Principal Component Analysis of Yield and Its Attributing Traits in Recombinant Inbred Lines of Rice Under Submerged Condition (*Oryza sativa L.*)

Lakshmeesha R [a*], Mahesh, H.B. [b], Basavaraj M Pattanashetti [c], K.M. Harinikumar [a] and Veena S Anil [a]

[a] Department of Plant Biotechnology, University of Agricultural Sciences, Bangalore, India.
[b] Department of Genetics and Plant Breeding, University of Agricultural Sciences, Bangalore, India.
[c] Department of Genetics and Plant Breeding, University of Agricultural Sciences, Dharwad, India.

Original Research Article

ABSTRACT

Rice, *Oryza sativa L.*, is the world's most important staple crop, feeding more than half of the world's population. The phenotype of plant is the result of interaction of many factors and final yield is the sum of total effect of several component factors. Therefore evaluation of genetic variability forms the basis for any crop improvement programme, the success of which depends on sufficient

_____

*Corresponding author: Email: lakshmeeshar123@gmail.com;

genetic variability among genotypes so as to permit effective selection. Hence Evaluation of Principle Component Analysis (PCA)of recombinant inbred lines (RILs) was done at phenotypic level under submerged conditions to reduce a large series of data into smaller number of components by looking for groups that have very strong inter-correlation in a set of variables and each component explained *per cent* variation to the total variability. The RIL population was derived from an inter-specific cross between BPT5204 and HPR14 parents. A study was conducted using 1256 Recombinant Inbreed Lines submerged condition in the two seasons at College of Agriculture V.C. Farm, Mandya with nine agro-morphological traits and a principle component analysis was carried out. Out of nine principle components, four exhibited Eigenvalue more than one governing 77.74% variance and 69.86% variance in the *summer* and *kharif* seasons respectively. The highest positive Eigenvalue was observed for total number of tillers, productive tillers, non-productive tillers and fallowed by single plant yield in PC1 in the *summer* and *kharif* season respectively. The highest positive Eigenvalue was observed for five panicle weight, single panicle length, single plant yield and plant height in PC2 of *summer* and *kharif* season respectively. Indicating their pronounced effect on the overall variation in the Recombinant Inbreed Lines of Rice.

Keywords: Eigenvalues; PCA; rice; RILS; yield.

1. INTRODUCTION

Rice is the most important and dependable food crop in India, feeding more than two-thirds of the population. In India, rice is grown in an area of 43 million ha with a production of 118.43 million tons and average productivity of 2.75 t ha-1 (Anon, 2020). The Kharif rice production in India during 2021–22 is estimated to be 107.04 million tonnes (Mt) [1]. Rice yield increase of 1.0–1.2% per year beyond 2020 will be required to feed the world's still-growing population while keeping prices low [2]. Conventional breeding for yield improvement is time consuming and difficult process since important quantitative characters like yield is mainly governed by environmental factors. Therefore evaluation of genetic variability forms the basis for any crop improvement programme, the success of which depends on sufficient genetic variability among genotypes so as to permit effective selection. Plant breeders often assess numerous traits, some of which may lack sufficient discriminatory power for germplasm evaluation, characterization, and management [3]. In such cases, Principal Component Analysis (PCA) emerges as a valuable tool. PCA is a multivariate statistical technique employed to unveil patterns within datasets and streamline them by reducing redundancy. It aims to simplify and assess the interrelationships among a multitude of variables into a smaller set of components without compromising the essential information within the original dataset [4]. According to Clifford and Stephenson [5], the initial three principal components typically capture the most crucial variation patterns. With the above arguments an experiment was conducted to evaluate the Principal Components Analysis of the 9 traits in Recombinant Inbred Lines population was derived from an inter-specific cross between BPT5204 and HPR14 parents under submerged condition.

2. MATERIALS AND METHODS

2.1 Plant Materials

Parents with diverse genetic background *viz.,* BPT-5204 (good grain qualities and high yield) and HPR14 (high protein content) [6] were crossed to develop 1255 segregating lines and selections were carried out based on phenotypic parameters in RILs of rice.

2.2 Experimental Site and Layout

The RIL population along with parents was planted in an augmented design at college of agriculture V C Farm, Mandya, during summer 2020 and observations were recorded on individual lines and used for statistical analysis. Twenty one days seedlings were transplanted with 20cm x 10cm spacing and minimum of 15 plants maintained in each line. Recommended cultural practices for rice cultivation were carried out to ensure uniform crop stand as per the package of practices. Phenotypic data was collected for a total of 1255 RILs related to yield and yield contributing traits like days to 50% flowering (DFF) was recorded as the number of days from germination to 50% of plants with initiation of flowering based on visual observation of the each RIL, plant height (PH) was recorded at maturity stage (cm) from base of the plant at soil surface to the panicle tip of main tiller and averaged over 5 plants, number of tillers per plant (TN) was measured by counting total

number of tillers per plant at harvesting stage and averaged over five plants, panicle length (PL) (cm) measured from the panicle neck to the tip (excluding awn) at reproductive stage, panicle weight (PW) was calculated by total weight of 5 panicles and averaged, thousand grain weight (TW) (grams) measured by weighing 1000 filled grains from each plant [7]. Principal Component Analysis was performed on the mean data collected concerning all of the phenotypic quantitative traits using R programme (V1.4.1717).

3. RESULTS AND DISCUSSION

Principal component analysis (PCA) is a statistical procedure that used to identify the traits which contribute maximum to the most observed variation within a large group of genotypes and employed to reduce the complexity of the data while minimizing the variation within the data and increasing interpretability [4]. It is efficient to finding the components useful to group the genotypes that maximize the variation and it is also helpful in reducing the number of variables with a maximum variance but does not estimate the specific effect. The PCA was performed separately for each season using data of all nine

agro-morphological traits evaluated on 1256 recombinant inbred lines of rice genotypes. The resulted PCA from nine independent principal components (PCs) explained cumulative variance during summer and Kharif season 2020.

The importance of a character towards the PCs could be seen from the corresponding eigenvalues. The results of PCA revealed that, out of the 9 PCs, first four PCs with eigenvalue of greater than 1.00 accounted for 77.74 *per cent* of total variability for summer season (Table 1).The PC1 accounted for 29.19 *per cent* of the total variation in the population during summer and the results also showed that total number of tillers (0.584), productive tillers (0.528), non-productive tillers (0.447), single plant yield (0.344) contributed positively and significantly to PC1 fallowed by days to fifty *per cent* flowering, five panicle weight and single panicle length. The PC2 accounted for 21.23 *per cent* of the total variation in the population during summer and the results also showed that five panicle weight (0.591), single panicle length (0.457), single plant yield (0.386), test weight (0.364) contributed positively and significantly to PC2 fallowed by plant height and days to fifty per cent flowering (Table 2) similar reports was shown by Sahu et al. [8], and Christina et al. [9].

Table 1. Estimate of eigenvalues and per centage of variation contributed by each principal component (PC) to the single plant yield and yield related components under submerged conditions during summer season

PC	Eigenvalue	Per centage of Variance	Cumulative per centage of variance
1	2.62	29.19	29.19
2	1.91	21.23	50.43
3	1.48	16.54	66.97
4	1.06	10.77	77.74
5	0.75	8.38	86.13
6	0.52	5.79	91.92
7	0.38	4.24	96.17
8	0.32	3.55	99.73
9	0.02	0.26	100

Table 2. Contribution of each trait to the principle component during summer season 2020

	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9
DFF	0.135	0.129	-0.704	-0.123	0.015	-0.139	0.537	0.382	-0.004
PH	-0.004	0.282	0.66	-0.124	-0.173	0.104	0.557	0.345	0.013
TT	0.584	-0.147	0.112	0.092	0.147	-0.052	0.091	-0.082	-0.76
NPT	0.447	-0.175	-0.025	-0.343	0.229	0.692	-0.104	0.137	0.301
PT	0.528	-0.112	0.138	0.276	0.084	-0.432	0.193	-0.229	0.575
TW	-0.132	0.364	-0.013	0.607	0.634	0.264	0.076	0.06	0.007
5 P W	0.143	0.591	-0.152	-0.112	-0.261	0.244	0.144	-0.668	-0.018
1 P L	0.083	0.457	0.102	-0.532	0.468	-0.416	-0.302	0.088	-0.001
S P Y	0.344	0.386	-0.053	0.321	-0.447	0.009	-0.477	0.446	0.028

DFF- Days to 50 % flowering, PH- Plant height (cm), TT- Total number of tillers, NPT- Number of non-productive tillers, PT-Number of productive tillers, TW- Test weight (gm), 5PW- 5 Panicle weight (gm), 1PL- One panicle length, SPY- Single plant yield

The first four PCs with eigenvalue of greater than 1.00 accounted for 69.86 per cent of total variability for kharif season (Table 3). The PC1 accounted for 25.67 per cent of the total variation in the population during kharif and the results also showed that total number of tillers (0.642), productive tillers (0.598), non-productive tillers (0.378), single plant yield (0.201) contributed positively and significantly to PC1 fallowed by days to fifty per cent flowering, five panicle weight. The PC2 accounted for 20.00 per cent of the total variation in the population during kharif and the results also showed that single panicle length (0.512), plant height (0.496), five panicle weight (0.476), single plant yield (0.410) contributed positively and significantly to PC2 fallowed by days to fifty per cent flowering and test weight (Table 4) similar findings were reported by Manohara et al. [10]. The screen plot showed the variance percentage between Eigenvalues and the principal components during summer and Kharif season respectively (Fig. 1 and Fig. 2). A similar was shown by Shivani et al. [11-13].

Table 3. Estimate of eigenvalues and per centage of variation contributed by each principal component (PC) to the single plant yield and yield related components under submerged conditions during Kharif season

PC	Eigenvalue	Per centage of variance	Cumulative per centage of variance
1	2.31	25.67	25.67
2	1.80	20.00	45.68
3	1.12	12.46	58.14
4	1.05	11.72	69.86
5	0.83	9.24	79.10
6	0.71	7.90	87.01
7	0.64	7.17	94.18
8	0.52	5.78	99.97
9	0	0.02	100

Table 4. Contribution of each trait to the principle component during kharif season 2020

	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9
DFF	0.194	0.272	-0.542	-0.145	0.556	0.024	0.513	0.009	-0.001
PH	-0.043	0.496	0.456	-0.216	-0.085	-0.061	0.257	0.649	0.001
TT	0.642	-0.077	0.087	-0.031	-0.026	-0.162	-0.084	0.047	-0.733
NPT	0.378	-0.106	0.412	0.035	0.379	0.686	-0.083	-0.026	0.232
PT	0.598	-0.05	-0.05	-0.044	-0.164	-0.437	-0.067	0.062	0.64
TW	-0.028	0.026	0.131	0.894	0.292	-0.249	0.08	0.167	-0.003
5 P W	0.065	0.479	-0.415	0.126	-0.013	0.215	-0.682	0.26	0
1 P L	-0.047	0.512	0.356	-0.127	0.305	-0.319	-0.218	-0.592	-0.002
S P Y	0.201	0.41	-0.044	0.311	-0.578	0.314	0.365	-0.356	0.001

DFF- Days to 50 % flowering, PH- Plant height (cm), TT- Total number of tillers, NPT- Number of non-productive tillers, PT-Number of productive tillers, TW- Test weight (gm), 5PW- 5 Panicle weight (gm), 1PL- One panicle length, SPY- Single plant yield
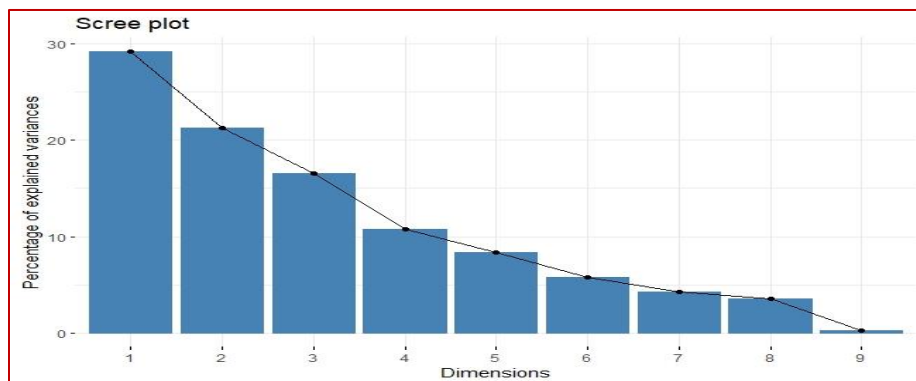

Fig. 1. Scree plot based on PCA indicating percentage of explained variance on Y axis by each individual principal component on X- axis (*Summer* season)
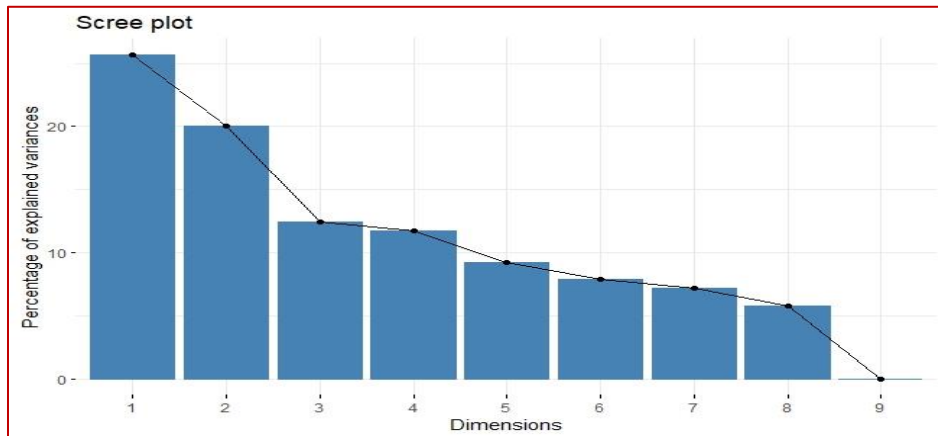
Fig. 2. Scree plot based on PCA indicating percentage of explained variance on Y axis by each individual principal component on X- axis (*Kharif* Season)

4. CONCLUSION

PCA can be used to extract all of the relevant components and emphasize their contributions to overall variability, making it an excellent tool for speeding up the breeding process. Out of nine principle components, four exhibited Eigenvalue more than one governing 77.74% variance and 69.86% variance in the *summer* and *kharif* seasons respectively. The highest positive Eigenvalue was observed for total number of tillers, productive tillers, non-productive tillers and fallowed by single plant yield in PC1 in the *summer* and *kharif* season respectively. The highest positive Eigenvalue was observed for five panicle weight, single panicle length, single plant yield and plant height in PC2 of *summer* and *kharif* season respectively. Indicating their pronounced effect on the overall variation in the Recombinant Inbreed Lines of Rice.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Anonymous. Second Advanced Estimates of Production of Major crops for 2021-22 Press Information Bureau, Delhi; 2021.
2. Anonymous. GRiSP (Global Rice Science Partnership 2013). Rice almanac, 4th edition. International Rice Research Institute, Los Baños, Philippines. 2013; 283.
3. Maji AT, Shaibu AA. Application of principal component analysis for rice germplasm characterization and evaluation. Journal of Plant Breeding and Crop Science. 2012;4(6):87–93.
4. Upadhyay S, Rathi S, Choudhary M, Snehi S, Singh V, Singh PK, Singh RK. Principal Component analysis of Yield and its attributing Traits in advanced Inbred Lines of Rice under Sodicity condition (*Oryza sativa* L.); 2022.
5. Clifford HT, Stephenson W. An Introduction to Numerical Classification. Academic Press, London. 1975;229.
6. Hittalmani SV, Inheritance of quantitative traits and exploitation of male sterility in rice (*Oryza sativa* L.) breeding. M.Sc Thesis ,Univ. Agril. Sci. Bangalore;1990.
7. Mohanty A, Marndi BC, Sharma S, Das A. Biochemical characterization of two high protein rice cultivars from Assam ric ecollections. Oryza. 2011;48(2):171-174.
8. Sahu H, Saxena RR, Verulkar SB, Rawte S. Association, principal component and genetic divergence study in Recombinant Inbred Lines (RIL's) Population of Rice. International Journal of Bio-resource and Stress Management. 2016;7(4):673–681.
9. Christina M, Jones MR, Versini A, Mézino M, Lemezo L, Auzoux S, Soulie JC, Poser C, gérardeaux E. Impact of climate variability and extreme rainfall events on sugarcane yield gap in a tropical Island. Field Crops Research, 2021;274:108-112.
10. Manohara KK, Mahajan GR. Sahoo RN. Spectroscopy based novel spectral indices, PCA-and PLSR-coupled machine learning models for salinity stress phenotyping of rice. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. 2020;229:117-121.

11. Shivani D, Neeraja CN, Cheralu C, Shankar VG. Multivariate analysis and character association studies for yield and nutritional characters in swarna and type 3 RIL population of rice (*Oryza sativa* L.). Journal of Cereal Research. 2021;13(2): 180–187.

12. Anonymous. Adika iluvarige adunika besaya paddathigalu package of practices for field vrops. Univ. Agril. Sci., Bangalore; 2020.

13. Meghana HS, Shailaja H, Gandhi RV, Andmeera N. Phenotypic screening for salt tolerance at germination and seedling stage and SSR marker validation in rice (*Oryza sativa* L.). Mysore J. Agric. Sci. 2015;49(4):686-692.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/114425