Principal component analysis of rice genotypes grown under normal and flash flood environments

D. Sathyaraj, T. Sabesan*, K. Sowmia, K. Saravanan and P. Satheesh Kumar

Department of Genetics and Plant Breeding, Faculty of Agriculture, Annamalai University, Annamalai Nagar 608 002, T.N., India

ABSTRACT

In the present study, Principal Component Analysis (PCA) was carried out to find the genetic divergence among 30 genotypes of rice grown under normal (E1) and flash flood environments (E2).They were evaluated for nine quantitative traits *viz.,* days to fifty per cent flowering, plant height, number of productive tillers per plant, panicle length, number of grains per panicle, hundred grain weight, grain length, grain breadth and grain yield per plant. Principal Component Analysis (PCA) under normal (E1) environment revealed that out of 9 components, only two PCs exhibited more than 1.0 Eigen value and showed about 62.66 per cent of total variability. Under flash flood (E2), three PC's exhibited Eigen value >1 with 73.92 per cent of total variability among the characters. The biplot analysis, which was employed to identify the best performing genotypes and to unravel patterns of genotypes both under normal and stressed conditions, resulted in identification of CR-1009 Sub-1, Swarna Sub-1, AU-1 GSR, *Mappilai samba* and *Seeraga samba* as potential parent for future submergence tolerance breeding programme.

Introduction

Rice is the most important human food crop in the world and is the staple food for more than half of the global population. Rice is the rich source of dietary energy (27%), and dietary protein (25%) which plays an pivotal role in Indian diet. However, global rice production is currently insufficient to meet the increasing demand for food, which is expected to rise by 1-2.4% by 2050 (Gautam and Shrestha, 2023). Global rice (paddy) production increased more than threefold between 1961 and 2019, from 215 million tonnes to 755 million tonnes; most of the production was from Asia. In India, the total area under rice cultivation was 46.38 million hectares with a production of about 130.29 million tonnes and productivity of 2809 kg / hectare. In Tamil Nadu, the area and production of rice were about 2.21 million hectares and 8.07 million tones and productivity of 3658 kg / hectare respectively (Directorate of Economics and Statistics, 2022). Globally, the major rice production constraints included several stress factors namely biotic and abiotic, which led to notable yield loss in rice production.

Among the abiotic stresses, submergence due to flooding ranks second after drought among the abiotic stresses causing great economic losses. For the past few years, flash flood stress is often seen in coastal ecosystem. Submergence is a condition where the whole plant is immersed in water. It deprives the plants for free atmospheric oxygen and reduces the rates of photosynthesis and respiration Bailey-Serres *et al.* (2012). It usually has harmful effects on crop plants; however, rice has capability to

withstand such conditions to some extent. Flash food negatively affects the plant growth and productivity by disturbing source–sink relationships and resource acquisition strategies. Incessant obstacles are faced by the local farmers of Cuddalore district every year due to flash flooding which is caused by massive unseasonal rain Sathyaraj *et al.* (2023). Therefore, in order to maintain rice production in coastal areas, it is necessary to find solutions to overcome these problems. To cope with the challenges of climate change, therefore, it is essential to develop climate-smart rice varieties with concerted research effort (Bin Rahman and Zhang, 2022).

Principal component analysis (PCA) helps in the identification of a set of genotypes capturing maximum genetic diversity of the collection. Principal component analysis (PCA) is an implausibly propitious tool for dissecting the correlation between traits, as well as their interactions, and determining the genotypic performance in crop plants. The original purpose of PCA was to extraction of principal component in large data sets, the detection of underlying data structures and the representation of complex biological patterns. In view of the importance of flash flood response traits in breeding for improved high-yielding rice varieties, the present investigation was undertaken to evaluate 30 genotypes and identify promising genotypes under flash flood.

Materials and Methods

The experimental material comprised of 30 genotypes (Table 1) including fifteen released varieties, two rice culture and eleven traditional varieties along with two submergence tolerant checks that included Swarna *Sub* 1 and CR 1009 *sub1* were used in this study. The experiments were carried out at the Plant Breeding Farm, Department of Genetics

Table 1. List of genotypes used in this study

Genotypes code	Varieties/ Culture	Source
G1	CR-1009 SUB-1	Plant Breeding Farm, Annamalai University, TamilNadu.
G2	SWARNA SUB-1	Plant Breeding Farm, Annamalai University, TamilNadu.
G3	ADT-49	Tamil Nadu Rice Research Institute, Aduthurai
G4	ADT-43	Tamil Nadu Rice Research Institute, Aduthurai
G5	ADT-36	Tamil Nadu Rice Research Institute, Aduthurai
G6	TKM-13	TNAU, Coimbatore, Tamilnadu
G7	TKM-9	TNAU, Coimbatore, Tamilnadu
G8	ASD-18	TNAU, Coimbatore, Tamilnadu
G9	BPT-5204	Bapatla Agriculture College, Andhra Pradesh
G10	CO-43	TNAU, Coimbatore, Tamilnadu
G11	IR-64	TNAU, Coimbatore, Tamilnadu
G12	IR-36	TNAU, Coimbatore, Tamilnadu
G13	AURC-4	Plant Breeding Farm, Annamalai University, TamilNadu.
G14	AU-1 GSR	Plant Breeding Farm, Annamalai University, TamilNadu.
G15	ANPRC-9	Rice Culture, Arunachal Pradesh
G16	NDLR-7	Nandyal Rice Research Station, Andhra Pradesh.
G17	NDLR-8	Nandyal Rice Research Station, Andhra Pradesh.
G18	MTU-7029	Rice research station, Marteru, AndhraPradesh
G19	JGL-384	Jagityala Rice Research Station, Andhra Pradesh
G20	White ponni	TNAU, Coimbatore, Tamilnadu
G21	Mappillai samba	Thiruvannamalai, Tamilnadu
G22	Kullakar	Sethurayan, Kudikadu Thanjavur, Tamilnadu
G23	Aruvathm kuruvai	Mandalakottai, Thanjavur Tamilnadu
G24	Kallimadayan	Thanjavur, Tamilnadu
G25	Kattu yanam	Vaduvur, Thanjavur Tamilnadu
G26	Poongkar	Avaniyapuram, Thanjavur
G27	Seeraga samba	Thuramadi, Thanjavur
G28	Salem sannam	Salem, Tamilnadu
G29	Mutrina sannam	Ulunthurpet, Tamilnadu
G30	Adukunel	Vaduvur, Thanjavur Tamilnadu

and Plant Breeding, Faculty of Agriculture, Chidambaram during the Samba season, (September- January) 2020, (11º 38´ N latitude, 79 º 72´ E longitude). This study consisted of two experiments, namely normal (E1) and natural flash flood (E2) conditions. The field evaluation of selected rice genotypes for their flash flood tolerance (E2) was carried out in a low lying field of the Plant Breeding Farm (11º 24´ N latitude, 79 º 44´ E longitude and + 5.79 m MSL) Department of Genetics and Plant Breeding, whereas normal (E1) grown in an elevated field of the Agronomy Farm, Department of Agronomy, Faculty of Agriculture, Annamalai University, Tamil Nadu, India during late Samba (September-January) 2020. The sowing date for our study was so planned that the flash flooding might coincide at active tillering phase and initiation of flowering. Nursery sowing was done in well prepared raised seed bed in first week of September, 2020. Twenty five days old seedlings were transplanted in well puddle field in Randomized Block Design with three replications. Each rice genotype was transplanted in two rows of 2 m row length. The plant to plant and row to row distance was maintained 15 cm and 20 cm, respectively. Recommended package of practices was adopted to raise a healthy crop.

Nine quantitative traits were measured and recorded at appropriate stages following, SES (IRRI, 2013). Multivariate analyses were performed on the following phenological (days to fifty per cent flowering), morphological (plant height, number of productive tillers per plant, panicle length, number of grains per panicle). Seed traits (length of grains, width of grains; length: breadth ratio of seeds, Hundred grain weight and seed yield per plant).

In order to identify the patterns of variation, Principal Component Analysis (PCA) was conducted. Those PCs with Eigen values greater than one were selected. Statistical analyses were done using R-4.0.2 for win (http://CRAN.R-project.org/) in Rstudio-1.3.1093 (https://rstudio.com/). Principal component analysis (PCA) was performed on different traits to identify the underlying sources of variability using the "factoextra" packages in R-studio. Additionally, multiple correlation analyses were analyzed using several R-studio packages, including readxl, ggplot2, dplyr, FSA, psych and performance analytics to examine the relationship between various yield component traits.

Results and Discussion

Principal component analysis is an useful technique in clarifying the relationship among traits in complex multi trait system such as the growth characters, quantitative traits and yield contributing characters. PCA explains the relationship of the Eigenvectors and Eigen-values with economic yield and helps in identification of principal components of yield in diverse breeding populations Yugandhar *et al.* (2018).

Principal component analysis helps in determining selection criteria and selection of superior lines in a diverse breeding population. Prafull Kumar *et al.* (2015) suggested that principal component analysis is valuable technique to quantify relationship among traits and for pattern of data by reducing dimensions. The parents for hybridization program should be selected based on the magnitude of genetic distance, contribution of different characters towards the total divergence and magnitude of cluster means for different characters performance having maximum heterosis. Genotypes of distantly located clusters were suggested to use in hybridization programs for obtaining a wide. The principal component analysis (PCA) was carried out using the mean data of the traits studied for the both the environments separately. Impact of each trait to the total variance and the degree of contribution of every original variable with which each principal component is associated were measured in both normal and flash flood environment. The result of the analysis was explained below environment wise. Principal components with Eigen values more than 1 and variation percent more than 4% were considered as main PC for both environments.

Genetic Variability

In the PCA under normal (E1) condition, a total of 9 principal components (PCs) were extracted and it revealed two most informative PCs with Eigen values of 2.98 and 2.65 which accounted for 62.65 % of the cumulative variance. The PC1 showed 33.14 % while, PC2 accounted 29.51 % variability, respectively among the genotypes for the characters under normal environment study explained by the first two axes with an Eigen value of greater than unity indicates that, the identified traits within the axes exhibited great influence on the phenotype of germplasm lines. Remaining components with

Eigen values <1 contributed 37.35 percent variability. The principal components with less than one were considered non- significant (Table 2).

In flash flood environment (E2) (Table 2), PCA analysis revealed that out of nine quantitative traits, three principal components contributed >1 Eigen value and exerted about 73.91 % of total variability. The PC1 showed 33.31% whereas PC2, PC3 accounted 23.24 % and 17.35 %. It was concluded that maximum variation was present in first PC. So selection of genotypes from this PC will be useful. Sruthi *et al.* (2023) reported that out of eight PCs, two exhibited more than 1.0 Eigen values and showed 67.50 per cent total variability among nine traits of forty rice germplasm lines. Venkata Ratnam *et al.* (2023) reported that out of fourteen traits studied in 37 rice genotypes, 3 PCs exhibited more than 1.0 eigen values and showed about 78.33% total variability among the characters studied.

Scree Plot

Scree plot elucidated the percentage of variance associated with each principal component obtained by drawing a graph between Eigen values and principal component numbers (Figure 1). Screen plot

maximum variation expressed in normal environment (E1), PC1 expressed the greatest variation compared to nine principle components, whereas PC1 and PC2 show the greatest variation as the Eigen value is more than one. Christina *et al.,* (2021) reported maximum variation of 38.72 per cent in the PC1 in their study on twenty five genotypes.

In flash flood environment (E2), it was clearly depicted in the graph that the highest variation was exhibited in PC1, PC2 and PC3.Similar results were obtained by Shivangi *et al.* (2021). Semi curve line is obtained after three PC tended to become straight with little variance observed in each PC. From the graph, it is clear that the maximum variation was observed in PC1 in comparison to other nine PCs. So, selection of lines from this PC will be useful. Those principal components having more than one Eigen value that showed more variation among the rice genotypes for the selection of the diverse parents. Characters with high variances were expected to provide high level of gene transfer during breeding program Gana *et al.* (2013).

Rotated component matrix presented in (Table 3 and Figure 2). In normal environment (E1), PCA revealed 62.66 per cent contribution of the first two

Table 2. Extracted Eigen values, percentage of variances and cumulative percentage of variances for each principal components under normal (E1) and flash flood environment (E2)

Traits	Principal components	Environment	Eigen Value	Percentage of Variance	Cumulative Percentage of Variance
DFF	PC1	E1	2.983	33.144	33.144
		E2	2.998	33.315	33.315
PH	PC2	E1	2.656	29.513	62.656
		E2	2.092	23.244	56.559
NPTPP	PC3	E1	0.948	10.533	73.189
		E2	1.562	17.355	73.914
PL	PC4	E1	0.730	8.115	81.304
		E2	0.844	9.376	83.289
NGPP	PC5	E1	0.629	6.985	88.289
		E2	0.637	7.083	90.372
HGW	PC6	E1	0.502	5.576	93.865
		E2	0.388	4.312	94.685
GL	PC7	E1	0.280	3.110	96.975
		E2	0.230	2.552	97.237
GB	PC8	E1	0.161	1.791	98.767
		E2	0.149	1.653	98.890
GYPP	PC9	E1	0.111	1.233	100.000
		E2	0.100	1.110	100.000

DFF -Days to fifty per cent flowering, **PH**- Plant Height, **NPTPP**- Number of Productive Tillers Per Plant, **PL** - Panicle Length, **NGPP** - Number of Grains Per Panicle, **HGW** - Hundred Grain Weight, **GL** -Grain Length, **GB** - Grain Breadth and **GYPP** - Grain Yield Per Plant.
E1- Normal Environment **E2**- Flash Flood Environment

Under Normal Environment (E1)

Under Flash Flood Environment (E2)
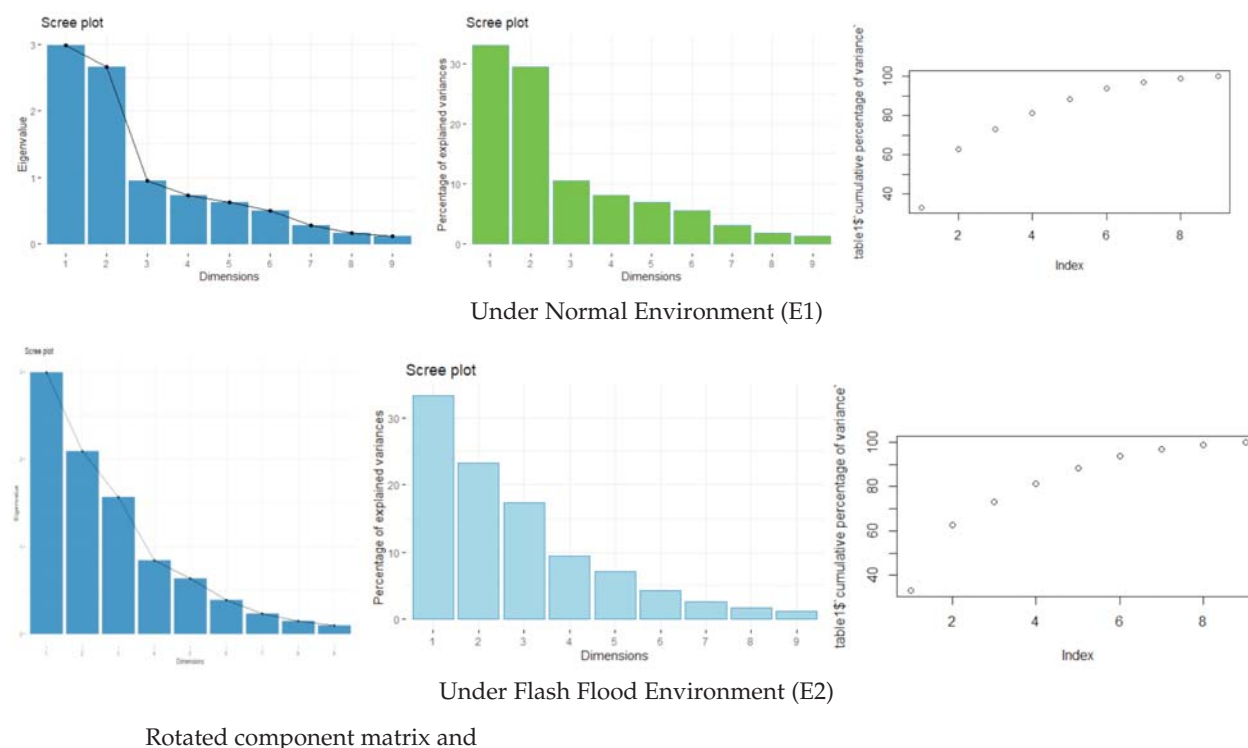
Rotated component matrix and

Fig. 1. Scree plot of principal component analysis for rice genotypes between their Eigen value, percentage of variance, cumulative percentage variance and the number of principal components under normal (E1) and Flash Flood Environments (E2)

principal components to the variability. PCA- biplot showed the PC1 contributed about 33.14% of the total variability and explained principally by 100 grain weight, grain length and grain breadth contributed positively, whereas grains per panicle (-0.520), grain yield (-0.445) and panicle length (-0.342) represented the maximum negative variance in this component. Hence, the selection of lines and traits from this PC will be useful. The second principal component (PC2) contributed to 29.52% of the total explained variance. The characteristics such as grain length (0.504), 100 grain weight (0.455) and grain breadth (0.438) described the highest loadings in this component.

In flash flood environment (E2), The PC1 accounts for days to fifty per cent flowering (-0.400), grain yield per plant (-0.391) and panicle length (-0.355). PC2 was highly loaded with traits such as grain yield per plant (0.396), number of grains per panicle (0.398) and number of productive tillers per plant (0.248). The PC3 was mostly dominated by yield contributing traits *viz.*, number of productive tillers per plant (0.011), number of grains per panicle (0.315), 100 grain weight (0.247), grain length

(0.052), grain breadth (0.282) and grain yield per plant (0.225).

PCA- Biplot

Biplot diagram analysis between PC1 and PC2 explains the distribution and nature of the diversity based on the better performing variables as well as genotypes. The length of the vector is based on the contribution of the character to the principal component (Table 3 and Figure 3). Moreover, the angle of the character vectors is reflecting the Correlation of variables. The angle between the trait vectors indicates the direction of association between the traits. An acute angle (less than 90º) indicates high positive correlation while obtuse angle (more than 90º) indicates high negative correlation. Orthogonal vectors indicate no correlation.

In the normal environment (E1), out of nine traits studied, the characters *viz.*, days to fifty per cent flowering, plant height, number of productive tillers per plant showed a positive correlation with grain yield per plant. The genotypes that are present close to the trait vector of the same quadrant would be the best performing for those traits. Based on biplot PC1
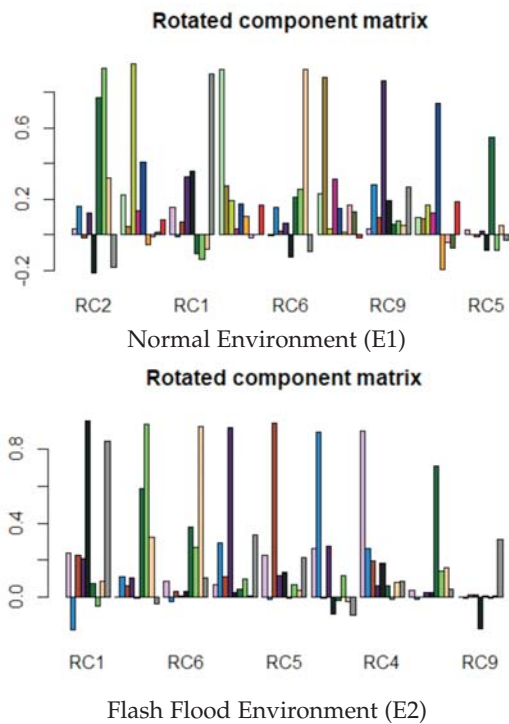
Fig. 2. Graphical and Box plot representation of Rotated component matrix for different traits of 30 rice genotypes under normal (E1) and Flash Flood Environments (E2)

and PC2 it was observed that G1, G2, G14 and G21 had high principal component score for PC1. The genotypes G23, G16 and G13 were formed three separate clusters. G27 and G26 are some of the poor performing genotypes for the traits under consideration as they are present in the opposite direction to the trait vector. It is concluded the genotypes with high principal component score can be selected for above said characters for breeding purpose in normal environments (E1).

In flash flood environment (E2), all the nine traits showed positive correlation with grain yield per plant. The genotypes namely G1and G2 and others in the particular quadrant perform better for the traits like number of grains per panicle and grain yield per plant whereas G27, G30 and G22 are some of the poor performing genotypes for the traits under consideration as they are present in the opposite direction to the trait vector (Figure 3). Similar results were reported by Christina *et al.* (2021) and Lakshmi *et al.* (2022). Studies on flash flood tolerance under coastal eco system are scarce. This study has brought out the hidden genetic diversity in the genotypes taken for the study. In the form of two PC axes under normal (E1) and three PC axes under flash flooded (E2) condition. Furthermore, correla-

Table 3. Rotated component matrix (Eigen vectors) for the different characters for the principal components under normal (E1) and Flash Flood Environments (E2)

Variables	Environment	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
DFF	E1	-0.339	0.224	-0.444	0.521	-0.367	0.249	-0.045	-0.178	0.365
	E2	-0.400	0.090	-0.324	0.428	-0.451	0.085	-0.219	-0.508	-0.166
PH	E1	-0.216	0.427	0.092	0.528	0.432	-0.165	0.000	0.105	-0.506
	E2	-0.183	-0.138	-0.688	-0.042	-0.200	0.097	0.195	0.523	0.355
NPTPP	E1	-0.360	0.110	-0.578	-0.513	0.122	-0.165	-0.335	-0.178	-0.268
	E2	-0.333	0.248	0.011	0.615	0.613	-0.089	-0.004	0.254	0.018
PL	E1	-0.342	0.309	0.538	-0.140	0.742	-0.221	-0.434	-0.092	0.475
	E2	-0.355	0.116	-0.360	-0.564	0.345	-0.288	-0.132	-0.104	-0.427
NGPP	E1	-0.520	-0.026	-0.067	-0.189	0.124	-0.082	0.583	0.518	0.234
	E2	-0.330	0.398	0.315	-0.121	-0.299	0.372	0.330	0.346	-0.403
HGW	E1	0.239	0.455	-0.045	-0.111	-0.548	-0.149	-0.180	0.588	-0.139
	E2	-0.329	-0.447	0.247	-0.088	0.009	0.301	-0.674	0.279	0.006
GL	E1	0.199	0.504	0.006	-0.179	-0.118	-0.295	0.557	-0.509	0.044
	E2	-0.312	-0.493	0.052	-0.057	0.297	0.381	0.527	-0.372	0.068
GB	E1	0.156	0.438	-0.006	-0.255	0.337	0.767	0.011	0.082	0.000
	E2	-0.321	-0.373	0.282	0.097	-0.285	-0.716	0.229	0.142	-0.055
GYPP	E1	-0.445	-0.054	0.403	-0.138	-0.461	0.356	0.115	-0.189	-0.476
	E2	-0.391	0.396	0.225	-0.288	-0.039	-0.080	-0.063	-0.193	0.712

DFF -Days to fifty per cent flowering, **PH**- Plant Height, **NPTPP**- Number of Productive Tillers Per Plant, **PL** - Panicle Length, **NGPP** - Number of Grains Per Panicle, **HGW** - Hundred Grain Weight, **GL** -Grain Length, **GB** - Grain Breadth and **GYPP** - Grain Yield Per Plant.

E1- Normal Environment   **E2**- Flash Flood Environment**Figure 2.** Graphical and Box plot representation of Rotated component matrix for different traits of 30 rice genotypes under normal (E1) and Flash Flood Environments (E2)
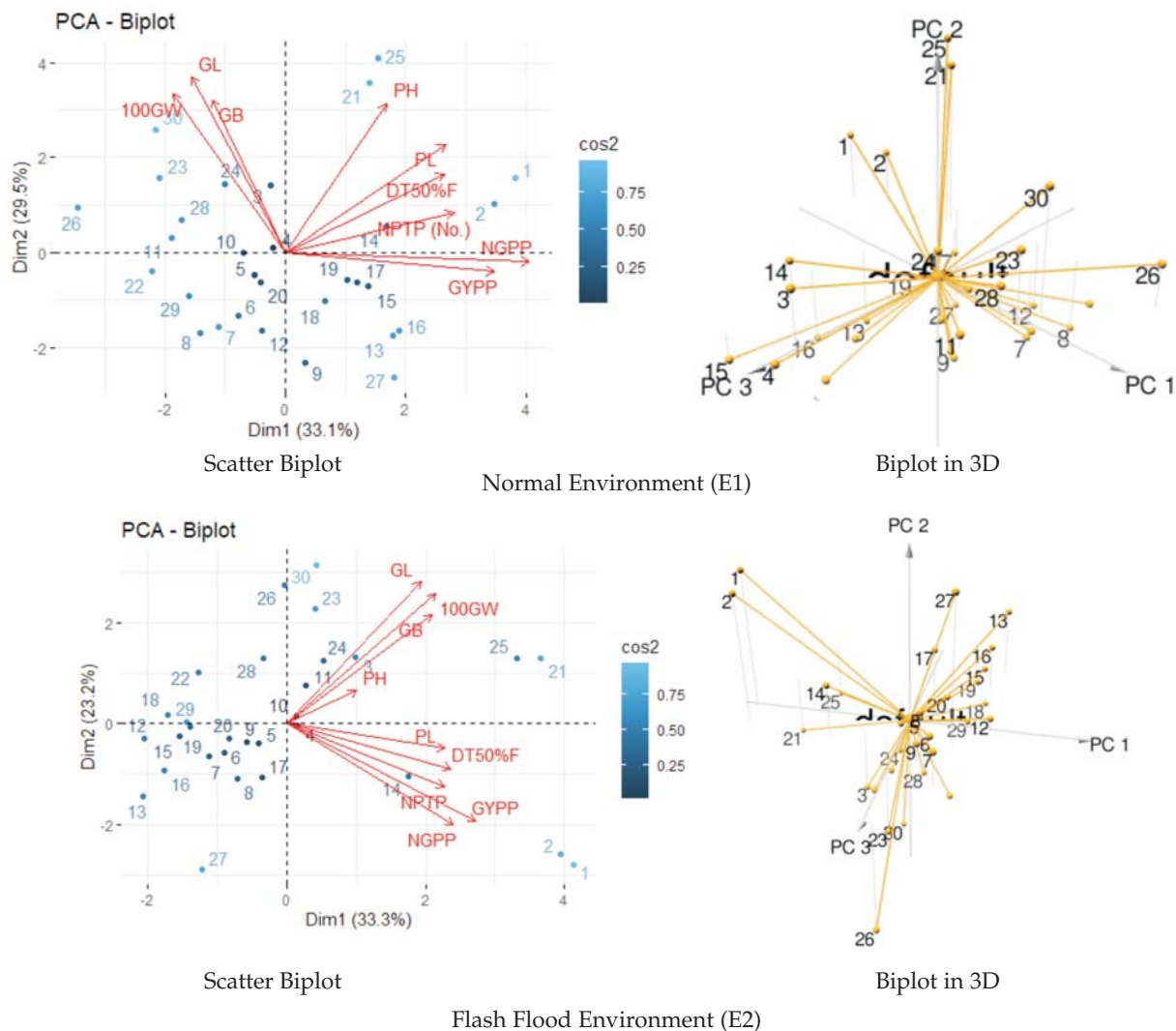
Scatter Biplot   Biplot in 3D

Normal Environment (E1)

Scatter Biplot   Biplot in 3D

Flash Flood Environment (E2)

Fig. 3. Scatter and 3D Visualization biplots of the various rice genotypes and characters represented in two major principal components under normal (E1) and Flash Flood Environments (E2)

tion studies under normal (E1) and flash flooded conditions revealed the hidden relationship of characters in maximizing yield under stress situations.

Conclusion

The present study identify the variability contributing parameters and to select flash tolerant genotypes for use as donors for the improvement of traits in future breeding programs. Has resulted in identification of five rice genotypes namely CR-1009 Sub-1, Swarna Sub-1, AU-1 GSR, Mappilai samba and Seeraga samba as potential donors for flash flood tolerance in rice.

Conflict of interest

The authors declared that there is no conflict of interest.

References

Bailey- Serres, J., Lee, S. C. and Brinton, E. 2012. Waterproofing crops: effective flooding survival strategies. *Plant Physiol.* 160: 1698– 1709.

Bin Rahman, A. N. M. R. and Zhang, J. 2022. The coexistence of flood and drought tolerance: an opinion on the development of climate-smart rice. *Frontiers in Plant Science*. 13. https://doi.org/10.3389/fpls.2022.860802.

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. *Elect. J. Pl. Breed.* 12(3): 907-911.

Directorate of Economics and Statistics. 2022.

Gana, A.S., Shaba, S.Z. and Tsado, E.K. 2013. Principal component analysis of morphological traits in thirty nine accessions of rice (*Oryza sativa* L.) under controlled (PVC pipes) condition. *Elect. J. Pl. Breed.* 1(4): 1016-1020.

Gautam, A. and Shrestha, A. 2023. Evaluation of panicle architecture traits in rice genotypes using PTRAP. *Nepal Agriculture Research Journal*. 15(1): 115–124.

https://CRAN.R-project.org/package=biotools.

IRRI, 2013. Standard evaluation system for rice, 5th edn. International Rice Research Institute, Los Baños.

Lakshmi, M., Shanmuganathan, M., Jeyaprakash, P. and Ramesh, T. 2022. Genetic variability and diversity analysis in selected rice (*Oryza sativa* L.) varieties. *Elect. J. Pl. Breed.* 13(3): 959-966.

Kumar, P., Sao, A., Gupta, A. K. and Kumar, M. 2015. Principal component analysis for assessment of genetic diversity in upland paddy for Bastar plateau. *Elect.*

J. Pl. Breed. 6(4): 1052-1059.

Sathyaraj, D, T. Sabesan, K. Sowmia, K. Saravanan and P. Sathish Kumar. 2023. Genetic variability of rice genotypes under screening of normal and flash flood condition. DOI: 10.31830/2454-1761.2023.CR-840. *Crop Res.* 2023; 58: 76-82.

Shivangi Rahangdale, Yogendra Singh, Upadhyay, P. K. and Koutu, G. K. 2021. Principal component analysis of JNPT lines of rice for the important traits responsible for yield and quality. *Indian J. Genet.* 81(1): 127-131. DOI:10.31742/IJGPB.81.1.14.

Sruthi, S. R., J. Johnny Subakar Ivin, Greetty Williams and Anbuselvam, Y. 2023. Genetic variability and diversity analysis in traditional rice (*Oryza sativa* L.) varieties of Tamil Nadu. *Environment and Ecology.* 41 (2): 746—753.

Yugandhar, P., Yafei Sun, Y., Liu, L., Neg, I.M., Nallamothu, N. and Sun, S. 2018. Characterization of the loss-of- function mutant NH101 for yield under phosphate deficiency from EMS-induced mutants of rice variety Nagina 22. *Plant Physiol. Biochem.* 130: 1-13.

Venkata Ratnam, T., B.N.V.S.R, Ravi Kumar, L.V. Subba Rao, T. Srinivas and A.D.V.S.L.P. Anand Kumar. 2023. Principal component analysis of yield and quality traits in Zinc rich landraces of rice (*Oryza sativa* L.). *Elect. J. Pl. Breed.* 13(4): 1162-1169. https://doi.org/10.37992/2022.1304.169.