rice genotypes

Source of variation	Degrees of freedom	Days to 50% flowering	Effective bearing tillers/m^2	Plant height	Panicle length	Test weight	Number of grains /panicles	Head rice recovery	Seed yield
Replication	2	3.62**	3238.3	10.60	3.26	2.79**	212.3	14.27	634492
Treatments	14	215.26**	2716.9*	376.41**	11.84**	82.925**	7343.6**	119.69**	6060407**
Error	28	0.075	986.5	7.50	1.13	0.378	631.9	11.975	468032
CD (5%)		0.456	52.53	4.58	1.78	1.027	42.041	5.787	1144.217

Table 3: Eigen values, percentage of variance and cumulative variance of first three principal components of rice.

Principal components	**PC1**	**PC2**	**PC3**
Eigen value	2.942	2.309	1.172
Percentage of variance	36.783	28.865	14.651
Cumulative percentage	36.783	65.648	80.000
Individual characters PC score			
Days to 50% flowering	-0.477	-0.052	-0.388
Effective bearing tillers/m^2	-0.230	-0.370	-0.497
Plant height	-0.179	0.466	-0.313
Panicle length	0.045	0.549	-0.180
Test weight	0.530	-0.032	-0.300
Number of grains/panicles	-0.343	0.197	0.594
Head rice recovery	-0.361	0.360	-0.140
Seed yield	0.395	0.413	-0.087

Table 4: PC scores of 15 rice genotypes in first three principal components

Entry Name	Designation	Seed yield (kg/ha)	PC1	PC2	PC3
KNM 12472	G1	5152	-2.259	-0.0585	-1.176
BPT 5204	G2	4390	-0.856	-2.653	-0.533
KNM 12469	G3	5446	-2.432	0.789	-1.852
JGL 38206	G4	7526	2.656	0.656	-0.081
MTU 1001	G5	6990	1.216	-0.833	-2.214
RNR 31753	G6	7277	3.051	-0.238	-0.199
JGL 38889	G7	8815	2.361	0.588	0.162
Local Check	G8	3934	-1.682	-0.009	1.386
JGL 38903	G9	6570	-1.385	1.168	0.859
HMT Sona	G10	4096	-0.197	-3.884	1.163
KNM 12510	G11	6925	-0.999	1.459	1.007

WGL 1380	G12	7674	0.164	2.040	-0.269
Krishna	G13	6550	0.322	0.120	0.466
WGL 1590	G14	6699	-0.203	0.361	0.328
RNR 35197	G15	6854	0.244	0.494	0.954

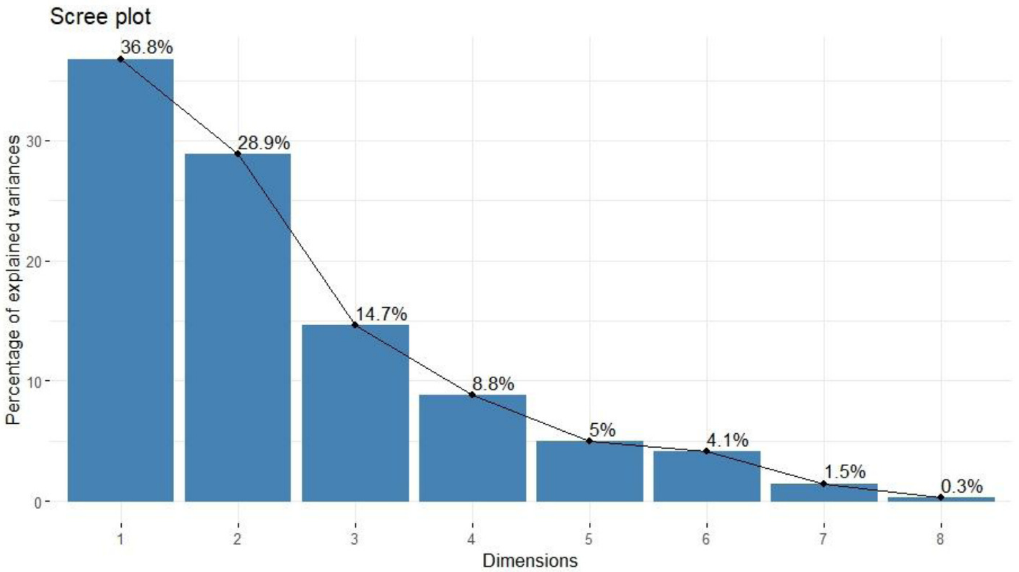

Fig. 1: Scree plot illustrating percentage of variance explained by 8 principal components

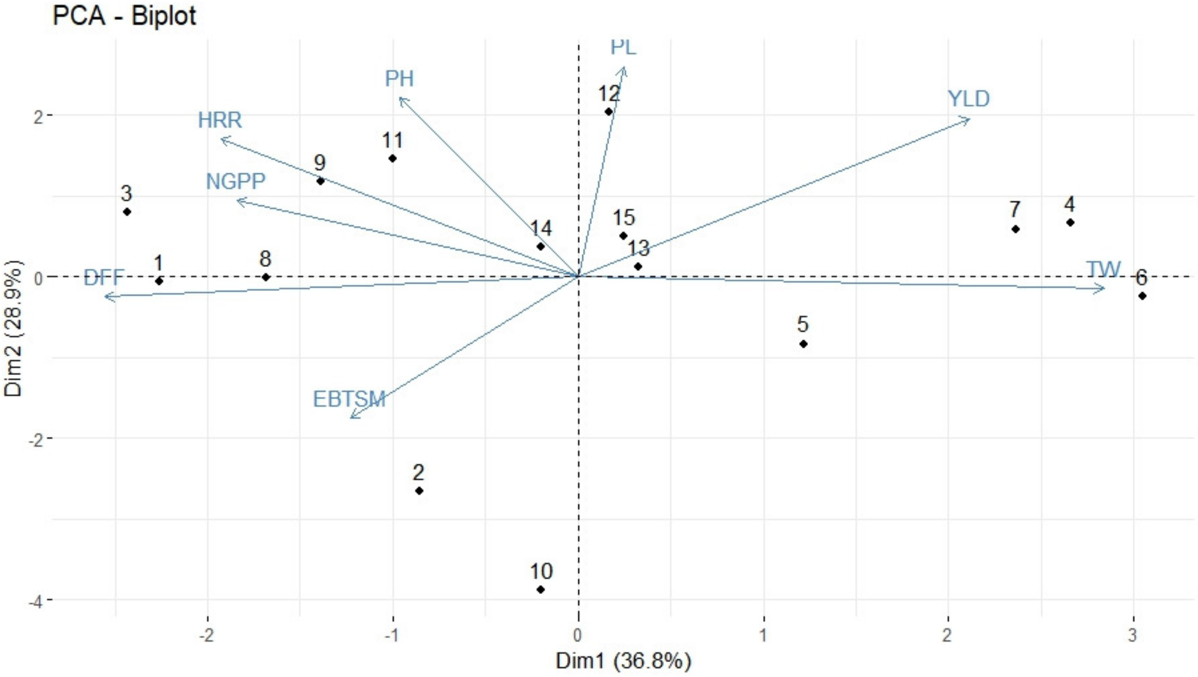

Fig. 2: Biplot explaining the distribution of 15 rice genotypes and contribution of 8 variables in first two principal components

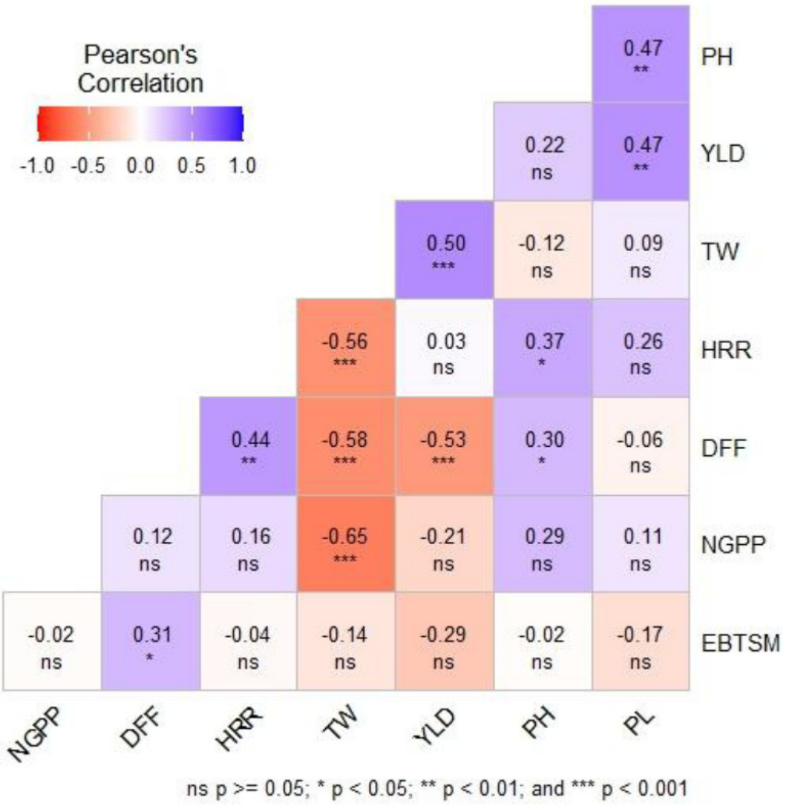

Fig. 3: Correlation coefficients among 8 characters in 15 rice genotypes

Acknowledgement

The authors are thankful to Regional Agricultural Research Station, Polasa, Jagtial, Professor Jayashankar Telangana Agricultural University for providing the fund and research facilities to carry out the present research work.

References

Alshugeairy, Z.K., Alogaidi, F.F., Hadi B.H. and Hassan W.A. (2023). Path coefficient analyses of introduced rice varieties under different planting distances. *SABRAO J. Breed. Genet*., **55(2),** 516-524.

Barde, Y.D., Wandhare, M.R., Madke, V.S., Manapure, P., Bagade, V.H., Kunte S.G. and Pardhi K.C. (2024). Correlation and path analysis studies in rice (*Oryza sativa* L.). *Int. J. Adv. Biochem. Res.*, **8(9),** 973-979.

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. *Electron. J. Plant Breed.,* **12(3),** 907-911.

Das, S., Sawarkar, A., Saha, S., Raman R.B. and Dasgupta T. (2024). Principal Component and Cluster Analysis in Mungbean [*Vigna radiata* (L.) Wilczek]. *Legume. Res.,* **47(8),** 1-10.

Dewey, D.R and Lu K.H. (1959). A correlation and path coefficient analysis of components of crested wheat grass seed production. *Agron. J.,* **51**, 515-518.

Jadhav, R.A., Mehtre, S.P., Patil D.K. and Gite V.K. (2023). Multivariate analysis using D2 and principal component analysis inmung bean [*Vigna radiata* (L.) Wilczek] for study of genetic diversity. *Legume Res.,* **46(1),** 10-17.

Johnson R.A. and Wichern D.W. (1998). Applied multivariate statistical analysis. Madison: *Prentice Hall International*, 816p.

Kalpande, P., Meshram M. and Borkar V. (2024). Genetic Variability, Correlation and Path Analysis Studies in Early Transplanted Rice. *Int. J. Plant & Soil Sci.,* **36(7),** 128-134.

Mohan, Y.C., Thippeswamy, S., Bhoomeshwar, K., Madhavilatha B. and Samreen J. (2015). Diversity analysis for yield and gall midge resistance in rice (*Oryza sativa* L.) in Northern Telangana Zone, India. *SABRAO J. Breed. Genet.* **47,** 160-171.

Palaniyappan, S., Arunachalam, P., Banumathy S. and Muthuramu S. (2023). Principal component analysis (PCA) as a genetic diversity tool to understand the variation of rice mutant culture. *Electron. J. Plant Breed.,* **15(4),** 979-985.

Panse V.G. and Sukhatme P.V. (1967). Statistical Methods for Agricultural Workers. Indian Council of Agricultural Research, New Delhi. 1967, 103-108.

Prashanth, M., Gireesh, C., Lakshmi, S., Laha, G.S., Kumari, P.L., Anantha, M.S., Sundaram R.M. and Senguttuvel P. (2024). Correlation and path analysis of association among yield and yield contributing traits under aerobic condition in F2 population in rice. *Plant Arch.,* **24(2),** 1967-1973.

Reddy, V.V.S., Rani, S.Ch., Kumar, C.V.S. and Basavaraj K. (2024). Study of Variability and Correlation for Yield Attributing and Grain Quality Characters in Aromatic

Short Grain Rice. *J. Experiment. Agri. Int.,* **46(9),** 164-173.

Saha, I., Morat, D.A., Nalley, L.L., Alam M.J. and Nayga R. (2021). Rice quality and its impacts on food security and sustainability in Bangladesh. *PLoS One.,* **16(12),** 1-17.

Saketh, Shankar, V.G., Srinivas B. and Hari Y. (2023). Correlation and Path Coefficient Studies for Grain Yield and Yield Components in Rice (*Oryza sativa* L.). *Int. J. Plant & Soil Sci.,* **35(19),** 1549-1558.

Sandhya, Mukul, Kumar M., Kumar, P., Gaurav, A., Meena D. and Soni R.K. (2022). Assessment of Principal Component Analysis and Genetic diversity among linseed (*Linum usitatissimum* L.) germplasm under Humid South Eastern plain of Rajasthan. *Agri. Mechanization in Asia.*, **53(2)**, 5809-5818.

Singh R.K. and Chaudhary B.D. (1977). *Biometrical methods in quantitative genetic analysis.*, 1-304.

Sruthi, S.R., Kumar, L.N., Kishore, D., Subakar J.I. and Anbuselvam Y. (2024). 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.*, **11(4),** 634-639.

Vaidya, N., Nayak, G., Parida, M., Parameswaran, C., Sharma, R.S., Verma, R., Katara, J.L., Samantaray, S., Sharma, S., Nema, S., Gaur, V.S., Kumar A. and Ramakrishnan R.S. (2024). Principal Component Analysis of Grain Yield and its Attributing Traits in Recombinant Inbred Lines of Rice (*Oryza sativa* L.) Developed for Submergence Tolerance. *Int. J. Plant & Soil Sci.*, **36(2),** 143-148.

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

Yan W. and Tinker N.A. (2006). Biplot analysis of multi-environment trial data: Principles and applications. *Canadian J. Plant Sci.,* **86**(3), 623-645